UNIVERSITY OF PADOVA

DEPARTMENT OF INFORMATION ENGINEERING
Ph.D. Course in Information Engineering
Information and Communication Science and Technologies Curriculum
XXXV series

Exploiting mmWave radios for indoor environmental sensing

Ph.D. Candidate
Enver Bashirov

Ph.D. Supervisor
Professor Michele Rossi

Ph.D. Coordinator
Professor Andrea Neviani

Academic Year
2022–2023

To all life

Abstract

The concept of contactless detection of human activity has the potential to revolutionize our interaction with technology and surroundings, enabling remote monitoring systems that are unobtrusive. Millimeter-Wave (mm-Wave) reflected radio signals have drawn significant attention from academia and industry due to their ability to detect, track, and analyze human movement and other objects using the Radio Detection and Ranging (RADAR) principle. The high sensitivity and reliability of mm-Wave frequencies in detecting small-scale movements of the human body, coupled with the non-intrusiveness of these signals, make them a promising technology for applications where privacy is a concern. However, there exist challenges in practical application of this technology, such as the complexity of reflections in real-life environments and limited Field-of-View (FOV) of the RADARs, requiring tedious investigation.

This thesis delves into the mm-Wave localization and sensing fields, exploring various applications and developing integrated sensing and communication platforms. The research begins with RADAR-based person tracking using supervised and unsupervised Deep Learning (DL) algorithms, that are compared to the literature's cornerstone algorithm, Kalman filter (KF). An integrated sensing and communication platform using IEEE 802.11ay Access Points (APs) is then built for Human Activity Recognition (HAR), extracting the Micro-Doppler (μD) signatures from the Training (TRN) units embedded in the data packets for beam tracking. Finally, a RADAR network testbed is implemented for sensing applications with multiple objectives, exploiting the strengths of multiple RADARs and shared field-of-view for distributed and real-time tasks.

The thesis provides a comprehensive perspective on mm-Wave environment sensing, utilizing various hardware and combining standard signal processing techniques and data-driven machine learning techniques to develop algorithms. The results are supported by extensive experimentation with state-of-the-art mm-Wave RADAR and Integrated Sensing And Communication (ISAC) testbeds.

Sommario

Il concetto di rilevamento contactless dell'attività umana ha il potenziale per rivoluzionare la nostra interazione con la tecnologia e l'ambiente circostante, consentendo lo sviluppo di sistemi di monitoraggio remoto non invasivi. I segnali radio nelle frequenze Millimeter-Wave (mm-Wave) hanno attratto l'interesse dei ricercatori e dell'industria grazie alla loro capacità di rilevare, tracciare e analizzare i movimenti umani e di altri oggetti utilizzando il principio del Radio Detection and Ranging (RADAR). L'alta sensibilità e affidabilità delle frequenze mm-Wave nel rilevare i movimenti del corpo umano, combinata con la non invasività di questi segnali, li rendono una tecnologia promettente per le applicazioni in cui il la privacy dei soggetti va preservata. Tuttavia, esistono diverse sfide nell'applicazione di questa tecnologia, come la complessità delle riflessioni negli ambienti reali e il limitato campo visivo dei RADARs, che richiedono lo sviluppo di nuove soluzioni.

Questa tesi approfondisce i temi della localizzazione e della rilevazione di persone con segnali RADAR mm-Wave, esplorando diverse applicazioni e sviluppando piattaforme integrate di rilevazione e comunicazione congiunti. La ricerca inizia con il tracciamento di persone basato su RADAR utilizzando algoritmi di apprendimento supervisionato e non supervisionato Deep Learning (DL), che vengono confrontati con l'algoritmo di riferimento della letteratura, ovver il Kalman filter (KF). Viene poi costruita una piattaforma integrata per effettuare rilevazione e comunicazione congiuntamente, utilizzando Access Points (APs) IEEE 802.11ay per il riconoscimento delle attività dei soggetti, estraendo le firme Micro-Doppler (μD) dalle unità Training (TRN) incorporate nei pacchetti dati. Infine, viene implementata una rete di RADAR di test per applicazioni di rilevazione con obiettivi multipli, sfruttando i punti di forza di più RADARs e il campo visivo condiviso.

La tesi fornisce una prospettiva completa sulla rilevazione di persone in ambienti indoor con segnali mm-Wave, utilizzando vari hardware e combinando tecniche di elaborazione del segnale standard e tecniche di apprendimento automatico basate sui dati. I risultati sono supportati da un'ampia sperimentazione con RADARs mm-Wave all'avanguardia e con testbed Integrated Sensing And Communication (ISAC).

x

Contents

Listing of figures

Listing of tables

Listing of acronyms

Symbols

μ**D** micro-Doppler

2D 2-Dimensional

3D 3-Dimensional

3GPP Third Generation Partnership Project

4G fourth generation

5G fifth generation

A

A-BFT Association Beamforming Training

ABT Asymmetric Beamforming Training

AD Angle-Doppler

ADC Analog-to-Digital Converter

ADoA Angle Difference-of-Arrival

AE Autoencoder

AoA Angle of Arrival

AoD Angle of Departure

AP Access Point

API Application Program Interface

AWV Antenna Wave Vector

B

BI Beacon Interval

BP Beam Pattern

BPTT Back-Propagation Through Time

BRProt Beam Refinement Protocol

BS Base Station

C

CA-CFAR Cell-Averaging Constant False Alarm Rate

CBAP Contention Based Access Period

CDF Cumulative Distribution Function

CFO Carrier Frequency Offset

CIR Channel Impulse Response

CNN Convolutional Neural Network

COM Center of Mass

COTS Commercial-Off-The-Shelf

CSI Channel State Information

D

DAC Digital-to-Analog Converter

DAE Denoising Autoencoder

DBSCAN Density-Based Spatial Clustering of Applications With Noise

DFT Discrete Fourier Transform

DKF Discrete Kalman Filter

DL Deep Learning

DSP Digital Signal Processing

DT Doppler-Time

DTI Data Transfer Interval

E

ECD Expand-Contract Dilation

EKF Extended Kalman Filter

ELU Exponential-Linear Unit

EM Expectation Maximization

EN-DC Enhanced UTRA-Dual Connectivity

ESPRIT Estimation of Signal Parameters via Rotational Invariance Techniques

F

FC *Fusion Center*

FCon Fully Connected

FCOS Fully Connected One-Stage

FFT Fast Fourier Transform

FMCW Frequency-Modulated Continuous-Wave

FOV Field-of-View

FPGA Field-Programmable Gate Array

FSCN Fractionally Strided Convolutional Network

FTM Fine Time measurement

G

GAN Generative Adversarial Network

GRU Gated Recurrent Unit

GSCM Geometry-Based Stochastic Channel Model

GT Ground Truth

H

HAR Human Activity Recognition

HDC Hybrid Dilated Convolution

HR Heart Rate

HVRAE Hybrid Variational RNN Autoencoder

I

IF Intermediate Frequency

IFFT Inverse FFT

IIR Infinite Impulse Response

IoT Internet of Things

ISAC Integrated Sensing And Communication

ITU-R International Telecommunication Union – Radiocommunication Sector

J

JCR Joint Communication & Radar

K

KF Kalman filter

L

LEO Localization Error Outage

LFMCW Linear Frequency-Modulated Continuous-Wave

LM Levenberg-Marquardt

LMS Least Mean Squares

LOS Line-of-Sight

LSTM Long-Short Term Memory

M

MAC Medium Access Control

MAE Mean Absolute Error

MCU Micro-Controller Unit

MF Matched Filter

MIMO Multiple-Input Multiple-Output

MISO Multiple-Input Single-Output

ML Machine Learning

mm-Wave Millimeter-Wave

MMSE Minimum Mean-Square Error

MOTA Multi Object Tracking Performance Accuracy

MOTP Multi Object Tracking Performance Precision

MPC Multipath Component

MSE Mean-Square Error

MUSIC MUlti SIgnal Classification

N

NLOS Non-Line-of-Sight

NN Neural Network

O

OKS Object Keypoints Similarity

P

PA Power Amplifier

PDP Power-Delay Profile

PHY Physical Layer

PRS Positioning Reference Signal

PSD Power Spectral Density

PW Pulsed Wave

R

RA Range-Azimuth

RADAR Radio Detection and Ranging

RCS Radar Cross Section

RD Range-Doppler

RDA Range-Doppler-Azimuth

ResNet Residual Network

RF Radio Frequency

RGB Red-Green-Blue

RMS Root Mean Square

RMSE Root Mean Square Error

RNN Recursive NN

RSS Received Signal Strength

RSSI Received Signal Strength Indicator

RT Range-Time

RTT Round-Trip Time

RX Receive/Receiver

RZF Regularized Zero-Forcing

S

S2S Sequence-to-Sequence

S2SAE Sequence-to-Sequence Autoencoder

SAF Spatial Attention Fusion

SAGE Space-Alternating Generalized Expectation Maximization

sDAE Supervised Denoising Autoencoder

SDR Software Defined Radio

SLAM Simultaneous Localization and Mapping

SLS Sector Level Sweep

SNR Signal-to-Noise Ratio

SoC System on a Chip

SP Service Period

SPI Subsample Peak Interpolation

SRS Sounding Reference Signal

sS2SAE Supervised Sequence-to-Sequence Autoencoder

SSW Sector Sweep

STA Station

STFT Short-Time Fourier Transform

SVM Support Vector Machine

T

TDoA Time Difference-of-Arrival

TF Time-Frequency

ToA Time of Arrival

ToF Time of Flight

TRN Training

TX Transmit/Transmitter

TXSS Transmit Sweep

U

UAV Unmanned Aerial Vehicle

uDAE Unsupervised Denoising Autoencoder

UE User Equipment

UKF Unscented Kalman Filter

URLLC Ultra-Reliable Low-Latency Communications

uS2SAE Unsupervised Sequence-to-Sequence 3Autoencoder

USRP Universal Software Radio Peripheral

V

V2X Vehicle-to-Everything

VCO Voltage-Controlled Oscillator

W

WLAN Wireless Local Area Network

Z

ZF Zero-Forcing

"Be yourself; everyone else is already taken."

Oscar Wilde

1
Introduction

1.1   Background and Motivation

With the expanding capabilities of modern-day passive sensing, the application areas in which such technologies are used are becoming more accurate and robust. Modern-day technology areas, such as autonomous driving or vital sign monitoring, are thriving due to this fact. Usually, camera-based systems attempt to replicate, or in some cases even enhance, what human beings are capable of perceiving visually. However, such systems are prone to such environmental conditions as lighting and air quality; not to mention these systems are constricted by what is possible to realize within the optical spectrum of electromagnetic waves.

Millimeter-Wave (mm-Wave) sensors have emerged as a promising alternative to traditional camera-based sensors, as they offer similar capabilities without the constraints of the latter. In recent years, mm-Wave sensors have gained significant attention in the field of indoor environment sensing. The reason for this is their ability to enhance human comfort and productivity. Furthermore, the demand for more accurate and comprehensive information about the environment is on the rise, which has led to the recognition of the untapped potential of mm-Wave sensors waiting to be harnessed.

Achieving high-quality indoor environmental sensing requires advanced sensing technologies that can provide precise and dependable measurements under various environmental conditions, including lighting levels (light, dark, etc.) and air quality (foggy, dusty, etc.). Traditional wireless sensors based on low-frequency radio technologies and camera systems, such as surveillance cameras and smartphones, have limitations in terms of resolution quality, accuracy, and sensing range.

To overcome these limitations, researchers have turned to mmWave sensors due to their unique properties such as high bandwidth, high data rate, and high spatial resolution. mm-Wave radios operate at frequencies above 30 GHz and can provide more accurate and detailed information

about the environment compared to traditional wireless sensor networks and popular camera systems. mm-Wave sensors are not affected by lighting conditions, and can even operate in dusty and foggy environments. This makes them an excellent choice for applications that require high-quality indoor environmental sensing, such as smart homes, hospitals, shopping centers and office buildings.

1.2 Environment Sensing with mmWave Sensors

Sensing in the mm-Wave band has become an increasingly important research topic in recent years due to the unique properties of the mm-Wave spectrum, which includes frequencies between 30 GHz and 300 GHz. Compared to lower frequency bands, mm-Wave signals have shorter wavelengths, which allows for higher resolution and accuracy in sensing applications [1]. In particular, mm-Wave sensing is well-suited for applications that require high-resolution imaging [2], [3] and sensing, such as environmental sensing, Human Activity Recognition [4]–[6], and medical imaging [7], [8]. mm-Wave sensors can penetrate through various materials, including walls and clothing, making them suitable for imaging and sensing through obstacles. Additionally, mm-Wave sensing can be used for accurate localization and tracking of objects and people, as well as for gesture recognition and vital sign monitoring. However, mm-Wave sensing also presents unique challenges, including high path loss, low penetration through some materials, and susceptibility to interference [9], [10]. Therefore, new signal processing and sensing techniques are required to overcome these challenges and fully exploit the potential of mm-Wave sensing.

Although there are various types of Radio Detection and Rangings (RADARs) present in this field, most often Frequency-Modulated Continuous-Wave (FMCW) RADARs [11] are seen in the sensing fields such as indoor monitoring and automotive applications. When diving deeper into the activity recognition domain, we see a widely used technique called Micro-Doppler (μD) [12]. This technique allows us to differentiate the micro-movements within the targets body from its overall trajectory.

However, mm-Wave sensing devices unfortunately do not have infinite Field-of-View (FOV). To cope with this drawback, we should cover as much 3D ground as possible. Developing and using better sensors could be one of the options however we already have commercial off-the-shelf RADARs that are affordable and are tailored for this very purpose. We can create a *RADAR network* [6], [13] which can be used as a system of nodes working together in a concurrent and real-time fashion to achieve greater FOV thanks to the shared vision from each device. In the rest of this section we will introduce some of these aspects from mm-Wave RADAR technologies which are the underlying factors behind the works carried out throughout the thesis.

Frequency-Modulated Continuous Wave Radar

A FMCW RADAR [11] is a type of RADAR that uses a continuous wave with a frequency that is modulated by a linear or nonlinear ramp signal. This modulation of the frequency allows the RADAR to measure the range and velocity of targets in its field of view. In a FMCW RADAR,

the frequency of the transmitted signal (which is often called *chirp*) is continuously varied over a range during each transmission. The received signal is then mixed with a copy of the transmitted signal to produce a beat frequency, which is proportional to the range of the target. The modulation of the frequency allows for a range resolution that is proportional to the bandwidth of the transmitted signal. FMCW RADARs are commonly used in applications such as automotive collision avoidance systems, motion detection, and distance measurement. They are also used in a variety of other applications where accurate range and velocity measurements are required. Compared to other RADAR systems, FMCW RADARs are typically less expensive, lighter in weight, and consume less power, making them well-suited for many different applications [14]. Thus, the FMCW RADARs are thos that are mainly used for mm-Wave environment sensing nowadays and as such in this thesis, the types of RADARs used are also FMCW. FMCW RADAR is detailed in Section 2.6.3.

Micro-Doppler Signature

Specifically, Micro-Doppler (μD) signature [12] refers to the unique RADAR signal produced by the movement of small objects or parts of larger objects, such as moving limbs or rotating blades. It is a Doppler RADAR signal that results from the movement of individual components within a larger object. The micro-Doppler effect is caused by the movement of objects that have a sub-wavelength size relative to the wavelength of the RADAR signal. This movement results in a Doppler shift in the RADAR signal that is proportional to the velocity and direction of the object's movement. The micro-Doppler signature can be used to identify the type of object that is producing the signal, as different types of movement produce unique signatures. For example, the μD signature of a person walking is different from that of a person running, and the signature of a helicopter rotor blade is different from that of a wind turbine blade. An explanation of μD can be found in Section 2.6.4. The μD is the baseline for the works discussed in Chapters 3 and 4.

Radar Networks

Radar network is a system consisting of multiple RADAR sensors that work together to provide a comprehensive view of an area of interest. Such a system is usually composed of different types of RADARs, such as ground-based, airborne, or space-based, that operate at different frequencies and have different capabilities. The main purpose of RADAR networks is to increase situational awareness and improve the accuracy and reliability of RADAR data. By using multiple sensors, the network can detect and track objects from different angles, reducing the likelihood of false alarms and improving the accuracy of object position and velocity estimates. Additionally, RADAR networks can provide redundancy in case of sensor failure, and they can cover larger areas than a single RADAR sensor. Overall, RADAR networks provide a powerful tool for monitoring and managing the environment and enhancing situational awareness for a wide range of applications. In Chapter 5, a system for indoor application is designed and developed where the underlying idea is specifically the RADAR networks.

1.3 Research Objectives

This thesis explores the potential indoor environmental sensing by truly exposing mm-Wave sensor capabilities for indoor applications. More specifically, this thesis is aimed on:

1. To investigate the performance of mm-Wave radios for indoor environmental sensing and compare it with traditional sensing techniques;

2. To develop novel sensing techniques based on mm-Wave radios to enhance the accuracy and reliability of indoor environmental sensing;

3. To explore the potential of mm-Wave radios for other applications, such as indoor localization, tracking, and Human Activity Recognition (HAR);

4. To design and implement a mm-Wave RADAR network for environmental sensing that is robust, expandable, and easy to use, and to evaluate its performance in a real-world application.

1.4 Thesis Outline

The following chapters explore the state of the art mm-Wave techniques and applications, Human Activity Recognition (HAR) aspect of mm-Wave RADARs, usage of traditional WiFi Access Points (APs) as RADAR-like tool and designing a flexible RADAR network environment for future real-world applications. The rest of the thesis is organized as follows.

Chapter 2 reviews the latest developments in device-based and device-free indoor localization and sensing using mm-Wave communication and RADAR devices [1]. It covers important concepts regarding mm-Wave signal propagation and system design, and provides an overview of various approaches, algorithms, and applications.

Chapter 3 presents a Deep Learning (DL)-based tracking algorithm for accurate indoor human tracking using mm-Wave RADAR [5]. The proposed model-free tracking approach based on Denoising Autoencoder (DAE) and Sequence-to-Sequence Autoencoder (S2SAE) Neural Network (NN) outperforms state-of-the-art methods and can be trained in a supervised or unsupervised manner. Backscattered mm-Wave radio signals are used to track humans as they move within indoor environments. The proposed system is tested on our own measurements, obtaining sufficiently low tracking errors and outperforming the common Extended Kalman Filter (EKF) method.

Chapter 4 presents RAPID [4], a human detection and sensing system based on retrofitting IEEE 802.11ay access points for zero-cost monitoring of human movement and activities in indoor spaces. RAPID performs activity recognition and person identification using existing wireless networks with no modifications to the transmitted packet structure specified by the standard. The system uses the beam training and tracking mechanisms of IEEE 802.11ay to accurately localize and track multiple individuals and extract micro-Doppler signatures of the subjects. RAPID is

implemented on an IEEE 802.11ay-compatible Field-Programmable Gate Array (FPGA) platform with phased antenna arrays and achieves reliable positioning and tracking of multiple subjects and significant accuracy for activity recognition and person identification.

Chapter 5 presents the design and implementation of RadNet [6], an experimental testbed for mm-Wave RADAR networks. The chapter discusses the limitations of single mm-Wave RADAR sensors and the need for RADAR networks for medium to large indoor spaces. The chapter presents the architecture and functionality of RadNet and demonstrates its ability to deploy and test RADAR network algorithms. The chapter concludes by showing experimental results of a multi-RADAR people tracking algorithm implemented on the RadNet experimental platform.

Finally, the Chapter 6 presents the conclusions of the thesis and provides an overview of the contributions of the research. The chapter also discusses the limitations of the research and suggests future of mm-Wave environment sensing.

"The fool doth think he is wise, but the wise man knows him-self to be a fool."

William Shakespeare

2

A Review Of Millimeter Wave Device-Based Localization and Device-Free Sensing Technologies and Applications

2.1 Introduction

Millimeter-Wave (mm-Wave) communications in the 28–300 GHz band are looked at with great interest, as they may be able to quench –at least temporarily– the ever-increasing bandwidth requirements of such applications as massive Internet of Things (IoT), virtual/augmented reality, mobile cloud services and ubiquitous ultra-high definition multimedia streaming [15]–[17]. This would cover the shortcomings of sub-6 GHz technologies such as WiFi and Fourth generation (4G) cellular networks, which currently cannot support the massive bandwidth and number of users the above applications imply.

   The potential of mm-Wave technology, however, is not limited to higher-rate communications: rather, mm-Wave devices can become a proxy for high-resolution device-based localization as well as device-free sensing. These capabilities follow from the physics of mm-Wave propagation. First, the shorter wavelength of mm-Wave (compared to sub-6 GHz signals) enables accurate location estimates and lower location error bounds [18], [19]. Second, mm-Wave have well-known and peculiar propagation characteristics [20], [21] which yield higher spatial scanning resolution. For example, mm-Wave propagate quasi-optically, meaning that a Line-of-Sight (LOS) Multipath Component (MPC) is predominant over Non-Line-of-Sight (NLOS) contributions to the received signal [22]. Scattering also has a limited impact off typical non-rough reflecting surfaces such as walls, furniture, metal plates as well as glass layers [23], [24].

   Another consequence of mm-Wave propagation is that mm-Wave signals undergo much higher

path loss with respect to microwaves. To compensate for this attenuation, and still enable long-reach wireless links, mm-Wave devices resort to large or massive antenna arrays. Via beamforming, they can focus their transmitted energy towards a confined portion of the 3-Dimensional (3D) space, and thus achieve greater directionality. While this requires specific protocols for initial access [25]–[27] and beam training such as the IEEE 802.11ad [28], [29] and 802.11ay [30], [31] standards, it also means that a reduced amount of power is typically directed towards secondary multipath components. In addition with the quasi-optical propagation patterns discussed above, the main consequence is that the received angular spectrum of a mm-Wave signal is sparse: in typical conditions, one can identify one LOS MPC along with a number of NLOS MPC corresponding to signal reflections off the surrounding environment. The above features of mm-Wave communications have significant implications for localization and sensing [32]. For example, being able to separate MPC in the angular domain enables angle-based localization schemes that are not normally used in sub-6 GHz systems due to limited angular resolution when using small antenna arrays. Fingerprinting-based algorithms can also be enhanced by incorporating angle-based features to improve location discrimination. From the point of view of device-free sensing, mm-Wave propagation also implies typically clearer reflections off sensed targets and parts thereof. For example, a) quasi-optical mm-Wave propagation along with b) the large mm-Wave bandwidth available at typical mm-Wave radar frequencies respectively imply that reflections off targets are usually separate in the a) angle and b) time domains. This makes it possible to measure features that point to each reflection's movement velocity (e.g., the Doppler shift) and use this data to precisely localize and identify different targets.

In this work, we focus on indoor mm-Wave device-based localization and device-free sensing, and provide a comprehensive review of approaches, technologies, schemes and algorithms to estimate a device or object's location in an indoor environment. The objective of our survey is to shed light on indoor applications of localization and sensing using mm-Wave signals.

Location information can be extremely useful in different indoor setups [33], [34]. For example, in factories and industrial environments, location information can be exploited to enhance Ultra-Reliable Low-Latency Communications (URLLC) for industrial IoT and smart manufacturing [35], [36]. Accurate localization and sensing can benefit healthcare scenarios for patient tracking and lifesign/behavior monitoring, help people navigate in indoor areas, provide trajectory suggestions through relevant waypoints in museums, malls, and company headquarters, as well as support mission-critical applications such as disaster relief and indoor security. Location systems are also crucial for network performance optimization. Accurate location information can support the fast alignment of transmit and receive antenna arrays, optimize the association between clients and Access Point (AP), and prevent blockage of high-power LOS paths via predictive handovers to provide seamless coverage. This can result in low-latency communications as needed for augmented reality, virtual reality, and tactile Internet applications.

In the following, we start with an overview that touches on mm-Wave signal structure and propagation characteristics that make this domain unique with respect to other radio communication and sensing technologies. We consider practical constraints that define the applicability of algorithms and processing schemes to mm-Wave devices operating indoors. We then delve into

a detailed description of device-based indoor localization algorithms, explaining the main localization techniques employed in the literature, and how they are practically implemented in real mm-Wave hardware whenever available. For device-free sensing, we list a number of relevant applications and technologies that leverage mm-Wave hardware and signals to detect, localize and track targets indoors, as well as to specifically identify features related to sub-sections of a target (e.g., a part of the human body). Because these device-free approaches are mainly based on mm-Wave radar devices, we will briefly discuss how mm-Wave radar bands are being standardized for different applications.

2.1.1 Differences with respect to previous surveys

Localization and sensing are topics of great interest for both current and future-generation wireless communication system engineering. The research on these topics has proceeded at a steady pace, considering aspects as diverse as localization techniques, heterogeneous technologies, different scenarios, and different kinds interactions between the device to be localized and the location server, among others. Several surveys cover these aspects, typically for sub-6 GHz technologies. For example, Zafari et al. [37] and Geok et al. [38] focus on localization techniques for wireless systems in general, and cover heterogeneous technologies. These works only tangentially consider mm-Wave, and instead survey geometric and signal processing-based localization methods for sub-6 GHz systems. Ngamakeur et al. [39] delve into device-free sensing of different human signatures using sub-6 GHz technologies indoors. Here, the focus is on the localization, tracking and identification of multiple subjects using Wi-Fi and other kinds of wireless sensors.

By leveraging similar technologies, Singh et al. [40] consider techniques and algorithms to localize IoT devices indoors. In this case, the focus of the survey is on a specific source of location information (received WiFi signal strength fingerprints) and on how machine learning works when applied to such datasets. By expanding into the concept of smart world, the work in [41] also surveys how sub-6 GHz technologies can help improve a variety of services via data collection and system automation using active and passive sensing techniques. Finally, the work in [42] touches on aspects related to the modeling and estimation of wireless channels in Fifth generation (5G) cellular systems. While the work touches on localization, the covered techniques apply to outdoor cellular systems, and can thus leverage the density and much higher computational power of their hardware.

Unlike our survey, none of the above works targets millimeter wave device-based and device-free indoor localization. This area is characterized by several interesting research works to date, but remains a very hot topic due to the inception of mm-Wave coverage for future 5G-and-beyond networks as well as wireless (indoor) local-area networks. The objective of our survey is to cover the most significant work in this area, while giving a comprehensive view of unsolved challenges and open research avenues.

Note that, in our survey, we are *not* seeking an analysis of the limits of mm-Wave localization and sensing technology based on purely theoretical arguments, or an operational description of well-known geometric localization algorithms, or even a coverage of the integration between

mm-Wave communications and 5G, beyond-5G, and future 6G networks. These are related yet tangential topics for which we rather refer the interested reader to one of the several excellent surveys that touch on these aspects, e.g., [32], [33], [36], [42]–[47].

2.1.2 Outline and organization of the manuscript

Rest of the chapter expresses three purposes: to cover the characteristics of mm-Wave propagation and communication/sensing hardware that impacts localization and sensing performance, including standardization efforts (Sections 2.2 through 2.4); to detail the state of the art in device-based mm-Wave localization (Section 2.5) and in device-free mm-Wave sensing (Section 2.6); and finally to discuss our findings, discuss promising research avenues, and draw concluding remarks (Sections 2.7 and 2.8).

In particular, Sections 2.5 and 2.6 constitute the core of our technological survey. Section 2.5 discusses device-based localization algorithms for indoor environments, whereas Section 2.6 presents several approaches for radar-based device-free localization. Each section is organized to first present the section topic, and then to add progressively more details related to the typical techniques appropriate for each section, the hardware typically used in testbeds, and the description of each surveyed approach. We also include summary tables to help the reader navigate the contents and extract key information. Both Sections 2.5 and 2.6 end with a summary of the most relevant aspects and findings.

Fig. 2.1 represents the organization of the survey as a mind map, starting from Section 2.2 (top right), proceeding clockwise, and concluding with Section 2.7.

2.2 Influence of mmWave channels

2.2.1 Impact of mmWave frequencies on propagation conditions

The propagation of a wave through any medium depends on its frequency: this basic property helps us predict the behavior of the channel for diffeangularrent carrier frequencies. When it comes to mm-Wave, considering the Friis equation under the assumption that the antenna gain G at both link ends is frequency-independent (by reducing the antenna aperture), the free space path loss increases with the square of the carrier frequency f. On the contrary, assuming a constant physical area A at both the transmitter (TX) and the receiver (RX), the antenna gains $G = A(4\pi/\lambda)^2$ increase on both sides, and thus the overall path loss *decreases* quadratically with increasing frequency f [48]. Specular reflections for dielelectric halfspaces (e.g., ground reflections) depend on frequency as long as the dielectric constant is itself a function of frequency. For reflections at a dielectric layer (e.g., building walls) the specular reflections depend on the electrical thickness of the wall, which in turn is also a function of frequency. Interestingly, we have no evidence that the reflection coefficient varies with frequency, although the transmission power decreases uniformly with increasing frequency due to the skin effect in lossy media [49].

Figure 2.1: Mind map showing the organization of this survey.

Two effects that have gained spotlight with the increased interest in the mm-Wave band are diffraction and diffuse scattering. The former reduces noticeably at high frequencies, and larger objects lead to "sharp" shadows. The latter effect is more significant as the surface roughness becomes comparable to mm-Wave wavelengths. As the surface roughness increases, the objects behave like a Lambertian radiator, which scatters the radiation. Foliage has a similar effect as scattering; with the decreasing wavelength relative to the size of the leaves, we observe more diffused scattering and less penetration. Another factor is atmospheric attenuation due to fog or rain[50] and may affect the mm-Wave frequencies in case of extreme weather.

Channel models used for localization need to account for the above mentioned phenomena, and are often based on ray tracing or cluster-based modeling with some Geometry-Based Stochastic Channel Model (GSCM)[51]–[53]. Moreover, for ray tracing approaches, high-resolution en-

vironment information is needed to account for such surface roughness, as different materials have different properties (e.g., glass windows vs. concrete walls). These effects also depend on the environment: the high concrete walls and glass surfaces of the urban areas lead to different propagation conditions, compared to the greener suburban areas with, e.g., stucco exteriors and shorter walls.

2.2.2 Measurement techniques and results

To model the properties of a channel, we need to perform the measurements for different propagation scenarios. A channel sounder, that helps to measure these properties is not only an expensive piece of equipment but as we move towards higher frequencies, the susceptibility to *phase noise* as well as *antenna spacing* errors start to increase. Similarly, the cost and energy consumption of up/down-conversion chains, in particular of the front-end mixed signal circuitry in analog-to-digital and digital-to-analog converters (ADCs/DACs) as well as Power Amplifier (PA) becomes of paramount importance. For up-to-Gbit/s sampling rates (as often required by best-in-class channel sounding), 12-15 bit resolution is required. To penetrate larger distances (and thus to maximize the forward link gain), PA typically need to operate with 6-10 dB backoff power efficiency and need to be continuously driven close to their 1 dB compression point limits.

Consequently, the channel sounders used often for measurements at high frequencies use omnidirectional antennas [54] or if directional [55], then the angular resolution is not taken into account. Directionality is achieved by mechanically rotating horn antennas in most cases and the angular resolution corresponds to the beamwidth, e.g., [56]–[58]. For indoor measurement scenarios, the directional information though can be enhanced by using switched antenna arrays along with super-resolution algorithms like Space-Alternating Generalized Expectation Maximization (SAGE)[59] and RIMAX [60]. It is possible to use electronically-switched horn arrays [61] as well, which additionally lets us evaluate the MPC and intra-cluster information.

Key outdoor results

When it comes to outdoor measurements, path loss is a key parameter. For channel modelling, we need to measure the pathloss coefficient, its mean and its variance. The pathloss coefficient for mm-Wave frequencies is close to that of microwaves, i.e., often there is no strong frequency dependence beyond the f^2 dependence of free-space path loss [62]. In LOS scenarios, the path loss coefficient lies between 1.6-2.1 (2 for pure free-space propagation) and in NLOS scenarios the value increases to 2.5 and 5 (e.g., [56], [57], [63]).

On the other hand, the *variance of the path loss around the distance-dependent mean* is higher at mm-Wave frequencies, which in turn increases the probability of outage [64]. The standard deviation as well is strongly dependent on the distance and its values increases from 5-10 dB to more than 20 dB as the distance increases from 30 m to 200 m [63]. This is due to the variation in power levels caused by location and orientation of a street in an urban macro cell [65] and not due to shadowing as one may expect.

Another parameter important for channel modelling is the Root Mean Square (RMS) delay spread. But it changes with frequency and thus it may not be the best parameter to model the delay dispersion. Instead, delay windows may be a better alternative as they define the time interval containing part of the energy of Power-Delay Profile (PDP). Delay spreads in an outdoor environment are measured or simulated by ray tracing [56]–[58], [66]. Beamforming can help with minimizing the delay spread [67]. The type of beamforming to be used depends on the angular dispersion properties. Angular spreads measured at the Base Station (BS) are more accurate than those measured at the User Equipment (UE) as the ray tracers used often do not include scattering objects such as street signs, parked cars, etc. in their geographic database [63], [68]. As observed in [57], [69], the RMS angular spread at the BS is of the order of 10° with one cluster only while at the UE, the angular spreads are in the range 30-70° [56], [57], [69], [70].

More information related to fixed wireless scenarios can be found in [71].

Table 2.1: Summary of channel models and their spatial parameter values

Parameter		mmWave Channel Models				
		3GPP [72] / ITU-R [73]	COST IRACON [74]	METIS [66]	QuaDRiGa [75]	NYUSIM [76]
f (GHz)		6	2.6	0.45-63	5.4	28
Type		2D GSCM	GSCM	3D Map-based & GSCM	3D GSCM	TCSL
K- factor	μ_K	7	N/A	7.9	-1.6	N/A
	σ_K	4	N/A	6	2.7	N/A
Delay Spread	μ_{DS}	-7.7	1.07	-7.42	-7.22	2.7
	σ_{DS}	0.18	0.93	0.32	0.08	1.4
AoA Spread	μ_{ASA}	1.62	3.94	1.65	1.67	19.3
	σ_{ASA}	0.22	3.91	0.47	0.15	14.5
AoD Spread	μ_{ASD}	1.60	0.71	1.64	1.54	23.5
	σ_{ASD}	0.18	0.59	0.43	0.1	16.0
ZOA Spread	μ_{ZSA}	1.22	3.73	1.28	1.61	7.4
	σ_{ZSA}	0.297	2.11	0.26	0.07	3.8
ZOD Spread	μ_{ZSD}	N/A	1.95	1.31	1.17	-7.3
	σ_{ZSD}	N/A	1.80	0.31	0.07	3.8
XPR (dB)	μ_{XPR}	11	15.59	29	13	N/A
	σ_{XPR}	4	10.39	6.5	1.6	N/A
Shadow fading	μ_{PL}	47.9	N/A	N/A	36.1	N/A
	σ_{PL}	3	N/A	3	1.6	N/A

Key indoor results

Measurements for indoor environments have picked up in recent years as we look at localization applications for 5G. The results are often from office and industrial environments, where different material densities can be studied. The path loss coefficient in this case ranges from 1.2-2 in LOS to 2-3 in NLOS scenarios [77], [78]. The frequency dependence of the path loss is more significant for indoor than outdoors, f^k with $k \approx 2.5$ was observed in [79]. Overall though, the values are closer to those at sub-6 GHz, with an increased probability of outage. Path loss in some cases is shown to follow a dual-slope model and is the same for both mm-Wave and sub-6 GHz. The floating intercept model is another alternative used in Third Generation Partnership Project (3GPP)

13

standards for indoor modelling at high frequencies. Human blockage can cause upto 10-20dB attenuation regardless of one or two people[80] and similar values in case of trucks in outdoor scenarios[81]. In [82], Fast Fourier Transform (FFT) based beamforming is used in conjunction with a very large virtual array (25×25×25 elements). It highlights the scattering caused by small objects specifically in NLOS case and the importance of small scale characterization. Further, it is shown that the indoor environment leads to enhanced diffused MPC energy.

Delay spread measured in office scenarios is usually less than 5 ns in LOS conditions, and 10–20 ns in NLOS conditions [78], [83]–[86]. Though these measurements were limited to under 100 GHz, recently [87] performed measurements at 142 GHz and observed delay spread values of 3 ns in LOS and 9 ns for NLOS. Further, the observed channels are much sparser at frequencies over 100 GHz and we notice higher partition loss compared to 28 GHz. It is worth noting for indoor measurements, the number of MPC is higher with more clusters than measured for outdoor with rotating horn antennas [88]. Here the angular spreads are often measured for clusters, with the intra-cluster azimuth and elevation angles are described as having a Laplacian distribution with a spread of $5°$ [89].

2.2.3 Models for mmWave channels

Because mm-Wave propagation channels differ from microwave channels, we need to redefine or rather add certain parameters for mm-Wave channel modeling. As mentioned in [71], mm-Wave channels require 3D modeling of azimuth as well as elevation spreads, inclusion of temporal/spatial/frequency consistency and multipath cluster based modeling. These have further impact when we consider positioning and localization. Prevalent models for mm-Wave are GSCM that imitate the propagation environment with stochastic processes, and create a 3D map. To correctly reproduce the wireless environment, parameter values need to be extracted from the channel impulse response of real time measurements done using a channel sounder. An extensive review of propagation chacteristics at mm-Wave frequencies is available in [90], which also provides a summary of channel sounder measurements and relevant channel models. The 3GPP defined different environments for mm-Wave channel modeling, these include Urban Macro, Urban Micro, Indoor Office and Rural Macro. Several outdoor and indoor measurements are available, but for this chapter we compare large-scale parameter values for an indoor office scenario listed in Table 2.1.

Prominent channel models have been developed for the above mentioned scenarios based on measurements done in each of them. Some key results have already been discussed, but we also observe that cluster-based multipath channel components have been modelled, in order to specifically account for an indoor office environment. Also, as can be seen from the table, the angular spread is no longer limited to the azimuth plane.

Static vs. dynamic modeling

Due to the high frequency and thus higher path loss, there is significant deterioration when the UE is stationary and more so when the UE is moving or is in a high movement zone and transitions

14
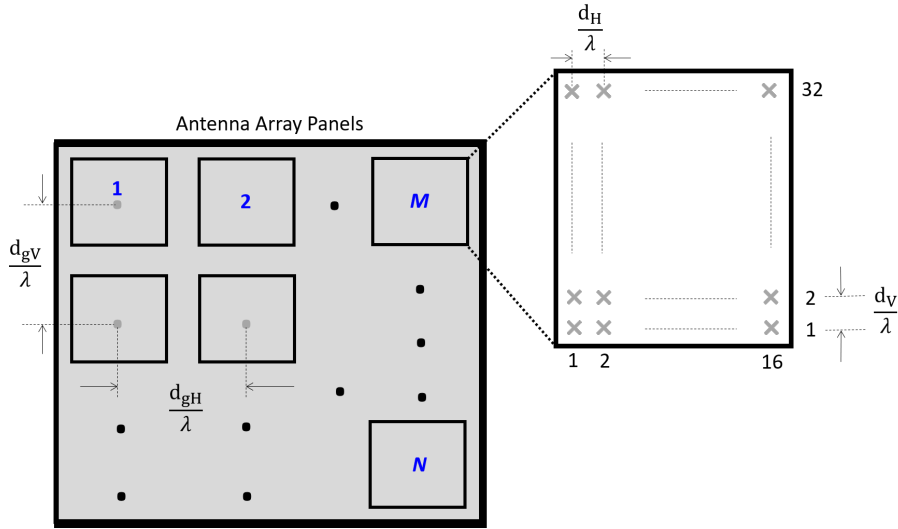
Figure 2.2: Cross-polarized antenna array panel [72].

from a LOS to NLOS scenario. This requires the dynamic modeling of the communication channel, as the moving objects in the vicinity also act as random blocking obstacles. The BS needs to transmit training beams more frequently so as to update the Angle of Departure (AoD)/Angle of Arrival (AoA) estimates, since the location of UE changes over time, and slight errors in the orientation of the beams can lead to significant performance loss [91]. So far, we have considered a fixed BS and slow moving UE, but with 5G and Vehicle-to-Everything (V2X) communications we expect high mobility scenarios[10]. Most mm-Wave channel models are still defined only for a fixed BS, but have added support for dynamic modeling scenarios for V2X.

Blockage

mm-Wave cannot penetrate obstacles such as human bodies, walls, foliage, etc. Thus, these blockage sources need to be modelled in the link budget itself. One such characterization study is found in [92], which measured power loss (in dB) when 70-GHz mm-Wave signals propagate through a brick wall, a PC monitor, and book shelves. Blockage does not affect just the total received power but also the angle or power of multipath signal components, due to varying sizes, positions and directions of the blocking object/human. Localizing the position of the UE with respect to these blockage sources becomes onerous, especially in a dynamic setting.

Spatial consistency and clusters

A new, previously unexplored requirement was added to 3GPP Release 14 [72]. When mm-Wave communications take place through narrow antenna radiation beams, the channel characteristics become highly correlated, especially when two UE are close and see the same BS. Also, for applications related to V2X communications, it is paramount that the channel evolves smoothly without

discontinuities during mobility [93].

Polarization

The radiation pattern of each antenna element of an array extends over both the azimuthal plane and the elevation plane, and should be separately modelled for directional performance gains. Moreover, as we consider indoor scenarios with higher number of reflections, the polarization properties of the multipath components also come into play.

Large bandwidth and large antenna arrays

Antenna arrays that are larger in size and also massive in the number of antenna elements are needed at mm-Wave, thus high resolution channel modeling includes propagation patterns both in the angular domain and in the delay domain. Massive Multiple-Input Multiple-Output (MIMO) channel models[94] have previously not considered these exceptions but at mm-Wave, accurately modeling of the higher number of multipath components and their AoA/AoD is paramount. Antenna elements in azimuth and elevation plane both need to be evaluated to consider all possible array structures (planar array, rectangular array, cylindrical array). Fig. 2.2 depicts an antenna array panel used for 3GPP/International Telecommunication Union – Radio-communication Sector (ITU-R) antenna modeling [72], [73]. Figs. 2.3 and 2.4 show the BS and UE array radiation pattern based on parameters as defined in [72, Table 7.3-1, page 22].
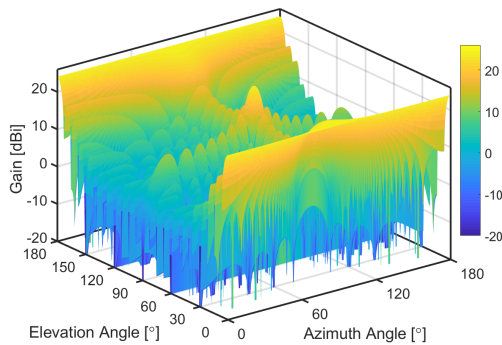
| **Figure 2.3:** BS antenna array pattern as a function of azimuth and elevation scan angles [93]. | **Figure 2.4:** UE antenna array pattern as a function of azimuth and elevation scan angles [93]. |
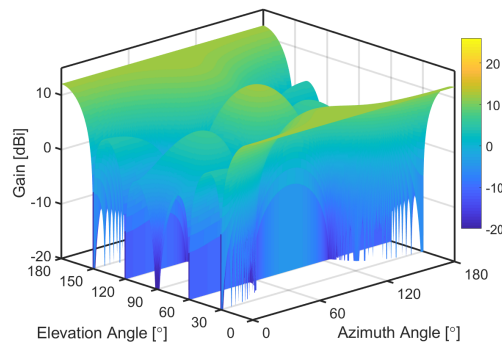
2.2.4 Summary

The mm-Wave channel when considered for indoor applications differs from the microwave channel in key aspects such as free space path loss, diffraction, and penetration loss with respect to different surfaces. This required the need for different measurements to be done for channel characterization. Some key results are presented in Section 2.2.2. Path loss equations and penetration loss for indoor scenarios can be found in [72, Tables 7.4.1-1 and 7.4.3-1]. Various channel models
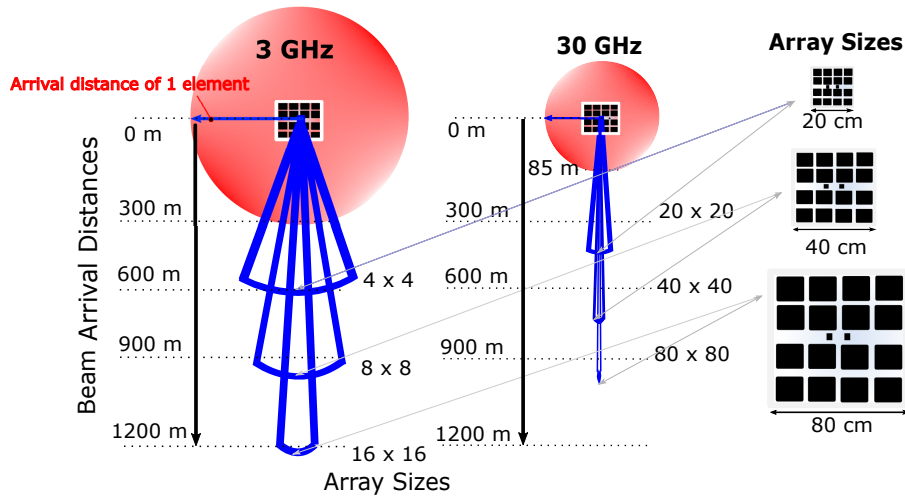
Figure 2.5: Effect of beamwidth relative to operating frequency and array sizes [48].

have been developed, these include those by 3GPP [72], ITU-R [73], METIS [66], MiWEBA [58], Fraunhofer HHI's QuaDRiGa [75], COST2100 [95], NYUSIM [96] which still has ongoing measurements for indoor scenarios. The channel models are all GSCM-based with added cluster based modeling. Small-scale parameter values are further available when considering indoor scenarios found in the documentations mentioned for corresponding models.

Several measurements have been done in the mm-Wave band for outdoor (urban macro and urban micro) scenarios but the indoor measurements are limited to the sub-6 Ghz band for the channel models developed with the exception of [97], where the authors propose an extension for an indoor channel model based on extensive measurements carried out at 28 and 140 GHz. We observe that indoor channel models are an extension of outdoor ones, and can be adapted easily based on the delay and angular spreads of any environment, as well as by adapting path loss modeling.

2.3 Implications of beamforming architectures for mmWave localization

It is a common misconception that for higher frequencies the free space propagation loss is higher. As explained in [98], [99], for given aperture area of the antennas used, shorter wavelengths propagate farther due to the narrow directive beams. This is further verified in [100] with a patch antenna operated at 3 GHz and an antenna array operated at 30 GHz of the same physical size. We observe equal amounts of propagation loss irrespective of the operating frequency. Thus, mm-Wave frequencies enable the use of antenna arrays that produce highly directional beams which lead to large array gains. This can be observed from Fig. 2.5, which shows not only the increase in array size with respect to the beam penetration distance, but also how the larger array size increases the coverage area [48].

2.3.1 Analog beamforming

Analog beamforming, sometimes also referred to as beam steering, is done by connecting a single Radio Frequency (RF) chain to a string of phase shifters that are both energy- and cost- efficient. Each phase shifter multiplies its input by $e^{j\frac{2\pi k}{2^N}}$, where $j = \sqrt{-1}$, N is the number of bits, and $k = 0, \ldots, 2^N - 1$ is used to control the phase shifters. Most commonly, codebook-based schemes are used to steer the beams in the direction of the UE/receiver. At the receiver, the Received Signal Strength Indicator (RSSI) is the most commonly used parameter to estimate the direction of arrival and delay, and thus localize the device. However, phase shifters have a constant amplitude constraint and limited phase resolution. It is also worth noting that analog beamforming converges to a single beam for multiple data transmissions, and in multi-user case the inter-user interference is very high. This is a drawback for localization applications, as the phase resolution for analog beamforming is low. The popularity of analog beamforming systems comes from the availibility of Commercial-Off-The-Shelf (COTS) devices, that are being used for research on mm-Wave positioning. The devices come with a pre-programmed codebook to generate beam patterns and with support for retrieving the RSSI and Channel State Information (CSI) which can be used to isolate the position of the UE. One such hardware front-end is available from TMYTEK, an analog correlator with beamformer chips and smart-antenna arrays [101]. Another company that provides beamformer integrated circuits and scalable antennas for mm-Wave is Anokiwave [102]. Siver Semiconductors provides transceiver modules for mm-Wave frequencies, i.e., 28 GHz and 60 GHz [103]. National Instruments (NI) also has the PXIe-5831, a mm-Wave vector signal transceiver that has beamforming capabilities and phased antenna arrays [104]. It has been used for channel measurements as mentioned above as well [105]. We discuss the hardware devices used in more detail in Section 2.5.3.

2.3.2 Hybrid beamforming

Hybrid beamforming is by far the most researched form of beamforming, as it provides a middle ground between complexity and cost. Here, the analog beamformer is used in the RF domain, along with a digital precoder at baseband. This can be either a fully connected structure or a partially connected one. Hybrid analog/digital beamforming structures provide balance between the beam resolution and cost and power consumption. By using multiple RF chains concurrently, beam sweeping can be done in a short time leading to shorter beam training time which leads to higher effective data rate. At mm-Wave frequencies the sparse channel behaviour is useful for beam training and higher array gains. Multiple hybrid beamforming techniques for mm-Wave have been proposed in the last ten years which broadly fall under codebook dependent, spatially sparse precoding, antenna selection and beam selection [106]. [107] first gave the idea of what we call hybrid beamforming today. It was a combination of a digital baseband precoder and an RF precoder which falls under spatially sparse precoding. The work in [108] first proposed the idea of baseband beamforming, or "hybrid beamforming" as the authors named it, that chooses the best RF beam based on a capacity maximization criterion, and then derives a Zero-Forcing (ZF)-based weighing matrix for digital precoding. Also, both [109] and [110] suggest codebook-based pre-

coding solutions. Recent works have proposed compressive sensing, least squares- and Discrete Fourier Transform (DFT)-based solutions for hybrid beamforming with use cases in car-to-car scenarios and high speed trains. In most cases, hybrid beamforming is seen to perform as well as fully-digital beamforming, and as being both cost-effective and spectrally efficient.

2.3.3 Digital beamforming

Digital beamforming adjusts the amplitude and phase of the transmitted signals using precoding. Linear precoding algorithms such as Matched Filter (MF), ZF, and Regularized Zero-Forcing (RZF) methods were classically used for single-antenna user systems. For multiple-antenna users, block diagonalization is a feasible approach. Digital beamforming can be considered as the best option for mm-Wave positioning. With the possibility of huge antenna arrays (256×128 upwards) a beam resolution of the order of centimeters can be achieved. The calibration accuracy of digital systems allows us to use high-resolution parameter estimation algorithms that can estimate not only the Time of Arrival (ToA) and AoA but also the Doppler frequency offset in case of mobility, making it possible to update the position of a UE in real-time. The issue here arises from the use of a RF chain per antenna, which leads to a complex, non-cost-effective hardware system for massive MIMO structures.

As digital beamforming offers higher beam resolution, it is a viable candidate where multi user mm-Wave or rather mm-Wave massive MIMO systems are considered. However, commercial hardware for a fully digital system is still in its infancy, and only laboratory results exist. Several authors have proposed alternative techniques for the realization of a digital system that is power efficient. For instance [111] gives an option for digital beamforming that employs switches to bypass the hardware constraint of using multiple RF chains. In [112]–[114], the authors propose different ways to form an antenna array using waveguides and printed circuit boards that support digital beamforming. Alternatively, [111], [115] propose novel frameworks to do digital beamforming for a mm-Wave setup using linearization to help with power amplifier loss and improved quantization.

2.3.4 Performance vs. complexity overview

In localization applications, the requirement for mm-Wave indoor systems is to isolate the position of the receiver inside a room, while taking into account blockage caused by humans and objects alike, with LOS being the dominant component. The presence of pillars, metal and glass surfaces affects the channel impulse response and thus make it difficult to extract position information. Presence of antenna arrays greatly enhances the accuracy of the position coordinates. Whereas digital systems have cleaner isolated beams and can potentially yield centimeter-level pointing accuracy, analog setups have a limit to the number of beam patterns they can generate: when trying to increase the resolution, these beam patterns eventually start to overlap. As stated above, the number of *beams* is proportional to the number of available RF chains, thus increasing the complexity hundred-fold for digital systems. Calibration issues also prevent analog systems from performing high-resolution parameter estimation which could improve the localization accuracy.

19

Hybrid beamforming seems a promising tradeoff as of now, due to the easier availability of COTS devices, and to a performance almost as good as that of fully digital systems.

2.4 Progress in standardization of cellular mmWave systems

The frequency bands used for 5G systems were proposed at the 2015 World Radio Conference (WRC) by ITU-R and approved during WRC 2019. The frequency bands standardized by 3GPP in Release 15-17 [116]–[118] for 5G systems are classified as FR-I region (below 7.125 GHz) and FR-II region (between 7.25 GHz and 71 GHz). The approved FR-II bands are (in GHz): 24.25–27.5; 31.8–43.5; 45.5–50.2; 50.4–52.6; 66–71. FR-I bands act as the key bands for cellular communications, while the FR-II are more suited to short-range communications. The FR-II bands also provide increased bandwidths compared to FR-I, and are managed via licensed access mechanisms such as Enhanced UTRA-Dual Connectivity (EN-DC). As some bands overlap with other services, coexistence management is needed for terrestrial access in overlapping satellite communication channels and for fronthaul and backhaul in fixed wireless systems. 5G commercial deployments have already been taking place since the end of last year, and some spectrum congestion was observed initially amongst multiple operators. Since then, some novel forms of spectrum access/coordination mechanisms have been implemented.[1] When it comes to localization of UE, it was the focus of 3GPP Release 16 [117] especially for the use case of URLLC. In the past, Global Navigation Satellite Systems assisted by cellular networks have been mostly used for UE positioning, but their accuracy is high only in outdoor environments, as they rely on satellites to localize UE. As we move towards higher frequencies, we require localization indoors as well, and we can accomplish it in 5G networks using the location server, as it was for long-term evolution-advanced (LTE-A) systems. The location server collects and provides position estimates and assistance data and measurements to the other devices. Various localization methods are used, based on downlink or uplink communications, either separately or in combination, to meet the accuracy requirements for different scenarios. The overall architecture is as depicted in Fig. 2.6.

As shown in Fig. 2.7a, downlink-based localization is performed when each of multiple BS/AP send a different reference signal, known as the Positioning Reference Signal (PRS). The UE receives the different PRS and reports the ToA difference for PRS received from multiple distinct BS/AP to the location server. The location server can use the reports to determine the position of the UE. Compared to LTE-Advanced, the PRS has a more regular structure and a much larger bandwidth, which enables a more precise correlation and ToA estimation.

The canonical 3GPP Release 15 Sounding Reference Signal (SRS) with Release 16 extensions added uplink-based localization or BS/AP centric localization as shown in Fig. 2.7b. In this case, the UE sends the reference signal. Based on the received SRS, the BS/AP can measure and report (to the location server) the arrival time, the received power and the AoA from which the position of the UE is estimated. The time difference between downlink reception and uplink transmission can also be reported, and used in Round-Trip Time (RTT)-based positioning schemes, where the

---

[1]We note that these are operator- and vendor-specific, since frequency band combinations vary depending on the specific country.
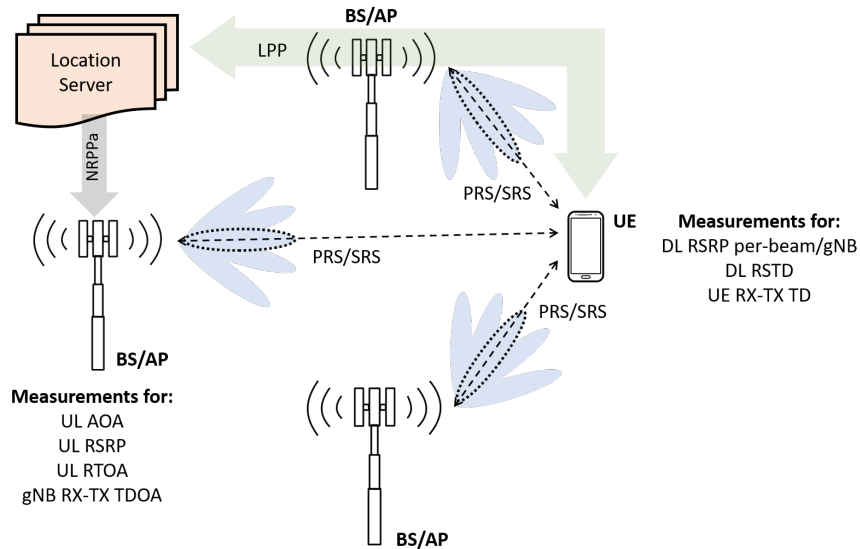
Figure 2.6: 3GPP Release 16 radio access type-dependent architecture standardized for UE localization in URLLC scenarios. All BS/AP are interfaced with a centralized unit enroute to a URLLC core network.

distance between a BS/AP and a UE can be determined based on the estimated RTT. By combining several such RTT measurements, involving different BS/AP anchors, it becomes possible to estimate the location of the UE.

We note that these methods do not utilize the full-dimensional nature of the propagation channel (azimuth and elevation domains), and do not fully take into account the phase information needed to estimate the underlying MPC with high resolution. While this is an ongoing topic for research in many study items of 3GPP Releases 17 and 18, we refer the reader to [118], [119] for further details. Along this same line, a steady stream of work is also conducted in academia, see e.g., [120].

2.5 Device-based mmWave localization algorithms for indoor communication systems

2.5.1 Introduction

In this section, we introduce algorithms and methods that leverage lab-grade and commercial-grade mm-Wave hardware to localize devices indoors. We start with a brief recap on classical methods for indoor radio localization. The standard techniques designed for localization involve exploiting the parameters of radio signals from existing wireless infrastructure. These have been well explored and surveyed in, e.g., [37], [43], [121]–[125]. With reference to Fig. 2.8, localization algorithms typically make use of signal parameters related to received signal power (RSSI and SNR, SNR), time-information such as Time of Flight (ToF) and Time

(**a**) Downlink based, UE-centric   (**b**) Uplink based, BS/AP-centric
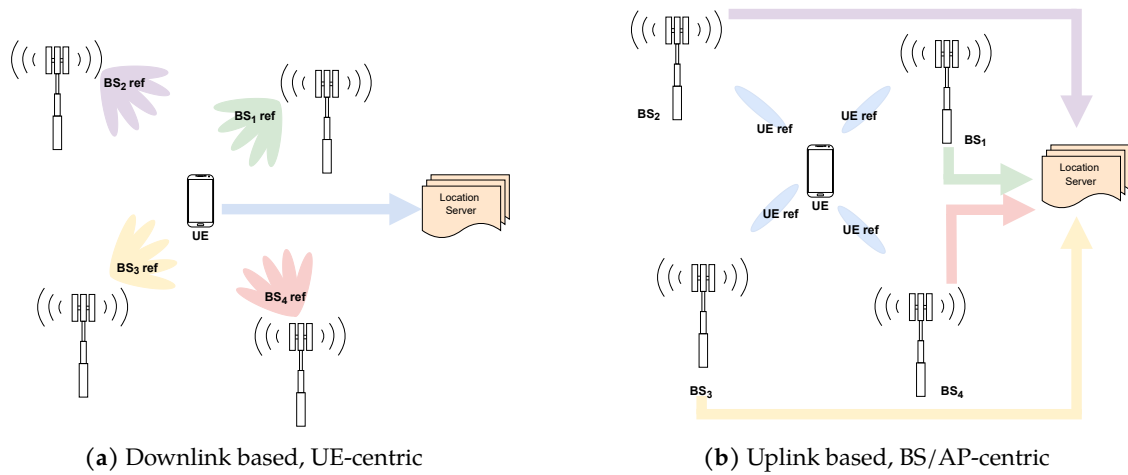
Figure 2.7: Architecture of BS/AP-centric vs UE-centric localization.

Difference-of-Arrival (TDoA), and angle information (AoA and AoD) in order to obtain distance and direction estimates, which enable a device or group of devices to estimate either their own location, or the location of another device in their proximity, or both.
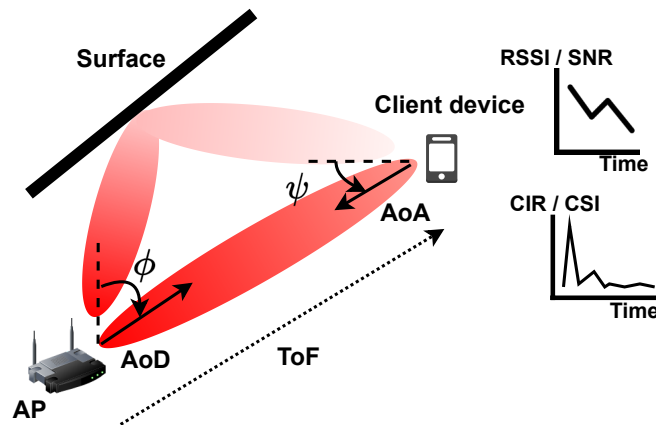
Figure 2.8: Illustration of the signal measurements obtained from mmWave propagation. The color gradient of the beam represents the decreasing signal strength due to path loss.

Fig. 2.9 offers a general view of this process considering the papers on mm-Wave localization surveyed in the literature. After a device has extracted location-dependent features from a received signal, such features are either used directly for localization, or further processed to extract additional information, or joined into a global map of the environment along with other measurements. The device then applies geometric or Machine Learning (ML)/Deep Learning (DL) algorithms to derive location information.

The most typical localization techniques rely on geometric algorithms. For example, *trilateration* and *triangulation* utilize distance and angle measurements from fixed reference points to com-
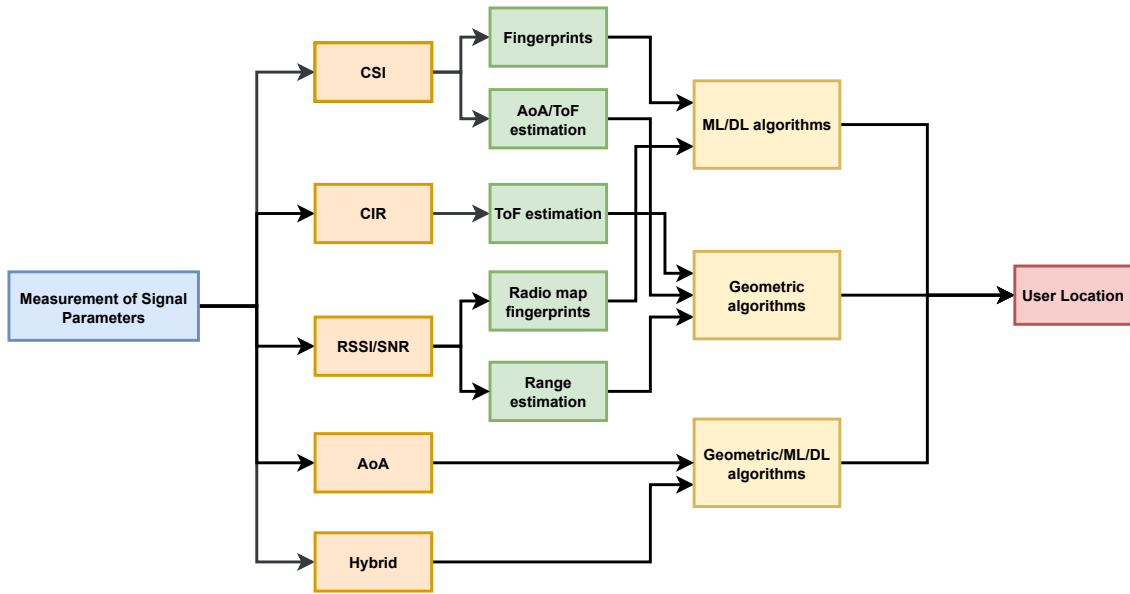
Figure 2.9: General flow chart of the steps of a mmWave localization algorithm from the surveyed literature.

pute an intersection, which yields the estimate of a device's location [37]. The reference points are usually the location of the access points, and the localized device is typically a client. The distances between the AP and the client are measured by exploiting either the ToF of the signal or by mapping the RSSI information to absolute distance using path-loss models. Fig. 2.10a shows an illustration of trilateration using ToF to estimate distances.

AoA (the angle at which the received signal strikes the receiver antenna or antenna array) and Angle Difference-of-Arrival (ADoA) (the difference between two AoA), are estimated by applying signal parameter estimation algorithms (like MUlti SIgnal Classification (MUSIC) [126] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [127]) on the received signal. The AoA from different AP are then triangulated to localize the client device. Fig. 2.10b illustrates the triangulation-based technique, whereas Fig. 2.10c depicts ADoA-based localization.

Wireless channel characteristics, e.g., in the form of the Channel Impulse Response (CIR) between a transmitter and a receiver, also provide valuable information for localization purposes, including the ToF of the received signal. The CSI can also be extracted from the receiver antennas to obtain rich information about multipath signal components [128]. As a result, one can separate the LOS propagation path from NLOS paths, or detect that only NLOS components reached the receiver, thus improving the accuracy of the signal parameter measurements.

The advent of bandwidth-hungry applications such as augmented reality, virtual reality, etc., and the ever-increasing demand for high data rates, has made mm-Wave communication technology a popular potential replacement for existing WLAN systems. This is mainly due to the availability of large bandwidth in the frequency range of 30-300 GHz, resulting in multi-Gbit/s

(a) Trilateration   (b) Triangulation   (c) Angle-difference of arrival
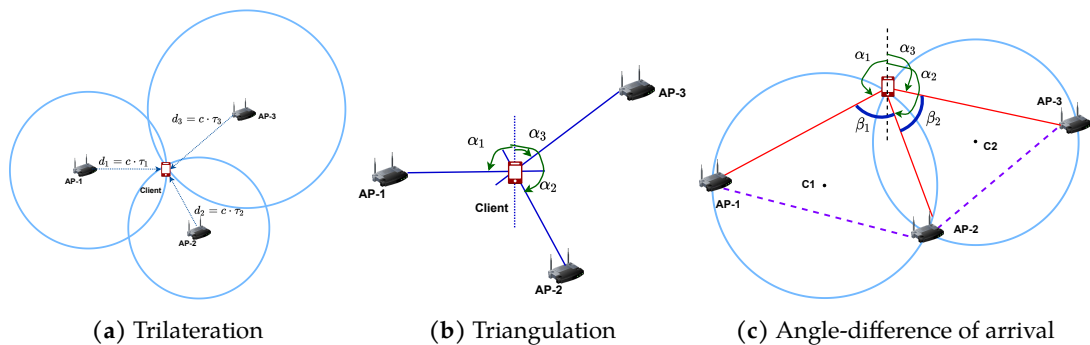
Figure 2.10: Illustration of the (a) trilateration, (b) triangulation, and (c) angle-difference of arrival processes using ToF, AoA, and ADoA localization geometries, respectively. Note that d_i and τ_i respectively denote the distance and propagation delay between AP i and the client, c is the speed of light in air, α_i denotes the AoA of the signal from AP i, and β_i is the ADoA, i.e., the difference of the AoAs from APs i and $i+1$.

data rates. mm-Wave propagate quasi-optically, thus reflecting crisply off indoor surfaces and obstacles with limited scattering just like light rays [20]. This makes finer measurements of signal parameters such as RSSI, AoA, SNR, and ToF, more feasible and more accurate. Moreover, we remark that wireless devices typically collect location-dependent signal features through the interaction between a client and one or more AP. Such interactions naturally take place in mm-Wave networks, e.g., during standard-compliant link establishment and beam refinement procedures (see also Section 2.5.3). Therefore, in principle the measurement of signal features does not require the devices to implement localization-specific message exchange protocols. This makes localization an almost-inherent feature of mm-Wave communication systems [44], [129].

As remarked in Section 2.3, however, mm-Wave devices have peculiar characteristics that differentiate them from commonplace WiFi equipment. Specifically, mm-Wave arrays can incorporate a large number of antennas. The presence of large arrays enable mm-Wave devices to output low-level physical layer measurements from each antenna separately. Once the device has locked onto a signal, each antenna receives the same signal with a different phase, corresponding to the delay incurred by the signal due to its spatial position in the array. These measurements can be made available as CSI and localization algorithms can exploit them to localize a device, either by converting them into AoA estimates (e.g., [130], [131]) or by directly inferring the location of a device by exploiting the CSI as a location-dependent feature.

Whenever CSI measurements are not available, a device can still retrieve angle information by post-processing the output of standard-compliant beam training procedures. Typically, each mm-Wave has a number of pre-programmed beam patterns that provide it with the necessary flexibility to focus energy towards different directions. Each beam pattern ideally covers a well-defined portion of the 3D space, so that observing each beam pattern separately makes it possible to implement a scan of all azimuthal and elevation angles that the mm-Wave array can cover. Therefore, measuring the power received through each beam pattern configuration would implement a sweep of lookout angles. By identifying the beam pattern that leads to the largest received power,

a mm-Wave device could easily estimate angles of arrival. We now proceed to discuss each type of location-dependent feature separately in the context of mm-Wave communications.

2.5.2 Pros and cons of location-dependent measurements for mmWave localization

Angles of arrival and departure, angle difference-of-arrival [132]–[138] — The term angle of arrival (AoA) refers to the angle at which radio signals illuminate the antenna array at the receiver. The transmitter-based counterpart, the AoD, refers to the angle at which the radio signals emanate from the antenna array at the transmitter front-end in order to reach the receiver. In most cases, more than one antenna elements are required to compute angle information. Other methods to extract AoA information from the receiver array involve the use of CSI, beamforming methods, or subspace approaches such as the well-known MUSIC [126] and ESPRIT [127] algorithms. We cover angle-based approaches in Section 2.5.4.
Pros: Relatively accessible information in mm-Wave systems, thanks to the large number of antennas in transmitter and receiver arrays.
Cons: If not associated to some range information, can only yield location estimates in a relative coordinate system. Multipath propagation can distort angle estimates, if not properly modeled or compensated for.

CSI [139]–[141] — CSI refers to the measurable properties of a received mm-Wave signal that relate to the propagation channel linking two devices, e.g., the AP and the client. Different mm-Wave hardware may provide different forms of CSI. For example, patching TP-Link's Talon routers [142] with special firmware makes it possible to extract receiver-side CSI in the form of one complex gain coefficient per receiving antenna, expressing the attenuation and phase shift that affect the strongest propagation path at each antenna. Post-processing CSI yields different signal parameters, including path attenuation and angle information. If CSI values are sufficiently precise (e.g., no coarse quantization affects the amplitude or phase), collecting receiver-side CSI from multiple antennas also enables the estimation of AoA. We cover CSI-based approaches in Section 2.5.5.
Pros: Rich information that can be readily used for ranging or as an input to learning-based approaches.
Cons: Typically not available straightforwardly on all devices. Different devices may provide different types of CSI.
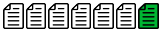
RSSI [143]–[150] — RSSI is one of the simplest proxies for the range of a device in an environment. It is measured at a receiving device as the power or amplitude of the received RF signal. mm-Wave Received Signal Strength (RSS) measurements can be extracted from the physical or Medium Access Control (MAC) layer of a device and used to measure the distance of a client from the AP, based on the knowledge of a path loss model. The client is believed to lie on the circumference of the circle centered on the AP and having the estimated range as the radius.

Such estimates from more than two AP can be trilaterated to approximate the location of the client. We cover RSSI-based approaches in Section 2.5.6.

Pros: Simple ranging method, typically available on communication devices.

Cons: Error-prone, typically requires an extensive tuning of the path loss model. RSS measurements are often affected by the losses in the front-end receiver architecture of the client and by the number of quantization bits in its ADC circuitry.

Table 2.2: Visual representation of the distribution of research efforts for device-based mmWave localization. Green icons represent recent papers that employ some form of machine learning.

	INDOOR MMWAVE LOCALIZATION				
	Traditional methods		**Tailored methods**		
	RSSI and SNR (e.g. [144], [147])	**Time information** (e.g. [151])	**Angle information** (e.g. [132], [134])	**CSI-based** (e.g. [139], [140])	**Hybrid approaches** (e.g. [152], [153])
Client-centric	▤▤▦▦		▤▤▤▤▤▦		▤▤▤▤▤▦
AP-centric	▦▦▦		▦	▤▤▤	▤▤▤▤▤▤
AP-client cooperation	▦	▤			▤▤▤▤

Time information [151] — Time information is another common proxy for the distance between two devices. Typical measurements used for this purpose involve ToF and TDoA measurements. ToF (also known as ToA) measurements exploit the time taken for a signal to propagate from the AP to the client in order to estimate the distance between them. The client intuitively lies on the circumference of the circle with the AP as the center and the distance estimate as the radius. Multilateration methods can be used to estimate the location of the client. It is important to note that ToF measurements require a tight synchronization between the AP and the client. mm-Wave signals offer better ToF estimation accuracy (thus better ranging resolution), owing to the large bandwidth available, especially in the unlicensed bands. We cover time-based approaches in Section 2.5.6.

Pros: ToF information is usually accurate when directly extracted from a device's physical layer, which helps accurate localization. Such protocols as the Fine Time measurement (FTM) protocol, when available on a device, can provide very accurate timing estimates.

Cons: Requires sub-nanosecond sampling times in a device's ADC in order to yield a sufficiently fine range resolution.

Hybrid approaches [152], [154]–[167] — Several solutions propose to fuse information from multiple sources in order to improve localization accuracy. For example, several works merge AoA and RSSI, or AoA and ToF estimates. We cover hybrid approaches in Section 2.5.7.

Pros: Hybrid schemes usually achieve better accuracy. In some purely angle-based algorithms, side information such as RSSI and ToA can help resolve geometric translation, rotation, and scaling ambiguities.

Cons: The algorithms become more complex, and rely on the estimation of multiple quantities.

In ill cases, errors compound and may make the location system more inaccurate than non-hybrid ones.

According to our survey of the literature on mm-Wave localization algorithms and to the above discussion, we identify two broad categories in the available literature:

1. Algorithms *tailored* to mm-Wave communication protocols and schemes, that exploit protocol operations to extract geometric scenario information and infer the location of the devices;

2. *General* algorithms that apply well-known range-based or range-free localization approaches to mm-Wave communications.

The algorithms in the first category are mainly angle-based or CSI-based: they infer the angle of arrival structure by leveraging, e.g., sector measurements in communication protocols. Then, they use angle information to localize a device. By way of contrast, the algorithms in the second category are not necessarily mm-Wave-specific. These works can be further subdivided by considering where the algorithm mainly runs:

1. In *client-centric* algorithms, the intelligence mainly resides on the client, which may collect location-dependent measurements by receiving signals from one or multiple AP, and by estimating its own location locally. This approach is useful for systems that need to scale to up a large number of devices, as each device runs the algorithm independently. Literature surveyed: [132]–[134], [136]–[138], [143], [145]–[147], [152], [153], [157]–[159]

2. In *AP-centric* algorithm, the intelligence resides in a computing entity connected to one or multiple APs, which coalesce their measurements from multiple clients in order to estimate the location of each client. These schemes are ideal for seamless network management purposes (e.g., to optimize client-AP associations) but scale less than client-centric approaches when the number of clients increases. Literature surveyed: [135], [139]–[141], [148]–[150], [154], [156], [161], [164], [166], [167]

3. Schemes based on *AP-client cooperation* are based on a shared intelligence, where both one or more AP and the client run portions of the localization algorithm, and possibly exchange information to finally estimate the client location. Literature surveyed: [144], [151], [155], [162], [163], [165]

In our scan of the literature, we observed a comparatively small number of works that employ a form of machine learning to compute location estimates. We believe this is due partly to localization being a somewhat understood problem (whereby the community prefers the use of understandable and optimizable signal processing algorithms rather than training black-box machine learning models) and partly to the sometimes daunting collection of training data. Yet, these prove a feasible solution in some cases, e.g., when a huge database of different location-dependent features is available, and the complexity of the considered indoor environment prevents straightforward modeling.

Table 2.2 summarizes the above preliminary subdivision pictorially, and conveys in what category most of the research efforts has concentrated so far. We observe that a few approaches have considered baseline RSSI, SNR and time measurements to localize mm-Wave devices. However, most of the research moved to exploit the fine angle resolution that large mm-Wave antenna arrays enable. A significant number of works also consider hybrid approaches, which mix good angle resolution with the extra information yielded by time- or RSSI-based measurements, and thus achieve greater accuracy. Finally, we observe that a few recent works (from 2017 to the time of writing) rely on ML techniques, typically to process RSSI and SNR measurements and predict the location of a device. We highlight these works in green in Table 2.2, in order to emphasize the emergence of this paradigm, previously unobserved in indoor mm-Wave localization.

In *client-centric* algorithms, the client collects signal measurements thanks to the interaction with different AP. The client then trains an ML model and employs it to estimate its own location. For example, in [147], the client collects SNR information to train ML regression models. In [168], instead, the client resorts to AoA information to train shallow neural networks and estimate its coordinates of the client. Other works in this survey that employ client-centric machine learning algorithms are [145] and [152].

AP-centric algorithms rely on AP collecting location-dependent signal features that relate to the location of each client in a given environment. These radio fingerprints are then used to train models to localize the client. For example, in [148], [149], the AP use the spatial beam SNR measurements collected during the beam training process in order to create a radio map of the environment. DL models are then trained to estimate the location and orientation of the client devices. Other works in this survey that employ AP-centric machine learning algorithms are [150] and [135].

Other algorithms rely on some form of *AP-client cooperation* to collect location-dependent signal features and train machine learning models. In these schemes, the features can be collected either by the AP and the client separately and then exchanged, or through possibly multi-step procedures requiring AP-client cooperation. The only work in the literature that uses this technique for ML models is [144]. Here, RSSI and beam indices obtained both at the client and at the AP after the beam alignment process are used to generate radio fingerprints at different client locations.

Notably, Table 2.2 clearly shows that ML-based algorithms are mostly AP-centric or hinge on a cooperation between AP and clients. The main reason is most that AP are infrastructured devices, and have easier access to compute power in local servers through fast cabled connections.

2.5.3 Evaluation tools for mmWave localization

We now look into the tools that have been used so far to evaluate mm-Wave localization algorithms. From the surveyed literature, we observe both *experimentation-based* and *simulation-based* performance evaluation, depending on whether a proposed scheme is evaluated using mm-Wave hardware- or software-based setups.

Experimentation-based performance evaluation

Localization experiments so far have been carried out using either laboratory-grade or commercial-grade equipment. Laboratory-grade equipment typically includes Software Defined Radio (SDR) for signal generation and a mm-Wave up-converter, with a directional antenna to drive signal emission. For example, the above setup is used in [138], where the authors employ horn antennas to emulate narrow beam patterns. A similar setup is part of the work in [169] and [151], where the authors employ the Zynq 7045-based SDR and the Universal Software Radio Peripheral (USRP) X310-based SDR, respectively, in addition to a 60-GHz analog front-end to emit the mm-Wave signals. The authors of [170], [171] have used an NI SDR with a 60-GHz transceiver that enables the user to fully program of the Physical Layer (PHY), MAC, and network layers, especially for Wireless Local Area Network (WLAN) applications. It also incorporates a 24-element Sibeam reconfigurable antenna array. A Field-Programmable Gate Array (FPGA)-based setup is discussed in [172], where the authors have used the XCKU040 Kintex UltraScale FPGA for the baseband processing of a 60-GHz reconfigurable phased antenna array. PEM-003 60 GHz transceivers were used as the RF front-end for the experimentation. Recently, the New York Uuniversity spin-off Pi-Radio [173] developed dedicated SDR boards for mm-Wave wireless communications. The Pi-Radio v1 SDR consists of a 4-channel fully-digital transceiver board with a Xilinx's ZCU111 RF System on a Chip (SoC) [174], and operates over a bandwidth of about 2 GHz in the 57-64 GHz band.

Other platforms currently in use in experimental work include the open source mm-Wave experimentation platform proposed in [175]. It consists of a Xilinx Kintex Ultrascale FPGA with a 60-GHz front-end. The FPGA is integrated on an AMC599 board that implements hardware signal processing and storage for real-time frame processing. It can also provide antenna array reconfigurability for fast beam switching, e.g., for high mobility scenarios. Moreover, Polese et al. [176] propose a 60-GHz SDR, fully digital experimentation platform. It uses a Xilinx KC705 and has 4 independent streams.

Alternatively, commercial-grade equipment can be leveraged for localization purposes, usually by substituting the provided operating system image with a custom build that embeds Application Program Interface (API) to access the output of the beam training procedure. For example, the work in [154] realizes a geometric 3D localization system using a 4×8 phased array within a router that embeds a Qualcomm QCA9006 tri-band chipset for AoA and ToF measurements. The work in [161], instead, taps into the output made available by the Talon AD7200 [142] routers' firmware. In the latter case, the hardware and the interface require significant adaptations of the angle estimation algorithms. For example, the firmware and operating system used in [161] returned coarsely quantized power measurements for each beam pattern and sometimes incomplete measurement outputs, which required to re-cast the angle estimation algorithm to be robust against quantization noise and missing values. The proprietary setup used in [154] returns the raw CIR measurements, which are then sanitised to extract the azimuth and elevation angles of arrival from the LOS paths, and the ToF information for distance estimation.

Other works such as [178] also employ COTS devices like the 802.11ad-enabled Airfide

Table 2.3: Summary of the hardware and software platforms used in mmWave localization algorithms

Hardware Platform	Related Literature
Vubiq 60 GHz development system	[132], [133], [138], [161], [165]
Zynq 7045 based SDR with 60 GHz analog front-end	[151]
4×8 phased array AP with QCA9006 triband chipset	[154]
TP-Link Talon AD7200	[139], [144], [147], [148], [149], [150]
QCA6320 baseband module with QCA6310 RF front-end	[156]
USRP X310 and TwinRX daughterboard with 60 GHz analog front-end	[163]
MicroTik wAP 60G	[157]

Software Platform	Related Literature
NYURay Ray tracer	[162],[177]
S_5GCHANNEL simulator	[159]

AP [179] to enhance the antenna array performance for omni-directional coverage and to improve link resilience in mobile and dynamic environments. Table 2.3 summarizes the above discussion by relating the works in our survey with the hardware platforms used to validate mm-Wave localization algorithms. We observe that software-defined platforms are still preferred, due to their greater versatility and to the availability of multiple digital receiver chains. COTS hardware is starting to appear in experimental evaluations, although this typically requires system management (and sometimes hacking) skills to flash the hardware with firmware and custom operating systems that give access to information from the radio receiver chain.

From a practical standpoint, the manufacturers of commercial-grade mm-Wave devices typically define a codebook of antenna weights that drive beam patterns to cover the largest set of lookout directions. As a result, the corresponding beam patterns are not necessarily narrow, nor do they necessarily present a single direction where the gain is maximum [180].

Yet, standard-compliant beam training procedures still help retrieve location-dependent measurements through an automated process that is typically implemented in every device. For example, the 802.11ad standard [29] presents a two-phase beam training process:

- *Sector Level Sweep* (*SLS*): During this phase, the transmitter (or *beamformer*) periodically transmits Sector Sweep (SSW) frames using the different beam patterns defined in the sector codebook. The receiver (or *beamformee*), receives these frames omnidirectionally and sends back an acknowledgment with the transmit sector yielding the highest signal quality. Subsequently, the two devices swap roles, and the receiver selects its best transmit sector. This phase provides coarse-grained beam patterns that are best suited for the two communicating devices.

- *Beam Refinement Protocol* (*BRProt*): This optional phase can be used to refine the beam patterns chosen after the SLS phase. The BRProt process is iterative. The two devices exchange special BRProt packets requesting and acknowledging the transmit (TX) and receive (RX) training requests (TX-TRN and RX-TRN). The result is fine-grained beam patterns for the transmission and reception of the data, resulting in not just better directivity and therefore
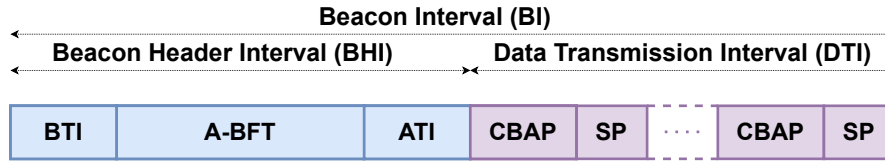
Figure 2.11: Beacon Interval frame of the IEEE 802.11ad standard [29]. It is important to note that after beam training process, Station (STA) contend for the channel during the contention based access period (Contention Based Access Period (CBAP)) and access it contention-free during the service period (Service Period (SP)).

> higher-throughput links, but also in a higher correlation between the beam pattern used and the AoA of a signal.

These phases occur during the Association Beamforming Training (A-BFT) subinterval of the Beacon Interval (BI), as part of the channel access mechanism. The beamforming process during the Data Transfer Interval (DTI) is to handle device mobility, blockage, etc. The BI frame for channel access is shown in Fig. 2.11 and the two beamforming phases are illustrated in Fig. 2.12.

The more recent 802.11ay standard [181] formalized beam training procedures that enhance those of 802.11ad, namely the beam refinement protocol transmit sweep (BRProt TXSS) and the Asymmetric Beamforming Training (ABT) [31]. These procedures rely on a channel reciprocity assumption to speed up beam training (through the *BRProt TXSS* scheme) and slightly improve the process to compensate for the possibly different antenna gains at the AP and at the client.

In addition, 802.11ay speeds up training in the presence of several clients through group beamforming, which extends beam training to manage multiple clients simultaneously.

When run with generic beam patterns, the above procedures do not yield a one-to-one relationship between the angle of arrival or departure of a mm-Wave signal and the antenna configuration that leads to the highest received power. Yet, if the beam patterns of the codebook are known, a mm-Wave device can still estimate angles of arrival via signal processing techniques involving compressive sensing [180], or linear programming and Fourier analysis [161]. Knowing angles of arrival enables angle-based localization techniques, with the additional advantage that angle estimation hinges on standard beam training procedures, with no need for external hardware components. In other words, localization becomes an embedded feature of mm-Wave communications.

Simulation-based performance evaluation

Simulation is the performance evaluation tool of choice if mm-Wave hardware is not available or if the available platforms do not offer sufficient flexibility to measure location-dependent features. A common practice observed in the literature is to employ ray tracers to mimic the propagation of mm-Wave signals. These ray tracing simulators are typically designed based on the channel models described in Section 2.2. The main idea is to simulate the mm-Wave wireless channel

(a) Sector Level Sweep

(b) Beam Refinement Protocol
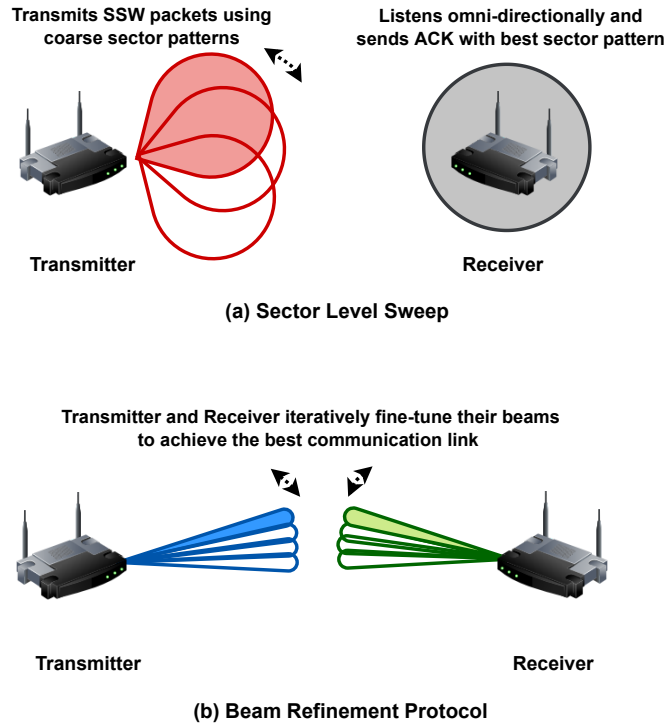
Figure 2.12: A simple illustration of the sector level sweep and beam refinement protocol as proposed in the IEEE 802.11ad standard [29].

characteristics at various indoor locations. Besides allowing the experimenter to measure channel features, ray tracers help create a radio map of the environment, and can thus substitute costly and time-consuming measurement campaigns [182].

Two examples of such simulators are NYURay, a 3D mm-Wave ray tracer developed by New York University [182], and S_5GChannel, developed by Siradel. NYURay was initially conceived as a geometry-based 2-Dimensional (2D) ray tracer and was used in [177] to investigate indoor positioning algorithms based on AoA, combined path-loss and AoA, or RSSI values. NYURay was later extended in [162] to support 3D ray tracing by combining the shooting-and-bouncing rays (SBR) technique [183] and the geometry-based technique. NYURay found extensive use, not just in indoor environments, but also outdoors [184], [185]. Siradel developed the S_5GChannel [186] 5G channel simulator to address the challenges of 5G signal propagation at mm-Wave frequencies indoors and outdoors. S_5GChannel's ray model has been used in [159] to develop a framework for joint localization and mapping.

A few additional works in the literature evaluate their proposed schemes using custom simulation software typically written in MATLAB or Python. The general purpose of such software is to generate synthetic datasets with realistic mm-Wave propagation characteristics, although typically restricted to the specific signal properties required for each study (e.g., AoA values, ToA measurements, etc.).

In the following subsections, we explain the details of each surveyed work, and provide a synopsis of the main results of each paper and of the main enabling techniques in the form of summary tables at the end of the section.

2.5.4 Angle-based algorithms

AoA measurements, alongside the quasi-optical nature of mm-Wave signal propagation, facilitate high-accuracy localization based on triangulation. This is the simplest approach to localization using AoA, wherein the angle information from the transmitting AP and simple geometric principles are used to compute the client's position. In a 2-D plane, such position can be estimated using just two AP [187].

Geometric methods are the simplest methods for localization when using AoA estimates. In [133], the authors present three lightweight single-anchor algorithms based on the AoA measurements. These algorithms are based on triangulation, ADoA, and fingerprinting, respectively. The algorithms have been simulated and also experimentally validated on pre-standard mm-Wave hardware operating at 60 GHz, showing that they achieve sub-meter accuracy with high probability, given the AoA estimate errors are low.

The simplicity of these algorithms motivated the authors of [132] to generalize the schemes in [133] for any number of AP. These algorithms are extensively simulated as well as experimentally validated on 60 GHz COTS devices, in different indoor scenarios against two benchmark algorithms based on fingerprinting and AoA. The two algorithms provide sub-meter accuracy in most indoor environments with multiple antennas. Triangulation-based scheme performs slightly better than the ADoA-based one in most scenarios, but independence of orientation and compass bias makes ADoA more preferable.

The ideas proposed by [133] have also been used by the authors of [136] for context inference and obstacle detection. They use the TV and ADoA algorithms for receiver localization using one AP, estimate the locations of virtual anchor nodes, and thus infer the presence of obstacles.

AoA measurements have also been used for Simultaneous Localization and Mapping (SLAM). For example, in [137], the authors propose a joint access point and device localization (JADE) algorithm that jointly maps the location of the client and of the physical and virtual AP, while mapping the indoor environment, without any prior information (i.e. number of access points, boundaries of the room, etc.). The algorithm measures AoA from the beam training procedure and leverages ADoA to estimate the location of the AP and then of the client. Environment mapping follows by matching physical and virtual anchors and by predicting reflection points on surrounding surfaces. Simulation results show sub-meter accuracy in 90% of the cases, even for erroneous AoA estimates. JADE outperforms the approaches in [132] in almost all scenarios.

A similar algorithm that exploits AoA information to derive ADoA estimates and fuses multiple measurements at different locations is CLAM [138]. Like in [137], the algorithm proceeds by first estimating the location of the anchor AP, then of the client, and finally of the environment's boundaries. The algorithm is simulated and experimentally evaluated, showing sub-meter device localization errors in about 90% of the cases.

A recent work explores deep learning-based localization scheme. The authors of [168] propose a shallow neural network model to estimate the coordinates of the client device in an indoor environment, using ADoA measurements. The network is trained with imperfect location estimates from the JADE algorithm [137], which jointly estimates the location of the AP and the clients with zero knowledge of the environment. This relieves the burden of explicitly collecting the training dataset. The performance evaluation of the proposed scheme results in sub-meter client localization accuracy in $\approx 90\%$ of the scenarios, even with large AoA errors.

In [134], the authors present mobile device positioning scheme in an indoor mm-Wave massive Multiple-Input Single-Output (MISO) scenario. The two-fold scheme utilizes coarse-grained AoD information from mobile clients with a single antenna to estimate the position of each client via downlink transmissions using adaptive beamforming.

We can observe that angle-based algorithms usually rely on geometric approaches for device localization. However, ML and neural network regression models can also be used to learn a non-linear mapping between AoA measurements and client locations.

2.5.5 Channel information-based algorithms

In recent works, mm-Wave CSI is also used to estimate the location of the client. The definition of CSI varies from work to work. Typically, the term refers to the complex amplitude of the channel gain perceived at a receiving antenna, or to the vector of such gains measured by all elements of an antenna array. A work exploiting CSI for localization is [140], where the authors present a channel parameter estimation method that transforms the mm-Wave uplink training signal into a higher-dimensional tensor using the canonical polyadic model. Tensor factorization using the proposed generalized structured canonical polyadic decomposition results in time delay, AoA, and path fading coefficient estimates. These parameters are used to localize and track a mobile device.

A different way to exploit uplink CSI estimates [139] requires that the AP convert the LOS CSI measurements into angle information and then localize the client. The system is implemented on Talon AD7200 routers (without interfering with 802.11ad operations), and the authors propose to employ the location estimates to optimize AP–client associations. The system achieves sub-meter localization accuracy in about 80% of the cases.

With a focus on localizing passive objects, in [141] the authors use the CIR captured after reflection from different objects and surfaces in an indoor environment to detect objects and also model the indoor environment in 2D. The proposed method has been evaluated using a testbed developed specifically for this purpose.

The use of CSI for localization is comparatively new for mm-Wave indoor device-based localization, most likely because retrieving full CSI or CIR data requires low-level hardware access, and only a few experimental firmware versions provide it. However, CSI and CIR can map to angle and time information, and therefore represent a promising and practical research direction, especially as feature-richer mm-Wave hardware and firmware emerges.

2.5.6 RSSI and ToF

RSSI and SNR based localization systems generally employ trilateration or fingerprinting-based techniques to localize the client. A number of works in the literature illustrate this concept. The authors in [143] investigate trilateration-based localization algorithm using RSSI measurements for 60-GHz IEEE 802.11ad WLAN. They modify the trilateration algorithm based on the concept of (weighted) center of mass. Simulations on randomly generated data points and the RSSI measured based on the IEEE 802.11ad channel model result in an average positioning error of about 1 m. This is among the earliest works on mm-Wave-based indoor localization that leverages RSSI measurements.

RSSI is also the foundation of several fingerprinting-based localization schemes, especially in sub-6 GHz wireless networks. The authors of [144] propose a localization system that generates fingerprints of transmit beam indices and the corresponding RSS measurements between a pair of mm-Wave devices. Probabilistic location models are generated based on the fingerprint data and are leveraged for location estimation. The algorithm is experimentally evaluated using 60 GHz COTS devices. Many times, SNR-based fingerprinting is also at the core of some mm-Wave localization works, especially in combination with machine learning and deep learning techniques. The authors of [35], [147] propose machine learning regression models for localization in warehouses. SNR information is collected from Talon AD7200 routers. The supervised regression models are trained offline and then deployed for localization at run time. The proposed method achieves sub-meter accuracy in 90% of the cases.

Similar machine learning regression models have been used for location estimation in [148], where the authors use spatial beam SNR values, typically available during the beam training phase, in order to generate a location- and orientation-dependent fingerprint database. Deep learning techniques are also the main enablers for localization in [149] and [150], where the authors proposed ResNet-inspired models [188] for device localization in LOS and NLOS scenarios. To tackle the challenges imposed by NLOS conditions, the authors use spatial beam SNR values in [149], whereas they employ multi-channel beam covariance matrix images in [150].

One example of how ToF measurements have been used in the mm-Wave context is presented in [151]. Here, the authors present a two-way ranging based on round-trip ToF (RTToF) information. The scheme estimates the distance between master and slave nodes, and then trilaterates the position of the slaves. The authors implement their algorithm on an SDR with a 60 GHz SoC. The proposed system achieves an average distance estimation of 3 cm and an average positioning error below 5 cm.

Although conventional wireless localization schemes relying on RSSI or SNR measurements employed trilateration, machine learning-based fingerprinting algorithms are gaining more popularity for mm-Wave-based localization systems. This is due to the availability of mid-grained channel measurements from the beam training procedures of 5G and IEEE 802.11ad/ay systems [150]. These techniques provide higher-accuracy location estimates compared to conventional techniques.

We also observe that mm-Wave systems do not rely on purely time-based measurements for

35

localization. Even though the large bandwidth of mm-Wave signals can provide fine time measurements, such measurements tend to be fully available only on custom high-end mm-Wave transceivers. Therefore, many schemes tend to collect other signal measurements as well.

2.5.7 Hybrid approaches

A combination of two or more techniques mentioned above can be used to build systems that achieve better localization or mapping accuracy, with respect to stand-alone techniques. Coupling different sources of information is useful in challenging environments, where some mm-Wave parameter measurements may fail.

Angle information along with RSSI-based ranging are the foundation of several mm-Wave localization approaches in the literature. The authors in [164] propose a positioning algorithm using RSS and AoA measurements. These measurements are derived from a channel compression scheme designed for a mm-Wave mMIMO scenario with only one AP. The RSS and AoA estimates from the above methods are employed for position estimation. The system provides decimeter-level accuracy even at low SNR, and even lower errors as the SNR increases.

As opposed to ranging, the algorithms proposed in [152] and [145] are based on location fingerprinting. In particular, the authors measure RSSI and AoA information at various reference points in an indoor environment to generate location fingerprints. *K* fingerprints nearest to the client measurement are selected from the dataset, and the location estimate corresponds to the weighted average of these *K* reference points. The algorithm has been simulated with 2.4 GHz and 60 GHz, showing that the average position error is much lower for mm-Wave signals than lower-frequency signals. To solve the problem of collecting a sufficiently large dataset, [145] generates 3D beam fingerprints using RSSI and beam information. Weighted K-NN was used to localize an Unmanned Aerial Vehicle (UAV) in GPS-denied indoor environments. Particle filters were used along with the imperfect location estimates to track the motion UAV. The proposed scheme was experimentally validated, and the results showed sub-meter positioning accuracy on average.

RSS jointly with AoA information enables *mmRanger* [156] to autonomously map an indoor environment without infrastructure support. The mmRanger scheme senses the environment and uses time domain RSS sequences to reconstruct the path geometry via a path disentanglement algorithm. Then, AoA and RSS information from the reflecting surfaces are exploited to reconstruct the geometries of each surface. Moreover, a robot pedometer assembles all estimated fragments to form a complete map of the environment. The results of the proposed system implementation show a mean estimation error of 16 cm for reflection points, and a maximum error of 1.72 m.

In [161], the authors leverage coarse-grained per-beam pattern SNR measurements provided by a modified operating system flashed on multiple TP-Link Talon AD7200 802.11ad-compliant COTS mm-Wave router. The AoA estimation problem is formulated using linear programming, and the location is estimated using a modified particle filter and a Fourier analysis-based goodness function. The proposed scheme is experimentally validated and the system achieves sub-meter accuracy in 70% of the cases. AoD and SNR information were used in [153] to design beam-based

midline intersection and beam scaling-based positioning algorithms. These were evaluated using both ray-tracing simulation and a WiGig SoC transceiver. The experiments, carried out under LOS conditions, yielded centimeter-level location estimation errors.

Time-based measurements are often enriched with angle information in order to achieve better positioning accuracy, especially for mm-Wave systems. For example, in [154], the authors propose *mWaveLoc*. The proposed system uses measured CIR to calculate AoA and ToF data. The system is implemented on IEEE 802.11ad off-the-shelf devices leveraging the OpenWRT operating system, and achieves centimeter-level distance estimation and decimeter level 3D localization accuracy (median error 75 cm and sub-meter error in 73% of the cases) in a realistic indoor environment. The system has also been evaluated in various experimental conditions.

The author of [162] propose a map-assisted positioning technique using the fusion of ToF and AoD/AoA information. A 3D map of the environment is either generated on the fly or assumed to be known a-priori. The scheme measures a set of possible user locations by fusing the estimated ToF values with angle information. These estimated locations are clustered, and the cluster centroid is output the final location estimate. The algorithm is simulated on the data collected at 28 GHz and 73 GHz by NYURay 3D ray tracer. The best-case and the worst-case mean localization error is found to be about 12 cm and 39 cm respectively.

Instead of explicitly fusing ToF and AoA information, the authors of [165] propose a pseudo-lateration protocol, that enacts the three following steps: i) sector sweeping for tracking LOS and NLOS paths to compute physical and virtual anchors, respectively; ii) angular offsets measurements using extended sector sweeping; and iii) ToF measurements for distance estimation. A post-processing stage is employed for position estimation. The protocol has been simulated and implemented using a 60 GHz mm-Wave testbed. The protocol implementation achieves centimeter-level location estimation accuracy within 1.5 m and decimeter accuracy beyond 1.5 m.

The authors of [166] explored adaptive filters for motion-assisted indoor positioning. An improved LMS filter estimates the AoA of the client by using the client location, velocity and measured ToF as the inputs. AoA and ToA estimates are fed to an Unscented Kalman Filter (UKF) to track the client's position. The two-stage algorithm is simulated in an office environment with one AP and achieves centimeter-level positioning accuracy.

Because mistaking LOS for NLOS paths may offset location estimates significantly, the authors of [167] propose a scheme to tell apart mm-Wave LOS and NLOS MPC having incurred up to one reflection. For this, they use TDoA and AoA information and apply the mean shift clustering technique. Then, they apply an AoA-based localization scheme that computes least-squares estimates. The methods show a 98.87% accuracy in path identification and positioning error of less than 75 cm in 90% of the cases. NLOS scenarios have also been exploited in [155], where the authors propose a positioning scheme that relies on differential angle information, which is independent of angular reference. This scheme has been evaluated in an indoor environment with a geometric ray tracer based on an IEEE 802.11ay channel model, and achieves sub-30 cm position estimation errors in 90% of the cases.

In [158], the authors present schemes for localization, mapping, obstacle detection and classification. Localization and mapping make use of AoA and ToA measurements to estimate the

location of the receiver and of virtual anchors. The latter are used to detect obstacles by estimating reflection points. Snell's law and the relationship between the RSS and the reflection coefficient are used to classify the obstacles based on material composition. The presented algorithms have been simulated in an indoor environment.

Besides locating a client, the schemes presented in [158] have been integrated into a SLAM framework in [159]. This framework involves algorithms for localization, obstacle mapping and tracking. Extended Kalman Filter (EKF)-based tracking helps improve obstacle detection and mapping. The framework has been simulated in an indoor environment, yielding sub-meter errors in 90% of the cases. In the same context, the EKF improves the obstacle mapping accuracy to sub-centimeter.

In [163], the authors present a device localization scheme, where the AP and the client are equipped both with sub-6 GHz and with mm-Wave technology. Sub-6 GHz antennas are used for AoA estimation and mm-Wave antennas are fed with the AoA estimates for subsequent beam training and two-way ranging. The proposed method has been experimentally validated using SDR platforms, both in an anechoic chamber and in an office environment. Results show 2° AoA errors and centimeter-level ranging accuracy in the anechoic chamber, and 5° AoA error with an average 16-cm range error in the second one.

In [157], the authors propose to track the changes in the CIR measured at the station, that is equipped with an FPGA-based platform with IEEE 802.11ad, in order to localize a device-free object in an indoor industrial environment. The station uses the estimated CIR to measure the AoD and ToF of the signal reflecting off a moving object. Tracking CIR changes over time helps classify the reflections as static or mobile. Then, a Kalman filter smooths the trajectory of the mobile object. The results show sub-meter location errors in all scenarios, and a mean accuracy of 6.5 cm.

From the literature surveyed above, we can observe that most localization schemes use angle information along with RSSI/time information, and often rely on geometric algorithms to compute high-accuracy location estimates. The use of adaptive filters such as Least Mean Squares (LMS) and Kalman filters helps mitigate location estimate errors, especially with mobile clients.

2.5.8 Summary, highlights, and challenges

We now summarize the surveyed literature in order to highlight the main pitfalls and lessons learned from the methods.

Geometry-based algorithms – The algorithms based on geometric techniques mostly rely on angle information (AoA/AoD) for localization. As mm-Wave signals propagate quasi-optically, angle information becomes a reliable means to estimate the direction of the source. RSSI and ToF information help estimate the distance between a mm-Wave source and its receiver; thus, applying geometric methods such as triangulation and trilateration can help localize a client. However, the accuracy of such algorithms depends upon the accuracy of angle and time measurements, and most of them require accurate indoor floor plan information to work reliably.

From the perspective of COTS devices, angle measurements are obtained either by decomposing CSI measurements using parameter estimation techniques or from the beam patterns chosen after beam training. However, imperfect beam patterns with broad main lobes and non-negligible sidelobes can lead to angle estimation errors. Similar issues affect the estimation of time information through CIR or packet exchange means. We can obtain fine time measurements thanks to the large bandwidth of the mm-Wave signals. However, this requires a very tight synchronization between the AP and the client devices.

ML-based algorithms – Owing to recent contributions, we observe a paradigm shift towards self-learning location systems that exploit the information from mm-Wave signals. In these works, ML and DL-based models use information extracted from mm-Wave signals at different locations to form a dataset and eventually learn an accurate model to estimate client positions. However, most of these models are specific for the location the training data comes from, and do not translate well to other locations. Most of the algorithms in the literature train ML and DL models through RSSI or SNR fingerprint maps. Recent works have showed how AoA and CSI information also help ML models learn a non-linear function either to estimate the location of the client or to associate a client to the best AP.

Although these systems provide good localization accuracy, they also face several challenges: the collection of large training dataset; the computational complexity which limits their application to COTS or embedded devices; their dependence on the training environment. ML methods have thus found comparatively limited application to date. A valuable contribution to the community would be a collaborative effort towards a public benchmark dataset, that different authors would use to feed different machine learning approaches.

Error mitigation in mm-Wave localization – Errors in signal measurements due to imperfect signal parameter estimation limit the performance of localization systems [44]. These errors are often due to the unpredictable interference between multiple propagation paths and the fading that results, or to NLOS arrivals reaching a device [189]. A detailed mathematical analysis for error mitigation is presented in [44]. In the case of mm-Wave, measurement errors may affect angle and time measurements. The works in the literature resort to adaptive filter-based techniques mostly to mitigate the location estimation errors resulting from localization algorithms fed with error-prone data [33]. For example, the approaches in [145], [161] resort to particle filters to mitigate client location errors. Different types of Kalman filters are another typical choice to smooth out client location estimates and trajectories [190]. The authors of [167] use LMS filters to mitigate large errors in AoA and ToF measurements. A detailed account of the tools and techniques employed in each surveyed paper is provided in Table 2.6 at the end of this section.

Table 2.4 supports the above discussion by summarizing the techniques employed in each of the surveyed works. We observe a preference for geometry-based localization approaches, with different supporting signal processing algorithms.

Table 2.4: Summary of the main techniques used in the surveyed papers

Analytical Tools	Related Literature
Beamforming techniques	[134]
Clustering methods	[140], [167],[156]
Deep learning	[135], [149], [150], [168]
Fourier analysis	[161]
Geometry	[132],[133],[136], [143],[146],[153],[154], [162], [158],[160]
Kalman filters	[139], [166], [157], [159]
Least mean square filters	[166]
Levenberg-Marquardt (LM) method	[155]
Linear programming	[161]
Machine learning models	[135], [145], [147], [148], [152]
MUSIC	[163]
Particle filters	[145], [161]
Probabilistic data modelling	[144]
Tensor analysis	[140]

Table 2.5: Summary of the evaluation methods used in the mmWave localization algorithms

Evaluation	Related Literature
Experimentation	[132], [133], [138], [141], [144], [145], [147], [149], [150], [151], [153], [154], [161], [163], [165], [156], [157]
Simulation	[134], [136], [140], [143], [146], [152], [162], [164], [166], [167], [155], [158], [159], [160], [168]

2.5.9   Discussion and future research directions

We summarize the findings of our survey in Table 2.6. The table succinctly conveys the main proposition of each paper, the main techniques used among those outlined in Sections 2.5.4 to 2.5.7, the tools employed, and a highlight of the performance attained. We observe a number of mm-Wave localization approaches exploiting different features of mm-Wave signals as well as different properties of mm-Wave propagation. While some schemes rely on well-known techniques, e.g., based on ToF and RSSI measurements, even these techniques have been further developed to leverage the sparsity of mm-Wave multipath patterns in order to collect more precise measurements with a finer time resolution. In some environments, typically with special-purpose lab-grade mm-Wave hardware, the corresponding localization schemes often yield decimeter- or sub-decimeter-level accuracy, but require specific protocols to exchange the data the algorithms need.

With respect to such approaches, angle-based localization schemes relying on AoA and ADoA measurements still prove accurate, and yield the additional benefit that AoA measurements can directly result from beam training operations at link setup time. Hybrid solutions that leverage both angle information and time/RSSI information tend to show even better performance, although only a minority of them has been tested in operational environments with COTS equipment.

Finally, Table 2.5 summarizes the performance evaluation approach taken in each surveyed paper. We observe a majority of evaluations based on experiments with real hardware, although simulation is still used in several contexts, e.g., as a tool to quickly and affordably generate large datasets.

mm-Wave technologies are expected to keep gaining momentum as part of the 5G-and-beyond ecosystem, and there exist a wealth of promising research directions to realize the vision of embedding localization as a feature of mm-Wave communications. According to our analysis, we identify the following key research directions. The community needs more low-cost mm-Wave platforms that implement standard-compliant operations while still providing API for researcher and developers to access low-level physical layer measurements, such as per-beam pattern CSI, or even better, full CIR. This would democratize the research on practical algorithms that fully integrate localization as part of standard communications operations. In particular, such platforms would help research better algorithms to manage scenarios featuring multiple AP, which are expected to be common in indoor mm-Wave deployments. Moreover, there is ample space for the design of zero-initial knowledge algorithms that require no input data from the user, and autonomously bootstrap the algorithm by finding the location of all anchors (e.g., all AP), localizing the clients, and using the joint location information of all clients and AP to estimate the floor plan of indoor environments in a SLAM fashion, both as a stand-alone solution and as a complement to the device-free radar-based approaches described in Section 2.6.

From the point of view of ML schemes, we observe that most approaches still require lengthy training data collection operations before achieving practical accuracy levels. Moreover, a trained ML algorithm remains specific to the area where training data was collected. Therefore, further research is needed on machine learning approaches that work with less training data, federate training results from different clients and AP in order to speed up the training phase, and can be transferred across different, even previously unseen environments.

Table 2.6: Summary of the literature on indoor mmWave localization

Proposition	Techniques	Tools Used	Performance
Localization Algorithms			
Round-trip time based localization [151]	ToF	Geometry	Distance estimation error within 3 cm and location estimation error within 5 cm.
Accurate 3D indoor localization using a single AP [154]	AoA-ToF	Geometry	3-D localization with median accuracy of 75 cm with sub-meter accuracy in 73% of the cases.
Improving localization accuracy [161]	AoD, RSSI	Linear programming, Fourier analysis, Particle filters	Sub-meter accuracy in 70% of the cases.

Table 2.6: SUMMARY OF THE LITERATURE ON INDOOR MMWAVE LOCALIZATION (CONTINUED)

Proposition	Techniques	Tools Used	Performance
Improving the accuracy of device localization [132], [133]	AoA	Geometry	Sub-meter localization accuracy in 70% of the cases.
Improving location estimation accuracy and network performance [139]	CSI	Regression trees, Extended Kalman filter	Sub-meter localization accuracy in 80% cases and throughput improvement between 8.5% and 57%, lesser outage probability, SNR within 3 dB of optimum.
Fingerprinting based indoor localization [144]	RSSI	Probabilistic models	Mean and median localization error of ≈30 cm.
Indoor localization for intelligent material handling [147]	SNR	Multi layer perceptron regression, Support vector regression, Logistic regression	Centimeter-level accuracy with Root Mean Square Error (RMSE) of 0.84 m and MAE of 0.37 m.
Fingerprinting based indoor localization [148], [149]	SNR	Machine learning algorithms, Deep learning	Avg. RMSE is 17.5 cm with coordinate estimates within 26.9 cm in 95% of the cases. Median and mean RMSE of 9.5 cm and 11.1 cm respectively.
Fingerprinting based indoor localization in NLOS environments [150]	SNR	Deep learning, Machine learning algorithms	Location classification accuracy greater than 80%. Median location estimation error of about 11 cm.
Map-assisted indoor localization [162]	AoA, AoD, ToA	Geometry	Mean localization accuracy of 12.6 cm and 16.3 cm in LOS and NLOS respectively.
Sub-6 GHz-assisted device localization [163]	ToF-AoA	MUSIC, Geometry	AoA estimation error less than 5° and about 16 cm distance estimation error.

Table 2.6: Summary of the literature on indoor mmWave localization (continued)

Proposition	Techniques	Tools Used	Performance
Single-antenna client localization using downlink transmissions [134]	AoD	Adaptive beamforming	60% improvement in the accuracy in the downlink scenario as compared to in the uplink scenario.
Indoor positioning for wideband multiuser millimeter wave systems [140]	CSI	Tensor decomposition, Clustering methods	Decimeter-level position estimation errors.
Indoor network localization [143]	RSSI	Geometry	Mean positioning error around 1 m.
3D indoor positioning for mm-Wave massive MIMO systems [164]	AoA-RSSI	Geometry	Low complexity channel compression and beamspace estimation developed. Decimeter-level positioning errors achieved in NLOS scenarios.
Location fingerprint-based localization [152]	AoA-RSSI	K-nearest neighbours	Average positioning error for mmWave is 4 times less compared to lower frequency signals.
UAV positioning in GPS-denied environments [145]	RSSI	Weighted K-NN, Particle filters	Sub-meter 90th-percentile location errors in different cases.
Beam-based UE positioning in indoor environment [152]	SNR, AoD	Geometry	Centimeter-level estimation error in all cases. Experimental results approach the simulations results with MSE difference of 0.1 m.
Single RF chain-based localization [165]	ToF-AoA	Geometry	Centimeter accuracy in location estimation within 1.5 m and decimeter accuracy beyond 1.5 m.
Motion feature-based 3D indoor positioning [166]	AoA-ToA	Least Mean Squares and Kalman filters	Centimeter-level positioning accuracy.

Table 2.6: Summary of the literature on indoor mmWave localization (continued)

Proposition	Techniques	Tools Used	Performance
LOS and NLOS path identification and localization [167]	TDOA, AOA	Mean shift clustering, Geometry	98.87% accuracy in path identification and positioning accuracy ≤ 0.753 m in 90% of the cases.
NLOS mmWave indoor positioning [155]	AoA, AoD, ToA	Geometry, Levenberg-Marquardt (LM) method	Positioning accuracy within 30 cm in 90% of the observations with differential angle information along with time information.
5G mmWave indoor positioning [146]	RSSI	Geometry, Beamforming	Single-beam geometric model for indoor positioning. Mean error of 0.7 m for stationary in LOS and 2.4 m for a mobile user in LOS/NLOS scenario.
Data-driven indoor localization [135]	AoA	Multi-layer perceptron, Deep learning	Sub-meter localization accuracy.
Indoor localization with imperfect training data [168]	AoA	Deep learning	Sub-meter localization accuracy in $\approx 90\%$ of the cases, when trained with client location estimates from JADE algorithm.
Localization and Mapping Algorithms			
Autonomous environment mapping [156]	RSSI, AoA	Geometry, K-means clustering	Reflection point mean estimation error of 16 cm with a max error of 1.72 m.
Passive object localization [157]	AoD, ToF	Kalman filters	Sub-meter accuracy in all cases with 6.5 cm mean error accuracy.
Localization and obstacle detection [136]	AoA	Geometry	Sub-meter accuracy in 70% of the cases. High accuracy obstacle detection and obstacle limits estimation.
Localization and mapping [137], [138]	AoA	Geometry	Sub-meter localization accuracy in 90% of the cases. SLAM without any prior knowledge.

Table 2.6: SUMMARY OF THE LITERATURE ON INDOOR MMWAVE LOCALIZATION (CONTINUED)

Proposition	Techniques	Tools Used	Performance
Accurate object detection [141]	CIR	Geometry	Accuracy of about 2 cm achieved in most experiments.
Simultaneous localization and mapping without a-priori knowledge [158], [159]	AoA, ToF, RSSI	Geometry, Extended Kalman filters	Sub-meter device localization accuracy in 90% of the cases. Sub-centimeter obstacle mapping accuracy.
3-D localization and mapping [160]	RSSI, AoA, ToA	Geometry	Perfectly maps the environment for AoA errors $\leq$ 5°.

2.6 Device-free radar-enabled mmWave localization and sensing

2.6.1 Introduction

In this section, we focus on mm-Wave-based radar systems that operate over short distances (a few tens of meters), which have recently emerged as a low-power and viable technology for environment sensing. These devices are expected to find extensive use in a number of relevant applications, by replacing standard camera-based systems, either fully or in part. Survey papers have been recently published on mm-Wave sensing, with a focus on signal processing [191] (both with traditional and machine learning-based approaches) and applications [3]. In the following review, we emphasize the main signal processing algorithms that are being successfully exploited for indoor sensing, discussing their pros and cons. In doing this, we especially focus on Neural Network (NN) algorithms, discussing their different flavors, and exposing the most promising directions for research and development. We also comment on the level of maturity of this technology, i.e., about whether the proposed techniques are robust and work without requiring environment-specific and manual calibration. Our analysis will also discuss on the role of the supporting architecture, which should provide communication and computing/processing capabilities, and on the opportunity of implementing networks of radar devices. This would extend current systems, which often involve a single radar, to the large-scale monitoring of physical spaces. An illustrative overview of the main techniques used for sensing applications that exploit mm-Wave radars is provided in Fig. 2.13.

Two main components of a radar device are TX and RX RF antennas, which are combined with an ADC, Micro-Controller Unit (MCU), Digital Signal Processing (DSP) and a clock. The main idea behind such a system is to transmit a properly shaped radio wave (e.g., pulses or continu-
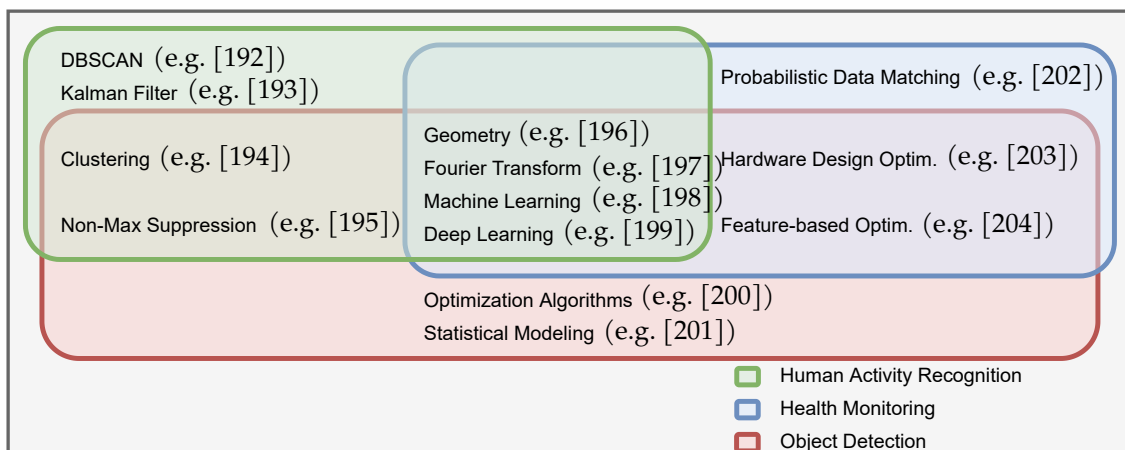
Figure 2.13: Overview of techniques used for mm-Wave radar sensing applications.

ous waves) and estimate the modifications that occur in the back-scattered copy of such wave, i.e., which is returned as a reflected signal from the surrounding environment. Through some signal processing algorithms (usually, Fourier transform-based), it is then possible to estimate the distance, angle, velocity, and, to some extent, the shape of the targets. TX and RX are usually co-located within the same device: the transmitter sends a first version of the modulated signal and the receiver detects its back-scattered copy from the surrounding environment, after a very short time delay.

Modern radar systems utilize two main *wave functions*; Pulsed Wave (PW) and Frequency-Modulated Continuous-Wave (FMCW). While radars are traditionally used to detect and track objects that move in the far field, such as vehicles and airplanes, here we are concerned with indoor or city environments where the objects to be tracked may be cars or humans. Moreover, besides the tracking, vital signs can also be monitored such as the breathing rate and the heart rate. Although these recent applications share common features with traditional (far field) approaches, they also exhibit major differences due to the short distance of the radar from the targets (near field).

2.6.2 Pulsed Wave Radar

With PW radars the electromagnetic waves from an antenna are emitted in short bursts. The logic behind PW is to wait for the reflections from the previously transmitted signal to reach to the antenna before sending the next burst. Thus, the reflected signal from the initial emitted sequence of pulses are sampled via a secondary sequence of pulses with a different repetition time. In PW radars, energy of the transmitted pulse is relatively small due to the limited peak amplitude. This limitation in amplitude together with sequential sampling limits the dynamic range and results in a relatively poor SNR at larger distances. For these reasons, PW-type radars have fallen out of favor, and are not used for the applications that will be discussed next, which are mainly about object and people detection, tracking/identification, and vital sign monitoring. From the next

section onward, we thus concentrate on FMCW systems, which typically are the technology of choice for medium and larger ranges. Still, for short range applications, such as gesture tracking, PW-type radars might still be a viable alternative.

2.6.3 Frequency-Modulated Continuous Wave Radar

As the name implies, FMCW radars transmit a frequency-modulated signal in a continuous fashion. Due to the larger temporal duration of continuous-wave signals, FMCW yields a much larger energy on the emitted signal as compared to PW. In order to cover the desired frequency band, the signal is linearly modulated over time starting from the lower frequency within the band to the higher frequency (or vice-versa). This type of signal is most frequently referred to as a chirp, and the linear modulation of the signal is called frequency sweep. An analogue continuous-wave signal can be generated with a Voltage-Controlled Oscillator (VCO), providing flexible adjustments to the sweep duration independent of the bandwidth. A frequency synthesizer together with a VCO can be used to provide a digital alternative. This technique also provides a higher spectral purity which makes it possible to avoid accidental emission of frequencies adjacent to the desired band, and thus to comply with given regulations. In FMCW radars, the received signal is multiplied by the TX signal. The intermediate-frequency signal component that results is then isolated via low-pass filtering. Additionally, a low-cost ADC can be used to convert the received signal into the digital domain. Due to the recent developments on radar hardware, the wider operating frequency range and the above mentioned advantages, FMCW radars are currently preferred over PW ones, especially in millimeter-wave band applications.

FMCW Signal: As previously mentioned, a chirp is a linearly modulated FMCW signal: it is a sinusoidal function formulated as $\mathbf{x}_{tx} = \sin(\omega_{tx}t + \varphi_{tx})$, where the frequency $f_{tx} = \omega_{tx}/(2\pi)$ increases linearly over time. After transmission, the reflected chirp signal from an object is collected at the RX antenna and can be written as $\mathbf{x}_{rx} = \sin(\omega_{rx}t + \varphi_{rx})$. The Intermediate Frequency (IF) signal is produced by mixing RX and TX signals in the mixer component of the radar as $\mathbf{x}_{if} = \sin\left((\omega_{tx} - \omega_{rx})t + (\varphi_{tx} - \varphi_{rx})\right)$. The time delay between the RX and the TX signals is $\tau = 2d/c$, where d is the distance to the objects and c is the speed of light in air. The start of the IF is at τ, which is also when RX chirp is realized and ends when the TX chirp is entirely transmitted. Time delay is the foundation for computing the range and velocity of a target in an environment. While the given introductory FMCW concepts are sufficient for the purpose of this chapter, further details on FMCW radars can be found in [11].

Range Measurement and Resolution: Range resolution is defined as the ability of a radar to identify closely packed objects. When the distance separating two objects is smaller than the resolution, the radar becomes unable to distinguish between them, returning a single range reading. The range measurement is carried out by computing the phase difference between TX and RX chirps, yielding the initial phase of an IF signal, that is formulated as $\varphi_0 = 2\pi f_c \tau$, where

$f_c = c/\lambda$ stands for the frequency, c is the speed of light and λ is the wavelength. Hence, the distance d to an object, the so called range $d = \tau c/2$, can be retrieved as $d = \varphi_0 c/(2\pi f_c) = \varphi_0 \lambda/(4\pi)$. When multiple objects are present, a single TX chirp results in the reception of multiple reflected (RX) signal copies. According to the different time delays (τ) between the TX and each of the RX chirps, multiple IF signals are computed, and range measurements for each corresponding object are derived. The range resolution $d_{\mathrm{res}} = c/(2B)$, highly depends on the bandwidth B of the radar [205]: it can be improved by increasing the bandwidth swept by the chirp, yielding a longer IF signal and, in turn, leading to a more precise reading of the environment.

Velocity Measurement and Resolution: In an FMCW radar the velocity computation (commonly referred to as Doppler) can be achieved using two TX chirps. Initially, the object range is calculated by applying a FFT to the RX chirps. This range calculation is commonly called range-FFT. The range-FFT of separate chirps at the same location will yield different phases. The object velocity is then derived according to the phase difference between the two chirps as $\mathbf{v} = \lambda \Delta\varphi/(4\pi T_c)$, where $\Delta\varphi$ is the phase difference and T_c is the chirp duration. However, in the case of multiple moving objects having the same distance from the radar, the above method no longer works. To overcome this, the radar needs to transmit N chirps with equal separation, i.e., a so called *chirp frame*. When the chirp frame is passed through the range-FFT, it yields a phase difference containing combined phase differences of all the moving objects. The result of the range-FFT is passed through a second FFT called Doppler-FFT to identify specific phase differences ω of each object. In the case of two objects, the corresponding phase differences, ω_1 and ω_2, can be used to derive two velocities as $\mathbf{v}_1 = \lambda \omega_1/(4\pi T_c)$ and $\mathbf{v}_2 = \lambda \omega_2/(4\pi T_c)$. The velocity resolution, $\mathbf{v}_{\mathrm{res}}$, of the radar is inversely proportional to the duration of a single frame, $T_f = NT_c$. By knowing the frame duration T_f, the velocity resolution is $\mathbf{v}_{\mathrm{res}} = \lambda/(2T_f)$ [205].

Angle Measurement and Resolution: In radar sensing applications, most often the "angle" refers to the horizontal-plane AoA at the receiver (or *azimuth* in a spherical coordinate system). It is calculated by observing the phase changes occurring on the range-FFT or Doppler-FFT peaks. In order to observe these changes, there have to be at least two RX antennas. The difference between the readings of each antenna is what produces the phase change in the FFT peaks. The phase change is formulated as $\Delta\varphi = 2\pi\Delta d/\lambda$ s.t. $\Delta d = \ell\sin(\theta)$, where ℓ is the distance between the antennas. Accordingly, the angle can be estimated as $\theta = \sin^{-1}\left(\lambda\Delta\varphi/(2\pi\ell)\right)$. The closer θ is to zero, the more accurate the angle estimation becomes. In fact, the angle resolution $\theta_{\mathrm{res}} = \lambda/Nd\cos(\theta)$ is usually given assuming $\theta = 0$ and $d = \lambda/2$ which simplifies it to $\theta_{\mathrm{res}} = 2/N$. The field of view of the radar depends on the maximum AoA that can be measured [205].

We remark that the distance and angle resolution of a mm-Wave radar device are especially important as they characterise the density and the minimum separation of the points that are detected in the radar maps (see next section). This, in turn, has a major impact on the resolution of the clustering algorithms that are used to separate signals reflected by different subjects and objects in the radar maps (see, e.g., Density-Based Spatial Clustering of Applications With Noise (DBSCAN) in the following sections) and, on the final tracking performance of any signal

Table 2.6: Summary of the hardware platforms used in the literature

Hardware platform	Related literature
Google SOLI	[206]
Infineon SiGe	[207]
INRAS RadarLog	[208]
Keysight EXG N5172B	[209], [210]
Qualcomm 802.11ad device	[197]
Xilinx Kintex Ultrascale	[4]
TI AWR1243	[211]
TI AWR1443	[192], [199], [201], [204]
TI AWR1443BOOST	[212]
TI AWR1642	[213], [214]
TI AWR1642BOOST	[9], [196], [215], [216]
TI AWR1643BOOST	[195]
TI AWR1843BOOST	[217]
TI AWR2243	[203]
TI AWR6843	[198], [200]
TI IWR1443	[201], [218]

processing pipeline.

In Table 2.6, we summarize the types of passive mm-Wave radars employed in the literature covered by our survey. We observe that the availability of commercial evaluation boards from Texas Instruments (TI) and of software interfaces enabling the retrieval of raw radar data has made TI devices the platforms of choice in many of the works. However, others still prefer powerful but less commercial devices or come up with custom boards when commercial platforms are not sufficient to satisfy the requirements of the application.

2.6.4 Key Processing Techniques

Next, we describe some key processing techniques that are utilized in modern mm-Wave based radar systems. As detailed below, these are used for various purposes such as noise removal, object/people tracking, people detection and identification, vital signal estimation, etc. Note also that multiple techniques are often used concurrently as part of the same solution. By processing distance, velocity and angle information, it is possible to get two or three dimensional data points, such as *RD* (RD), *RA* (RA) or *RDA* (RDA) maps. This type of data shape, with temporal information between the data frames, can be further processed to provide valuable information about objects and users in positioning, tracking and identification applications.

micro-Doppler — In addition to the main (bulk) movement of an object, it is possible to have mechanical vibrations within the object body as well. These internal vibrations are called *micro motions*. The micro-Doppler phenomenon is observed when these micro motions from the object cause a frequency modulation on the returned signal [12]. An example for this would be the

individual movements of the legs and the arms of a person while walking, or rotations of the propellers of a fix-winged aircraft while flying. Assuming that the scalar range from the radar to an object is $r(t)$ and that the carrier frequency is f_c, then the phase of the baseband signal is defined as $\varphi(t) = 2\pi f_c \frac{2r(t)}{c}$. With this, it is possible to obtain the micro-Doppler frequency shift caused by the motion of an object. Taking the time derivative of the phase yields $\frac{d}{dt}\varphi(t) = 2\pi f_c \frac{2}{c}\frac{d}{dt}r(t)$. We manipulate this equation by introducing the Doppler shift induced by the rotation of the object and referring to vector $\mathbf{p}$ as the location of an arbitrary point on it. Thus, the micro-Doppler frequency shift equation is obtained as $f_D = \frac{2f}{c}(\mathbf{v} + \times \mathbf{p})^T \cdot \mathbf{n}$. The first term of the equation is the Doppler shift due to the object's translation $f_{\text{trans}} = \frac{2f}{c}\mathbf{v} \cdot \mathbf{n}$, where $\mathbf{v}$ is the bulk velocity of the object and $\mathbf{n}$ is the radar's line of sight direction. The second term is the Doppler shift due to the object's rotation $f_{\text{rot}} = \frac{2f}{c}(\times \mathbf{p})^T \cdot \mathbf{n}$, where is the angular velocity of the object. In order to get time-varying frequency distribution of micro-Doppler modulation, the Short-Time Fourier Transform (STFT) [219] is used. STFT is a moving window Fourier transform where the signal is examined for each window interval in order to generate a time-frequency distribution. This process can be pictured as a DFT multiplied by the sliding window's spectrum, which yields a spectrogram of time-varying micro-Doppler modulation [220], [221]. Due to the different characteristics in micro-Doppler, it is possible to detect a moving body and even to identify it, by capturing the particular modes of motions of the body parts.

Kalman filter (KF) — It is a key tool for the analysis of time-series containing noise or inaccuracies, providing a precise understanding on how the signal changes with time. The KF estimates the *state* of the monitored process through time, by removing random noise. It is commonly used in movement control, navigation and activity recognition, and it is as well widely employed in radar applications. The Discrete Kalman Filter (DKF) was initially developed in [222]. It is a two-step recursive algorithm. The first step of the recursive loop is the *prediction* step, where a projection from the current state of the model and corresponding uncertainties into the next time-step is made. Second, the *correction* (or update) step is where adjustment of the projection is made by taking the weighted average of the projected state with the measured information. In linear systems with additive Gaussian noise, DKF works as an optimal least-square error estimator. While for non-linear systems, the most common KF variants are the EKF and the UKF. One of the possible ways of obtaining state estimations in non-linear models is converting the system into a linear one. At each time step, the EKF uses a first-order partial derivative matrix for the evaluation of the next predicted state starting from the current one. This essentially forces the system to use *linearized versions* of the model in the correction step [223]. However, when the model is highly non-linear, the EKF could experience very slow convergence to the correct solution. In such non-linear models, the KF is used with an unscented transformation and hence the derivation of UKF. In order to carry out predictions, the UKF picks a finite set of points (called sigma points [224]) around the mean and generates a new mean by passing this set through the non-linear function that describes the system. Thus, the new estimate is obtained. While the computational complexity of both filters are same, for most cases the UKF practically runs faster as compared to the EKF, as it does not calculate partial derivatives [225].

In radar systems, he KF makes it possible to reliably estimate the trajectory of the targets, which is achieved by filtering the temporal sequences of points in the RD, RA or RDA maps, by identifying the Center of Mass (COM) of the moving target(s) and estimating its (their) trajectory (trajectories) over time. KF allows coping with random noise, obtaining robust trajectories, and to also estimate tracks for the targets in those cases where some temporal RD/RA/RDA frames are lost due to occlusions see, e.g., [193]. Given the sampling time of radar applications and the typical speed of movement of people, linear KF models are usually appropriate for human trajectory tracking. Also, most prior works use KF to track the COM of an object or person, treating it as an idealized point-shaped reflector.

A recent solution for mm-Wave indoor radars [194] uses an extended object tracking KF, which makes it possible to jointly estimate the COM and the *extension* of the target around it. In [194], such extension is mapped onto an ellipse around the COM, whose shape and orientation matches those of the target. This KF technology has similar performance as standard KF assuming point-shaped reflectors in terms of tracking accuracy for the COM, but *additionally it makes it possible to track the object extension over time*. In the case of human sensing, the ellipse represents the way the torso is oriented within the monitored environment. This information, combined with the target trajectory, reveals where the target is steering at, which may be a valuable information for some applications, e.g., for the retail industry.

2.6.5 Main learning techniques

Nowadays, ML and especially DL is successfully being applied to many different fields and applications. Although most of these techniques have been developed for a long time, they are recently emerging due to hardware advancements. ML methods are used for regression, classification and clustering tasks. A more comprehensive analysis and discussion of ML and DL techniques can be found in [226] and [227], respectively. Just like many other fields, these techniques are being successfully and abundantly exploited within mm-Wave radar sensing systems.

In some cases, it is required to group sets of objects into categories, i.e., to perform *clustering*. This technique is widely used in such areas as pattern recognition, image analysis and machine learning. This is particularly relevant when there are scattered data points in the observed space, and the information about which point belongs to what category is non trivial. In our setup, it is used for the analysis of radar images. After the cluster analysis, if the results are good, the clustering method could be exploited to compute labels on the dataset, and it could even be used as a part of a more sophisticated system, e.g., for a subsequent identification of the subject or of the human that has generated each data cluster within an image. Often, the clustering task is carried out in an *unsupervised* fashion. Over the years, many researchers have designed clustering algorithms tailored for a variety of models. Some of the well known of clustering algorithms are *k-means* (based on partitioning), *AUTOCLASS* (based on Bayesian statistics), *Expectation Maximization* (*EM*) (based on parametric statistics) and also unsupervised neural networks and DBSCAN [228] (density based). More on the existing clustering models and algorithms, their categorization and discussion can be found in [229].

DBSCAN — Considering the data gathered by mapping the radar signal on the environment are tightly packed points in range, angle and velocity dimensions, one algorithm stands out in the field of radar sensing, DBSCAN [228]. DBSCAN is a density-based clustering technique where the points belonging to a high density region are grouped discarding those that are recognized to be isolated, in accordance with precise definitions of the neighborhood of a point and of its local density. The algorithm starts at an initial point featuring a dense neighborhood and tags it as a *core* point. The remaining points within the *core* point's neighborhood, i.e., within a preset radius from it, are referred to as *reachable*. Upon initializing the first *core* point, DBSCAN evaluates the neighborhood density of each *reachable* points within its neighborhood, and the ones residing within a dense neighborhood are chosen as the new *core* points. The *density connected* region of the cluster is thus extended by connecting dense neighborhoods, constructing clusters of generic shapes and only containing densely connected points. This process is continued in a recurrent fashion until there are no more *reachable* points whose neighborhood exceeds the minimum density. Finally, a cluster is defined as the collection of all points that are either density-connected or density-reachable. Multiple clusters are possible and represents density-connected regions of points. Points that are not contained within a high-density cluster are referred to as *outliers* (these are termed noise points, and are rejected).

In mm-Wave based radar applications, DBSCAN has been extensively and successfully used to extract the clusters of data points in the RD, RA or RDA maps associated with the tracked humans and/or objects (e.g., vehicles) in the monitored environment [230]. This technique was found to be very robust and efficient due to the following reasons: i) most importantly, DBSCAN is an unsupervised method, its simplest version only needs two parameters to work (a density threshold and a radius for the density neighborhood), while it does not need one to know in advance the number of clusters (objects/humans) to be tracked. The DBSCAN parameters are to be set at training time and, for given hardware (mainly, working frequency, distance and angle resolution) and environment choices (e.g., indoor vs outdoor), their set values remain rather effective across a large number of scenarios [193], ii) DBSCAN runs fast, with a time complexity of $\mathcal{O}(n \log n)$, where n is the number of points to be evaluated, iii) the clusters do not have to be spherical, DBSCAN works well with clusters of any shape and it is very effective in rejecting random noise, which is quite common in radar maps. Further discussion on how DBSCAN is used in radar systems and applications from the state-of-the-art is presented in Section 2.6.6.

Neural networks — The term *neural network* (NN) comes from biological processes where a collection of neurons create a network. In the modern sense, NN are the technology counterpart of the brain. They try to achieve learning by identifying the relationships in a set of data similarly to how brain does [231]. The most basic NN is the perceptron originally devised by Rosenblatt [232]. It only has a single layer and performs a classification task based on taking the input and multiplying it by given weights, summing the resulting signals, and passing the result through a non-linear decision function. Essentially, this is the idea behind the whole DL field. Below, we talk briefly about some state-of-the-art DL architectures, which have captured

the attention of researchers working on radar sensing applications.

Convolutional Neural Network (CNN) — One of the most common NN models is the CNN [233]. CNN usually consist of back to back convolutional and pooling layers with a final fully-connected layer. The convolutional layers take the input and process it via a kernel function (a filter) where the feature detection is carried out. These feature maps are then fed to the pooling layer where dimension reduction of the domain is performed. This process is continued until a fully-connected layer, where a flattened feature map is computed and used to obtain the classification output (either via a single non-linearity or a softmax layer). CNN is a feedforward NN where information can only move in the forward direction from the input to the output layer, without cycles nor loops. While the convolution operation is naturally invariant with respect to rigid translations of input patterns, it does not work with other types of transformations, such as rotations. For this reasons, in the radar sensing field dedicated CNN-based approaches have been specifically proposed for radar point clouds, which are discussed shortly below.

Use with mm-Wave radar signals: due to the lack of mathematical models to describe RD/RA and RDA maps from mm-Wave radars and to the presence of strong noise components (e.g., from ghost reflections and metallic objects), CNN have been extensively used to automatically obtain meaningful feature representations from radar readings. Usually, CNN are applied to the RD/RA/RDA clusters found by a preceding clustering algorithm, e.g., DBSCAN, assuming that each cluster represents a target object within the monitored space. These representations can be then utilized to detect objects within an environment [234], to assess the type of activity a person is carrying out [213], or to even track their identity [193].

Recursive NN (RNN) — Unlike feedforward NN, RNN [235] utilize their internal memory to retain information from previous input samples. This allows temporal sequences to be used as input and thus the learning process can extract temporal correlation. Hence, RNN remember the information during the learning process, while feedforward NN cannot. This is especially relevant for radar data, as it makes it possible to extract temporal features from a sequence of radar maps (i.e., a trajectory). For example, such features can describe the way a person moves his/her limbs while performing a certain activity. Vanishing or exploding gradients are commonly seen during the back-propagation [227] based training of an RNN. This prevents the NN to effectively learn, leading to a premature stopping of the training process. Long-Short Term Memory (LSTM) cells, or alternatively Gated Recurrent Unit (GRU) cells [236], extend the original RNN neuron to effectively cope with vanishing or exploding gradients [237], by intelligently redefining the structure of a memory unit. This solves the gradient vanishing problems at the cost of a greater complexity.

Use with mm-Wave radar signals: activity recognition usually cannot be achieved on data coming from a single time step, e.g., from a single RD/RA or RDA map. For an activity to be determined, analysis of a sequence of such radar maps should be carried out. Combining this with the micro-Doppler effect observed in humans, it is possible to estimate the identity of a person based on the specific micro motions of their body parts [193], [194], [238].

Autoencoder (AE) — Autoencoders [239] encode the input and then decode it to generate the output. While an AE is trained to copy the input onto the output, the main rationale behind this is to learn internal representations (features) that describe the manifold where the high-dimensional input signal resides. That is, the AE features should be highly representative of the input and can be used to classify it with high accuracy. For a proper training of the AE, the encode/decode functionalities are constrained, e.g., by limiting the number of neurons in the inner layer or forcing some sparsity for the inner representation. This forces the AE network to approximate the output by preserving the *most* relevant features. Denoising autoencoders [239] are trained to denoise the input signal and reconstruct, at their ouput, the original (noise-free) signal version. This was found to produce better features in the AE inner layers. In addition, the denoising capability of such NN architectures is valuable for RD/RA/RDA radar maps due to their noisy nature.

Use with mm-Wave radar signals: radar system are prone to noisy data and can be significantly affected by unwanted or fake reflections (e.g., ghost reflections). Due to this, many radar applications use the AE encode/decode functionalities as a middle ground for the reconstruction of the desired observation such as anomaly analysis for human fall detection [217], person detection for surveillance systems [240] and indoor person identification [193].

Generative Adversarial Network (GAN) — In general, GAN [241] are divided into two sub-models called the *generator* and the *discriminator*. In the *generator* network the expected outcome is a newly generated sample which should reflect the features found in the input data/domain. Conversely the task of the *discriminator* network is to classify an input to detect whether it is a fake (generated) or a real example. Learning proceeds as a game, where the generator becomes progressively better in generating fakes, and the discriminator improves at detecting them. The final goal is again to learn meaningful representations (features) of a (usually) high-dimensional input signal.

Use with mm-Wave radar signals: Because of the competitive nature of the generator and discriminator networks, jumping from one to another during training makes them better at their respective tasks. Most often, algorithms exploit this fact to generate the required data and use this newly generated input whenever it fits. For instance, GAN have been used in [204] to generate dense maps from sparse inputs (also referred to as super resolution imaging) for the purpose of environment mapping in a low-visibility environment. In this case, the generator network is used to improve the image resolution and the discriminator to train the generator better.

Residual Network (ResNet) — ResNets [188] use shortcuts to skip layers. Typically, the skipped layers include activation and batch normalization [242]. The reason behind using shortcuts is to overcome vanishing gradients and/or gradient degradation problems. Despite the seemingly simple architectural change, this leads to a major change in terms of learning paradigm, which preserves the correct propagation of the error gradients across the whole network, allowing one to build very deep networks with hundreds of layers and with a remarkable representation (feature

extraction) effectiveness.

Use with mm-Wave radar signals: Due to the large number of layers that can be stacked, ResNets are exploited in complex scenarios where the input signal contains a high number of patterns to be concurrently classified. Examples include human skeletal posture estimation [216], where the detection of more than 15 joints and the subsequent tracking of the person are carried out, or real-time object detection for autonomous driving [234], where real time obstacle detection is performed.

PointNet and PointConv — Images are represented through dense regular grids of points, whereas point clouds are irregular and also unordered. For these reasons, using the convolution operation with them can be difficult. Pointnet [243] is a deep neural network which uses unordered 3D (graph) point clouds as input. The applications of PointNets are object classification and semantic segmentation. An extension of this network is Pointnet++ [244], where the PointNet architecture is recursively applied on a nested partitioning of the given point cloud. PointNet++ can identify local features on a greater contextual scale. The key reason of using such architecture is to make the extracted features *permutation invariant* with respect to the input signal. Along the same lines, in [245] the convolution filter is extended into a new operation, called Point-Conv, which can be effectively applied to point clouds to build convolutional neural networks. These new network layers can be used to perform translation-invariant and permutation-invariant convolutions (and obtain invariant features) on any point set in the 3D space. These qualities are especially important for radar point clouds. When tracking people or objects from radar data, being rotation invariant is relevant as the traits that we want to capture about the target (movement of limbs, body shape, etc.) do not depend on their orientation in space.

Use with mm-Wave radar signals: In the recent papers [194] and [246], novel Pointnet based NN are presented to track and identify people from point clouds obtained by mm-Wave radars. We remark that mm-Wave systems can either operate on dense radar Doppler maps, or on point clouds which can be derived from these dense maps by only retaining the most significant (strongest) reflections. Point clouds are less informative, as some information is lost when moving from dense to sparse representations, but are on the other hand easier to store, transmit and their processing is also lighter. For these reasons, algorithms that operate on sparse point clouds are particularly appealing and are gaining momentum.

2.6.6 Selected Applications

Some of the works that we review in this section adopt a custom design for the whole sensing system, from the radar hardware to the implementation of the software. Others, instead, use off-the-shelf radar devices and present new algorithms. Most of the applications deal with human activity recognition, object detection and health monitoring, but other use cases are emerging such as vibration detection, environment mapping and even speech recognition.

Human Activity Recognition

For the purpose of tracking and identity recognition of humans moving in a room, the authors of [193] use micro-Doppler signatures obtained from back-scattered mm-Wave radio signals. An off-the-shelf radar is used to gather RDA maps and noise removal is carried out. Hence, DBSCAN is applied to the pre-processed RDA maps to detect the data points (signal reflections) generated by each of the human targets in the monitored environment. With DBSCAN, a cluster of RDA points is obtained for each subject and updated as the targets move, across subsequent time steps. Trajectory detection is carried out by applying a KF to the clusters and, as a final step, identity recognition is carried out using a CNN with inception layers. The average accuracy is of 90.69% for single targets, 97.96% for two targets, 95.26% for three targets and 98.27% for four targets. Similarly, authors in [4] have designed RAPID in order to use off-the-shelf IEEE 802.11ay devices for person detection and activity recognition. Underlying techniques for human activity recognition are similar to the previous work (e.g micro-Doppler signatures, KF, CNN). However, RAPID uses CIR estimation and TRN fields to expose targets movement information from the radio signals. As a result, the authors have achieved person detection accuracy between 97.8% (for 2 subjects being the highest) and 90% (for 7 subjects being the lowest). In addition, activity recognition rates for walking, running, sitting, and waving hands are 92.9%, 71.6%, 99.8%, and 89.9% respectively.

Similarly, in [213] micro-Doppler signatures are extracted and exploited for human motion detection, where both RD data cubes as a whole, and RD point clouds are considered. The real-time information is received by passing RD data through Doppler-time extraction. DBSCAN is used to group the RD point cloud data from each of the tracked users in the monitored space. The movements of arms, torso and legs of a walking person are then identified via a CNN model. Tests were carried out for walking and leaving, waving hands, sitting to walking transition, walking back and forth, and combining all behaviors. An average accuracy of 96.32% (walking), 99.59% (leaving), 64% (waving hands), 91.18% (sitting to walking), 97.84% (walking back and forth) and 95.19% (all) was observed for each scenario, respectively.

In the same vein, movement pattern detection for of one or two patients is the key result in [215]. Together with DBSCAN, Kalman filtering has been applied to track the trajectory of each patient. Walking, falling, swinging, seizure and restless movements are the movement patterns which are classified by the proposed CNN model. For these movement types, the authors have obtained accuracy values ranging between 82.77% and 95.74%.

The authors of [209] have proposed a framework called "mmSense". It uses an LSTM-based classification model for localization. Initially, environment fingerprinting is carried out with and without human presence. Hence, the presence of people and their location within the environment are estimated using an LSTM model. Moreover, an approach combining human outline profile and vital sign measurements gathered from 60 GHz reflected signal strength series is devised to identify the targets. mmSense was tested on five people concurrently sharing the same physical space, achieving an accuracy of 97.73% for classification and of 93% for identification tasks, respectively.

With the purpose of preventive decision making in autonomous driving applications, the authors of [216] propose "mm-Pose", a model for estimating the posture of a person in real-time. To achieve this, RDA data is used to obtain 3D cloud point representations and Red-Green-Blue

(RGB) projections of depth-Azimuth and depth-Elevation are used. CNN is used to cope with noise and unwanted reflections and also to detect skeletal joint coordinates. The final model was able to locate 17 human skeletal joints with errors of 3.2 cm, 2.7 cm and 7.5 cm on the depth, elevation and azimuth axes, respectively.

A similar application is presented in [195] for human skeletal pose estimation. In this model, range-angle heatmaps are initially fed to a CNN followed by a Fractionally Strided Convolutional Network (FSCN). To exactly locate the target points, the non-max suppression algorithm was used and the obtained key points were combined, implementing and testing the proposed solution on single user scenarios. The evaluation metrics used in this work are Object Keypoints Similarity (OKS) and Mean Average Precision (AP) over different OKS thresholds. The authors obtained an average OKS of 70.5 over eight different body parts. As a comparison, camera based pose estimators achieve higher performance, i.e., Openpose (avg. OKS: 93.3) and Leave One Out (avg. OKS: 66.6).

In [217], a fall detection framework, called "mmFall" is presented. 4D cloud points are used, i.e., range, azimuth angle, elevation angle and Doppler. To perform fall detection, the authors exploit a sequence-2-sequence Hybrid Variational RNN Autoencoder (HVRAE) model that utilizes an encoder/decoder logic. They use a tailored loss function along with a simplified version (HVRAE_SL) for testing purposes. They also test the model on vanilla Recurrent Autoencoders (RAE). Overall, HVRAE achieved 98%, HVRAE_SL had 60% and vanilla RAE had 38% probability of detecting a fall.

The authors of [212] designed a system to classify static hand gestures, namely, palm facing the radar, hand perpendicular to radar and thumbs-up gesture. In addition to the real data, artificial reconstructions of the gestures were used to gather synthetic data. Tests were performed both on range and RA maps and, 85% and 90% accuracy were respectively achieved with them, while with the addition of synthetic data, the accuracy increased up to 93.1% and 95.4% for range and RA maps, respectively.

A framework for human detection and tracking by using radar fusion is presented in [201]. Ground truth data is obtained via a camera system. DBSCAN is used for clustering and temporal relationships between clusters are exploited to obtain the probability distribution of the new positions to perform tracking, similar to Kalman filtering. The concurrent use of two radars allows improving the accuracy from 46.9% to 98.6%.

The "GaitCube" algorithm was proposed in [199]. It utilizes so called gait data cubes, i.e., 3D joint-feature representations of micro-Doppler and micro-Range signatures for human recognition purposes. The idea behind this algorithm is to exploit the radar's multi-channel capabilities to improve the recognition accuracy. Their proposed system achieves 96.1% accuracy with a single antenna, 98.3% when using all antennas and an average accuracy of 79.1% when tested in an environment not seen at training time.

Akin to the objectives of the above paper, [198] develops a posture estimation algorithm using DBSCAN to cluster and single out real targets. The authors generate their dataset by installing the radar on the ceiling, and receiving data at 10 frames per second. The data processing model is based on CNN, and the CNN network is trained on lying, seated and upright moving postures.

Classification results demonstrate a mean accuracy of 99.1% and an average processing time of 0.13s.

Another work in [214] performs the classification of 7 fitness exercises. CNN and LSTM neural network models are utilized for the classification task, by training them on RD, RA, Angle-Doppler (AD) and joint-image data. For these data types, a classification accuracy of 92.08%, 98.65%, 97.7% and 99.27% is attained, respectively. In [192], fitness activities were tested both in offline and also in online scenarios. Classification was performed on 5 human activities achieving an accuracy of 93.25% and 91.52% for the offline and online operation modes, respectively. The system was also tested on multiple locations and the obtained average accuracy is 88.83%.

The authors of [201], [218] have created a human detection and tracking algorithm by using two radars simultaneously. In both of these works, Kalman filter and DBSCAN were used for tracking and identifying the location of the person, and the synchronization of the radars were carried out in an offline fashion. The results in [201] show a significant improvement when a two-radar setup was used with an accuracy of 98.6% compared to 46.9% from single-radar setup in human detection. In [218], radars were placed so that one had a top-view and the other had side-view angle. This work in addition to prior work proposes an alarm system and a posture estimation method. An alarm system is triggered upon a positive evaluation in the change of cluster number, number of points in the cluster or the center point of the cluster. The posture estimation is done for standing, sitting and lying poses by analysing the height of a person at a particular instance and the accuracy of estimation is from 92.5% to 93.7%.

Towards performing human activity recognition, any combination of range and Doppler (or in some cases of range, Doppler and angle) is used. RDA is typically used with DBSCAN and/or Kalman filtering to identify the clusters within the environment. After extracting micro-Doppler data, a NN architecture (i.e., CNN, RNN, AE etc.) is employed to perform activity/sensing applications. If properly designed, DL models are generally the preferred way to identify movement patterns of RDA clusters, as this is the common theme among most of the surveyed material above. Deterministic algorithms often fail to provide good performance due to the need of a careful parameter tuning (which is very sensitive to the monitored scenario) and to the lack of mathematical models that accurately represent the signals at the receivers.

Object Detection

The authors of [234] propose a method called Spatial Attention Fusion (SAF) for obstacle detection with mm-Wave radar and vision sensors. A Fully Connected One-Stage (FCOS) NN is used for the detection of objects. For the training of this neural network, radar data is converted into radar maps (images) and during the feature extraction step, the SAF block within the FCOS network is used for combining radar with vision features. The proposed SAF-FCOS model is trained and tested on the nuScenes dataset with a ResNet-50 backbone, achieving an average precision of 90.0% with an intersection-over-union of 0.50 or higher.

The detection of concealed objects implies additional challenges, as it becomes necessary to single out hidden objects from rest of the scenery. In [247], the authors employ EM to fit a Gaussian mixture model of the image acquired: through a two-step image segmentation procedure,

they first extract the body area from the image and then detect concealed objects. The model is evaluated in terms of average probability of error and the authors report that multi-level EM has an increased performance of up to 90.0% when compared to conventional EM.

A real-time outdoor hidden object detection model is proposed in [248]. This work also utilizes EM, Bayesian decision making and Gaussian mixture model for image segmentation, with an architecture similar to that of [247]. However, vector quantization is adopted for the first segmentation level to achieve faster computation times. The authors also state that EM can be avoided as a whole to reduce computation time (and complexity) significantly, but this causes an error increase as well. As a result, [248] achieves a computation time of 1.11 s (with EM) and 0.134 s (without EM) per frame.

Along a similar line, the authors of [249] propose a writing object (e.g., a pen) tracking system called "mTrack," that uses dedicated mechanisms to suppress interference from background reflections. After this, the RX antenna is steered according to the peak response observed on the reflected received signal. In other words, the antenna is adaptively steered to face and track the direction of the pen. Finally, the target movement tendency is evaluated by the trend and amount of phase shifting. The system can detect the location of the pen at a 94% accuracy, with a tracking error smaller than 10 mm across 90% of the trajectory.

In [250], a non-imaging sensor for hidden object detection is developed. The authors test their device both in an outdoor scenario with a gun and in an indoor scenario where plastic sticks of diameter equal to 2 cm are covered by a fabric. Final results of the model are evaluated by applying the Fourier transform to IF chirps to get the range map on horizontal and vertical scans of the environment compared with a captured image. In [251], an improved version of the sensor is proposed, using a horn antenna integrated with a focusing dielectric lens operating in the 80–100 GHz frequency range. This sensor can be operated with any preferred movement (e.g., up-down) and the authors claim that the probability of detection can go up to 100%.

In [197], an IEEE 802.11ad device is used as a pulsed wave radar to perform passive handwriting tracking. Slow- and fast-time dimension analysis of the complex CIR, Cell-Averaging Constant False Alarm Rate (CA-CFAR) and Subsample Peak Interpolation (SPI) are the underlying techniques used in their algorithm. After applying digital beamforming, the authors could extract Doppler maps and by choosing the bins with higher Doppler power, could localize the writing tool (a pen). Finally, the pen is tracked by picking the lowest elevation angle of its lower part at each time-step. With this, they achieved a tracking accuracy between 3 mm (at a distance of 20 cm) and 40 mm (at a distance of 3 m) and a character recognition accuracy ranging from 72% to 82%.

The authors of [252] perform object classification considering three classes: humans, drones and cars. The feature set used in their algorithm consists of radial range, area under the peak, width of the peak, height and standard deviation of the peak in the range-FFT domain. Logistic regression and Naive Bayes led to a classification accuracy of 86.9% and 73.9%, respectively.

In [202], authors have developed a new deep learning model called Hybrid Dilated Convolution (HDC) for concealed object detection. HDC uses two-class semantic segmentation network for keeping a high resolution in order to detect small objects. As a design rule and assign-

ment strategy, Expand-Contract Dilation (ECD) assignment is applied. In this assignment stage, the first dilation rate group forms the "rising edge" (increasing dilation) and the rest forms the "falling edge" (decreasing dilation) of dilation rates. As a result, their average precision with intersection over union of 0.5 is at 0.69% which outperforms rest of the existing techniques.

As it may be apparent from our discussion, a wide variety of algorithms have been used for object detection. Initially, signal processing with DFT or FFT is performed to distill signal features. Next, such features are either converted into images such as radar maps, or further data processing is applied, e.g., CIR analysis. ML and DL methods, or decision making algorithms such as EM, are then applied to obtain the final results. In general, there is not a single winning methodology. Rather, the optimal approach depends on hardware, software, environment and application limitations.

Health Monitoring

The authors of [253] propose a model for remote Heart Rate (HR) monitoring and analysis. They use the Levenberg-Marquardt (LM) algorithm for the extraction of heart-rate information. The sum of heartbeat complex, respiration, body movement, background noise, and electronic system noise is gathered by expressing the received in-phase and quadrature-phase components from LM as the cosine and sine of the received signal. One distinctive advantage of this method is that there is no phase unwrapping as the fitting of the HR signal is directly carried out on the cosine and sine of the received phase modulated signal, which is important for low SNR scenarios. The method is able to estimate beat-to-beat HR and individual heartbeat amplitude, both having a critical role in the diagnosis of heart diseases.

The authors of [211] demonstrate a remote breathing and sleep position monitoring system over multiple people at the same time. High resolution AoA detection is used to identify closely located targets. A Support Vector Machine (SVM) is used for finding the sleep position, and an optimal filter to estimate the breathing rate. The designed system achieved an accuracy of 97% for breathing rate estimation and of 83% for sleep position detection.

In the same vein, the work in [210] proposes vital sign and sleep monitoring system. Initially, the location of the person is detected by using the reflection loss as the classification parameter, performing a 360° sweep of the environment. After localizing the human, reflected signal strength samples from the reflected signal directed at the human are gathered. For heart rate detection, FFT, customized band-pass filter, Inverse FFT (IFFT) and peak detection are applied, while for breathing rate detection only IFFT and peak detection were sufficient. The achieved accuracy was 98.4% and the mean estimation error in breathing rate and heart rate estimation for an incident angle of 70° was smaller than 0.5 bpm and 2.5 bpm, respectively.

A similar purpose is found in [200], which designs a robot for human detection and heart rate monitoring. The Hungarian Algorithm and Kalman filtering are used to detect and track the user position. Once the person is located, the robot approaches him or her and starts the scanning process. The biquad cascade Infinite Impulse Response (IIR) filter is used to extract the heartbeat waveforms from the signal, whereas a NN is used for predicting the heart rate. The proposed system achieved an accuracy between 91.08% and 97.89% across eight different poses.

For the purpose of remote glucose level monitoring, the authors of [206], [207] observe reflected multi-channel signal signatures collected through the SOLI mm-Wave sensor [254]. The signal is analyzed by obtaining average Power Spectral Density (PSD) of each gated signal vector by applying DFT and FFT. With this, they were able to sense the change in dielectric constant due to a varying glucose level in the blood.

The authors of [203] use the radars' multi-channel capabilities to improve the estimation of vital signs (heart rate). Experiments are performed on 4 different scenarios by changing the location of the radar and the posture of the subject. Authors claim that using multi-channel Maximal Ratio Combining (MRC) outperforms single channel estimates in most cases, quantifying the benefits for each scenario.

Although an increasing number of articles is appearing on health monitoring via mm-Wave radars, this field of application deserves significant additional work. In fact, prior art presents results for rather ad-hoc and artificial scenarios, where people are still, positioned at known locations, etc. A fully automated monitoring system should instead operate in free living conditions, where users are free to move and no prior location information is available. Further research is thus needed to enable multi-user tracking of vital signs, by also compensating for people movement, which has a detrimental effect on the estimation of breathing and heart rate.

Other Applications

In [196], a system namely "mmVib" for micrometer level vibration detection is presented. The authors propose a multi-signal consolidation model to understand In-phase and Quadrature (IQ) domain and in turn exploit the consistency among the two obtained signals to estimate the vibration characteristics of an object. With this, they can detect vibrations at micrometer level.

The authors of [204] propose an indoor mapping system called "milliMap", designed for low-visibility environments. A lidar is used for environment mapping as a ground truth data collector. A GAN is used to construct the grid map by recognizing obstacles, free spaces and unknown areas. Finally, semantic mapping is applied for the classification of obstacles.

A noise-resistant speech sensing framework, "WaveEar," is proposed in [255]. Directional beamforming is used to make the system robust to noise. After localizing the throat and receiving the data, voice reconstruction is achieved by a neural network based on an encoder-decoder (autoencoder) architecture. As a result, WaveEar achieves a stable 5.5% word error rate at a distance of about 1.5 m from the user. The authors also point out that joint optimization speech-to-text and WaveEar would further enhance the capabilities of their system.

In [9], a mm-Wave radar device is mounted on a robot to estimate its position. This is achieved by exploiting the interference produced by other mm-Wave radars located in the same environment (with known positions), and by only estimating the angle of arrival of each other radar interference. The proposed system attains position errors for the robot ranging from 14 cm (with three radars) down to 6 cm (ten radars).

The applications presented in this section vary from micrometer-level activity recognition to speech recognition. We observe that radio sensing enables new and unforeseen use cases, such as vibration detection [196], indoor navigation [204] and speech reconstruction [255]. However, it

Table 2.7: Summary of the environments in which the experiments have been carried out

Evaluation	Related literature
Indoors	[4], [9], [192], [195]–[204], [206]–[213], [215], [217], [218], [247], [249]–[251], [253], [255], [256]
Outdoors	[195], [214], [216], [234], [248], [250]–[252]

Table 2.8: Summary of the main techniques used in the surveyed papers

Analytical Tools	Related Literature
DBSCAN	[192], [198], [201], [208], [213], [218]
Deep learning	[4], [192], [195], [198]–[201], [204], [208], [209], [212]–[217], [234], [255]
Fourier transform	[9], [196], [197], [206], [207], [210], [212], [250], [252]
Hungarian algorithm	[200], [208]
Kalman filter	[4], [200], [208], [218]
Levenberg-Marquard method	[253]
Machine learning	[198], [211], [252]
Non-max suppression	[195]
Signal processing	[4], [197], [203], [209], [210], [249], [255]
Statistical modeling	[201], [247], [248]

is still unclear whether these signals can be reliably detected in an environment with mobility and other sources of noise. Additional experimental data would be required to check the performance of these solutions in general settings and to possibly improve their robustness.

2.6.7 Summary

In this section, we have summarized the recent advances and trends in signal processing for passive mm-Wave radar systems for indoor spaces. These systems are rapidly gaining momentum as radar devices become commercially available, at a low cost. A number of applications are emerging, targeting diverse scenarios such as people detection, tracking and identification, estimation of biosignals such as respiration and heart rate, detection of gestures/activities/falls, vibrations, speech or environmental mapping. Table 2.9 summarizes these application-oriented propositions, while Table 2.7 categorizes them based on the environment where the experiments were carried out. While early works used standard machine learning algorithms such as expectation maximization and support vector machines, latest developments have been dominated by neural networks. This is clearly evident from Table 2.8, which presents a summary of the analytical tools discussed in the survey. These are being implemented in their many flavors, and are allowing researchers to obtain good results in scenarios where no analytical models are available. As far as human data monitoring is concerned (e.g., people tracking, activity monitoring, etc.), the key processing algorithms are DBSCAN clustering for the separation of user data in the radar RD/RA/RDA maps and Kalman filtering to reliably track their trajectories. Neural network architectures are evolving from standard CNN to more advanced convolutions (PointConv and PointNets) that were specifically designed for radar point clouds. Some solutions then use RNN to capture and exploit the temporal correlation of radar signals. Advanced architectures, such as GAN based, are also being exploited to extract features from radar images.

Although many applications and uses of this technology have emerged lately, a lot of research and implementation work is still required. As far as research is concerned, vital sign monitoring is still in its infancy as more robust algorithms are to be developed, capable of working in free living conditions, i.e., in the presence of user mobility and other noise sources. In addition, while advanced user tracking and positioning techniques are available for single radar systems, no substantial work can be found for multi-radar setups, i.e., *radar networks*. With multiple radar devices, many additional problems have to be tackled, including time synchronization, data fusion among radar signals, distributed calibration and means to quantify whether and to which extent radar devices share a common portion of their field of view. For what concerns implementation, much work still has to be performed architecturally, e.g., where to place the ML based intelligence, which messages are to be exchanged between the radars and the computing units, which protocols are to be exploited to synchronize multiple devices along time and data dimensions, etc. Lastly, experimental work is key to the development of robust algorithms, as analytical or simulated models often fail to accurately represent all the noise sources. Hence, the collection of experimental data and its publication along with the code of the developed solutions are vital to make progress.

Table 2.9: Summary of the mmWave radar sensing works in the literature

Proposition	Tools Used	Band (GHz)	Performance
Human Activity Recognition Algorithms			
Multi-person tracking and identification [208]	Micro-Doppler, DBSCAN, Kalman filter, Hungarian algorithm, CNN	77	Continuous identification of multiple persons with up to 98% accuracy.
Indoor human detection and sensing [4]	CIR, micro-Doppler, Kalman filter, CNN	60	Person detection accuracy of 97.8% to 90%. Walking, running, sitting and waving hands accuracy of 92.9%, 71.6%, 99.8% and 89.9% respectively.
Multi-person detection and identification [209]	LSTM-based model, RSS series analysis	60	Multi person classification and identification accuracy of 97.73% and 93% respectively.
Gait-based human recognition [199]	CNN	77	Classification accuracy of 96.1% and 98.3% with single gait cycle, when using single and all receive antenna respectively.

Table 2.9: SUMMARY OF THE MMWAVE RADAR SENSING WORKS IN THE LITERATURE (CONTINUED)

Proposition	Tools Used	Band (GHz)	Performance
Human detection and tracking [201]	DBSCAN, probability distribution matching, Kalman filter-like algorithm	77-81	Human detection sensitivity and precision of 90% and 98.6% respectively.
Real-time human activity recognition [192]	DBSCAN, CNN, RNN	77	Offline and real-time activity recognition accuracy of 93.25% and 91.52% respectively, over five different human activities.
Hand gesture classification [212]	Deep learning, Signal processing (FFT)	77-81	Hand gesture classification accuracy of 93% and 95% on range and range-angle data respectively.
Human motion behavior detection [213]	Micro-Doppler, DBSCAN, CNN	77	Accuracy of over 90% in detecting various human motion behaviours.
Activity recognition and fitness tracker [214]	Deep learning, CNN	77-81	Classification accuracy of 92.08%, 98.65%, 97.7%, and 99.27% for RD, RA, Angle-Doppler (AD), and joint-image evaluation respectively.
Real-time patient behaviour detection [215]	Micro-Doppler, STFT, CNN	77	Over 84.31% prediction accuracy for different behaviors for a single patient. Around 80% prediction accuracy for different behaviors for two patients.
Human skeletal pose estimation [216]	CNN	77	Detection of 17 human skeletal joints with 3.2 cm, 2.7 cm and 7.5 cm localization error on depth, elevation, and azimuth axes respectively.
Human pose estimation [195]	CNN, Fractionally strided CNN	77	Average object keypoints similarity of 70.5 over 8 different parts.
Fall detection system [217]	LSTM, RNN	77	Proposed scheme achieves 98% fall detection rate and outperforms the baseline techniques.

Table 2.9: SUMMARY OF THE MMWAVE RADAR SENSING WORKS IN THE LITERATURE (CONTINUED)

Proposition	Tools Used	Band (GHz)	Performance
Real-time posture estimation system [198]	DBSCAN, CNN, LSTM, Decision trees	77	Posture estimation with an accuracy of 99.1% at a processing time of 0.13s
Human detection and tracking [218]	DBSCAN, Kalman filters	76-81	Human detection sensitivity of over 90%. Two-radar setup improves precision from 46.9% to 98.6%. Posture estimation precision from 92.5% to 93.7%
Object Detection Algorithms			
Handwriting tracking [197]	STFT, CIR, Cell averaging-constant false alarm rate (CA-CFAR)	60	Tracking accuracy of 3 mm to 40 mm and character recognition accuracy of 72% to 82%.
Obstacle detection for autonomous driving [234]	Deep learning, CNN	77	Average precision of 90% with intersection of unions greater than 0.5.
Concealed object detection [247]	Gaussian smoothing filter, expectation-maximization, Bayesian	37.47	Usage of multi-level EM increased performance up to 90% compared to conventional EM.
Real-time concealed object detection [248]	EM, Bayesian decision making, Gaussian mixture model	94	Computation time of 1.11 s and 0.134 s with reduced processing.
Writing object tracking (mTrack) [249]	RSS, phase change analysis	60	The system tracks/locates a pen with sub-centimeter accuracy in 90% of the cases.
Concealed object detection [250], [251]	FFT	80-100	Object detection accuracy up to 100%.
Object classification [252]	FFT, Logistic regression, Naive Bayes	77-81	86.9% and 73.9% classification accuracy using Logistic Regression and Naive Bayes respectively.

Table 2.9: SUMMARY OF THE MMWAVE RADAR SENSING WORKS IN THE LITERATURE (CONTINUED)

Proposition	Tools Used	Band (GHz)	Performance
Hidden object detection [202]	Semantic segmentation, CNN	60	The proposed expand-contract dilation (ECD) scheme has an average precision (AP@0.5) of 0.69, and outperforms all the existing techniques.
Health Monitoring Algorithms			
Blood glucose level monitoring [206]	DFT, FFT	57-64	Remote detection of blood glucose levels by sensing the change in dielectric constant and loss tangent.
Glucose level detection [207]	Energy-density comparison, DTFT	57-64	Demonstrates accurate identification of blood glucose levels.
Vital sign and sleep monitoring [210]	RSS, IFFT	60	Human finding accuracy of 98.4% and the mean estimation error in breathing rate and heart rate is less then 0.43 Bpm and 2.15 Bpm.
Breathing and sleep position monitoring [211]	FFT, DOA, optimum filter, SVM	77-81	Accuracy of 97% and 83% for breathing rate estimation and sleep position detection respectively.
Vital sign monitoring [203]	Arctangent demodulation (AD), Maximal ratio combining (MRC)	77-81	Proposed signal processing chain significantly improves the heart rate estimation accuracy in all cases.
Heart rate sensing [200]	Neural networks, Hungarian algorithm, Kalman filter	60-64	Accuracy of 91.08–97.89% over 8 different human poses.
Heart rate analysis [253]	Non-linear Levenberg-Marquardt	94	Capability of estimating beat-to-beat heart rate and individual heartbeat amplitude.
Other Algorithms			
Indoor mapping [204]	GAN	77	Map reconstruction error within 0.2 m. Obstacle classification accuracy of 90%.

Table 2.9: Summary of the mmWave radar sensing works in the literature (continued)

Proposition	Tools Used	Band (GHz)	Performance
Vibration detection [196]	FFT, AoA	77	Median amplitude error of 3.4 μm for the 100 μm amplitude vibration.
Robot position estimation [9]	AoA, Range-Doppler FFT	77	Position estimation of the robot with an error below 20 cm.
Speech sensing [255]	Speaking SNR, Neural Network	77	5.5% word error rate around 1.5 m distance

2.7 Discussion and open research directions

Our comprehensive review of the state of the art in mm-Wave localization and sensing shows that a sizeable set of contributions have already covered significant work in this research area. Such works show that current mm-Wave equipment, even COTS devices, already offer sufficient opportunities to incorporate localization as part of communication processes. Moreover, commercial implementations of mm-Wave radars are currently very compact, and cater for precise device-free localization and sensing. However, additional efforts are required to democratize these tasks and make them natively available to vertical applications that rely on mm-Wave connectivity.

At the current stage of hardware development, fully-custom signal processing algorithms only apply to software-defined radio platforms, where fully-digital transceiver architectures can be available upfront. Conversely, commercial-grade hardware does not give full access to internal signal samples and measurements, requiring more complex processing and often yielding limited performance. For example, while theoretical analysis predicts millimeter-level device localization accuracy and fully digital architectures achieve centimeter-level accuracy, algorithms for commercial-grade mm-Wave devices typically achieve decimeter-level 3rd-quartile localization errors.

In this perspective, we conclude that promising research directions in the above field would greatly benefit from new-generation standard-compliant mm-Wave transceiver hardware that also exposes channel state information to external algorithms. While some efforts in this direction have been announced, there is still no such platform available on the market. The same observation holds for hybrid beamforming architectures. While preliminary works exist that exploit hybrid beamforming to improve beam pattern directivity and adaptivity, or to make the 802.11ay SLS operations faster, these architectures could also help localize mm-Wave devices faster, e.g., by enabling faster angular spectrum scanning. Moreover, the field still needs scalable algorithms that flexibly manage the presence of multiple AP or of several clients in the same area. These algorithms should work, if possible, with zero initial knowledge of the floor plan and surrounding area, and ideally estimate the whole environment, including the location of the AP and of all reflective surfaces automatically in a SLAM fashion, in order to relieve the need of input from

the user. Significant research opportunities also exist for integrating ML algorithms into location systems. Here, the main challenges relate to: relieving the need for extensive training datasets, whose collection requires important efforts; creating models that transfer well across different environments, especially indoors; speeding up the convergence of the trained models, e.g., through federated learning, particularly when involving heterogenous clients.

All of the above would be important enablers of a fully integrated device-based sensing and localization system, for which significant research is still needed. The benefits of such a system would be enormous, as the seamless integration of device-based localization and communications would enable advanced location-based services in multiple domains (including but not limited to healthcare, massive IoT, industrial scenarios, safety, and mission-critical applications), as well as multiple network optimizations (such as optimal client-AP associations, predictive re-association before link breakage due to movement or obstacles, or location-aided beam training and tracking).

Regarding passive radar sensing, a number of major advancements are envisaged. First, most commercial low-cost radars incorporate linear antenna arrays, which have limited detection and tracking capabilities. Bi-dimensional antenna arrays would make it possible to detect higher resolution radar images in the 3D space, enabling new uses of this technology (e.g., human gait analysis). Even though commercial mm-Wave radars with enhanced capabilities and 2D antenna arrays are becoming available, e.g., TI AWR/IWR radars [257] and TI cascaded imaging radar MMW-CAS [258] with relatively large antenna array size, very little work is available to date exploiting massive MIMO radars. These would allow high resolution sensing, which makes it possible to detect finer movements and shapes. Also, most of the available research only involves a single radar sensor, whereas radars could be as well co-deployed, allowing for large-scale monitoring applications. This will give rise to new opportunities and technical challenges to face, such as new techniques to perform sensor fusion from multiple radar views, self-calibration algorithms for the distributed radar sensors, transmission and compression of radar features from multiple sensing units. Architecturally, no clear approach was found on where the supporting computing facilities are to be located, which messages should be sent to them and what is the preferred interaction model between the field sensors and the computing units. All of this is of key importance especially for large deployments involving multiple sensors. Additional opportunities concern the combination of mm-Wave radar systems with camera-based ones (including thermal cameras), to perform data/feature extraction and fusion across different sensing domains. Finally, a promising research avenue is to modify commercial off-the-shelf communication technology, such as the forthcoming IEEE 802.11ay, so that it can double as a passive mm-Wave radar. This would enable joint communications and passive sensing, potentially without having to deploy a dedicated mm-Wave radar network. The recent creation of the TGbf task group (working on research and standardization of WLAN sensing towards the IEEE 802.11bf amendment) testifies the interest of the community on these emerging topics.

As a general observation, the research on machine learning methods applied to device-based localization remains limited compared to device-free radar-based sensing. For device-based localization, machine learning methods find their typical application in fingerprinting approaches. Yet, these schemes require a typically lengthy preliminary measurement effort, which is often

deemed excessive or impractical. Conversely, modern mm-Wave radar systems are both compact and affordable, and expose a number of features that can be more easily passed on to complex learning and clustering algorithms to map environments, track movement, or estimate the occurrence of some events of interest. The applicability of machine learning algorithms to to either field could change if more features become available, e.g, from multiple digital transceiver architectures integrated in the same client. For example, this would make it possible to use machine learning to increase the speed of intermediate localization algorithm steps (e.g., angle computations, ranging and simultaneous distance estimation among multiple mm-Wave devices, or joint angle/distance estimates based on radio features).

2.8   Concluding Remarks

mm-Wave communication devices will soon become a fundamental component of 5G-and-beyond communication networks. This survey put the lens on recent research advances in localization and sensing algorithms for indoor mm-Wave communication and radar devices. After introducing the most important properties of mm-Wave signal propagation and communication chain architectures that enable mm-Wave channel measurements, we presented a thorough account of localization algorithms for mm-Wave devices. These are based on a broad range of techniques, that include both traditional methods based, e.g., on timing and RSSI information, and more specific methods that exploit the properties of mm-Wave devices and signal propagation, e.g., by processing CSI.

Then, we turned our attention to consumer-grade mm-Wave radar devices, which are becoming extremely cost-effective sensing platforms. After introducing the basic structure of such radar architectures, we discussed different approaches that tackle applications such as human activity recognition, object detection and health monitoring. We unveiled that several research directions remain open in both fields, including better algorithms for localization and sensing with consumer-grade devices, data fusion methods for dense deployments, as well as an educated application of machine learning methods to both device-based localization and device-free sensing.

3

Deep Learning for Accurate Indoor Human Tracking with a mm-Wave Radar

3.1   Introduction

Radar devices for indoor environments are gaining a growing interest. Recent studies have demonstrated the possibility of exploiting the properties of the reflected radar signal to infer the presence, position, and activity of human targets in indoor spaces [238], [259], [260]. This approach is a sound alternative to traditional camera-based sensing systems, as it preserves the privacy of the users, i.e., no visual representation of the scene is collected, and is robust to poor light conditions [261]. The use of radio waves in the Millimeter-Wave (mm-Wave) frequency band allows the estimation of the target distance with high resolution, in the order of a few centimeters. Although this increased sensitivity makes mm-Wave prone to disturbances and clutter effects from the radio environment, the use of data-driven Deep Learning (DL) methods has recently emerged as a viable solution to these problems, enabling person identification [259] and activity recognition [260] tasks.

In the present article, we focus on the problem of tracking people as they move within an indoor environment, using the backscattered signal from a mm-Wave Frequency-Modulated Continuous-Wave (FMCW) radar. Our aim is to obtain accurate positioning information of the targets in the physical space. So far, the few available solutions to this problem [238], [262], have relied on a Bayesian approach using the Extended Kalman Filter (EKF) method [263]. Kalman filter (KF), however, is suitable for systems that follow a linear evolution model with Gaussian noise. The extension to the non-linear and non-Gaussian case (i.e., the EKF or the Unscented Kalman Filter (UKF), [263]) is often problematic, especially in highly non-linear models.

In this work, we alternatively use Neural Network (NN) architectures to sequentially estimate

71

the location of human targets in indoor spaces: we leverage Denoising Autoencoders (DAEs) [239] and Sequence-to-Sequence Autoencoders (S2SAEs) [264] to sequentially learn the best parameters from the data, not requiring any preliminary assumptions on the nature of the system evolution, nor on the noise process. S2SAE architectures, moreover, are capable of modeling long time dependencies.

Our main contributions are summarized next.

1. We propose two novel deep learning architectures for the task of tracking human targets in indoor spaces with a mm-Wave FMCW radar, based on a DAE and a S2SAE, respectively. The average tracking error is as low as 0.12 m for the single target case and 0.21 m for the multi-target one.

2. We evaluate our position tracking system on a challenging and realistic dataset collected in a room including furniture, metallic objects, and other people, emulating real-life conditions.

3. We train the proposed tracking system in supervised and unsupervised manners. For the former, the ground truth positions of the targets are provided at training time, while in the latter only the radar measurements are used. In both cases, our approach outperforms Bayesian methods such as EKF and UKF under several metrics.

The rest of the article is organized as follows. In Section 3.2, the FMCW radar signal model and the detection procedure are described. In Section 3.3 we present the tracking problem discussing the novelty of our approach. The signal processing workflow is presented in Section 3.4. In Section 3.5, experimental results are discussed, while concluding remarks are given in Section 3.6.

3.2 FMCW radar model

A Multiple-Input Multiple-Output (MIMO) FMCW radar allows the joint estimation of the distance, the angular position and the radial velocity of the target with respect to the radar device. This is achieved by transmitting sequences of *chirps*, i.e., sinusoidal waves with frequency that varies in time, and measuring the frequency shift of the backscattered signal at the multiple receiver antennas.

In this chapter, we use a Linear Frequency-Modulated Continuous-Wave (LFMCW) radar (the INRAS RadarLog) with one transmitting antenna and $M = 16$ receiving antennas. The frequency of the Transmit/Transmitter (TX) chirp signal is linearly increased from a base value of $f_o = 77$ GHz to a maximum $f_1 = 81$ GHz in $T = 180$ μs. We define the bandwidth of the chirp as $B = f_1 - f_o = 4$ GHz. The sinusoidal signals are transmitted every $T_{\text{rep}} = 250$ μs in sequences of $P = 256$ chirps each.

At the receiver, for each of the 16 antenna elements, a mixer combines the Receive/Receiver (RX) signal with the transmitted one, generating the Intermediate Frequency (IF) signal, i.e., a sinusoid whose instantaneous frequency amounts to the difference between the frequencies of the TX and RX signals.

Then, the IF signal is sampled along three different dimensions. First, *fast time* sampling allows obtaining $N = 1024$ points from each chirp. For the *slow time* sampling, P samples, one per chirp from adjacent chirps, are taken with period T_{rep}. Finally, the *spatial* sampling is related to the M receiving channels, spaced apart by a distance d, and enables the localization of the targets in the physical space.

A Discrete Fourier Transform (DFT) is applied along each sampling dimension to extract the frequency components. In the resulting signal, referred to as Range-Doppler-Azimuth (RDA) map, the position of the peak along the fast time dimension reveals the *beat* frequency of the IF signal, f_b, and the peak along the slow time gives the Doppler frequency, f_d. The DFT along the spatial dimension reveals the phase shift due to the angular displacement of the target, f_a. Denoting the speed of light by c, the distance, velocity and angular position of the target by R, v and θ, respectively. Their estimates are [265]

$$\hat{R} = \frac{f_b T c}{2B}, \quad \hat{v} = \frac{f_d c}{2f_o}, \quad \hat{\theta} = \sin^{-1}\left(\frac{f_a c}{2\pi d f_o}\right). \tag{3.1}$$

Using the parameters described above, our radar has a nominal range resolution of 3.75 cm.

3.3 Problem outline

3.3.1 Notation

We consider a discrete time system, where time steps have a fixed duration of Δt seconds, corresponding to the radar frame period. Boldface lowercase letters refer to vectors, e.g., $\boldsymbol{x}$. Symbol $\odot$ denotes the elementwise product between vectors and $\boldsymbol{x}^T$ denotes the transpose of vector $\boldsymbol{x}$. The concatenation between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is denoted by $[\boldsymbol{x}, \boldsymbol{y}]$. A sequence of vectors in subsequent frames from time n to time m is indicated with the subscript $\boldsymbol{x}_{n:m}$. $||\boldsymbol{x}||$ is the Euclidean norm of vector $\boldsymbol{x}$.

3.3.2 Tracking

Indoor person tracking can be described as *sequentially* estimating the position of a person as she/he moves in the environment, tracing the trajectory of the subject and possibly predicting her/his future location [266]. We denote by $\boldsymbol{x}_k = [x_k, y_k]^T$ the state of a target at time step k, with x and y being respectively the coordinates along the reference system axes, centered in the radar device location. The information provided by the radar at step k is called an *observation* and is referred to as $\boldsymbol{z}_k = [R_k, \theta_k, v_k]^T$, where R_k, θ_k and v_k are respectively the distance, the angular position and the radial velocity of the target with respect to the device at time k. A key problem in the tracking procedure is the *filtering*, i.e., the estimation of the current (unknown) target state $\boldsymbol{x}_k$ given a sequence of present and past observations $\boldsymbol{z}_{1:k}$.

Non-linear Bayesian tracking methods, e.g., EKF and UKF, provide an approximate solution to this filtering problem using the assumptions that (*i*) the state at time k only depends on the

state at time $k-1$ (first-order Markov assumption) and (*ii*) the observations are conditionally independent of one another, given the corresponding state value at a specific frame (conditional independency assumption). Bayesian tracking requires one to specify a transition and observation model, in the form of the probability distributions $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ (modeling transitions) and $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$ (observations). The model is used in a recursive fashion, starting from a prior $p(\boldsymbol{x}_0)$, to form a predictive distribution on the state at the next time step as [266]

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1})d\boldsymbol{x}_{k-1}, \tag{3.2}$$

and then compute the filtering distribution as

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})}{\int p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})d\boldsymbol{x}_k}. \tag{3.3}$$

From Eq. (3.3), an estimate of the current state $\hat{\boldsymbol{x}}_k$ can be obtained using, for instance, the Minimum Mean-Square Error (MMSE) criterion, returning the expected value of the state under the filtering distribution.

In contrast with the above approach, in this chapter we use NNs capable of modeling long term and non-linear dependencies among the samples of a time sequence, approximating the posterior filtering distribution in Eq. (3.3) and the state estimates through a deterministic function $\hat{\boldsymbol{x}}_k = F_\vartheta(\boldsymbol{z}_{1:k})$. This function is parameterized by the NN weights, ϑ, which are learned directly from the data (observations) by minimizing a loss function, and without requiring the definition of a system model. We stress that no assumptions, neither on the nature of the system (e.g., first order Markov process for the state evolution), nor on the noise distribution (e.g., Gaussian noise) are necessary. In particular, we use two types of NN architectures that have recently become the state-of-the-art in time series analysis and forecasting, i.e., the Denoising Autoencoder and the S2SAE with Long-Short Term Memory (LSTM) NN.

3.3.3 Neural networks for time series analysis

In the following, a brief introduction on LSTM networks and DAEs is provided.

Denoising autoencoder

The Autoencoder (AE) is a model trained to encode the input into a lower dimensional or sparse representation (called the *code* or *latent* representation). Hence, the code is inputted into a decoder, from which the input is reconstructed with the lowest possible error. In order to learn more meaningful latent representations, *denoising* AEs are often employed, where the training input is corrupted with artificial noise before being fed to the DAE [239]. We implement the DAE using a cascade of two Fully Connected (FCon) NNs: the first acts as an encoder by extracting relevant features (code) form the input, the second performs the decoding by reconstructing the original input starting from the code. The encoder and the decoder are *jointly trained* to minimize the
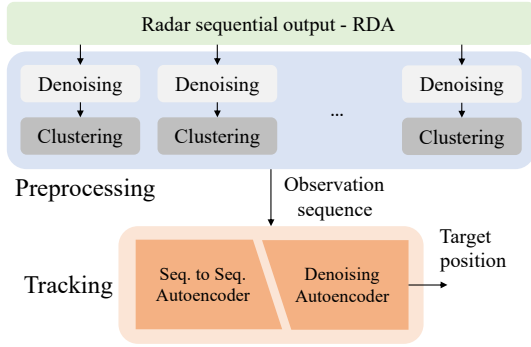
Figure 3.1: Proposed workflow.

(a) Denoising Autoencoder

(b) Sequence-to-Sequence Autoencoder
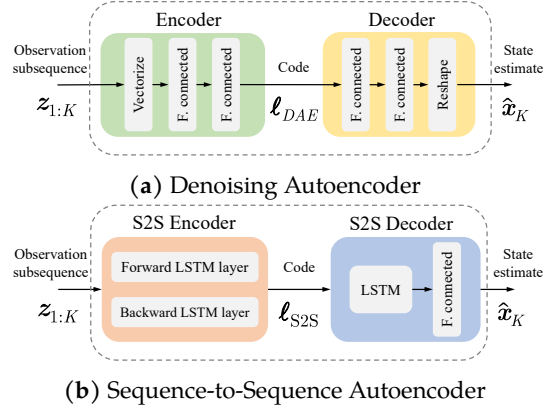
Figure 3.2: Proposed DAE and S2SAE architectures.

reconstruction error. The encoder-decoder structure of the DAE is shared by the S2SAE model (see Section 3.4.2). However, the latter is capable of representing the temporal features of the input more effectively thanks to the use of LSTM layers, instead of FCon layers.

Long short-term memory

the LSTM [236] is a Recursive NN (RNN) architecture suited to efficiently learn long term dependencies in time series. It has been successfully applied to different fields, such as natural language processing and speech recognition. The LSTM basic processing unit is the *memory cell*, which is characterized by two state vectors called the *cell* state c_k and the *hidden* state, h_k. The cell state propagates through different time steps, carrying long term information about the process, while the hidden state controls the behavior of the cell in the current step and is renewed at every frame. h_k is also the output of the cell at time k.

At each time step, the input vector z_k, corresponding to the current observation, is processed by the cell together with h_k through four FCon layers, whose parameters are shared across the time steps. The LSTM network training is carried out using the Back-Propagation Through Time (BPTT) algorithm [267].

3.4 Proposed approach

In Fig. 3.1, we show a high-level overview of the proposed signal processing workflow. The system operates on sequences of K RDA maps, obtained from consecutive radar frames. The maps are first preprocessed individually (and in parallel), without taking into account their temporal relation. A denoising procedure is carried out to reduce the effects of unwanted random disturbances and of reflections from static objects. Hence, a clustering algorithm is applied to identify the reflections of the target user(s) among the residual points after denoising. The centroids of the obtained clusters are observations of the real positions of the target(s) in the RDA space across time, and

are denoted by $\boldsymbol{z}_k$. The observation sequence is then fed to the tracking block, where either a DAE or a S2SAE is utilized to filter the observed time series, by jointly processing sequences of length K, $\boldsymbol{z}_{1:K}$, and outputting the estimated target positions at time $1, \dots, K$, namely, $\hat{\boldsymbol{x}}_{1:K}$. The process is iteratively repeated by applying a sliding window that selects the last K RDA maps that are to be processed each time a new map is acquired from the radar. During training, the whole sequence $\hat{\boldsymbol{x}}_{1:K}$ is compared with the target sequence $\boldsymbol{y}_{1:K}$ (see Section 3.4.2) and the error is backpropagated trough the network. When trained, the network provides a position estimate (i.e., $\hat{\boldsymbol{x}}_K$) for each new map acquired from the radar.

In the following sections we detail each processing block.

3.4.1 Pre-processing

The pre-processing involves two different phases, namely the removal of static reflections/denoising and the clustering.

Denoising and removal of static reflections

The contribution to the reflected signal of the static objects in the environment is removed by deleting the samples with an estimated velocity close to 0 in the RDA maps. Specifically, we remove the values in the interval $[-0.135, 0.135]$ m/s.

For the noise removal, we apply two thresholds. The first one, along the range dimension, is decreased linearly in the logarithmic domain as the range increases, going from -70 dBm at minimum range to -95 dBm at maximum range. The second, along the angular dimension, is set at a level of -15 dB with respect to the peak value. Only the points with received power greater than both thresholds are maintained and represent candidate reflections from the targets.

Target clustering in the RDA space – DBSCAN

To identify the reflections from the targets, we group the residual points into clusters, using the Density-Based Spatial Clustering of Applications With Noise (DBSCAN) algorithm [228]. The algorithm requires one to specify two input parameters, ε and m_{pts}, respectively representing the range around each point and the minimum number of other points inside this range that must satisfy a given density condition. In this work, we use $\varepsilon = 0.04$ and $m_{\text{pts}} = 40$.

We select the centroid of the cluster, $\boldsymbol{z}_k$, with the highest number of points as a noisy observation of the true coordinates (range, velocity and angle) of the person. This centroid is computed from a weighted average of the cluster points, where the weight assigned to each point is the corresponding normalized reflected power value. In this way, the centroid moves towards those points with higher power, as they are more likely to represent the actual target position.

In the case of multi-target tracking, there are multiple retained clusters at each time step, which produce multiple observations. We use the Hungarian algorithm with Euclidean distance scores to assign observations to tracks [268]. As our current focus is on the tracking step rather than

on the data association procedure, we refer to [193] for a detailed description of the multi-target association technique.

3.4.2 Target tracking – NN

Next, we present two different approaches to the tracking problem using the DAE and the S2SAE models. The first one is **supervised** (Unsupervised Denoising Autoencoder (uDAE), Unsupervised Sequence-to-Sequence 3Autoencoder (uS2SAE)) and consists of training the network by setting the target $\boldsymbol{y}_{1:K}$ as the true state sequence $\boldsymbol{x}_{1:K}$. This method requires one to know the ground truth values for the state sequence. To this end, we employ a motion-tracking system based on infra-red cameras that can measure the position of the targets in the physical space with an error in the order of the millimeter (see Section 3.5). The values obtained from the motion tracking system are used as the ground truth $\boldsymbol{x}_{1:K}$ during training. The second approach is **unsupervised** (uDAE, uS2SAE), and entails a much more challenging learning task. In this case, the target values $\boldsymbol{y}_{1:K}$ are set to be the observations $\boldsymbol{z}_{1:K}$ and the network is trained to reconstruct the input introducing smoothness and regularity in the sequence, by adding a specifically designed term to the loss function (this will be detailed shortly in Eq. (3.4)). The reconstructed sequence is then mapped onto the estimated state sequence $\hat{\boldsymbol{x}}_{1:K}$ by transforming the polar coordinate representation into the Cartesian space. The encoder-decoder structure retains only the meaningful properties of the input, discarding variations due to noise. The unsupervised method is appealing as it does not require labeled data, which is usually difficult and costly to obtain.

Both the supervised and the unsupervised approaches rely on the same NN architectures, detailed in the following.

Denoising Autoencoder

The proposed DAE is shown in Fig. 3.2a. The input sequence of observations, $\boldsymbol{z}_{1:K}$, is flattened onto a one-dimensional input vector and passed to the encoder. The encoder block has two FC layers with 20 and 10 units respectively and outputs a code, ℓ_{DAE}, of dimension 10×1. The code is inputted into the decoder with a hidden layer of 20 units and an output layer of 30 units. All units use the hyperbolic tangent activation function. The output, one-dimensional vector, is reshaped into the output time sequence $\hat{\boldsymbol{x}}_{1:K}$ and compared against the sequence of target values $\boldsymbol{y}_{1:K}$, measuring the error in the reconstruction through the Mean Absolute Error (MAE) loss, $L = \sum_{k=1}^{K} ||\hat{\boldsymbol{x}}_k - \boldsymbol{y}_k||/K$.

Sequence-to-Sequence (S2S) LSTM Autoencoder

The S2SAE includes an encoder and a decoder, both based on LSTM layers, see Fig. 3.2b.

(*i*) The encoder takes as input the observation subsequence $\boldsymbol{z}_{1:K}$, and outputs a latent representation of the whole input sequence, ℓ_{S2S}. To encode all the information available from the input sequence into the final state, we use a bidirectional LSTM layer that processes the observations $\boldsymbol{z}_{1:K}$ in the forward time direction, outputting the forward hidden and cell states $\overrightarrow{\boldsymbol{h}}_K, \overrightarrow{\boldsymbol{c}}_K$, and in the

(a) Measurement device setup. (b) Example trajectory, supervised case. (c) Example trajectory, unsupervised case.
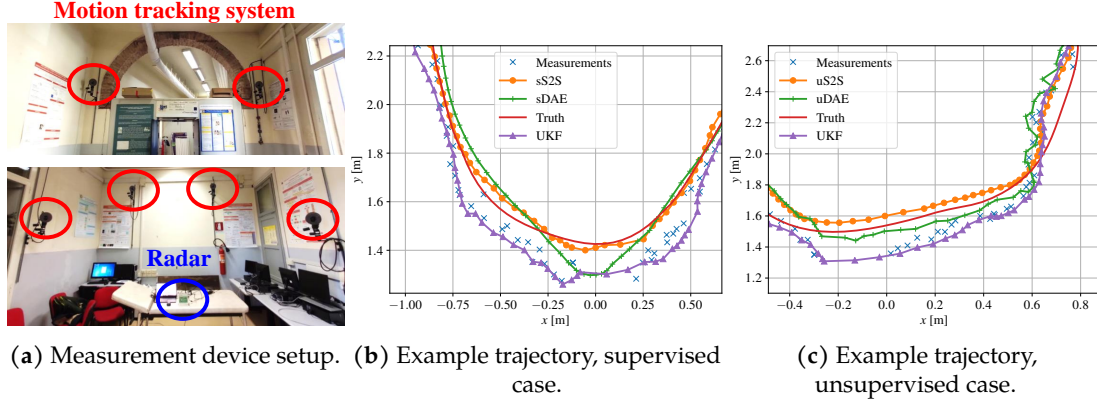
Figure 3.3: Measurement setup and example trajectories for supervised and unsupervised tracking techniques.

backward direction, outputting the backward hidden and cell states $\overleftarrow{h}_1$, $\overleftarrow{c}_1$. The FC layers in the encoder cells and the states $\overrightarrow{h}_k$ and $\overrightarrow{c}_k$, have dimension 8×1. The code is obtained concatenating the forward and backward states in the last processed time step $\ell_{S2S} = \left[\left[\overrightarrow{h}_K, \overleftarrow{h}_1 \right], \left[\overrightarrow{c}_K, \overleftarrow{c}_1 \right] \right]$.

(*ii*) The decoder is a unidirectional LSTM layer initialized with the cell and hidden states contained in ℓ_{S2S}. The input at time step 1 is a zero vector with the same dimensionality of the system state, (i.e., 2×1). Each output h_k is processed with a FCon layer with linear activation function to obtain an estimate of the target state $\hat{x}_k$, which becomes the input for the cell at the next time step. The state estimate at the last time step, $\hat{x}_K$, is the final output of the network and represents the value of interest for the filtering task.

The S2SAE model is trained by comparing the output vectors $\hat{x}_k$ with the target values y_k in each time step $k = 1, \ldots, K$, and backpropagating the reconstruction error. The loss function has two components: the first component is a standard Mean-Square Error (MSE) term, that measures the quadratic reconstruction error between the target sequence $y_{1:K}$ and the output subsequence $\hat{x}_{1:K}$, while the second is a *smoothness* term, which evaluates the regularity in the predicted subsequence by computing the square norm of the first-order differences between subsequent time samples of $\hat{x}_{1:K}$. Denoting the the first order differences by $\Delta_k = ||\hat{x}_k - \hat{x}_{k+1}||$, the expression of the loss function L is

$$L = \frac{1}{2K} \sum_{k=1}^{K} ||\hat{x}_k - y_k||^2 + \frac{\alpha}{K-1} \sum_{k=1}^{K-1} \Delta_k^2. \tag{3.4}$$

The trade-off between the two components (MSE and smoothness) is regulated by the hyperparameter α. The use of the smoothness term is an original contribution of this work and in Fig. 3.4 Root Mean Square Error (RMSE) distribution among varying α values are provided. Its effectiveness and how the value of α is determined is further discussed in Section 3.5.

3.5 Experimental results

3.5.1 Measurements setup

To evaluate the proposed tracking methods we conducted several measurement campaigns in two different rooms. The first room is an empty 20 m × 4.3 m corridor. A total of 20 minutes of RDA data were collected using the radar device from each of 4 different subjects walking freely in the environment. The unsupervised models were exclusively trained using the data from this setup. The second room is a 8 m × 4 m research laboratory equipped with a motion tracking system with 6 infra-red cameras, see Fig. 3.3a. Here, we collected the training data for the supervised models, along with some test sequences used to evaluate the tracking precision of all architectures: a total of 4 minutes training data and 2 minutes test data were collected, of which 1 minute with 2 targets walking simultaneously in the room. Subjects are allowed to move inside a 2 m × 4 m rectangle which is the working area of the motion tracking system. Ground truth data are collected by the motion tracking system, synchronized with the radar, through the detection of markers placed on the subjects by the six infra-red cameras. The measurements are acquired with a rate of 60 frames per seconds and include the location information of the subjects, in Cartesian coordinates, at any given time. To compute the orientation angle with respect to the motion tracking reference axes, and to correct the bias in the angular position values, markers were placed on the stationary radar device as well.

3.5.2 Evaluation metrics

The proposed tracking procedures are evaluated on the full test sequences of length I through the following metrics:

(i) RMSE, defined as RMSE $= \sqrt{\sum_{i=1}^{I} ||\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i||^2 / I}$

(ii) MAE, defined as MAE $= \sum_{i=1}^{I} ||\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i|| / I$

(iii) Localization Error Outage (LEO), corresponding to the probability that the error exceeds some threshold δ (expressed in meters), defined as LEO(δ) $= P(||\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i|| > \delta)$.

3.5.3 Training and tracking results

The proposed DAE models were trained using the Adam optimizer with Nesterov momentum, while for the S2SAE models we used RMSprop [227]. Training was carried out until convergence of the loss function on a validation set, which required around 50 epochs for all models. In the S2SAE models, regularization was applied using *dropout* with retain probability of 0.8 on the layers that output the hidden state and the cell state. Data augmentation was applied to extend the training sets to four times their original size: new samples were obtained by adding Gaussian noise to the collected ones.

The best results (see Section 3.5.3) were obtained setting the windows size of the input sequences to $K = 15$ for both uDAE and Supervised Denoising Autoencoder (sDAE), and to $K = 8$
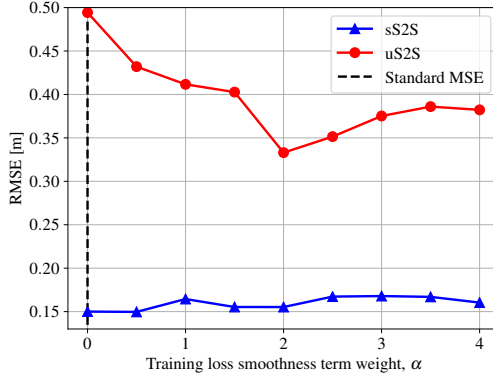
Figure 3.4: RMSE obtained varying the α parameter in Eq. (3.4).

and $K = 5$ for uS2SAE and Supervised Sequence-to-Sequence Autoencoder (sS2SAE), respectively, sliding the window one step from left to right at each new frame. For the S2SAE models we used $\alpha = 2$ for uS2SAE and $\alpha = 0.5$ for sS2SAE. In Fig. 3.4, we show the effect of varying parameter α on the resulting error on the test sequences, motivating our settings. The performance of uS2SAE is improved by tuning α and obtains better results with a higher value with respect to sS2SAE. Indeed, the enforcement of smoothness through the loss function is beneficial in this case, as no true data is available as a reference. sS2SAE works better with slight or no smoothness enforcement and is overall less affected by the parameter α.

The EKF and UKF methods are used as benchmarks and are implemented using a constant velocity model, which is customary in the person tracking application literature [238], [262], [269]. More advanced methods, such as switching KFs [270], have not yet been applied to the addressed problem, so we did not include them in the evaluation. Note that the numerical values for the process noise and the measurement covariance matrices have been empirically estimated on our dataset. We found that a proper estimation of these parameters is key towards obtaining good performance with EKF and UKF.

For a visual comparison between the approaches, in Fig. 3.3b and Fig. 3.3c, we show the same test trajectory estimated with UKF and with our NN models. Our approaches outperform UKF, especially when the subject changes direction.

The numerical results of the different methods obtained in the single and multi-target cases (two subjects) are reported in Tab. 3.1. For a single target, our approach based on NN architectures outperforms Bayesian tracking. The gap is particularly evident in the supervised case (sDAE, sS2SAE) where it reaches an RMSE of 15 cm, a MAE of 12 cm and zero empirical outage probability in a range of $\delta = 75$ cm (for a single target). With unsupervised training, our method reaches an RMSE of 30 cm with the uS2SAE model. This demonstrates the high generalization capabilities of the proposed architectures, given that training is carried out on data taken from a different measurement room, and without using the reference ground truth positions.

In the multi-target case, the unsupervised methods perform similarly to the single target one. Supervised methods instead show lower precision, most likely due to having been trained with

	Single target – multi target		
Method	RMSE [m]	MAE [m]	LEO(0.75) [%]
EKF	0.41 – 0.44	0.28 – 0.32	5.02 – 6.93
UKF	0.37 – 0.38	0.27 – 0.30	4.71 – 6.42
uDAE	0.35 – 0.36	0.25 – 0.25	5.40 – 5.92
uS2SAE	0.30 – 0.33	0.20 – 0.22	4.51 – 3.29
sDAE	0.22 – 0.29	0.19 – 0.26	0 – 0.3
sS2SAE	0.15 – 0.24	0.12 – 0.21	0 – 0.4

Table 3.1: Comparison between the proposed models, EKF and UKF methods. $X - Y$ in this table refers to the performance metric for the single target (X) and the multi-target (Y) cases (two subjects).

a small amount of data. This causes difficulties on very noisy measurements such as in the multi-target case, where two targets concurrently move within a small space. Having more labeled data would solve this issue.

3.6 Concluding Remarks

In this work, we have presented two neural network architectures for indoor person tracking from a mm-Wave radar signal, namely a DAE and a S2SAE. The proposed processing pipeline features a preprocessing phase with a threshold-based denoising step, a density-based clustering step and the final tracking procedure via the neural network models. The system has been tested on real measurements collected in a realistic indoor space, including furniture, obstacles and other people. With the best algorithm, we obtained an average tracking error of 0.12 m with a single target and 0.21 m with two targets. Our approach can be effectively applied in a smart home scenario, e.g., to detect anomalies in elderly people movements, and it is non-intrusive, as people are not require to wear any device. Future research directions include the joint estimation of the state and the covariance matrix of the monitored process, providing a richer statistical description of the motion.

4

Retrofitting IEEE 802.11ay Access Points for Indoor Human Detection and Sensing

4.1 Introduction

In this chapter, we introduce integration of a Millimeter-Wave (mm-Wave) radar sensing technology in a communication system. Specifically, we are concerned with the design of pervasive radio sensing systems that will extend the capabilities of upcoming WiFi technology operating in the 60 Ghz spectrum. Our target is to retrofit IEEE 802.11ay hardware so as to natively offer human and environment sensing services to end users. Thanks to their large available bandwidth, millimeter-wave signals allow localization and tracking with decimeter-level accuracy. Most emerging systems are based on *dedicated* mm-Wave radar devices and reconstruct the Micro-Doppler (μD) effect induced by human motion (*signature*) with high accuracy via specifically designed chirp signals [259], [271]. Radar μD signatures contain detailed information about the movement velocity of the different human body parts across time, and enable fine-grained sensing applications such as person identification from gait features [193], [238], [259], [272], activity recognition [273], gait disorder diagnosis [271] and fall detection [217], among others. The main drawback of mm-Wave radars is the need for installing bespoke sensors, which limits their scalability and ease of deployment. In several scenarios (e.g., smart homes/buildings, offices, etc.), retrofitting standard communication devices with human and environment sensing capabilities is of great value: this is precisely the purpose of RAPID, the technology that we present and validate in this chapter.

The ubiquitous deployment of WiFi devices has sparked research interest towards developing joint communication and Radio Frequency (RF) sensing technology, to avoid the cost of installing dedicated hardware while at the same time benefiting from the communication capabilities. The
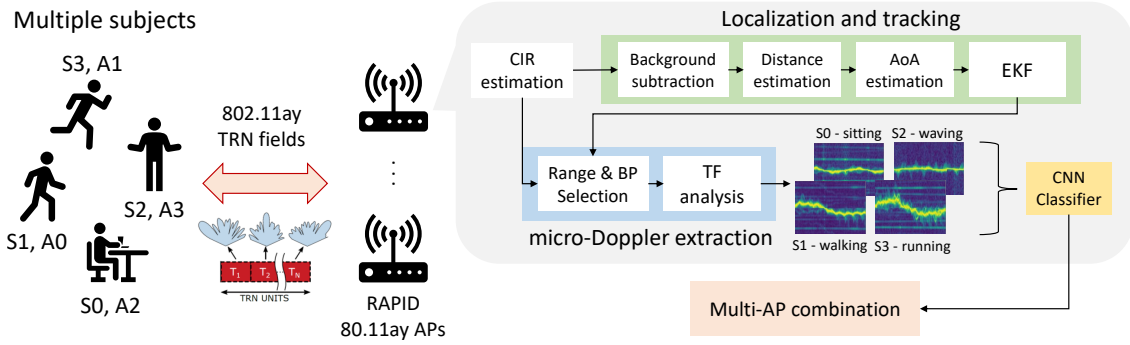
Figure 4.1: Overview of the RAPID system.

effort of enhancing WiFi Access Point (AP) with environment sensing features has recently led to the IEEE 802.11bf standardization group [274]. This initiative tackles the design of an amendment to legacy IEEE 802.11 aimed at integrating sensing techniques into WiFi AP. While legacy WiFi technology based on the IEEE 802.11n and ac standards provides a viable means for environment sensing [275], human activity recognition [276]–[279] and vital signal monitoring [280], it suffers from intrinsic limitations due to its relatively low bandwidth. This prevents highly accurate distance measurements and multi-person localization and tracking in realistic scenarios.

In this chapter, "RAPID", a Joint Communication & Radar (JCR) platform that performs radar-like sensing of human movement based on the next generation IEEE 802.11ay WiFi standard in the 60 GHz band has been explained. Previous works based on the precursor standard IEEE 802.11ad exploit the Channel Impulse Response (CIR) estimation procedure for localizing people [281], [282], but they are not fully compliant with the communication packet structure specified by the standard and they cannot achieve the sensing accuracy of radars, as no μD information is captured. In contrast, RAPID works without modifying the packet structure by leveraging the *in-packet* beam training and beam tracking features of IEEE 802.11ay. This leads to very low implementation and deployment cost, while achieving very high accuracy in extracting human movement information from the radio signals. IEEE 802.11ay uses highly directional antennas for communication. It specifies both efficient *in-packet* beam training and tracking procedures [283] based on Training (TRN) fields consisting of repetitions of complementary Golay sequences [284]. These fields are transmitted with different beam patterns, which allows to determine which of the beam patterns is best for communication. By exploiting beam training packets, RAPID can accurately localize multiple human subjects within the same indoor space. Then, the μD signature associated with the movement of each subject is extracted by relying on the TRN units embedded in the data packets used for beam-tracking, and analyzing the phase differences of the CIR across subsequent packets. For such radar-like operation, IEEE 802.11ay AP have to enable their transmit and receive chains simultaneously. However, note that this does not require complex self-interference cancellation for full-duplex communication, since the receiver needs to only detect the highly robust Golay sequences of the TRN fields. We show that the spatial separation achieved by transmit and receive beamforming from different antenna elements of closely spaced

sub-arrays is sufficient for this purpose.

To the best of our knowledge, RAPID is the first system that successfully extracts the μD from multiple concurrent subjects using standard WiFi transmission technology, achieving radar level accuracy. This is challenging, as it involves (*i*) striking a good balance between the packet transmission rate and the Doppler frequency resolution required to capture the μD of human movement, while (*ii*) ensuring sufficient phase coherence across adjacent packets. The obtained μD spectrograms with radar-level resolution are processed using a deep learning classifier to simultaneously carry out activity recognition and person identification.

Furthermore, RAPID allows the seamless integration of multiple AP to boost detection and tracking performance, as well as increase activity recognition and person identification accuracy by combining the information from different viewpoints. We implement 2 RAPID-AP using an Field-Programmable Gate Array (FPGA)-based Software Defined Radio (SDR) platform equipped with phased antenna arrays, which transmits IEEE 802.11ay-compliant packets and operates in a full-duplex fashion. The main contributions of our work are:

1. We design and implement RAPID, a fully standard compliant multi-AP JCR system that exploits IEEE 802.11ay TRN fields to achieve radar-like human sensing, including simultaneous multi-person tracking, activity recognition and person identification. RAPID reuses existing fields in the communication packets and avoids the need for a dedicated sensing infrastructure. It can combine information from multiple AP for improved performance.

2. We propose a novel method to extract μD signatures of human movement from IEEE 802.11ay CIR estimates obtained from a sequence of packets. To the best of our knowledge, this is the first work to do so from 60 GHz communication waveforms.

3. We implement RAPID on an FPGA-based testbed including multiple IEEE 802.11ay compliant APs which support full-duplex operation, so that each AP can listen to its own transmitted signal and act as a monostatic JCR device.

4. We conduct an extensive measurement campaign in an indoor space to evaluate the proposed system. RAPID achieves continuous tracking of up to 5 subjects concurrently moving in the environment, with an activity recognition accuracy of almost 93% and person identification accuracy of 90%.

This chapter is organized as follows. First, we present a summary of the related work in Section 4.2. In Section 4.3 we introduce RAPID and present the constituent processing blocks. A brief summary of how IEEE 802.11ay can be used for environment sensing is given in Section 4.4, while in Section 4.5 we discuss the implementation of RAPID on FPGA hardware. Section 4.6 contains an in depth evaluation of RAPID on real measurements. Finally, we provide concluding remarks in Section 4.7.

4.2 Related work

Sub-6 GHz sensing: Legacy WiFi technologies such as IEEE 802.11n and IEEE 802.11ac, respectively working at 2.4 or 5 GHz, have been extensively used for human sensing, including activity/gesture recognition [276]–[279], [285], vital sign monitoring [280] and person identification [286], [287].

Due to the rich multipath environment at lower frequencies, existing approaches have reached good accuracies by leveraging OFDM transmission and analyzing the CIR amplitude obtained at the different subcarriers, as done in [278]. The performance of such systems can be further improved by exploiting the phase components of the CIR [277], [285], but this entails using complex algorithms for the removal of random phase offsets.

Although there is a large body of work that exploits these technologies, they have two main drawbacks: either (*i*) they are effective for single-person scenarios, as the small available bandwidth only allows for coarse localization and tracking of the subjects, or (*ii*) they are highly sensitive to changes in the environment and hardly generalize to new scenarios (never seen at system calibration/training time), which can significantly worsen their performance. Addressing problem (*i*), in [286], multi-person identification using IEEE 802.11n is achieved in a through-the-wall setting, but the subjects still need to be well separated in space (e.g., by at least $20°$ in azimuth angle at a distance of several meters). To mitigate the dependence on the environment, more elaborate deep learning and optimization approaches have been proposed in [277], [288], [289].

mm-Wave frequencies offer a natural solution to the above issues, by providing decimeter-level accuracy in distance measurements and high sensitivity to the μD effect, due to their small transmission wavelength. In addition, due to the sparsity of the mm-Wave channel, higher robustness to environmental changes is achieved.

mm-Wave radar: mm-Wave radars have been intensively studied in the past few years as an effective means to achieve fine grained environment sensing [191]. Typical operating frequencies for these devices are the 60 or the 77 GHz bands. Centimeter-level accuracy in measuring distances is achieved thanks to the use of very large transmission bandwidths, up to 4 GHz, as dedicated radar devices are not constrained by communication requirements. Radars allow accurate activity recognition [273], [290] and have been used to perform person identification on small to medium-sized groups of people (up to a few tens), due to their very high resolution in obtaining the μD signatures of the subjects [194], [272]. In these works, the separation of the reflections from subjects concurrently moving in the environment is achieved through Multiple-Input Multiple-Output (MIMO) radars, which enable high angular resolution and allow tracking the users with errors below 0.2 m even in realistic scenarios where people walk and move freely [238]. However, these results are obtained within relatively small distances from the radar device, ranging from 4 [238] to $6 - 7$ m [272].

Despite the advanced sensing capabilities, mm-Wave radars entail high deployment costs to cover large indoor areas, even more considering their limited working range. For this reason, multi-radar networks to cover wider areas and avoid occlusion are seldom considered in the literature. Reusing existing mm-Wave communication links, as we do in this work, allows avoiding

the costly deployment of additional hardware, while maintaining radar-like human sensing and detection performance.

• **802.11ad 60 GHz sensing:** Commodity 60 GHz radios have been utilized for client device localization [154], [291], people tracking [281], fine-grained human gesture recognition [292], [293], vital sign monitoring [294] and RF imaging [282], [295]. All the works addressing human sensing are based on the IEEE 802.11ad standard and leverage the CIR estimation to obtain information about the environment. However, they typically do not address the problem of *joint* communication and sensing, which requires to reuse the packet structure specified by the communication standard. [292] and [293] address fine grained hand gesture tracking. In [292], pulsed radar-like operations are performed to detect and track a human hand, reconstructing handwriting with centimeter-level accuracy. Notably, [293] performs similar processing using the IEEE 802.11ad CIR estimated by a mobile device for gesture classification. In [282], a commodity 60 GHz radio equipped with a 6×6 antenna array is used to obtain the silhouette of a person moving directly in front of the device. This is achieved with an angular super-resolution algorithm derived from MUSIC [296]. However, the device needs to be operated in a *radar mode* for transmission, which may not comply with the communication standard.

In [281], the estimated CIR amplitude is used along with receiver beamforming to localize and track multiple people, achieving a median localization error of 9.9 cm. This work focuses on people tracking and does not exploit the phase of the CIR to extract the μD signature of the subjects, which is necessary to carry out activity recognition and person identification tasks. Moreover, the extension to the case of multiple APs is not considered.

While dedicated mm-Wave technology has been successfully exploited for human and environment sensing, and new platforms are appearing regularly. To the best of our knowledge, RAPID is the first system that extracts radar-like μD signatures of human movement from IEEE 802.11ay APs, by retrofitting them with channel sensing and Doppler extraction capabilities. This is obtained by preserving the IEEE 802.11ay packet structure, obtaining a joint radar-communication platform that is fully standard compliant.

4.3   RAPID design

RAPID enables indoor human sensing in IEEE 802.11ay networks, by leveraging the network's *in-packet* training structure.

4.3.1   Overview

From a high-level perspective, the system performs the following operations, as shown in Fig. 4.1. **(1) IEEE 802.11ay CIR estimation:** 802.11ay specifies the transmission of a variable number of TRN units for in-packet beam training, each using a (possibly) different Beam Pattern (BP). From the CIR, which is estimated from each TRN unit (see Section 4.4), RAPID obtains a scan of the whole angular Field-of-View (FOV), which contains accurate information about all the surrounding objects and people.

(2) People localization and tracking: the individuals are detected by performing background subtraction from the CIR amplitude and applying a thresholding algorithm to detect candidate reflection paths from humans, see Section 4.3.3 and Section 4.3.3, respectively. Subsequently, a correlation based algorithm is utilized to estimate the angular position of the subjects, as described in Section 4.3.3, and an Extended Kalman Filter (EKF) is exploited to sequentially track and refine the positions of the individuals across time (Section 4.3.3). By combining more than one AP, RAPID can significantly boost its human detection capabilities, while effectively coping with occlusion problems, as quantified in Section 4.6.2.

(3) μD spectrum extraction: here, the μD spectrum of each detected person is extracted. This is implemented by utilizing the CIR model as a radar return signal, and using the estimated positions from point (2) to single out the CIR portions (the paths and the BPs) containing the contributions of each subject, see Section 4.3.4. The μD signature of each individual's movement is then extracted by computing the power spectrum of the corresponding complex-valued portion of the CIR over windows of suitable length, employing Time-Frequency (TF) analysis.

(4) Activity recognition and person identification: the spectrograms from step (3) are fed to a deep learning classifier based on a residual Convolutional Neural Network (CNN) [188]. Thanks to the separation of the CIR, and to the subsequent computation of the μD for each individual, RAPID is capable of performing activity recognition for multiple subjects performing different activities within the same indoor space. Moreover, it is able to recognize the identity of a person, by extracting and analyzing their gait features from the μD signature. With multiple APs, this classification is refined by selecting the best AP to make the decision, according to the confidence of the classifier output.

In this work, we aim at localizing and tracking the people within a given physical space, by identifying which person is performing which activity. This requires person identification, tracking and activity recognition capabilities. The person identification task is carried out by extracting and analyzing the μD associated with the human gait, as this is an effective (soft) biometric signature, which has been successfully used in many works [297]. Hence, we first detect when a person is walking, then we get his/her identity from the μD gait signature and, finally, we keep tracking the person by also recognizing their activities. This also works the other way around, i.e., if a person is at first sitting and doing other activities, and then starts walking later on; as long as tracking works, we can later determine who was sitting earlier on. This also explains why tracking a person is critical, so that it is still clear which person is where, even when he/she performs other activities than walking.

We now present in detail each RAPID processing function, following the workflow of Fig. 4.1.

4.3.2 CIR model

CIR estimation is a key component of most communication systems and is used to obtain information about the environmental reflections of the signal, such as their associated angle of arrival and delay at the receiver. A key aspect to our design is that the large transmission bandwidth of mm-Wave systems leads to CIR containing fine-grained information about the environment. In

our system, the transmitter and the receiver units are co-located: the signal sent by the former, after bouncing off nearby reflectors (objects or humans), is collected at the receiver that retrieves information for each reflector, such as its distance and angle from the receiver, its velocity and micro-Doppler.

The CIR is represented as a vector of complex channel gains, also denoted by *paths* in the following, and indicized through letter ℓ. Each collected path at the receiver carries information on the associated time delay of the reflected signal, τ_ℓ. The L reflected paths are detected by correlating the received signal with pre-defined Golay sequences, identifying the L strongest multi-path components and their associated delays τ_ℓ, $\ell = 0, \ldots, L-1$, using standard techniques [298], see also Section 4.4. Path ℓ is mapped onto the corresponding reflector distance using $d_\ell = c\tau_\ell/4B$, where B and c respectively represent the transmitted signal bandwidth and the speed of light. The vector containing all the distances of interest is defined as $\boldsymbol{d} = [d_0, d_1, \ldots, d_{L-1}]$, with L being the number of paths in the CIR.

If multiple CIR estimations are performed over a single packet, using different BPs, the reflections from the environments are amplified differently due to the different BP shapes, as each BP steers the transmission signal towards a specific direction (beam steering). In addition, the CIR estimation is repeated for each packet k, which can be seen as sampling the CIR in time, with sampling period corresponding to the inter-packet transmission time T_c. The expression of the ℓ-th path of the CIR obtained using beam-pattern p at time (packet) k is

$$h_{\ell,p}(k) = a_{\ell,p}(k) \exp\left[j\varphi_\ell(k)\right], \qquad (4.1)$$

where $a_{\ell,p}(k)$ and $\varphi_\ell(k)$ are the complex gain of path ℓ at time k and its phase, respectively. The path gain depends on the contribution of the BP used for the transmission and on the reflectivity of the target, whereas the phase depends on the delay τ_ℓ.

4.3.3 People localization and tracking

CIR estimation is followed by people localization and tracking. This can be further split into (i) background subtraction, to remove the reflected paths due to static objects, (ii) the estimation of the subjects' distances, (iii) the estimation of the angular positions of the subjects with respect to the device, and (iv) their joint processing using a Kalman filter (KF) to track each person's trajectory across time.

Activity recognition and identification require CIR readings at a rate $1/T_c$, whereas localization and tracking use a time granularity of $\Delta t > T_c$ seconds, denoting the localization/tracking time-steps by index t. The choice of setting $\Delta t > T_c$ stems from the fact that performing localization and tracking for every transmitted packet is unnecessary, as the packet transmission rate $1/T_c$ is much larger than the speed of human motion. So, the system computes estimates at different rates, according to the specific resolution that is required by each. This allows for an additional flexibility in the selection of the type of BPs that are used for each packet: as we explain shortly below in Section 4.4 and Section 4.5, we can modulate how many TRN units are included in a packet

according to the type of sensing function that is being performed, i.e., localization/tracking versus activity/identity recognition.

Background subtraction

To infer the positions of the subjects it is key to remove the reflections due to static (background) objects, as these typically have a much higher intensity than those generated by humans and may impact the localization accuracy. The background-related CIR is estimated by computing the time average of the CIR amplitude within a window of K_{static} samples, as static reflections are constant across time,

$$\bar{h}_{\ell,p} = \frac{1}{K_{\mathrm{static}}} \sum_{k=0}^{K_{\mathrm{static}}-1} |h_{\ell,p}(k)|. \tag{4.2}$$

Then, the foreground CIR amplitude component is obtained as $|\tilde{h}_{\ell,p}(t)| = \max\left(|h_{\ell,p}(t)| - \bar{h}_{\ell,p}, 0\right)$, i.e., removing the amplitude of the static paths and setting to zero the amplitude of those paths that would be present in the reference background CIR, but that are shielded by the presence of a person. We remark that, through different BPs we perform beam steering at the transmitter. Hence, the peaks in $|\tilde{h}_{\ell,p}|$ correspond to the strongest propagation paths, as seen at the receiver when beam-pattern p is used at the Transmit/Transmitter (TX) side. Changing the BP p allows scanning the environment by varying the transmission angle and, in turn, sweeping the whole field of view. We use this to infer the distance and the angular position of each individual, as described next.

Distance estimation

The distance of each subject is obtained by applying a threshold on $|\tilde{h}_{\ell,p}|$ (the time index is omitted for better readability), selecting the strongest paths across all the used BPs. First, for each reflected path ℓ, we consider vector

$$\boldsymbol{h}_{\ell} = \left[|\tilde{h}_{\ell,0}|, |\tilde{h}_{\ell,1}|, \ldots, |\tilde{h}_{\ell,N_p-1}|\right]^T, \tag{4.3}$$

containing the CIR values of path ℓ for each of the N_p BPs that are used at the transmitter. We collect the L_2-norms of $\boldsymbol{h}_{\ell}$, with $\ell = 0, 1, \ldots, L-1$, obtaining a new vector $\boldsymbol{h}$, as

$$\boldsymbol{h} = [||\boldsymbol{h}_0||_2, ||\boldsymbol{h}_1||_2, \ldots, ||\boldsymbol{h}_{L-1}||_2]^T, \tag{4.4}$$

containing the strengths of each path at the receiver. We locate the local maxima in $\boldsymbol{h}$, denoting them by $h'_0, h'_1, \ldots, h'_{n_{\mathrm{peaks}}-1}$. Hence, we discard those peaks with amplitude smaller than a dynamic threshold A_{th} computed from the maximum and average power of the paths in the current CIR. We introduce the following coefficients α_{max}, α_{mean} and α_{abs}, and compute the threshold value A_{th}, as

$$A_{\mathrm{th}} = \max\left(\alpha_{\mathrm{max}} \times \max_i h'_i, \alpha_{\mathrm{mean}} \times \bar{h}', \alpha_{\mathrm{abs}}\right), \tag{4.5}$$

with $\bar{h}' = \sum_i h_i'/n_{\text{peaks}}$. We empirically assessed that suitable values for the coefficients are $\alpha_{\text{max}} = 0.25$, $\alpha_{\text{mean}} = 2$ and $\alpha_{\text{abs}} = 2.5 \times 10^{-3}$ V^2. The peaks that exceed the threshold are selected as candidate targets of interest and used for the subsequent Angle of Arrival (AoA) estimation. Denoting by $\ell_1, \ell_2, \ldots, \ell_{N_s}$ the indices of the selected (candidate) paths ($0 \leq \ell_j \leq L - 1$), the subjects' distances are obtained as the corresponding components of vector $\boldsymbol{d}$, i.e., $d_{\ell_1}, d_{\ell_1}, \ldots, d_{\ell_{N_s}}$.

Angular position estimation

The following procedure is applied to each of the N_s candidate paths. Let vector $\boldsymbol{s}_{\ell_j} \in \mathbb{R}^{N_p}$ contain the squared CIR amplitudes from one of such paths, ℓ_j, for all used beam patterns, i.e., $\boldsymbol{s}_{\ell_j} = \left[|\tilde{h}_{\ell_j,0}|^2, |\tilde{h}_{\ell_j,1}|^2, \ldots, |\tilde{h}_{\ell_j,N_p-1}|^2 \right]^T$. $\boldsymbol{s}_{\ell_j}$ is normalized by dividing it by its L_2-norm $||\boldsymbol{s}_{\ell_j}||_2$, then a correlation measure is used to estimate the angular position of the target by exploiting the gains of each beam pattern along the azimuth angular FOV θ. Specifically, denoting by $g_p(\theta) \in [0, 1]$ the normalized gain of beam pattern p along direction θ (see Fig. 4.6a), the angular position for candidate path ℓ_j is estimated as

$$\theta_{\ell_j} = \underset{\theta}{\text{argmax}} \sum_{p=0}^{N_p-1} g_p(\theta) \frac{|\tilde{h}_{\ell_j,p}|^2}{||\boldsymbol{s}_{\ell_j}||_2}. \tag{4.6}$$

The rationale behind Eq. (4.6) is that if $|\tilde{h}_{\ell_j,p}|$ originates from the signal reflected off a subject, the corresponding angular direction is the one leading to the highest correlation between the CIR squared amplitude and the set of beam pattern gains. This is because each BP amplifies path ℓ_j differently, depending on the beam pointing direction.

Upon obtaining the distance and the angle estimates, an EKF is utilized to track the subjects' positions over time.

People tracking - extended Kalman filter

After the localization step, the candidate positions of the subjects are known in polar coordinates, and constitute our *observations* of the positions of the subjects, which we denote by $\boldsymbol{z}_t^j = [d_{\ell_j}, \theta_{\ell_j}]^T, \forall j = 1, 2, \ldots, N_s$. We employ an EKF [299] to track the physical position of each individual in the Cartesian space. Specifically, we define the true state of subject j at time t as vector $\boldsymbol{x}_t^j = \left[x_t^j, y_t^j, \dot{x}_t^j, \dot{y}_t^j \right]^T$, containing the coordinates along the $x - y$ horizontal plane and the movement velocity components along the same axes. We approximate the motion of the subjects with a constant velocity (CV) model [300]. As the observations $\boldsymbol{z}_t^j$ become available, we apply the predict and update steps of the EKF to follow the movement trajectories of the subjects [299]. The association between the observations from time $t + 1$ and the states from time t is done using the nearest-neighbors joint probabilistic data association algorithm (NN-JPDA) [301].

Using the EKF estimates $\hat{\boldsymbol{x}}_t^j$ of each person's state across subsequent time steps allows retrieving the path and the BPs in the CIR which contain his/her μD signature.

4.3.4 micro-Doppler extraction

CIR phase model

The CIR model in Eq. (4.1) is here expanded and related to Frequency-Modulated Continuous-Wave (FMCW) radar theory [265]. Using a typical radar terminology, we refer to the CIR samples $\ell = 0, 1, \ldots, L-1$ as the *fast-time* sampling dimension, as they are obtained at the highest available sampling rate. The CIR samples collected across different packets are instead referred to as the *slow-time* samples, indicized by variable k as in Section 4.3.2.

Next, we consider the case there is a moving object within the monitored indoor space; the transmitted signal is reflected off the object and the corresponding contribution is retrieved at the receiver in the ℓ-th path of the CIR. To extract the μD effect caused by the movement of this object, we analyze the phase of the ℓ-th path across time. The time-dependent phase term in Eq. (4.1) can be expressed as follows

$$\varphi_\ell(k) = -2\pi f_o \frac{2\left(d_\ell + v_{\ell,k} k T_c\right)}{c} = -2\pi f_o \bar{\tau}_\ell + 4\pi f_o \frac{v_{\ell,k}}{c} k T_c. \tag{4.7}$$

Here, $\bar{\tau}_\ell$ is the delay of the ℓ-th path due to the distance of the corresponding reflector from the device, while $v_{\ell,k}$ is the radial velocity of the reflector with respect to the device, which is assumed to be constant during the sampling interval T_c. Note that From Eq. (4.7) it can be seen that the velocity of the object at distance d_ℓ, if greater than zero, modulates CIR phase across the slow time dimension. Following a common convention [265], in this work objects moving away from the transmitter (AP) have *positive* velocity, while incoming objects have *negative* velocity.

Human body contains multiple moving parts that have different velocities and follow different trajectories. Thanks to the small wavelength of mm-Waves, in the μD we can observe these different contributions via TF analysis, as detailed in the next Section 4.3.4.

micro-Doppler spectrum

Human movement causes a frequency modulation on the reflected signal due to the small-scale Doppler effect produced by the different body parts. Using TF analysis of the received signal, it is possible to distinguish between different actions performed by a person or identify the individual based on his/her way of walking (*gait*) [259], [271]. mm-Wave radios are particularly suited for this kind of analysis, as their frequencies are sensitive to the μD effect due to their small wavelengths.

From Eq. (4.7), the μD effect of human movement can be extracted from subsequent estimates of the CIR, computed every T_c seconds. Specifically, one can compute the Short-Time Fourier Transform (STFT) of $h_{\ell,p}(k)$, across the slow-time dimension, for each path ℓ and each beam pattern p as

$$H_{\ell,p}(n,i) = \sum_{m=0}^{M-1} h_{\ell,p}(m+n\sigma)w(m) \exp\left(-j2\pi \frac{im}{M}\right), \tag{4.8}$$

where n is the time index, $i = 0, 1, \ldots, N_D - 1$ is the frequency index, w is a Hann window of appropriate size, σ is the time granularity of the mm-Wave and M is the (fixed) window length. The power spectrum of $h_{\ell,p}(k)$, computed as $\mu_{\ell,p}(n,i) = |H_{\ell,p}(n,i)|^2$, contains information on the phase modulation due to the velocity $v_{\ell,k}$, and can be used to analyze its non-stationary evolution in time.

Eq. (4.8) can not be used directly to extract the μD signature of a moving human in our setup, as it refers to a single fast time bin (a single path in the CIR) and a single BP, while people can be located in different positions across time. In addition, it would be inefficient to compute the mm-Wave for all the paths and all the BPs. Instead, the computation should only be performed for those physical locations where a person is detected. In the following, we leverage the localization and tracking process described in Section 4.3.3 to only extract the CIR portions that contain useful μD information.

μD separation

Assume that we want to extract the μD of a person that was detected and located by the previous algorithms at a certain distance from the device. Hence, we extract the CIR samples from the most useful BP, i.e., the one that points in the direction of the person and that, in turn, emphasizes the most the reflection from this target.

From the estimated state of this person (Section 4.3.3), his/her angular position is obtained as $\hat{\theta} = \arctan\left(\hat{y}_t / \hat{x}_t\right)$ and his/her distance from the device, as $\hat{R} = \sqrt{\hat{x}_t^2 + \hat{y}_t^2}$. The BP approximately pointing in the direction of this person, denoted by p^*, is thus selected as the BP having the highest gain along $\hat{\theta}$, that is

$$p^* = \underset{p}{\arg\max}\, g_p(\hat{\theta}). \tag{4.9}$$

Moreover, due to the high ranging accuracy of mm-Waves, humans typically produce reflections that influence more than a single CIR path. The CIR paths of interest are those that correspond to a neighbourhood of $\hat{R}$. In our analysis, we take the size of this neighborhood constant across all subjects, denoting it by Q. Specifically, we first select the path ℓ^* that best matches the subject's distance $\hat{R}$

$$\ell^* = \underset{\ell}{\arg\min}\, |d_\ell - \hat{R}|. \tag{4.10}$$

Then, from the original complex-valued CIR, we extract a window containing $Q + 1$ samples along the fast-time dimension, centered on ℓ^*, and use Eq. (4.8) to compute the μD spectrum components for our target at time step t as

$$\mu_i(t) = \sum_{\ell = \ell^* - Q/2}^{\ell^* + Q/2} \left|H_{\ell,p^*}(t,i)\right|^2, \quad i = 0, 1, \ldots, N_D - 1. \tag{4.11}$$

These μD spectrum components are collected through vector $\mu(t) = \left[\mu_0(t), \mu_1(t), \ldots, \mu_{N_D-1}(t)\right]^T$. Finally, we compute the μD vectors for a window of T subsequent time-steps and concatenate

them to obtain the spectrogram representing the μD signature of the target up to time t, as

$$Y_t = [\mu(t - T + 1), \mu(t - T + 2), \ldots, \mu(t)] \,. \tag{4.12}$$

The procedure described in this section is repeated for all the detected subjects. It should be noted that the above procedure is considered for Line-of-Sight (LOS) conditions. in a Non-Line-of-Sight (NLOS) problem, if Eq. (4.9) would be directed towards the BP that illuminates a subject with the smallest number of reflections, it theoretically would be possible to identify the static wall reflections from the reflections received from subjects since the latter would arrive later.

Human μD range and resolution

During everyday movement, the limbs of a person usually have velocities of up to $3 - 4$ m/s [259], [271]. To fully capture the μD signature of the subjects, we must ensure that our systems achieves a sufficient resolution. Recalling Eq. (4.7), we know that the Doppler frequency shift induced by a moving object on the ℓ-th path at time k (packet k), $f_{\ell,k}^{d}$, is

$$f_{\ell,k}^{d} = \frac{2f_o v_{\ell,k}}{c} \,. \tag{4.13}$$

Using time-frequency analysis to estimate the Doppler spectrum as in Eq. (4.8), the resolution that can be obtained on the Doppler frequency is $\Delta f^{d} = 1/(MT_c)$. The maximum measurable Doppler frequency is instead $f_{\max}^{d} = 1/(2T_c)$. These quantities can be mapped onto the velocity estimate resolution and the maximum measurable velocity using Eq. (4.13), obtaining

$$\Delta v = \frac{c}{2f_o M T_c}, \quad v_{\max} = \frac{c}{4f_o T_c} \,. \tag{4.14}$$

Given that we sample the CIR on a per-packet basis, to capture the μD effect of human motion we must ensure that the time T_c between the packets used in the μD estimation allows capturing the range of velocities of interest. See also Section 4.6 for the chosen values of M and T_c.

4.3.5 Activity recognition and person identification

The μD signature, obtained as in Eq. (4.12), contains information about the type of movement performed by the person.

To perform activity recognition and person identification, we use a deep neural network to classify each spectrogram. Specifically, once the μD signatures of each person have been separated, RAPID performs the following tasks: (*i*) classifies the activity carried out by the subject into: *walking* (A0), *running* (A1), *sitting* (A2) and *waving hands* (A3) and (*ii*) recognizes the subject's identity during a walking phase, among a known set of individuals, denoted by S0, S1, etc. In Fig. 4.2, we show μD signature examples for activities A0 $-$ 3, concurrently performed by 4 subjects within the same environment.
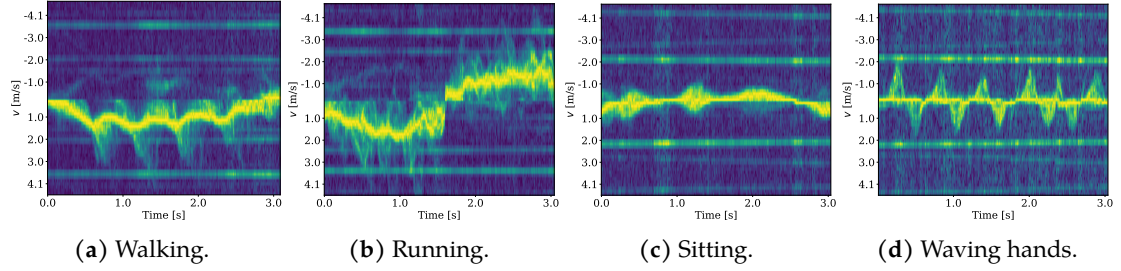
Figure 4.2: Example 3 s long μD spectrograms obtained by RAPID from 4 subjects concurrently moving in the environment and performing 4 different activities. The yellow and blue colors respectively represent high and low power values in the corresponding Doppler velocity bins (y axis).

As human μD is highly variable across different subjects, and we seek robustness to different environment conditions and noise, we employ deep learning to classify the μD signatures. Referring to a single subject, the μD spectrum Y_t is represented as an image and processed by a CNN architecture. As the subjects are continuously tracked over time, we adopt a sliding window approach, selecting μD spectrograms with T μD spectrum samples for each window (matrix Y_t). Subsequent windows partially overlap to increase the reactiveness of RAPID in obtaining predictions. The CNN is trained to extract features from the μD spectrograms and to classify them by learning a function $\mathcal{F}(\cdot)$ that maps a μD window, Y_t, of size $N_D \times T$, onto a vector $\boldsymbol{c}_t$ containing the class probabilities, i.e., $\boldsymbol{c}_t = \mathcal{F}(Y_t)$. The same network architecture, that is detailed in the following, is used for both activity recognition and person identification.

μD spectrogram pre-processing

Prior to feeding it to the CNN classifier, the μD spectrogram is pre-processed by removing the contributions from static reflections and normalizing it.

• **Static reflections removal:** a customary step when processing human μD signatures is the removal of static reflections, which appear as a strong power peak around the 0 m/s velocity bin. This can be done by either applying an high-pass filter to the signal or, if deep learning methods are used for classification, by directly removing the Doppler bins containing unwanted contributions, as done in [193], [259]. We adopt this latter method to remove the Doppler bins corresponding to the velocities in the interval $[-0.28, 0.28]$ m/s, as they contain very low, non informative velocities.

• **Normalization:** to compensate for differences in the strength of the reflections when subjects are far from the APs, we normalize each column of Y_t, $\mu(j), j = 0, 1, \ldots, T - 1$ in the range $[0, 1]$.

The plots in Fig. 4.2 were obtained by using the just described pre-processing steps.

Deep learning classifier

We use a CNN architecture based on deep residual networks [188]. This network consists of 4 consecutive residual blocks. Each residual block has two convolutional layers [227], the first of
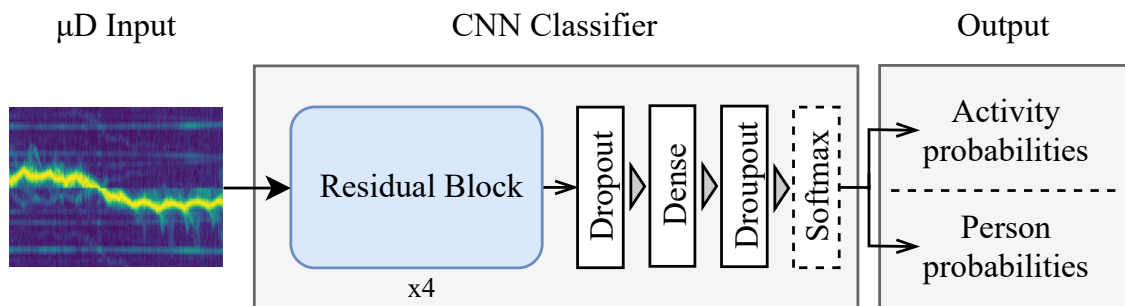
95

μD Input	CNN Classifier	Output

Figure 4.3: Block diagram of the CNN classifier used by RAPID.

which includes a down-sampling by a factor of 2 (*stride*). Each convolution is followed by an Exponential-Linear Unit (ELU) activation function [302] and batch normalization [242]. The output of the convolution is summed to the input (*skip connection*) and passed through another ELU activation and batch normalization. The 4 residual blocks use 8, 16, 32 and 64 filters, respectively, all having a kernel of size 3×3. After the last residual block, we apply Dropout [303] with a ratio of 0.5, and a fully-connected (or *dense*) layer with 64 units, then, a second Dropout operation with ratio of 0.2. Finally, the classification probabilities for activity recognition or person identification are computed via a Softmax activation function [227]. The network architecture is shown in Fig. 4.3.

Combining multiple APs

Using the different points of view provided by the different APs, RAPID can improve its activity recognition and person identification performance. Assuming that a person is independently detected and tracked by 2 or more APs concurrently, a slightly different μD signature of the person is obtained by each AP, according to the angular position and the distance of the device with respect to the person. At each time instant t, we adopt a simple decision fusion scheme including the following steps.

• If a single AP detects the person, the decision made by the classifier on the corresponding μD signature is used, i.e., $\mathrm{argmax}_j \, c_{t,j}$, where $c_{t,j}$ is element j of vector $\boldsymbol{c}_t$.

• If multiple APs detect the person, denote by $\boldsymbol{c}_t^a$ the probability vector predicted by AP a. The final decision is made by the AP that is most confident about its classification, i.e., the one that assigns the highest probability to the predicted class: $\mathrm{argmax}_j \left\{ \max_a c_{t,j}^a \right\}$.

4.4 Enabling sensing capabilities on IEEE 802.11ay systems

The high bandwidth of IEEE 802.11ay [304] not only provides high data throughput but also offers excellent accuracy for sensing applications. RAPID is able to extract highly accurate range, angle and μD information from CIR measurements. To this end, we take advantage of the beam training and beam tracking mechanisms of IEEE 802.11ay systems.
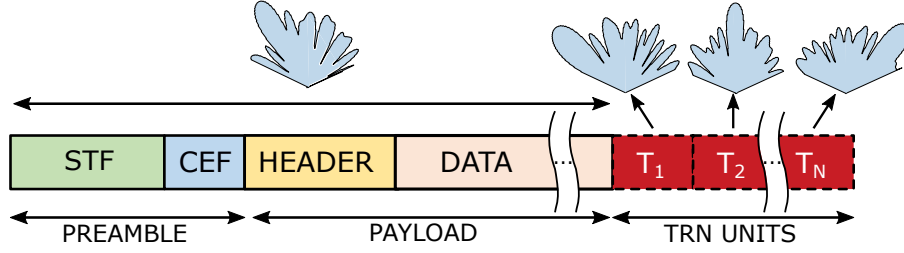
Figure 4.4: IEEE 802.11ay *in-packet* training.

Range and angle information are extracted from the beam training frames. Beam training is a two-step procedure: first, one device sends training frames using the available antenna configurations, while the second device listens using a quasi omnidirectional BP. Then, the devices exchange their roles to train the other device. After sending feedback, the devices can select the *best* combination of BP on both sides of the link. IEEE 802.11ay also introduces the concept of *in-packet* training [283], where different antenna configurations can be tested within a single packet. This is done by appending a TRN field to the packet as shown in Fig. 4.4. A TRN field is composed of multiple (variable) TRN units formed by complementary Golay sequences of type a ("Ga") and b ("Gb") with length 128 samples:

$$\{+Ga_{128}; -Gb_{128}; +Ga_{128}; +Gb_{128}; +Ga_{128}; -Gb_{128}\} \tag{4.15}$$

The excellent autocorrelation properties of the complementary Golay sequences and the availability fast hardware structures for the correlation [305] make them ideal for CIR estimation [284]. The high bandwidth (1.76 GHz) results in a range resolution of ≈ 17 cm directly from the CIR estimate. Considering the different BP shapes used during beam training, possible targets located in the FOV of the devices are *illuminated* by the respective BPs that focus energy in that direction. They appear as multi-path components in the CIR (Fig. 4.6c). Furthermore, we take advantage of the different amplification factors in the multi-path components (given by the different BPs). The angular positions of the subjects are estimated by applying the correlation based approach explained in Section 4.3.3 to the different multi-path components in the channel. APs typically send beacon frames via the different beam patterns every 100 ms to allow devices to discover the network. We use these for our estimation (note that a full beam training, which is carried out less frequently, is in fact not needed), and considering the common speed of human motion, this mechanism allows to accurately locate human targets in the FOV of the AP.

Extracting μD signatures from the CIR requires fine-grained frequency resolution, as detailed in Section 4.3.4. This cannot be achieved with the CIR estimates obtained from the TRN units of beam training packets only, as sampling the CIR with $T_c = 100$ ms would lead to an insufficient maximum measurable Doppler velocity of 6.25×10^{-3} m/s (see Eq. (4.14)). We addressed this by exploiting the *beam tracking* procedure defined in the standard [304], which adds a configurable number of TRN units to data packets. This allows to test different BP configurations to *quickly* correct possible misalignments without requiring a full beam training procedure.

After identifying the subjects' ranges and angles using beam training packets, we include a TRN field in *data packets* with a sufficient number of TRN units to illuminate all the subjects in the scene; each TRN unit uses a suitable BP that specifically points in the direction of a person. This steers the energy of the transmitted signal so as to best capture the μD signatures of the subjects, while maintaining low additional overhead for the data packets. Considering that data packets are sent much more frequently than beam training packets, our approach enables sampling the CIR with a much lower T_c compared to using beam training packets only, hence achieving higher frequency resolution.

4.5 Implementation

The available mm-Wave Commercial-Off-The-Shelf (COTS) devices support IEEE 802.11ad and offer very limited access to physical layer information [180]. To the best of the authors' knowledge, there are no COTS solutions for the new IEEE 802.11ay standard available yet. To address the lack of hardware, we turn a mm-Wave SDR system into a JCR experimentation platform. In this section, we cover the details and design decisions made to implement RAPID on such platform.

4.5.1 Hardware components

As a baseline to implement a RAPID AP, we use the mm-FLEX experimental platform [175]. This open platform is composed of a baseband processor including a Xilinx Kintex Ultrascale FPGA plus high-speed Analog-to-Digital Converter (ADC)/Digital-to-Analog Converter (DAC) and DDR memory banks. Besides, it is connected through a PCIe interface to a Core i7 processor card co-located within the same hosting chassis. The latter implements configuration and control tasks not only for the FPGA and converters, but also for the RF front-end.

The baseband processor is configured to fulfill the bandwidth requirements of IEEE 802.11ad/ay standards (1.76 GHz), using a sampling frequency of 3.52 GSPS for both AD/DA converters (with 2 samples per symbol).

The RF front-end includes a 60 GHz up/down converter and phased antenna arrays from Sivers [306]. The device is able to operate on all the channels defined in the IEEE 802.11ad/ay standards [28], [304]. As can be seen from Fig. 4.5, the device integrates two independent 16-element linear antenna arrays, one used for transmission and one for reception. Both can be freely configured with independent Antenna Wave Vector (AWV) belonging to different codebooks. The system can be configured in real-time using USB and SPI interfaces, as well as GPIO pulses for the quick BP changes require for beam training and tracking.

4.5.2 Full-duplex operation

To bring radar capabilities to the experimentation platform, it is necessary to enable simultaneous operation of the TX and Receive/Receiver (RX) chains. This is achieved by concurrently enabling
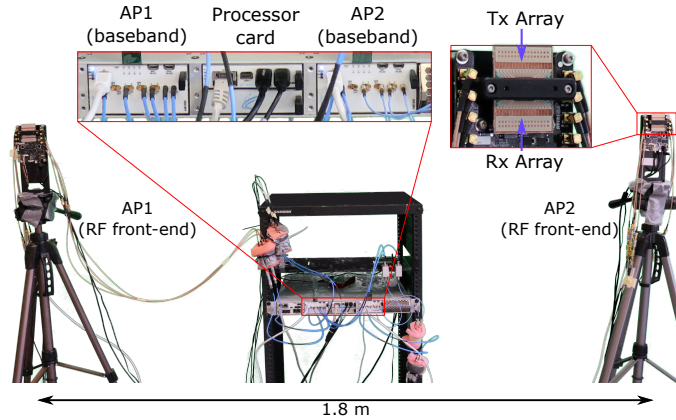
Figure 4.5: Two RAPID APs deployment

transmit and receive sub-systems in the RF front-end, and by enhancing the functionality of the baseband processor.

The 60 GHz front-ends used in this work [306] are laboratory equipment designed for *proof-of-concept* communication systems. In these devices, carrier frequency is generated from a 45 MHz clock, which introduces severe Carrier Frequency Offset (CFO) and destroys the phase coherence between the CIR estimates obtained from consecutive packets. This makes the extraction of μD signatures infeasible with two separate co-located antennas. Instead, we use both transmit and receive arrays from the same RF front-end (see Fig. 4.5). This way, up and down conversion sub-systems are fed by the same local oscillator which keeps CFO levels in the range of $[-40, 40]$ Hz, as shown in Fig. 4.6b. Although transmit and receive arrays are directly next to each other, no complex analog or digital self-interference cancellation techniques are required. Thanks to the directive BPs and the robustness of the Golay Sequences of the TRN units, the system only requires some power control. Specifically, we adjust the transmit amplifier gain to meet two requirements: (*i*) not to saturate the receive antennas and down-conversion stages because of the self-interference, which would destroy the channel measurements; (*ii*) to provide enough gain to receive sufficient signal strength from the reflections coming from human subjects in an indoor laboratory environment. In Fig. 4.6c, we show the CIR measurements obtained from multiple BPs within a packet, by marking the self interference path and the reflections from the test room, where the different amplitudes correspond to the different BP shapes towards the direction of the reflectors.

In the baseband processor, we implement a state-machine on the FPGA logic which controls the transmit and receive data-paths. Specifically, it handles the DDR memory that stores the transmit frames, performs multiple real-time antenna reconfigurations over the TRN field of the packet, triggers the DDR memory on the receive data-path, and sets the inter-frame spacing between multiple transmitted packets, among other configurations of the system. While here we focus on an AP-centric design, the same procedure can be applied to implement RAPID on any station in the network.

99

(a) Beam pattern shapes

(b) Measured CFO

(c) Example of CIR measurement
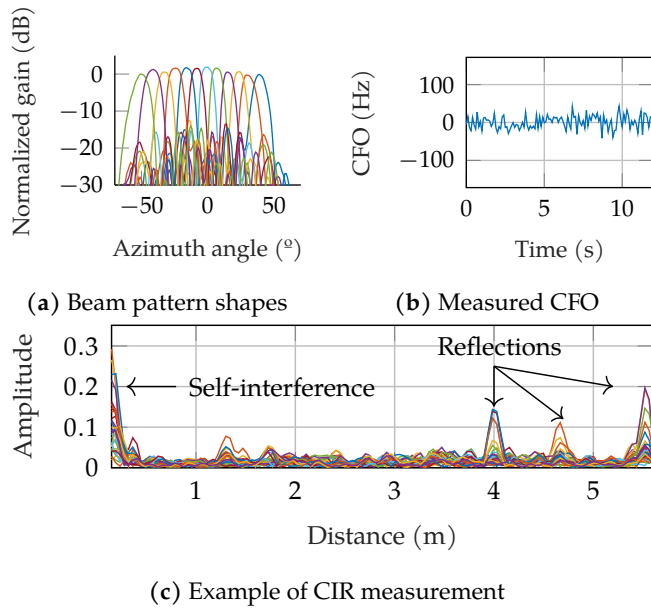
Figure 4.6: RAPID implementation
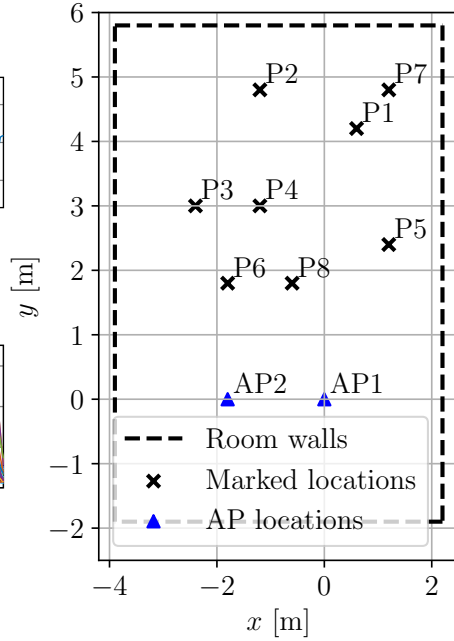
Figure 4.7: Experimental setup

Since our RAPID AP operates in a mono-static configuration, we enable the CIR extraction without requiring the use of packet detection and synchronization circuits. To do this, it is important to ensure deterministic latency between the transmit and receive data-paths. Considering that transmit and receive data-paths have their own independent clock structure, we use clock-domain crossing techniques to send the state machine signals across transmit and receive domains. Besides, latencies in the DDR controllers are variable, which requires the use of FIFO queues at the output/input of the TX/RX DDRs. Together, these solutions help to achieve the desired deterministic latency.

4.5.3 Multi-AP system

Since IEEE 802.11ay networks typically involve many AP and dense deployments, we extend the aforementioned testbed capabilities to handle multi-AP scenarios. To this end, we integrate a second baseband processor in the hosting chassis which is connected to an independent 60 GHz front-end. The FPGAs from both processors have their own clocking structure, i.e., they are not synchronized. Each AP can be freely configured with its own parameters. For the sake of simplifying the system management, we use different communication channels (58.32 and 60.48 GHz) for each RF front-end, avoiding cross interference. It is worth mentioning that the channels can be freely configured, making it possible to operate the two RAPID APs so that they share the same frequency band, by implementing carrier sensing mechanisms.

To deploy the system, we used 4 m cables to connect each baseband processor with its corresponding RF front-end. This is the maximum length to achieve good signal quality, and for a larger deployment multiple separate chassis would be needed. Furthermore, the USB and GPIO interfaces used to configure the 60 GHz kits are 1.5 m long. This allows to provide enough separation between the RAPID APs to cover different scenarios. Due to the hardware available to us, in this work we only use two RAPID APs, but it is straightforward to extend the system architecture to integrate more APs for larger scale scenarios.

4.6 Experimental results

In this section, we discuss the results of extensive measurement campaign with RAPID. Motivated by the discussion in Section 4.3.4 and Section 4.4, for the μD estimation we consider data packets (with TRN fields) spaced by $T_c = 0.27$ ms. This allows capturing velocities in the range $[-4.48, 4.48]$ m/s and leads to a resolution of $\Delta v = 0.14$ m/s when using a window of $M = 64$ samples in the DFT computation, see Eq. (4.14). These values are comparable to the ones achieved with radar devices [193], [238], [259]. Note that the even spacing of packets is just for convenience but is not a requirement, i.e., estimation can be done with random bursts of data packets with sufficiently small spacing. Moreover, we set to $Q = 4$ the size of the fast time window used to capture the contribution of the subjects in the CIR (see Section 4.3.4). The EKF time-step duration is set to $\Delta t = 16T_c$, which is also the time-granularity at which we obtain μD spectrum vectors. To extract range and angle information, we use in-packet beam training frames with 12 TRN units, using antenna beams covering a FOV range from $-45°$ to $45°$. With this configuration we achieve a mean accuracy of $2°$ for the angular position of a person standing in the room. We verify that this allows tracking multiple subjects reliably and with low localization error, as detailed in the following. In order to implement the angle estimation method from Section 4.3.3, we measured the BP shapes from the codebook using a motorized pan-tilt platform. In Fig. 4.6a, we show the 12 BPs we used to perform the experiments.

4.6.1 Experiment setup

We test RAPID in a research laboratory of dimensions 6.1×7.7 m containing whiteboards, windows, tables, computers and equipment, making it a challenging multi-path environment with a number of potential reflectors. We deploy two RAPID APs as shown in Fig. 4.5 close to one of the walls and separated by 1.8 m.

To test the localization and tracking capabilities of RAPID, we mark specific known positions across the test room to determine ground truth location as shown in Fig. 4.7, and perform our tests by having subjects move across these positions. The markers are denoted by Px, with x ranging from 1 to 8, while APs are represented as blue triangles. The room walls are represented with a black dashed line.

101
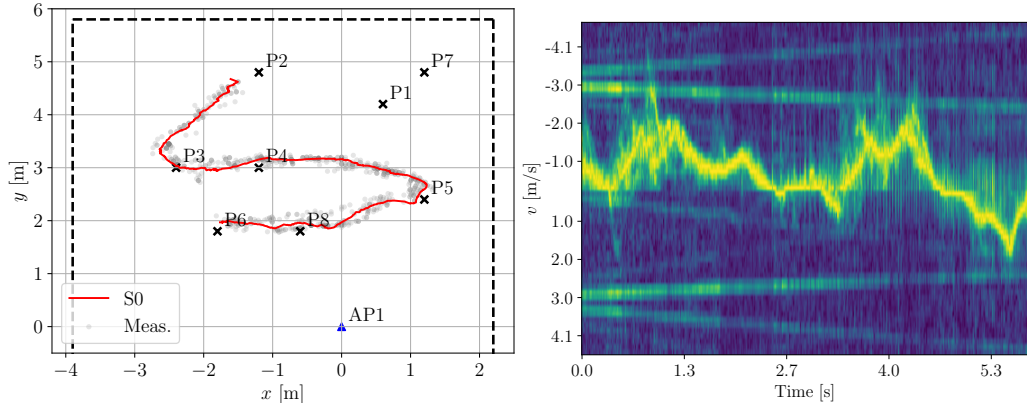
Figure 4.8: Subject walking trajectory (left) and a portion of the corresponding μD signature (right) extracted by RAPID.
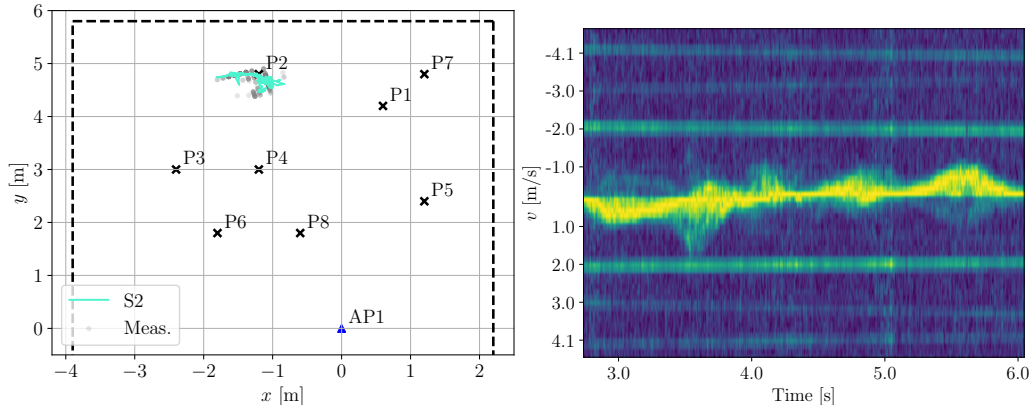
Figure 4.9: Estimated position of a subject sitting (left) and a portion of the corresponding μD signature (right) extracted by RAPID.

4.6.2 Baseline experiments

We first report the results obtained in two simple baseline experiments to verify the capability of RAPID to extract the μD signature of a moving person in an indoor scene. Here, we only use AP 1 and a single subject, performing different activities at different locations.

Fig. 4.8 shows the EKF estimated trajectory of the subject walking along the trajectory P2-P3-P4-P5-P8-P6 together with the corresponding μD spectrogram. The light grey points represent the raw measurements (observations) obtained as explained in Section 4.3.3, using Cartesian coordinates. The trajectory is correctly reconstructed with remarkable accuracy. The μD signature is extracted successfully and shows the different contributions of the torso and the limbs. The former reflects more power and follows a slightly oscillating motion, which is coherent with the direction changes in the walking trajectory, while the latter reflect lower power and are responsible for the higher velocity peaks.

Next, we test RAPID on a subject sitting at the marker P2, as shown in Fig. 4.9. Also in this case, RAPID correctly estimates the location of the subject, and the μD spectrum is coherent with the
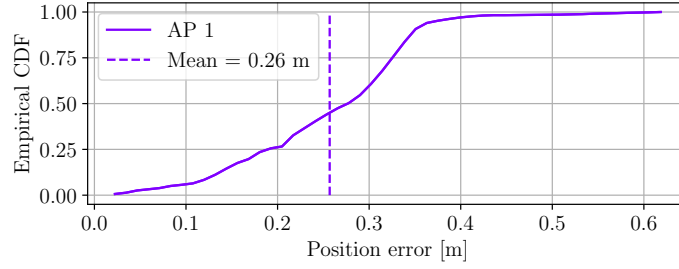
Figure 4.10: Empirical Cumulative Distribution Function (CDF) of the positioning error for a subject sitting in correspondence of marker P2.
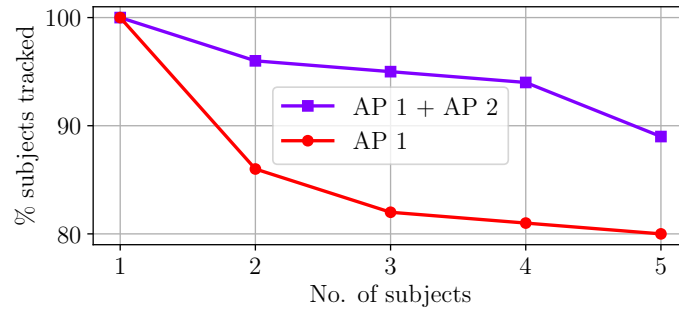
Figure 4.11: Rate of detection with a varying number of subjects using only AP1 and the combination of AP1 and AP2.

sitting activity. This is non-trivial, given that P2 is located at the edge of the experiment room.

The empirical CDF of the positioning error of the subject in Fig. 4.10 shows that we achieve a good localization accuracy. In this analysis, we included around 2000 position estimates made by the EKF. The median error is 26 cm, and the probability of the error being lower than 40 cm is close to 1. We stress that the subject in this case is not static, as the person alternates between sitting and standing up. This causes the estimated position to change slightly across time-steps, increasing the localization error.

4.6.3 Multi-person multi-AP scenario

In this section, we extend the scenario to analyze the impact of multiple subjects present on the scene, which we tackle using multiple APs. Here, all measurements are performed using AP1 and AP2. We first consider the results obtained solely by AP1 in Section 4.6.3, and then we combine AP1 and AP2 results in Section 4.6.3. Several experiments are carried out with 2 to 5 subjects, performing different activities. In total, we collect 28 such sequences each with duration ~ 10 s, of which 13 include 2 subjects, 5 include 3 subjects, 6 include 4 subjects and 4 include 5 subjects.

(**a**) AP1 estimated trajectories.   (**b**) AP2 estimated trajectories.
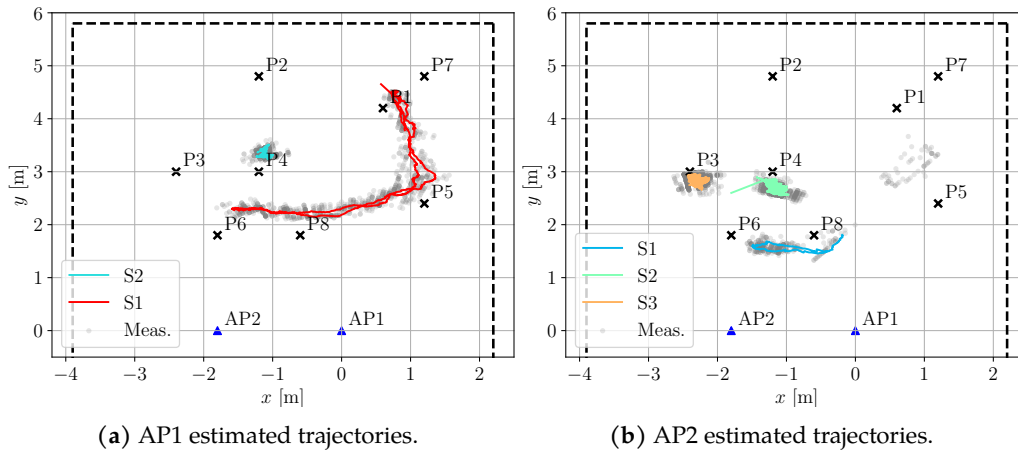
Figure 4.12: Impact of using multiple APs on the occlusion problem. Here, AP1 fails to detect and track S3, while AP2 can only partially reconstruct the trajectory of S1. The combination of the 2 APs successfully detects and tracks all subjects on the scene.

Presence of multiple subjects

Fig. 4.13 shows some example trajectories estimated by the EKF using the measurements from AP1. RAPID is able to successfully track the users with considerable accuracy in most cases, even for 5 subjects (see Fig. 4.13d). Note that this setup is extremely challenging, especially when more than 3 subjects are present, due to the small dimensions of the environment that lead to a high probability of occlusion happening, i.e., one subject covers the LOS path between the AP and another individual. mm-Wave signals do not propagate through the human body, and occlusion may cause missed detection and tracking errors. On the other hand, in real-life scenarios occlusion may happen frequently, and the system must be robust to these events. In Fig. 4.11, we report a quantitative analysis of the effect of increasing the number of subjects in terms of the percentage of subjects that are correctly detected and tracked by RAPID. Using only AP1 (red line) we observe that, despite achieving adequate tracking performance, the system capability of detecting the subjects decreases significantly as their number increases. In particular, on average one subject goes undetected when 5 individuals are present in the room.

Improvement with multiple APs

Combining the FOV of AP1 and AP2 effectively decreases the probability of occlusion events happening, as when the LOS between an AP and a subject is blocked, the other AP can exploit its own LOS path to detect the person. In Fig. 4.12 we report a qualitative example of this, by showing how combining the 2 APs, RAPID can recover from an occlusion event (in this case, 3 subjects are present in the environment). The EKF estimated trajectories from AP1 are shown in Fig. 4.12a: subjects S1 and S2 are successfully detected and tracked, while S3, who is waving hands in P3, is not. This is due to a combination of the occlusion caused by S1 and the fact that P3 is placed at the edge of the FOV of AP1. However, the position of AP2 enables it to detect S3 successfully, while
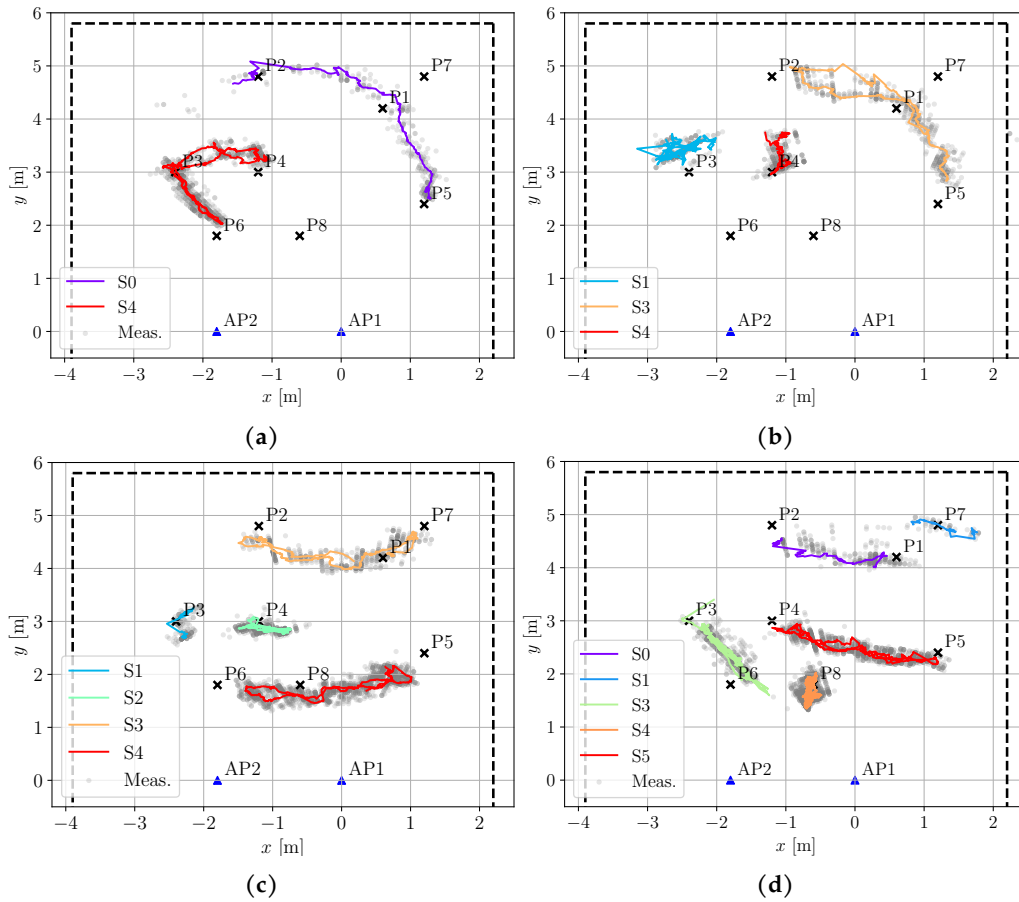
Figure 4.13: EKF trajectories obtained in the multiperson scenario. Here a single AP is used (AP1. We show four successful cases in which RAPID is able to reconstruct the movement trajectories of 2 (**a**), 3 (**b**), 4 (**c**) and 5 (**d**) people moving the the room.

(**a**) S1 running. (**b**) S2 sitting. (**c**) S3 waving.
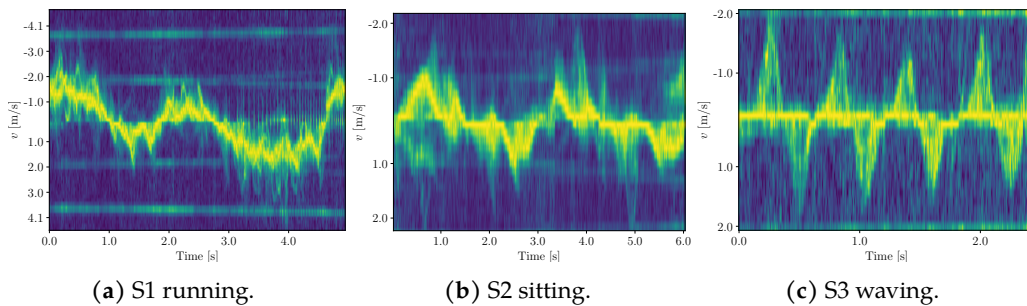
Figure 4.14: Extracted μD signatures of the subjects in Fig. 4.12.

the trajectory of S1 can only be partially reconstructed. Considering the trajectories estimated by both APs, RAPID can detect and track all subjects, successfully extracting their μD signatures, which are reported in Fig. 4.14.
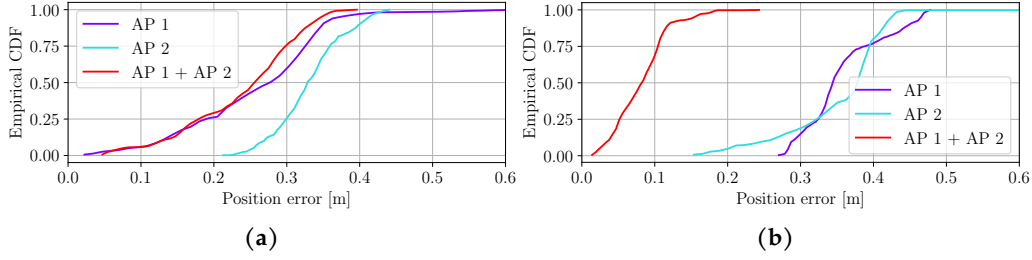
Figure 4.15: Localization error CDFs for a subject sitting in P2 (**a**) and in P4 (**b**). Combining multiple APs brings the largest improvement when their point-of-view of the subject is the most diverse.

The subject detection rate is also significantly improved by using multiple APs, as shown in the purple curve in Fig. 4.11. Despite AP1 and AP2 being placed along the same axis (x), and only 1.8 m apart, this is sufficient to increase subject detection probability by 11%, 16%, 16% and 11% for the cases of 2, 3, 4 and 5 subjects, respectively.

Finally, we show the impact of *averaging* the positions estimated by the two different APs on the localization error, see Fig. 4.15. In Fig. 4.15a we repeat the experiment described in Section 4.6.2 with a single subject sitting in position P2. Even using this simple fusion method, RAPID achieves a significant gain in the tail of the localization error distribution. A subject positioned in P2 represents a worst-case for this kind of analysis in our setting, as the locations of the APs with respect to this point are very similar in terms of distance and angle. The same experiment is repeated for position P4, showing a larger improvement from combining the APs. In this case, RAPID goes from an average localization error of 0.35 m using the single APs independently, down to an error of 0.08 m by averaging their estimates. This is due to the more favorable positions from which P4 is illuminated by the APs.

4.6.4 Activity recognition

Next, we evaluate the activity recognition performance of RAPID. For this, we collect a labeled training dataset including 6 subjects performing the 4 different activities A0 − 4 using a single RAPID-AP. The data are obtained in sequences of approximately 10 s, and split into windows of $T = 400$ time-steps (overlapping by 300 samples), which corresponds to 1.728 s, which are the input to the CNN. In total, the dataset contains around 2 minutes per activity per subject. In the training data all the activities are performed directly in front of the AP. This is intended to test the generalization capabilities of the network on more realistic measurements. The CNN classifier is trained using the cross-entropy loss function [227] and the Adam optimizer [307], with learning rate 10^{-4}, for 120 epochs.

We test the trained model on the same multi-person sequences used in Section 4.6.3, adding 6 additional sequences with a single subject, for a total of 34 sequences. Note that the subjects involved in the test are the 6 present in the training set plus 2 additional subjects never seen by the CNN classifier, in order to assess whether the learned activity features are *person independent*. We use the RAPID processing steps to extract the μD signatures of each subject's movement; when

Table 4.1: Confusion matrix for activity recognition obtained from all the test sequences, including single and multiperson measurements.

True [%]	**Predicted** [%]			
	Walking	Running	Sitting	Waving
Walking	**92.9**	0.8	6.2	0
Running	16.2	**71.6**	12.2	0
Sitting	0.2	0	**99.8**	0
Waving	3.3	0	6.8	**89.9**

using 2 APs, we implement the decision fusion scheme described in Section 4.3.5.

In Tab. 4.1 we report the confusion matrix obtained for all the predictions made by RAPID on the test sequences, when combining both APs. The average accuracy is 92.9%, with the most difficult activity to recognize being A1, running. This is mostly due to the small size of the test room that does not allow the subjects to run properly, especially in crowded scenarios with 4 or 5 subjects.

Tab. 4.2 shows the accuracy of RAPID for a varying number of people in the scene, and the gain obtained by combining the 2 APs with respect to using only AP1. We observe that the accuracy only slightly decreases when moving from 2 to 5 subjects. This shows that the proposed μD extraction process can reliably separate the contributions of the different individuals. In addition, combining multiple APs can bring a slight accuracy gain, by exploiting the different illumination angles of the devices for the subjects.

Table 4.2: RAPID activity recognition accuracy [%] vs. the number of concurrent users in the environment. We show the results obtained by the single AP1 and by the combination AP1 + AP2.

APs used	**No. of subjects present**				
	1	2	3	4	5
AP1	99.9	95.6	90.6	95.3	94.4
AP1 + AP2	100	96.1	92.7	95.4	94.4

4.6.5 Person identification

In this section we test the performance of RAPID on person identification, by building a dataset including the gait μD spectrograms of 7 subjects. We collect from 3 to 5 minutes of training data per subject. The input samples for the classifier are obtained using μD windows of the same length and overlap as for activity recognition, i.e., 1.728 s. The CNN classifier is trained using the same parameters and loss function used for activity recognition.

Table 4.3: Identification accuracy vs. number of subjects.

	No. of subjects to identify					
	2	3	4	5	6	7
Accuracy [%]	97.8	95.9	94.6	94.1	92.7	90.0

(**a**) S1 walking-waving. (**b**) S0 sitting-walking.
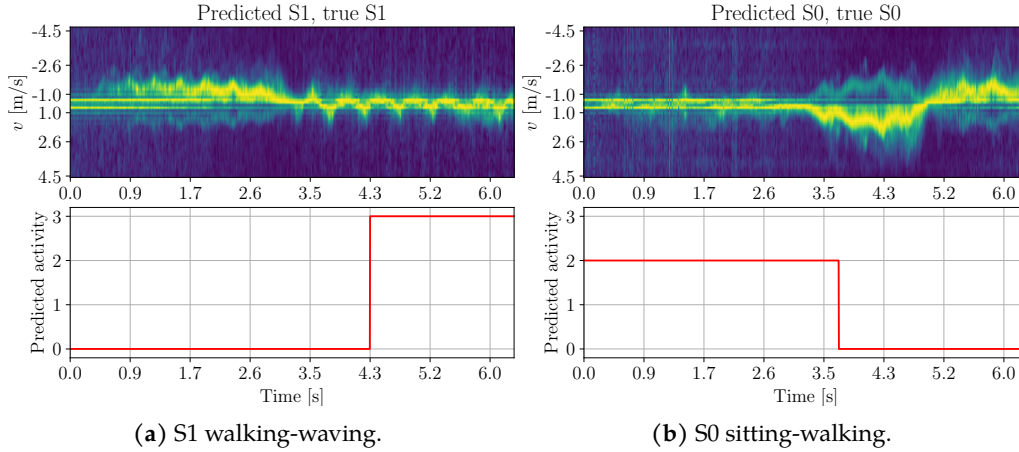
Figure 4.16: Example μD signature and corresponding CNN output for the multi-person, simultaneous activity recognition and person identification. In this case subject 0 is sitting (A2), then starts walking (A0), while subject 1 is walking and then starts waving hands (A3). RAPID is able to successfully track both subjects and recognize the activities they are performing, along with their identity.

Person identification accuracy

First, we evaluate the accuracy of person identification on a varying number of subjects to recognize. In Tab. 4.3 we report the accuracy values obtained by RAPID when increasing the number of subjects from 2 to 7. The obtained values are not significantly lower from those obtained with mm-Wave radars, and in some cases even superior, e.g., the 79% on 5 subjects in [259], the 98% with 4 subjects in [193] or the 89% with 12 subjects in [238]. This is even more valuable considering the few available training data and the short duration of the observation window used, compared to the windows used in the mentioned papers which vary between 2 and 3 s.

Simultaneous activity recognition and person identification

Finally, we show that RAPID is capable of simultaneously (*i*) tracking subjects, (*ii*) recognizing their activities, and (*iii*) identifying who is performing each activity from their gait. We perform several tests in which 2 subjects, *concurrently* present in the room, perform various activities sequentially, e.g., walking then sitting, etc. In this scenario, people tracking is of key importance to collect the temporal evolution of each subject's μD, so that all the activities performed by a person can be associated to that person's identity, obtained by RAPID when he/she is walking.

In Fig. 4.16 we show the results obtained by RAPID with 2 subjects, S0 and S1, behaving as

follows. S0 enters the scene walking, then after approximately 3.5 s S0 stops and starts waving hands, while S2 is sitting down and then starts walking after 3.5 s. We report the μD signature extracted after successfully tracking the subjects, along with the predicted activity using our moving window approach. We observe that RAPID detects the change in the activity performed by each subject; moreover, by applying the CNN on the spectrogram portion where the subjects are walking, it successfully identifies them as S0 and S1 among the 7 subjects in the training set.

4.7   Concluding remarks

In this chapter, we have designed and implemented RAPID, the first mm-Wave JCR system performing high resolution sensing of human μD signatures through standard-compliant IEEE 802.11ay packets. RAPID uses the in-packet TRN field, as specified by the 802.11ay standard, to estimate the channel impulse response. This makes it possible to perform joint tracking and localization of multiple people freely moving in an indoor environment. In addition, their μD signatures are extracted by analyzing the phase difference between subsequent packets, which allows RAPID to perform advanced sensing tasks such as simultaneous activity recognition and person identification, with radar-level accuracy. RAPID successfully combines the high resolution sensing capabilities of mm-Wave radars with the scalability and ease of deployment of existing communication hardware, allowing the seamless integration of multiple APs. We implemented multiple RAPID APs with full-duplex capabilities on an FPGA-based SDR platform equipped with phased antenna arrays, and we have thoroughly evaluated the system performance through an extensive measurement campaign. Our results show that 2 combined RAPID-APs can track up to 5 subjects concurrently moving in an indoor environment, achieving accuracies of up to 93% and 90% for activity recognition and person identification, respectively.

Future research includes the combination of our system with sub-6 GHz radios, to combine the points of strength of both frequency domains: while mm-Wave signals are ideal for localization, tracking and μD extraction, systems operating at lower frequencies can take advantage of the accurate location and tracking performance of mm-Wave radios to extract other vital parameters such as the heart rate. Other improvements are possible, such as devising additional data fusion strategies for the multiple AP setup.

5

Designing a Testbed for mm-Wave Radar Networks

5.1 Introduction and Related Works

Millimeter-Wave (mm-Wave) radars are emerging as valid alternatives to cameras for the pervasive monitoring of environments, as they combine highly accurate sensing capabilities with appealing properties of the mm-Wave spectrum such as insensitivity to extreme light conditions and to the presence of dust, smoke, or rain [261]. Furthermore, mm-Wave sensing preserves the users privacy, as no image of the subjects is captured [261]. Applications range from the estimation of human vital signs [7], to people tracking and identification [194], contact tracing [2], activity recognition [273], and many others. However, commercial mm-Wave radars have limited range (up to $6 - 8$ m) [238] and are impaired by occlusions [194], which may represent a significant drawback in crowded spaces or in the presence of cluttered furniture. For this reason, the deployment of multiple radars, i.e., *radar networks*, looks appealing to accurately monitor medium to large areas. Furthermore, the adoption of multiple sensors paves the way to a number of new sensing possibilities involving the fusion of radars data. While the first works regarding radar networks begin to appear [13], [308], in the literature there are still no experimental testbeds for the fast deployment and testing of radar network algorithms. Also, simulations have shown to be inapt for the development of human sensing algorithms, as reflections generated by humans moving in real life environments are often too complex to be accurately modeled. In [309], the authors propose a ray tracing model for radar signals propagation in radar networks, but they neither consider human targets nor the specific case of mm-Waves reflections. In [310], a simulator to estimate the Radar Cross Section (RCS) of people from computer animated human models is presented. The results are compared with measurements obtained from a real 2.4 GHz radar, but only

111

providing a qualitative analysis. Also, the authors use a primitive-based approach which works well only at low frequencies, as proven in [311], where a similar simulator is proposed. In [312], a RCS of pedestrian simulator is proposed and evaluated using a 77 GHz radar achieving acceptable results in the estimation of Range-Time (RT), Doppler-Time (DT), and Range-Doppler (RD) maps. However, all these works are tested with only one person performing simple movements (e.g., walking on a straight line) and single-input single-output radars, while at least a linear array of antennas is required for most applications. Furthermore, these simulators fail to consider some of the crucial aspects of radar measurements in real life such as the presence of furniture or walls, introducing noise, clutter, and ghost targets; the presence of multiple people, producing additional reflections and occlusions; typical real life movements, that produce a wide variety of reflections that are usually not generated by simple movements. In conclusion, currently available simulators are not suitable to develop mm-Wave radar sensing algorithms that will work in practice. For this reason, experimental testbeds are fundamentally important.

The main contributions of this work are:

1. the design of RadNet, an experimental testbed for the easy deployment and testing of mm-Wave radar networks;

2. the implementation of RadNet using commercial edge computers and mm-Wave radars;

3. the implementation of a multi-radar people tracking algorithm, to demonstrate RadNet capabilities.

The remainder of this chapter is organized as follows. Section 5.2 presents the RadNet architecture, its design/working principles and features. Section 5.3 formulates the multi-radar people tracking problem and provides the required preliminaries, while Section 5.4 illustrates the obtained results. Finally, concluding remarks are given in Section 5.5.

5.2 RadNet description

5.2.1 Architecture

RadNet adopts a *server-client* architecture [313]. Each radar is directly connected to an *edge* computer (a client node) that is capable of performing local processing and each edge computer is connected via a network connection to a *central* node (a server). The central node controls edge nodes, and the edge nodes serve as distributed computational resources to process radars data. However, users do not communicate directly with neither the central node nor the edge nodes. The user interface is provided by an additional *driver* node, which is a transparent entity designed to act as an intermediary for passing control messages to the central node. So, all control actions are provided via the driver node. Also, the driver node can be instantiated and terminated from any machine connected to the network, allowing a flexible control of the system, see Fig. 5.1.
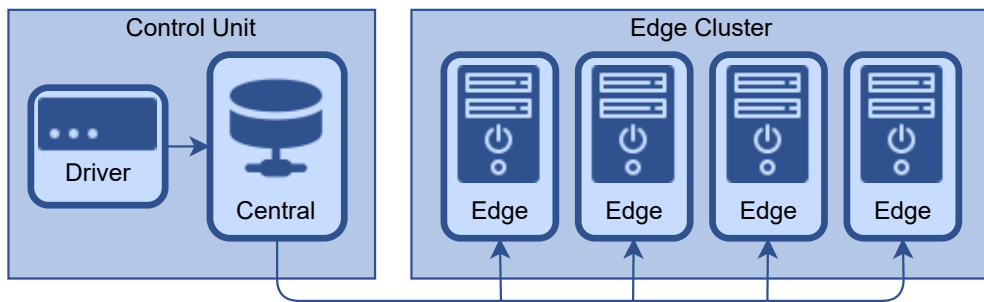
Figure 5.1: RadNet system overview.

5.2.2 System Control

RadNet works by tasks that are selected by the user through the driver node and assigned to the edge nodes by the central node.

Control Messages

The management of tasks is carried out through a message-passing interface. The aim of the system is to distribute the workload among the edge nodes, while having a centralized unit managing each of them. The connection between nodes is established and maintained by an internal exchange of messages, which is detailed later in this section.

Currently, available driver commands are:

- *ping*: to measure the latency between central and edge nodes;

- *speed*: to measure the communication speed between central node and an edge node;

- *timesync*: to verify the time synchronization between central node and an edge node;

- *open/close*: to enable/disable edge connections to the system;

- *start* and *stop*: to start and stop radar processing at the edges.

Note that additional parameters can be specified for each command, providing more flexibility. Also, the framework allows extending the functionalities provided by the commands and adding new ones.

As previously mentioned, the system works by exchange of messages between the nodes. The driver node only communicates with the central node and the central node distributes the messages to the edge nodes. An overview of the overall system is presented in Fig. 5.1. The control unit manages the system and consists of central and driver nodes. The edge cluster consists of edge nodes providing distributed processing capabilities.

Both the edge nodes and the central node send and receive messages. A message consists of a header, an action, and data, if available. After the retrieval of a message, the receiver performs a certain task and responds with another message that can contain some results related to the

current task or a simple acknowledgement. Communication between nodes is always kept alive by checking messages and responding.
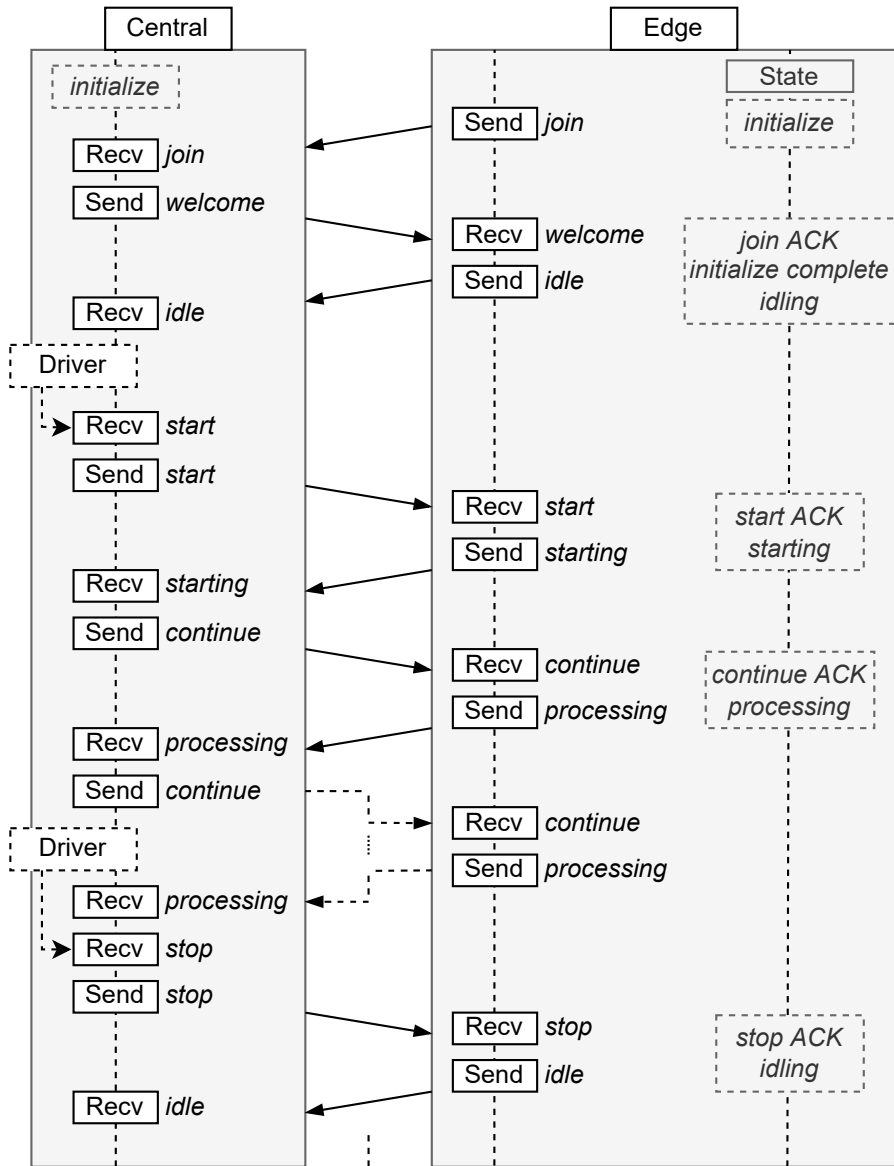
Figure 5.2: Example of communication between central and edge nodes. Send, Recv and ACK annotations in the figure correspond to sending, receiving and acknowledgement events, respectively.

Fig. 5.2 shows a typical exchange of messages from the system's startup, which will be thoroughly described next. At the beginning, the central node is initialized. Then, an edge tries to connect to the central node by sending a *join* message. Upon receiving the *join* message, the central node checks the message header and meta data which contains sender's information such as the IP address and the name of the edge node. Then, the central node proceeds to register the

edge and stores its information so that, later on, it will be able to communicate with it. If the registration is completed successfully, a *welcome* response is sent to acknowledge the edge about the success of the join operation. At this point, the edge node becomes idle and communicates this via an *idle* message. During the idle period, the edge node stays in receive mode and waits for the assignment of a new action by the central node. In the example, a *start* message is sent to the driver and forwarded to an edge device. This command tells the edge to start some radar processing (e.g., to start performing people tracking). Upon receiving *start* message, the edge starts its local process and sends back a *starting* confirmation message, which is received from the central node. Then, as long as the task is not stopped, the central node keeps sending *continue* messages to the edge node, which replies with *processing* messages. Note that *processing* messages sent by an edge node may also contain some data which comes from its processing operations, such as, for example, tracking information or raw point cloud data from the radar, depending on the required task. When a *stop* command is issued by the driver, it is forwarded to the intended edge node by the central node and the latter stops its processing operations and acknowledges the reception of the command by answering, in turn, with a *stop* confirmation message. Afterwards, the edge node goes into an idle state.

Parallelization and Data Management

Since applications based on radar networks often have to work in an online fashion, delays and latencies must be avoided or reduced as much as possible. For this purpose, RadNet is developed using multi-threading. All system modules which can work independently are deployed in a separate thread. This allows communication tasks and radar processes to continue in parallel, without interfering with one another. Also, data management is implemented in parallel to other activities. At an edge node, after retrieving and processing data from the radar, these data are placed in a temporary buffer. During processing, the communication module keeps checking the temporary buffer and, as soon as some data are available, they are sent to the central node. Similarly, at the central node, once data are received, a separate thread to save them is started, so that the system can keep working in parallel. Besides being sent, data can be stored locally at the edges or at the central node, or they can be consumed on-the-fly, without being saved.

5.2.3 Implementation Details

The backbone of RadNet framework is written in Python using socket programming and multi-threading. The communication between the nodes in RadNet is carried out via Transmission Control Protocol and Internet Protocol (TCP/IP) [314] as TCP allows data sequencing, guarantees the delivery of the transmissions and provides error checking. Thus, it is reliable and allows a simpler system implementation. In the future, we are also planning to allow users to choose between TCP and User Datagram Protocol (UDP) to send the data, as the latter may provide a faster data stream in certain situations. A dedicated (user defined) error correction protocol for data and commands may have to be implemented in this case. As already mentioned, RadNet has been designed to be modular, so that new functionalities can be added easily.

5.3 Application example

In order to test RadNet on a real application, we implemented a multi-radar multi-target tracking system on it and we used RadNet to evaluate its performance.

In the system, each radar is connected via USB to an *edge computer* (corresponding to a RadNet edge node) that performs a local target tracking, while all edge computers are connected via Ethernet to a *Fusion Center* (FC) (that is instantiated on the central node). The FC receives tracking information from the edges at regular time intervals and uses them to produce a central unified fused tracking. For the fusion, the radars relative position and orientation is required to refer all data to the same reference system. Here, we assume to have them and, in practice, we retrieve them from Ground Truth (GT) data. Methods for their automatic estimation can be found in [1], [308].

5.3.1 Preliminaries on mm-Wave indoor radar

A Multiple-Input Multiple-Output (MIMO) Frequency-Modulated Continuous-Wave (FMCW) radar jointly estimates the distance, the radial velocity, and the angular position of the targets with respect to the radar device [265]. To sense the environment, sequences of linear chirp signals with bandwidth B are sent by the radar. A full sequence is named "radar frame" and is repeated every T seconds, which is the period.

Distance, velocity and angle estimation

The targets distance, r, and velocity, v, are computed from the frequency shift induced by the delay of the reflections, usually applying Discrete Fourier Transform (DFT) processing. The range resolution of a FMCW radar is related to its bandwidth B by $\Delta r = c/(2B)$, where c is the speed of light. Thus, with a bandwidth of $2-4$ GHz, mm-Wave devices reach a resolution of a few centimeters [2], [194]. Moreover, a 2D array of multiple receiving antennas makes it possible to obtain the Angle of Arrival (AoA) of the reflections along the azimuth (θ) and elevation (φ) dimensions, by exploiting the phase shifts across different antenna elements. The azimuthal resolution depends on the number of antennas N as $\Delta\theta = \lambda/(Nd\cos\theta)$, where d represents the spacing between the antennas.

Radar point-clouds

The presence of a human in the environment generates a large number of reflections detected by the radar. These produce a set of points, termed *radar point-cloud*, that can be transformed into the 3-dimensional Cartesian space using the information about the distance, the azimuth, and the elevation angle. Each point is described by vector $[x, y, z]^T$, including its spatial coordinates x, y, z obtained transforming r, θ and φ.

116

People tracking

The common approach to people tracking from mm-Wave radar point-clouds [194], [238], [272] includes (*i*) a detection phase via density-based clustering algorithms (e.g., DBSCAN [228]) to separate the reflections from multiple subjects, and (*ii*) applying Kalman filter (KF) techniques [222] on each cluster centroid to track the movement trajectory of each subject in space. In this work, we estimate the subjects' trajectories in the (x, y) horizontal plane. The *state* of each subject at time k is defined as $\vec{\boldsymbol{x}}_k = \left[x_k, y_k, \dot{x}_k, \dot{y}_k \right]^T$, where $\dot{x}_k$ and $\dot{y}_k$ are the velocity components along the two axes.

5.3.2 Data Fusion

The FC operates in a time-slotted fashion: it considers subsequent time windows and, inside each window, it only processes the information produced during that particular time interval. The behavior of the track fusion algorithm differs in case it has to combine two sensor tracks (SS fusion) or one sensor track and a so-called central track (SC fusion), as we now explain. Considering a specific subject n, if the FC is currently not maintaining a track for it, but one or more radars are, a new central track has to be initialized based on the received information from the local sensors. Two cases may happen: (*i*) if n is tracked by one sensor only, then, the corresponding central track is initialized using its state and covariance; (*ii*) if n is tracked by more than one sensor, let us say, sensor 1 and sensor 2, then, their tracks are fused into a single central track (SS fusion). In case (*ii*), the fusion equations are

$$\vec{\boldsymbol{C}}_{0|0}^{c}(n) = \left[\vec{\boldsymbol{P}}_0^1(n) + \vec{\boldsymbol{P}}_0^2(n) \right]^{-1}, \tag{5.1}$$

$$\hat{\vec{\boldsymbol{x}}}_{0|0}^{c}(n) = \vec{\boldsymbol{C}}_{0|0}^{c}(n) \left[\vec{\boldsymbol{P}}_0^1(n) \hat{\vec{\boldsymbol{x}}}_0^1(n) + \vec{\boldsymbol{P}}_0^2(n) \hat{\vec{\boldsymbol{x}}}_0^2(n) \right], \tag{5.2}$$

where, here and in the rest of the paragraph, superscript c denotes quantities that refer to the FC. This fusion rule is typically used when the errors at the sensors are uncorrelated or the correlation can be neglected [315]. Note that, in case more than two radars are tracking the same target, the above process is sequentially repeated, firstly using sensors 1 and 2 to produce a central track, and then fusing the resulting track with the information from sensor 3 and so on until the information from all sensors (radars) is used.

If, instead, the FC has already initialized the track for a subject, the fusion has to be performed between the central track and a sensor track corresponding to the same subject. Denote by $\mathcal{T}_m^c$ and $\mathcal{T}_m^s$ the set of tracks maintained by the FC and by sensor s, respectively, at the central time step m. Upon receiving the local information $\mathcal{T}_m^s$, the FC runs a track-to-track association algorithm to find pairs of corresponding tracks $\{\mathcal{T}_m^c(n), \mathcal{T}_m^s(n)\}$. Once such pairs are available, each central track is updated with its corresponding sensor track using the *information decorrelation* method [315], [316] as follows

$$\vec{\boldsymbol{C}}_{m|m}^{c}(n) = \left[\vec{\boldsymbol{P}}_{m|m-1}^{c}(n) + \vec{\boldsymbol{P}}_m^s(n) - \bar{\bar{\boldsymbol{P}}}^s(n) \right]^{-1}, \tag{5.3}$$

117

$$\hat{\vec{x}}_{m|m}^{c}(n) = \vec{C}_{m|m}^{c}(n) \left[\vec{P}_{m|m-1}^{c}(n) \hat{\vec{x}}_{m|m-1}^{c}(n) + \right.$$
$$\left. + \vec{P}_{m}^{s}(n) \hat{\vec{x}}_{m}^{s}(n) - \bar{\vec{P}}^{s}(n) \bar{\vec{x}}^{s}(n) \right], \tag{5.4}$$

where $\bar{\vec{P}}^{s}(n)$ and $\bar{\vec{x}}^{s}(n)$ are the last communicated precision matrix and state estimate of track n from sensor s to the FC. Information decorrelation copes with the problem of correlated tracks between the FC and the radar sensors by removing the most recently received information about target n, as, otherwise, this would be accounted for twice by just summing $\vec{P}_{m|m-1}^{c}(u)$ and $\vec{P}_{m}^{s}(n)$. For more details about the fusion algorithm, including those regarding the time-slotted information processing and the tracks association, please refer to [13].

5.4 Experimental results
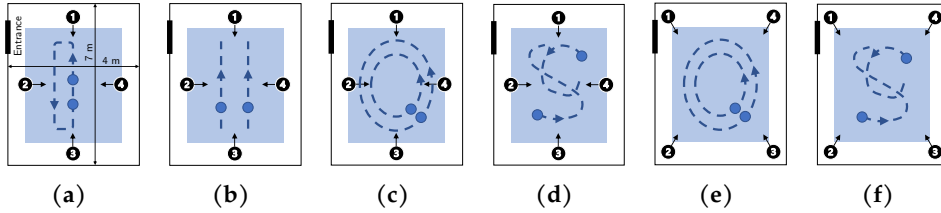
Figure 5.3: Illustration of the setups and trajectories used to test RadNet. The numbered dots represent the radar devices, with the arrow identifying their pointing directions. The blue dots represent the subjects, while the dashed blue lines represent subjects' trajectories. Fig. 5.3d and Fig. 5.3f represent free movement.

We tested RadNet using $4\times$ IWR1843BOOST Texas Instruments mm-Wave radars[1] operating in the 77-81 GHz band in real-time at a frame rate of $1/T = 15$ Hz with a Field-of-View (FOV) of $\pm 60°$ and $\pm 15°$ over the azimuth and elevation planes, respectively. Each radar was connected via USB to an edge computer. As edge computers, we used $1\times$ Jetson TX2 DevKit[2], $1\times$ Jetson Nano DevKit[3], and $2\times$ Jetson Xavier NX DevKit[4].

5.4.1 Measurement setup and Dataset

To assess the performance of the people tracking algorithm implemented using RadNet, we conducted tests in a 7×4 m research laboratory (see Fig. 5.3a) equipped with a motion tracking system featuring 10 cameras. This provides the GT 3-Dimensional (3D) localization of a set of markers placed on the radars and on the subjects with millimeter-level accuracy. We acquired 5 sequences with 1 target (1T) and 30 sequences with 2 targets (2T). Each sequence is 40 seconds long. People either moved according to predefined trajectories or moved freely, as depicted in Fig. 5.3. For single target sequences, only free movement was considered. Subjects were tracked

---

[1]https://www.ti.com/tool/IWR1843BOOST
[2]https://developer.nvidia.com/embedded/buy/jetson-tx2-developer-kit
[3]https://developer.nvidia.com/embedded/buy/jetson-nano-devkit
[4]https://developer.nvidia.com/embedded/jetson-xavier-nx-devkit

Table 5.1: Summary of tracking algorithm performance

Median error	1 radar	2 radars	3 radars	4 radars
1T-MOTA	96%	97%	96%	94%
1T-MOTP [m]	0.08	0.12	0.15	0.15
2T-MOTA	74%	77%	85%	86%
2T-MOTP [m]	0.10	0.29	0.26	0.22

by 4 radars simultaneously, deployed in 2 different setups. Then, tracking information have been fused in different ways to evaluate distinct scenarios. In particular, all possible combinations of 1 to 4 radars have been considered for the fusion, resulting in $\binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 15$ different experiments per sequence. Thus, a total of $35 \times 15 = 525$ experiments have been analyzed.

5.4.2 Evaluation metrics

To assess the tracking performance, we adopt the Multi Object Tracking Performance Accuracy (MOTA) metric, which accounts for the number of misses, false positives, and switches in the object detections, and the Multi Object Tracking Performance Precision (MOTP) metric, which represents the mean position error by considering only correctly tracked objects. More details about these metrics can be found in [317].

5.4.3 Tracking performance

Tab. 5.1 summarizes the median tracking performance, across all the configurations, see Fig. 5.3. Columns identify the various fusion cases: when only 1 radar is used, i.e., there is no fusion; when i radars are used, $i > 1$, it means that their tracks have been fused and the results refer to the fused track. Considering the 1T case, there is almost no difference between using a single radar or multiple radars. This result was expected, since the tracking of a single person in an empty room is a relatively simple task, as no occlusions can occur to impact the quality of radar tracks. A slight worsening of MOTA and MOTP is observed as the number of radars increases due to the fact that single radar target state estimations are not perfect and during the fusion process this can result in a small increment in the estimation error (the estimation noise is also fused). This is more evident in the MOTP values that go from 0.08 m to 0.15 m, which is still a small error. One may argue that, given these results, in the 1T case the use of only one radar is to be preferred with respect to more radars. This is reasonable for the very simple case of one target moving in a quite restricted area. If, instead, the target moves in a larger environment, the use of multiple radars is required to be able to track it in the entire space, as single sensors range is limited to a few meters. Considering the 2T case, it is possible to observe a clear increase in the MOTA as more radars are used for the tracking, going from 74% with 1 radar to 86% with 4 radars, providing an improvement of 12%. Situations with 2 targets are way more challenging, as they can occlude each other, possibly severely limiting the tracking capabilities of a single radar device. MOTP values remains always within a median value of 0.26 m, which is in line with state-of-the-art results [2].

5.5 Concluding Remarks

In this chapter we presented RadNet, the first testbed to facilitate the deployment and testing of algorithms for mm-Wave *radar networks*. We described its architecture and working principles, and implemented a first version of it using commercial mm-Wave radars and edge computers. As an example, we also deployed a distributed multi-radar multi-target people tracking algorithm, showing the improvement that exploiting multiple radars can bring, in terms of tracking accuracy (+12%).

6
Conclusion

The future of mm-Wave Radio Detection and Ranging (RADAR) sensing field looks promising, with the technology playing a critical role in a wide range of applications, from automotive safety systems to wireless communication. One of the significant advantages of mm-Wave RADAR sensing technology is its ability to provide high-resolution sensing in adverse weather conditions, such as heavy rain or fog, which can impede the performance of other sensing technologies such as low-frequency WiFi devices or optical cameras.

In the automotive industry, mm-Wave RADAR sensing technology is already being widely used for advanced driver assistance systems, such as collision warning, automatic emergency braking, and lane departure warning systems. As the automotive industry moves towards autonomous driving, mm-Wave RADAR sensing technology is expected to become even more critical, enabling the detection of obstacles, pedestrians, and other vehicles in real-time. Moreover, mm-Wave RADAR sensing technology is also gaining importance in the field of wireless communication, specifically for the development of Fifth generation (5G) networks. The technology enables high-bandwidth wireless communication over short distances, making it ideal for use in small cell networks and other applications that require high-speed data transfer rates. Overall, with the continued advancement in technology and increasing demand for reliable sensing solutions, the future of mm-Wave RADAR sensing technology looks bright.

In this thesis, we provide an overview of different aspect of mm-Wave indoor environment sensing. In Chapter 2, we review the literature to-date on mm-Wave technologies. In Chapters 3 and 4, we combine mm-Wave sensing technologies with Deep Learning (DL) models to provide certain Human Activity Recognition (HAR) applications. Chapter 5 provides a prototype of a communication system for RADAR networks where multiple RADARs could be utilized together. Thus, this thesis investigates the basics of mm-Wave signal systems, popular research areas for exploiting mm-Wave devices for indoor HAR applications and describes the design and implementation of a mm-Wave RADAR network.

6.1 Future of mm-Wave environment Sensing

Throughout this thesis, we analyze theoretical (including signal processing and Machine Learning), technological, and implementation (hardware and prototyping) aspects of mm-Wave sensing field for enabling a highly effective sensing of human parameters, such as position, movement and activity. Among many interesting findings, we observe that the following three research paths contain promising potential for future;

Integration of ML-based algorithms. Future mm-Wave sensing systems are expected to have the understanding of complex environments for certain applications such as object detection. While object detection is not a new field, it requires multi-modal learning experience which is what currently is missing in the mm-Wave sensing field. For this reason, developing new Neural Network (NN) architectures that are specifically designed for mm-Wave RADAR data should be investigated further. Additionally, there is a need to develop more efficient and scalable Machine Learning (ML)-based algorithms that can be implemented on resource-constrained hardware platforms. Finally, research on explainable AI methods will be crucial to ensure that the decision-making processes of ML-based algorithms are transparent and easily interpretable.

Joint Communication & Radar (JCR). The field of JCR is still in it's infancy. The main direction here would be to develop new signal processing and communication techniques that enable seamless integration of communication and sensing tasks in a single mm-Wave system. This could involve exploring joint design approaches that optimize both the communication and sensing performance, while accounting for the trade-offs between the two tasks. While doing so, the problem of minimizing the interference between various communication devices should be taken into account. Research on mm-Wave joint communication and sensing network architectures and protocols will be important to enable reliable and efficient operation of large-scale systems in real-world environments.

RADAR network. Final key direction is to improve the accuracy and robustness of sensing algorithms, particularly in complex indoor environments while using multiple RADARs. This could involve developing novel techniques for mitigating interference and noise, as well as exploring advanced synchronization and network topology methods for RADARs within such systems.

References

[1] A. Shastri, N. Valecha, E. Bashirov, *et al.*, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Communications Surveys & Tutorials*, 2022.

[2] M. Canil, J. Pegoraro, and M. Rossi, "Millitrace-ir: Contact tracing and temperature screening via mmwave and infrared sensing," *IEEE J. Sel. Top*, vol. 16, no. 2, pp. 208–223, 2022.

[3] B. van Berlo, A. Elkelany, T. Ozcelebi, and N. Meratnia, "Millimeter wave sensing: A review of application pipelines and building blocks," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10 332–10 368, 2021.

[4] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, "Rapid: Retrofitting ieee 802.11 ay access points for indoor human detection and sensing," *arXiv preprint arXiv:2109.04819*, 2021.

[5] J. Pegoraro, D. Solimini, F. Matteo, E. Bashirov, F. Meneghello, and M. Rossi, "Deep learning for accurate indoor human tracking with a mm-wave radar," in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, 2020, pp. 1–6.

[6] E. Bashirov, M. Canil, and M. Rossi, "Radnet: A testbed for mmwave radar networks," in *Proceedings of the 1st International Workshop on Emerging Topics in Wireless*, 2022, pp. 1–5.

[7] G. Paterniani, D. Sgreccia, A. Davoli, *et al.*, "Radar-based monitoring of vital signs: A tutorial overview," Mar. 2022.

[8] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Multi-resident non-contact vital sign monitoring using radar: A review," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4061–4084, 2021.

[9] D. F. Albuquerque, E. S. Goncalves, E. F. Pedrosa, F. C. Teixeira, and J. N. Vieira, "Robot self position based on asynchronous millimetre wave radar interference," in *Proc. IPIN*, 2019.

[10] D. Dupleich, R. Müller, S. Skoblikov, *et al.*, "Multi-band characterization of path-loss, delay, and angular spread in V2V links," in *Proc. IEEE PIMRC*, Sep. 2018, pp. 85–90.

[11] M. Jankiraman, *FMCW Radar Design*. Artech House, 2018.

[12] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and electronic systems*, vol. 42, no. 1, pp. 2–21, 2006.

[13] J. Pegoraro, M. Canil, A. Shastri, P. Casari, and M. Rossi, *Oracle: Occlusion-resilient and self-calibrating mmwave radar network for people tracking*, 2022.

[14] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 533–555, 2022.

[15] M. Xiao, S. Mumtaz, Y. Huang, *et al.*, "Millimeter wave communications for future mobile networks (guest editorial), Part I," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1425–1431, 2017.

[16]   X. Wang, L. Kong, F. Kong, *et al.*, "Millimeter wave communication: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1616–1653, 2018.

[17]   A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62 367–62 414, 2020.

[18]   Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Error bounds for uplink and downlink 3D localization in 5G millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 4939–4954, 2018.

[19]   F. Ghaseminajm, Z. Abu-Shaban, S. S. Ikki, H. Wymeersch, and C. R. Benson, "Localization error bounds for 5G mmWave systems under I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7971–7975, 2020.

[20]   T. S. Rappaport, S. Sun, R. Mayzus, *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[21]   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," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[22]   N. Peinecke, H. Doehler, and B. Korn, "Phong-like lighting for MMW radar simulation," in *Proc. SPIE*, Aug. 2008, pp. 1–10.

[23]   L. Subrt, P. Pechac, and S. Zvanovec, "New approach to modeling of diffuse reflection and scattering for millimeter-wave systems in indoor scenarios," *PIERS Online*, vol. 6, no. 8, pp. 719–722, 2010.

[24]   G. R. MacCartney, J. Zhang, S. Nie, and T. S. Rappaport, "Path loss models for 5G millimeter wave propagation channels in urban microcells," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3948–3953.

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

[26]   Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and J. C. Zhang, "Performance analysis of millimeter-wave cellular networks with two-stage beamforming initial access protocols," in *Proc. Asilomar conf. on signals syst. and comput.*, Nov. 2016, pp. 1171–1175.

[27]   F. Devoti, I. Filippini, and A. Capone, "Facing the millimeter-wave cell discovery challenge in 5G networks with context-awareness," *IEEE Access*, vol. 4, pp. 8019–8034, 2016.

[28]   T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. 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, 2014, invited paper.

[29]   IEEE, "IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band," *Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*, pp. 1–628, 2012.

[30]   P. Zhou, K. Cheng, X. Han, *et al.*, "IEEE 802.11ay-based mmWave WLANs: Design challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1654–1681, 2018.

[31] 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 Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[32] C. De Lima, D. Belot, R. Berkvens, *et al.*, "Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges," *IEEE Access*, vol. 9, pp. 26 902–26 925, 2021.

[33] Z. Xiao and Y. Zeng, *An overview on integrated localization and communication towards 6G*, Dec. 2021.

[34] F. Lemic, J. Martin, C. Yarp, *et al.*, "Localization as a feature of mmWave communication," in *Proc. IWCMC*, 2016, pp. 1033–1038.

[35] A. Vashist, M. P. Li, A. Ganguly, *et al.*, "KF-Loc: A Kalman filter and machine learning integrated localization system using consumer-grade millimeter-wave hardware," *IEEE Consum. Electron. Mag.*, 2022, in press.

[36] C. de Lima, D. Belot, R. Berkvens, *et al.*, *6G white paper on localization and sensing*, arXiv preprint 2006.01779, 2020.

[37] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[38] T. Kim Geok, K. Zar Aung, M. Sandar Aung, *et al.*, "Review of indoor positioning: Radio wave technology," *MDPI Applied Sciences*, vol. 11, no. 1, pp. 1–44, 2021.

[39] K. Ngamakeur, S. Yongchareon, J. Yu, and S. U. Rehman, "A survey on device-free indoor localization and tracking in the multi-resident environment," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–29, Sep. 2020.

[40] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127 150–127 174, 2021.

[41] R. C. Shit, S. Sharma, D. Puthal, *et al.*, "Ubiquitous localization (UbiLoc): A survey and taxonomy on device free localization for smart world," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3532–3564, 2019.

[42] F. Wen, H. Wymeersch, B. Peng, W. P. Tay, H. C. So, and D. Yang, "A survey on 5G massive MIMO localization," *Digital Signal Processing*, vol. 94, pp. 21–28, 2019.

[43] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A survey of enabling technologies for network localization, tracking, and navigation," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3607–3644, 2018.

[44] J. Yang, J. Xu, X. Li, S. Jin, and B. Gao, "Integrated communication and localization in millimeter-wave systems," *Frontiers of Information Tech. & Electronic Eng.*, vol. 22, no. 4, pp. 457–470, 2021.

[45] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[46] T. Wild, V. Braun, and H. Viswanathan, "Joint design of communication and sensing for beyond 5G and 6G systems," *IEEE Access*, vol. 9, pp. 30 845–30 857, 2021.

[47] J. Sanusi, O. Oshiga, S. Thomas, S. Idris, S. Adeshina, and A. M. Abba, "A review on 6G wireless communication systems: Localization and sensing," in *Proc. ICMEAS*, 2021, pp. 1–5.

[48] H. Tataria, K. Haneda, A. F. Molisch, M. Shafi, and F. Tufvesson, "Standardization of propagation models for terrestrial cellular systems: A historical perspective," *International J. of Wireless Inf. Netw.*, vol. 28, pp. 20–44, Mar. 2021.

[49] M. Shafi, J. Zhang, H. Tataria, *et al.*, "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5G cellular systems," *IEEE Communications Magazine*, vol. 56, pp. 14–20, Dec. 2018.

[50] H. J. Liebe, "An updated model for millimeter wave propagation in moist air," *Radio Science*, vol. 20, pp. 1069–1089, Sep. 1985.

[51] J. Kunisch and J. Pamp, "An ultra-wideband space-variant multipath indoor radio channel model," in *Proc. IEEE UWBST*, Nov. 2003, pp. 290–294.

[52] J. Poutanen, F. Tufvesson, K. Haneda, V.-M. Kolmonen, and P. Vainikainen, "Multi-link MIMO channel modeling using geometry-based approach," *IEEE Transactions on Antennas and Propagation*, vol. 60, pp. 587–596, Feb. 2012.

[53] A. Maltsev, A. Pudeyev, A. Lomayev, and I. Bolotin, "Channel modeling in the next generation mmWave Wi-Fi: IEEE 802.11ay standard," in *Proc. European Wireless*, May 2016, pp. 1–8.

[54] W. Keusgen, R. J. Weiler, M. Peter, M. Wisotzki, and B. Göktepe, "Propagation measurements and simulations for millimeter-wave mobile access in a busy urban environment," in *Proc. IRMMW-THz*, Sep. 2014, pp. 1–3.

[55] H. C. Nguyen, I. Rodriguez, T. B. Sorensen, L. L. Sanchez, I. Kovacs, and P. Mogensen, "An empirical study of urban macro propagation at 10, 18 and 28 GHz," in *Proc. IEEE VTC-Spring*, May 2016, pp. 1–5.

[56] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, 2015.

[57] S. Hur, Y.-J. Cho, T. Kim, *et al.*, "Wideband spatial channel model in an urban cellular environments at 28 GHz," in *Proc. EuCAP*, Apr. 2015, pp. 1–5.

[58] R. J. Weiler, M. Peter, W. Keusgen, *et al.*, "Quasi-deterministic millimeter-wave channel models in miweba," *EURASIP J. on Wireless Commun. and Networking*, vol. 2016, p. 84, Mar. 2016.

[59] C. Gustafson, F. Tufvesson, S. Wyne, K. Haneda, and A. F. Molisch, "Directional analysis of measured 60 GHz indoor radio Channels using SAGE," in *Proc. IEEE VTC-Spring*, May 2011, pp. 1–5.

[60] C. Schneider, M. Narandzic, M. Käske, G. Sommerkorn, and R. S. Thomä, "Large scale parameter for the WINNER II channel model at 2.53 GHz in urban macro cell," in *Proc. IEEE VTC*, May 2010, pp. 1–5.

[61] P. B. Papazian, C. Gentile, K. A. Remley, J. Senic, and N. Golmie, "A radio channel sounder for mobile millimeter-wave communications: System implementation and measurement assessment," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, pp. 2924–2932, Sep. 2016.

[62] K. Haneda, N. Omaki, T. Imai, L. Raschkowski, M. Peter, and A. Roivainen, "Frequency-agile pathloss models for urban street canyons," *IEEE Transactions on Antennas and Propagation*, vol. 64, pp. 1941–1951, May 2016.

[63] S. Hur, S. Baek, B. Kim, *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 454–469, Apr. 2016.

[64] S. Jaeckel, L. Jiang, V. Jungnickel, *et al.*, "Correlation properties of large and small-scale parameters from multicell channel measurements," in *Proc. EuCAP*, 2009, pp. 3406–3410.

[65] A. F. Molisch, A. Karttunen, S. Hur, J. Park, and J. Zhang, "Spatially consistent pathloss modeling for millimeter-wave channels in urban environments," in *Proc. EuCAP*, Apr. 2016, pp. 1–5.

[66] L. Raschkowski, P. Kyösti, K. Kusume, and T. Jämsä, *METIS deliverable D1.4: Channel models*, accessed: Sept. 2021, 2015.

[67] G. R. MacCartney, M. K. Samimi, and T. S. Rappaport, "Exploiting directionality for millimeter-wave wireless system improvement," in *Proc. IEEE ICCW*, 2015, pp. 2416–2422.

[68] T. A. Thomas, H. C. Nguyen, G. R. MacCartney, and T. S. Rappaport, "3D mmWave channel model proposal," in *Proc. IEEE VTC-Fall*, Sep. 2014, pp. 1–6.

[69] J. Ko, Y.-J. Cho, S. Hur, *et al.*, "Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5853–5868, Sep. 2017.

[70] M.-D. Kim, J. Liang, J. Lee, J. Park, and B. Park, "Directional multipath propagation characteristics based on 28 GHz outdoor channel measurements," in *Proc. EuCAP*, Apr. 2016, pp. 1–5.

[71] M. Shafi, A. F. Molisch, P. J. Smith, *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 1201–1221, Jun. 2017.

[72] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," *3rd Generation Partnership Project (3GPP), Tech. Rep. 38.901*, vol. 14, 2018.

[73] M. Series, "Guidelines for evaluation of radio interface technologies for IMT-2020," *Report ITU*, p. 2412, 2017.

[74] J. Poutanen, K. Haneda, L. Liu, C. Oestges, F. Tufvesson, and P. Vainikainen, "Parameterization of the COST 2100 MIMO channel model in indoor scenarios," in *Proc. EuCAP*, 2011, pp. 3606–3610.

[75] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "Quadriga-quasi deterministic radio channel generator, user manual and documentation," *Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2. 4.0*, 2020.

[76] S. Ju, Y. Xing, O. Kanhere, and T. S. Rappaport, "3-D statistical indoor channel model for millimeter-wave and sub-terahertz bands," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–7.

[77] P. F. M. Smulders, "Statistical characterization of 60-GHz indoor radio channels," *IEEE Transactions on Antennas and Propagation*, vol. 57, pp. 2820–2829, Oct. 2009.

[78] F. Erden, O. Ozdemir, and I. Guvenc, "28 GHz mmWave channel measurements and modeling in a library environment," in *Proc. IEEE RWS*, 2020, pp. 52–55.

[79] Y. Xing, O. Kanhere, S. Ju, and T. S. Rappaport, "Indoor wireless channel properties at millimeter wave and sub-terahertz frequencies," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.

[80] J. Huang, C.-X. Wang, H. Chang, J. Sun, and X. Gao, "Multi-frequency multi-scenario millimeter wave MIMO channel measurements and modeling for B5G wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2010–2025, 2020.

[81] R. J. Weiler, M. Peter, W. Keusgen, K. Sakaguchi, and F. Undi, "Environment induced shadowing of urban millimeter-wave access links," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 440–443, Aug. 2016.

[82] J. Medbo, H. Asplund, and J.-E. Berg, "60 GHz channel directional characterization using extreme size virtual antenna array," in *Proc. IEEE PIMRC*, Aug. 2015, pp. 176–180.

[83] W. Fu, J. Hu, and S. Zhang, "Frequency-domain measurement of 60 GHz indoor channels: A measurement setup, literature data, and analysis," *IEEE Instrumentation & Measurement Magazine*, vol. 16, pp. 34–40, Apr. 2013.

[84] K. Haneda, J. Järveläinen, A. Karttunen, M. Kyrö, and J. Putkonen, "A statistical spatio-temporal radio channel model for large indoor environments at 60 and 70 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 63, pp. 2694–2704, Jun. 2015.

[85] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 GHz for ultra-dense 5G wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.

[86] M. H. Tariq, I. Chondroulis, P. Skartsilas, N. Babu, and C. B. Papadias, "mmWave massive MIMO channel measurements for fixed wireless and smart city applications," in *Proc. IEEE PIMRC*, 2020, pp. 1–6.

[87] Y. Xing, T. S. Rappaport, and A. Ghosh, "Millimeter wave and sub-THz indoor radio propagation channel measurements, models, and comparisons in an office environment," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3151–3155, 2021.

[88] C. Cano, G. H. Sim, A. Asadi, and X. Vilajosana, "A channel measurement campaign for mmWave communication in industrial settings," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 299–315, 2020.

[89] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-Wave Multipath Clustering and Channel Modeling," *IEEE Transactions on Antennas and Propagation*, vol. 62, pp. 1445–1455, Mar. 2014.

[90] S. Salous, V. Degli Esposti, F. Fuschini, *et al.*, "Millimeter-wave propagation: Characterization and modeling toward fifth-generation systems," *IEEE Antennas and Propagation Magazine*, vol. 58, pp. 115–127, Dec. 2016.

[91] S. Hur, H. Yu, J. Park, *et al.*, "Feasibility of mobility for millimeter-wave systems based on channel measurements," *IEEE Communications Magazine*, vol. 56, pp. 56–63, Jul. 2018.

[92] F. Fuschini, S. Häfner, M. Zoli, *et al.*, "Item level characterization of mm-wave indoor propagation," *EURASIP J. on Wireless Commun. and Networking*, vol. 2016, no. 1, pp. 1–12, 2016.

[93] H. Tataria and F. Tufvesson, "Impact of spatially consistent channels on digital beamforming for millimeter-wave systems," in *Proc. EuCAP*, Mar. 2020, pp. 1–5.

[94] M. Steinbauer, A. Molisch, and E. Bonek, "The double-directional radio channel," *IEEE Transactions on Antennas and Propagation*, vol. 43, pp. 51–63, Aug. 2001.

[95] L. Liu, C. Oestges, J. Poutanen, *et al.*, "The COST 2100 MIMO channel model," *IEEE Wireless Communications*, vol. 19, pp. 92–99, Dec. 2012.

[96] K. Haneda *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *Proc. IEEE VTC-Spring*, May 2016, pp. 1–7.

[97] S. Ju, Y. Xing, O. Kanhere, and T. S. Rappaport, "Millimeter wave and sub-terahertz spatial statistical channel model for an indoor office building," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1561–1575, 2021.

[98] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300 GHz spectrum," in *Proc. IEEE Sarnoff Symposium*, May 2011, pp. 1–6.

[99] A. F. Molisch, *Wireless communications*. John Wiley & Sons, 2012, vol. 34.

[100] W. Roh, J.-Y. Seol, J. Park, *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, pp. 106–113, Feb. 2014.

[101] S. Wane, D. Bajon, P. Corrales, *et al.*, "Cognitive beamformer chips with smart-antennas for 5G and beyond: Holistic RFSOI technology solutions including ASIC correlators," in *Proc. EuRAD*, 2019, pp. 453–456.

[102] G. Menon, R. McMorrow, and D. Corman, "Active antennas for emerging x-band RADAR applications," in *Proc. IEEE RadarConf*, 2018, pp. 1334–1337.

[103] E. Öjefors, M. Andreasson, T. Kjellberg, *et al.*, "A 57-71 GHz beamforming SiGe transceiver for 802.11ad-based fixed wireless access," in *Proc. IEEE RFIC*, 2018, pp. 276–279.

[104] E. L. Bengtsson, "Massive mimo from a terminal perspective," Ph.D. dissertation, PhD thesis, Lund University, 2019.

[105] H. Tataria, S. Sangodoyin, A. F. Molisch, *et al.*, "Channel correlation diversity in MU-MIMO systems–analysis and measurements," in *Proc. IEEE PIMRC*, 2019, pp. 1–7.

[106] A. F. Molisch, V. V. Ratnam, S. Han, *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, pp. 134–141, Sep. 2017.

[107] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE ICC*, Jun. 2012, pp. 3724–3729.

[108] T. Kim, J. Park, J.-Y. Seol, S. Jeong, J. Cho, and W. Roh, "Tens of Gbps support with mmWave beamforming systems for next generation communications," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3685–3690.

[109] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *Proc. IEEE ICC*, Jun. 2015, pp. 1298–1303.

[110] S. Payami, M. Shariat, M. Ghoraishi, and M. Dianati, "Effective rf codebook design and channel estimation for millimeter wave communication systems," in *Proc. IEEE ICCW*, Jun. 2015, pp. 1226–1231.

[111] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmWave," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 756–770, 2019.

[112] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, pp. 3403–3418, Jul. 2018.

[113] Y. Hu, "A digital multibeam array with wide scanning angle and enhanced beam gain for millimeter-wave massive MIMO applications," *IEEE Transactions on Antennas and Propagation*, vol. 66, p. 11, 2018.

[114] P. Xingdong, H. Wei, Y. Tianyang, and L. Linsheng, "Design and implementation of an active multibeam antenna system with 64 RF channels and 256 antenna elements for massive MIMO application in 5G wireless communications," *China Commun.*, vol. 11, pp. 16–23, Nov. 2014.

[115] C. Yu, J. Jing, H. Shao, *et al.*, "Full-angle digital predistortion of 5G millimeter-wave massive MIMO transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, pp. 2847–2860, Jul. 2019.

[116] 3GPP, "Study on channel model for frequency bands above 6 GHz," *3rd Generation Partnership Project (3GPP), Tech. Rep. 38.900*, vol. 15, 2018.

[117] 3GPP, "Study on NR vehicle-to-everything," *3rd Generation Partnership Project (3GPP), Tech. Rep. 37.885*, vol. 16, 2019.

[118] 3GPP, "Study on NR positioning enhancements," *3rd Generation Partnership Project (3GPP), Tech. Rep.*, vol. 38.857, Mar. 2021.

[119] RP-213588, *Study on expanded and improved NR positioning*, accessed: Mar 2022, Dec. 2021.

[120] X. Li, E. Leitinger, M. Oskarsson, K. Astrom, and F. Tufvesson, "Massive MIMO-based localization and mapping exploiting phase information of multipath components," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 4254–4267, Sep. 2019.

[121] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[122] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–31, 2016.

[123] R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative sensor network localization: Algorithms and practical issues," *Proceedings of the IEEE*, vol. 106, no. 6, pp. 1089–1114, 2018.

[124] F. Morselli, S. Bartoletti, M. Z. Win, and A. Conti, "Localization in 5G ecosystem with Wi-Fi," in *Proc. IEEE SPAWC*, 2021, pp. 441–445.

[125] 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 Wkshps*, 2015, pp. 1–6.

[126] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[127] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques – ESPRIT," in *Proc. Asilomar conf. on signals syst. and comput.*, 1985, pp. 83–89.

[128] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, Dec. 2013.

[129] H. El-Sayed, G. Athanasiou, and C. Fischione, "Evaluation of localization methods in millimeter-wave wireless systems," in *Proc. IEEE CAMAD*, 2014, pp. 345–349.

[130] 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 *Proc. ACM MobiCom*, Oct. 2018, pp. 241–255.

[131] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "MD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proc. ACM MobiCom*, Aug. 2019, pp. 1–16.

[132] 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.

[133] 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*, 2016, pp. 1–9.

[134] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Low-complexity accurate mmwave positioning for single-antenna users based on angle-of-departure and adaptive beamforming," in *Proc. IEEE ICASSP*, 2020, pp. 4866–4870.

[135] M. Z. Comiter, M. B. Crouse, and H. T. Kung, "A data-driven approach to localization for high frequency wireless mobile networks," in *Proc. IEEE GLOBECOM*, 2017, pp. 1–7.

[136] A. Yassin, Y. Nasser, M. Awad, and A. Al-Dubai, "Simultaneous context inference and mapping using mm-Wave for indoor scenarios," in *Proc. IEEE ICC*, 2017, pp. 1–6.

[137] J. Palacios, P. Casari, and J. Widmer, "JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems," in *Proc. IEEE INFOCOM*, 2017, pp. 1–9.

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

[139] J. Palacios, P. Casari, H. Assasa, and J. Widmer, "LEAP: Location estimation and predictive handover with consumer-grade mmWave devices," in *Proc. IEEE INFOCOM*, 2019, pp. 2377–2385.

[140] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Channel estimation and indoor positioning for wideband multiuser millimeter wave systems," in *Proc. IEEE SAM*, 2020, pp. 1–5.

[141] H. Ajorloo, C. J. Sreenan, A. Loch, and J. Widmer, "On the feasibility of using IEEE 802.11ad mmWave for accurate object detection," in *Proc. ACM/SIGAPP SAC*, 2019, pp. 2406–2413.

[142] *Talon AD7200*, accessed: Sept. 2021, 2020.

[143] M. Vari and D. Cassioli, "mmWaves RSSI indoor network localization," in *Proc. IEEE ICC*, 2014, pp. 127–132.

[144] M. Pajovic, P. Wang, T. Koike-Akino, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmWave WiFi - part I: RSS and beam indices," in *Proc. IEEE GLOBECOM*, 2019, pp. 1–6.

[145] P. Hong, C. Li, H. Chang, Y. Hsueh, and K. Wang, "WBF-PS: WiGig beam fingerprinting for UAV positioning system in GPS-denied environments," in *Proc. IEEE INFOCOM*, 2020, pp. 1778–1787.

[146] O. AlHory, O. Shoushara, H. AlSuri, M. ALShunnaq, and F. Awad, "5G mmWave indoor location identification using beamforming and RSSI," in *Proc. ICICS*, 2020, pp. 91–95.

[147] A. Vashist, D. R. Bhanushali, R. Relyea, *et al.*, "Indoor wireless localization using consumer-grade 60 GHz equipment with machine learning for intelligent material handling," in *Proc. IEEE ICCE*, 2020, pp. 1–6.

[148] P. Wang, M. Pajovic, T. Koike-Akino, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmWave WiFi - Part II: Spatial beam SNRs," in *Proc. IEEE GLOBECOM*, 2019, pp. 1–6.

[149] T. Koike-Akino, P. Wang, M. Pajovic, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial MMWave WiFi: A deep learning approach," *IEEE Access*, vol. 8, pp. 84 879–84 892, 2020.

[150] P. Wang, T. Koike-Akino, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmWave WiFi: NLOS propagation," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–6.

[151] N. Maletic, V. Sark, M. Ehrig, J. Gutiérrez, and E. Grass, "Experimental evaluation of round-trip ToF-based localization in the 60 GHz band," in *Proc. IPIN*, 2019, pp. 1–6.

[152] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, "DoA-LF: A location fingerprint positioning algorithm with millimeter-wave," *IEEE Access*, vol. 5, pp. 22 678–22 688, 2017.

[153] T. .-. Tsai, L. .-. Shen, C. .-. Chiu, and K. .-. Feng, "Beam aod-based indoor positioning for 60 ghz mmwave system," in *Proc. IEEE VTC-Fall*, 2020, pp. 1–5.

[154] I. Pefkianakis and K.-H. Kim, "Accurate 3d localization for 60 ghz networks," in *Proc. ACM SenSys*, New York, NY, USA: Association for Computing Machinery, 2018, pp. 120–131.

[155] F. Fellhauer, J. Lassen, A. Jaber, N. Loghin, and S. Ten Brink, "Non-line-of-sight positioning for mmwave communications," in *Proc. IEEE SPAWC*, 2018, pp. 1–5.

[156] A. Zhou, S. Yang, Y. Yang, Y. Fan, and H. Ma, "Autonomous environment mapping using commodity millimeter-wave network device," in *Proc. IEEE INFOCOM*, 2019, pp. 1126–1134.

[157] D. Garcia, J. O. Lacruz, P. Jiménez Mateo, and J. Widmer, "POLAR: Passive object localization with IEEE 802.11ad using phased antenna arrays," in *Proc. IEEE INFOCOM*, 2020, pp. 1838–1847.

[158] A. Yassin, Y. Nasser, and M. Awad, "Geometric approach in simultaneous context inference, localization and mapping using mm-Wave," in *Proc. ICT*, 2018, pp. 159–164.

[159] A. Yassin, Y. Nasser, A. Y. Al-Dubai, and M. Awad, "MOSAIC: Simultaneous localization and environment mapping using mmWave without a-priori knowledge," *IEEE Access*, vol. 6, pp. 68 932–68 947, 2018.

[160] A. Yassin, Y. Nasser, Y. Corre, G. Gougeon, and Y. Lostanlen, "3D localization and mapping using mm-Wave: What are the opportunities in vehicular and indoor environments?" In *Proc. ICT*, 2018, pp. 324–330.

[161] 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 *Proc. IEEE INFOCOM*, 2018, pp. 2384–2392.

[162] O. Kanhere, S. Ju, Y. Xing, and T. S. Rappaport, "Map-assisted millimeter wave localization for accurate position location," in *Proc. IEEE GLOBECOM*, 2019, pp. 1–6.

[163] N. Maletic, V. Sark, J. Gutiérrez, and E. Grass, "Device localization using mmWave ranging with sub-6-assisted angle of arrival estimation," in *Proc. IEEE BMSB*, 2018, pp. 1–6.

[164] Z. Lin, T. Lv, and P. T. Mathiopoulos, "3-D indoor positioning for millimeter-wave massive MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2472–2486, Jun. 2018.

[165] J. Chen, D. Steinmetzer, J. Classen, E. Knightly, and M. Hollick, "Pseudo lateration: Millimeter-wave localization using a single RF chain," in *Proc. IEEE WCNC*, 2017, pp. 1–6.

[166] Y. Jia, H. Tian, S. Fan, and B. Liu, "Motion feature and millimeter wave multi-path AoA-ToA based 3D indoor positioning," in *Proc. IEEE PIMRC*, 2018, pp. 1–7.

[167] B. Hu, H. Tian, and S. Fan, "Millimeter wave LOS/NLOS identification and localization via mean-shift clustering," in *Proc. IEEE PIMRC*, 2019, pp. 1–7.

[168] A. Shastri, J. Palacios, and P. Casari, *Millimeter wave localization with imperfect training data using shallow neural networks*, *arXiv preprint 2112.05008*, accessed: Jan. 2022, 2020.

[169] M. Petri and M. Ehrig, "A SoC-based SDR platform for ultra-high data rate broadband communication, radar and localization systems," in *Proc. Wireless Days*, 2019, pp. 1–4.

[170] S. K. Saha, Y. Ghasempour, M. K. Haider, *et al.*, "X60: A programmable testbed for wideband 60 ghz wlans with phased arrays," *Computer Communications*, vol. 133, pp. 77–88, Sep. 2019.

[171] S. Aggarwal, U. S. Sardesai, V. Sinha, and D. Koutsonikolas, "An experimental study of rate and beam adaptation in 60 GHz WLANs," in *Proc. ACM MSWiM*, 2020, pp. 171–180.

[172] J. Zhang, X. Zhang, P. Kulkarni, and P. Ramanathan, "OpenMili: A 60 GHz software radio platform with a reconfigurable phased-array antenna," in *Proc. ACM MobiCom*, 2016, pp. 162–175.

[173] *Pi-radio*, accessed: Sept. 2021, 2020.

[174] *Zync ZCU111*, accessed: July 2021, 2020.

[175] J. O. Lacruz, D. Garcia, P. J. Mateo, J. Palacios, and J. Widmer, "Mm-flex: An open platform for millimeter-wave mobile full-bandwidth experimentation," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '20, Toronto, Ontario, Canada: Association for Computing Machinery, Jun. 2020, pp. 1–13.

[176] M. Polese, F. Restuccia, A. Gosain, *et al.*, "Millimetera: Toward a large-scale open-source mmwave and terahertz experimental testbed," in *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, ser. mmNets'19, Los Cabos, Mexico: Association for Computing Machinery, 2019, pp. 27–32.

[177] O. Kanhere and T. S. Rappaport, "Position locationing for millimeter wave systems," in *Proc. IEEE GLOBECOM*, 2018, pp. 206–212.

[178] S. Wang, J. Huang, X. Zhang, H. Kim, and S. Dey, "X-Array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays," in *Proc. ACM MobiCom*, 2020, pp. 1–14.

[179] *Airfide inc.* accessed: June 2021, 2019.

[180] 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 *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, 2017, pp. 414–425.

[181] C. R. C. M. Da Silva, J. Kosloff, C. Chen, A. Lomayev, and C. Cordeiro, "Beamforming training for IEEE 802.11ay millimeter wave systems," in *Proc. IEEE ITA*, 2018, pp. 1–9.

[182] T. S. Rappaport, Y. Xing, O. Kanhere, *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[183] S.-H. Chen and S.-K. Jeng, "An SBR/image approach for radio wave propagation in indoor environments with metallic furniture," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 1, pp. 98–106, 1997.

[184] O. Kanhere and T. S. Rappaport, "Millimeter wave position location using multipath differentiation for 3GPP using field measurements," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–7.

[185] O. Kanhere and T. S. Rappaport, "Outdoor sub-THz position location and tracking using field measurements at 142 GHz," in *Proic. IEEE ICC*, 2021, pp. 1–6.

[186] *S_5GCHANNEL*, accessed: Dec 2021, 2018.

[187] F. Gu, X. Hu, M. Ramezani, *et al.*, "Indoor localization improved by spatial context–a survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–35, Jul. 2019.

[188] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, Jun. 2016.

[189] O. Kanhere and T. S. Rappaport, "Position location for futuristic cellular communications: 5G and beyond," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 70–75, 2021.

[190] C. Fiandrino, H. Assasa, P. Casari, and J. Widmer, "Scaling millimeter-wave networks to dense deployments and dynamic environments," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 732–745, 2019.

[191] A. Davoli, G. Guerzoni, and G. M. Vitetta, "Machine learning and deep learning techniques for colocated MIMO radars: A tutorial overview," *IEEE Access*, vol. 9, pp. 33 704–33 755, 2021.

[192] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-Activity: Accurate and real-time human activity recognition via millimeter wave radar," in *Proc. IEEE ICASSP*, IEEE, 2021, pp. 8298–8302.

[193] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2021.

[194] J. Pegoraro and M. Rossi, "Real-time people tracking and identification from sparse mm-wave radar point-clouds," *IEEE Access 9*, vol. Early Access, pp. 78 504–78 520, 2021.

[195] G. Li, Z. Zhang, H. Yang, J. Pan, D. Chen, and J. Zhang, "Capturing human pose using mmWave radar," in *Proc. IEEE PerCom Workshops*, 2020, pp. 1–6.

[196] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "Mmvib: Micrometer-level vibration measurement with mmwave radar," in *Proc. ACM MobiCom*, 2020, pp. 1–13.

[197] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. J. R. Liu, "mmWrite: Passive handwriting tracking using a single millimeter wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 291–13 305, Sep. 2021.

[198] D. Nickalls, J. Wu, and N. Dahnoun, "A real-time and high performance posture estimation system based on millimeter-wave radar," in *Proc. MECO*, 2021, pp. 1–4.

[199] M. Z. Ozturk, C. Wu, B. Wang, and K. J. R. Liu, "GaitCube: Deep data cube learning for human recognition with millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 546–557, Jan. 2022.

[200] P. Zhao, C. X. Lu, B. Wang, *et al.*, "Heart rate sensing with a robot mounted mmwave radar," in *Proc. IEEE ICRA*, 2020, pp. 2812–2818.

[201] H. Cui and N. Dahnoun, "High precision human detection and tracking using millimeter-wave radars," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 1, pp. 22–32, 2021.

[202] C. Wang, K. Yang, and X. Sun, "Precise localization of concealed objects in millimeter-wave images via semantic segmentation," *IEEE Access*, vol. 8, pp. 121 246–121 256, 2020.

[203] T. K. Vodai, K. Oleksak, T. Kvelashvili, *et al.*, "Enhancement of remote vital sign monitoring detection accuracy using multiple-input multiple-output 77 GHz FMCW radar," *IEEE J. Electromagn., RF Microw. Med. Biol.,*, pp. 1–1, 2021, in press.

[204] C. X. Lu, S. Rosa, P. Zhao, *et al.*, "See through smoke: Robust indoor mapping with low-cost mmWave radar," in *Proc. ACM MobiSys*, 2020, pp. 14–27.

[205] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," Texas Instruments, Tech. Rep. SPYY005A, 2017, pp. 1–8.

[206] A. E. Omer, S. Safavi-Naeini, R. Hughson, and G. Shaker, "Blood glucose level monitoring using an FMCW millimeter-wave radar sensor," *Remote Sensing*, vol. 12, no. 3, pp. 1–25, 2020.

[207] A. E. Omer, G. Shaker, S. Safavi-Naeini, K. Murray, and R. Hughson, "Glucose levels detection using mm-wave radar," *IEEE Sensors Lett.*, vol. 2, no. 3, pp. 1–4, 2018.

[208] J. Pegoraro, F. Meneghello, and M. Rossi, "Multi-person continuous tracking and identification from mm-wave micro-doppler signatures," *arXiv*, no. 2003.03571, 2020.

[209] T. Gu, Z. Fang, Z. Yang, P. Hu, and P. Mohapatra, "Mmsense: Multi-person detection and identification via mmwave sensing," in *Proc. ACM mmNets*, 2019, pp. 45–50.

[210] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital sign and sleep monitoring using millimeter wave," *ACM Trans. on Sensor Networks*, vol. 13, no. 2, pp. 1–32, 2017.

[211] M. Alizadeh, G. Shaker, and S. Safavi-Naeini, "Remote health monitoring system for bed-bound patients," in *Proc. IEEE BIBE*, 2020, pp. 801–806.

[212] J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10 893–10 902, 2021.

[213] R. Zhang and S. Cao, "Real-time human motion behavior detection via CNN using mmWave radar," *IEEE Sensors Journal*, vol. 3, no. 2, pp. 1–4, 2018.

[214] G. Tiwari and S. Gupta, "An mmWave radar based real-time contactless fitness tracker using deep CNNs," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17 262–17 270, 2021.

[215] F. Jin, R. Zhang, A. Sengupta, *et al.*, "Multiple patients behavior detection in real-time using mmWave radar and deep CNNs," in *Proc. IEEE RadarConf*, 2019, pp. 1–6.

[216] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "Mm-pose: Real-time human skeletal posture estimation using mmWave radars and CNNs," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 032–10 044, 2020.

[217] F. Jin, A. Sengupta, and S. Cao, "mmFall: Fall detection using 4-D mmWave radar and a hybrid variational RNN autoencoder," *IEEE Transactions on Automation Science and Engineering*, pp. 1–13, 2020, in press.

[218] J. Wu, H. Cui, and N. Dahnoun, "A novel high performance human detection, tracking and alarm system based on millimeter-wave radar," in *Proc. MECO*, 2021, pp. 1–4.

[219] J. B. Allen and L. R. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.

[220] L. Cohen, *Time-frequency analysis*. Prentice Hall, 1995, vol. 778.

[221] V. Chen, D. Tahmoush, and W. Miceli, *Radar Micro-Doppler Signatures: Processing and Applications*. IET, 2014.

[222] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[223] G. Welch, G. Bishop, *et al.*, "An introduction to the kalman filter," University of North Carolina at Chapel Hill, Dept. of Computer Science, Tech. Rep. TR 95-041, 1995, accessed: Sept. 2021.

[224] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proc. IEEE ACC*, vol. 2, 2002, pp. 887–892.

[225] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE AS-SPCC*, 2000, pp. 153–158.

[226] T. M. Mitchell, *Machine learning*. McGraw-Hill, 1997.

[227] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[228] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of AAAI KDD-96*, Portland, Oregon, 1996, pp. 226–231.

[229] V. Estivill-Castro, "Why so many clustering algorithms: A position paper," *ACM SIGKDD explorations newsletter*, vol. 4, no. 1, pp. 65–75, 2002.

[230] D. Kellner, J. Klappstein, and K. Dietmayer, "Grid-based DBSCAN for clustering extended objects in radar data," in *Proc. IEEE IVS*, 2012, pp. 365–370.

[231] U. Hasson, S. A. Nastase, and A. Goldstein, "Direct fit to nature: An evolutionary perspective on biological and artificial neural networks," *Neuron*, vol. 105, no. 3, pp. 416–434, 2020.

[232] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[233] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[234] S. Chang, Y. Zhang, F. Zhang, *et al.*, "Spatial attention fusion for obstacle detection using mmwave radar and vision sensor," *Sensors*, vol. 20, no. 4, pp. 1–21, 2020.

[235] R. J. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural computation*, vol. 2, no. 4, pp. 490–501, 1990.

[236] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[237] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[238] P. Zhao, C. X. Lu, J. Wang, *et al.*, "Mid: Tracking and identifying people with millimeter wave radar," in *Proc. of IEEE DCOSS*, Santorini, Greece, 2019, pp. 33–40.

[239] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, pp. 3371–3408, Dec. 2010.

[240] S. Wagner and W. Johannes, "Target detection using autoencoders in a radar surveillance system," in *Proc. IEEE RadarConf*, 2019, pp. 1–5.

[241] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[242] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015, pp. 448–456.

[243] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE CVPR*, 2017, pp. 652–660.

[244] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, *Pointnet++: Deep hierarchical feature learning on point sets in a metric space*, arXiv preprint 1706.02413, 2017.

[245] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. IEEE CVPR*, 2019, pp. 9613–9622.

[246] Z. Meng, S. Fu, J. Yan, *et al.*, "Gait recognition for co-existing multiple people using millimeter wave sensing," in *Proc. AAAI Conf. on Artificial Intelligence*, 2020, pp. 849–856.

[247] D.-S. Lee, S. Yeom, J.-Y. Son, and S.-H. Kim, "Automatic image segmentation for concealed object detection using the expectation-maximization algorithm," *Optics express*, vol. 18, no. 10, pp. 10 659–10 667, 2010.

[248] S. Yeom, D.-S. Lee, J.-Y. Son, *et al.*, "Real-time outdoor concealed-object detection with passive millimeter wave imaging," *Optics express*, vol. 19, no. 3, pp. 2530–2536, 2011.

[249] T. Wei and X. Zhang, "Mtrack: High-precision passive tracking using millimeter wave radios," in *Proc. ACM MobiCom*, 2015, pp. 117–129.

[250] B. Kapilevich, Y. Pinhasi, M. Anisimov, B. Litvak, and D. Hardon, "FMCW mm-wave non-imaging sensor for detecting hidden objects," in *Proc. IEEE MTT-S IMWS*, 2011, pp. 101–104.

[251] B. Kapilevich, B. Litvak, and A. Shulzinger, "Passive non-imaging mm-wave sensor for detecting hidden objects," in *Proc. IEEE COMCAS*, 2013, pp. 1–5.

[252] J. Bhatia, A. Dayal, A. Jha, *et al.*, "Object classification technique for mmWave FMCW radars using range-FFT features," in *Proc. COMSNETS*, 2021, pp. 111–115.

[253] S. Bakhtiari, S. Liao, T. Elmer, A. Raptis, *et al.*, "A real-time heart rate analysis for a remote millimeter wave IQ sensor," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1839–1845, 2011.

[254] J. Lien, N. Gillian, M. E. Karagozler, *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. on Graphics*, vol. 35, no. 4, pp. 1–19, 2016.

[255] C. Xu, Z. Li, H. Zhang, *et al.*, "Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface," in *Proc. ACM MobiSys*, 2019, pp. 14–26.

[256] T.-Y. J. Kao, Y. Yan, T.-M. Shen, A. Y.-K. Chen, and J. Lin, "Design and analysis of a 60-GHz CMOS doppler micro-radar system-in-package for vital-sign and vibration detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1649–1659, 2013.

[257] *TI mmWave radar sensors*, accessed: Mar. 2022.

[258] *Mmwcas-rf-evm*, accessed: Mar. 2022.

[259] B. Vandersmissen, N. Knudde, A. Jalalvand, *et al.*, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[260]   M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709–1723, Feb. 2018.

[261]   S. A. Shah and F. Fioranelli, "Rf sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 11, pp. 26–44, 2019.

[262]   N. Knudde, B. Vandersmissen, K. Parashar, *et al.*, "Indoor tracking of multiple persons with a 77 ghz mimo fmcw radar," in *European Radar Conference (EURAD)*, Nuremberg, Germany, Oct. 2017.

[263]   V. Winkler, "Range doppler detection for automotive fmcw radars," in *European Radar Conference (EuRAD)*, Munich, Germany, Oct. 2007.

[264]   S. Amiriparian, M. Freitag, N. Cummins, and B. Schuller, "Sequence to sequence autoencoders for unsupervised representation learning from audio," in *Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, Munich, Germany, Nov. 2017.

[265]   S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, Mar. 2017.

[266]   D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1263–1278, 2015.

[267]   P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[268]   H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[269]   T. Wagner, R. Feger, and A. Stelzer, "Radar signal processing for jointly estimating tracks and micro-doppler signatures," *IEEE Access 5*, pp. 1220–1238, 2017.

[270]   K. P. Murphy, "Switching kalman filters," *Technical report, DEC/CompaqCambridgeResearchLabs*, 1998.

[271]   A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-doppler signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2629–2640, 2019.

[272]   Z. Meng, S. Fu, J. Yan, *et al.*, "Gait recognition for co-existing multiple people using millimeter wave sensing," 01, New York, USA, Apr. 2020, pp. 849–856.

[273]   A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. ACM mmNets*, Los Cabos, Mexico, 2019, pp. 51–56.

[274]   F. Restuccia, *Ieee 802.11bf: Toward ubiquitous wi-fi sensing*, 2021. arXiv: 2103.14918.

[275]   Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[276]   S. D. Regani, B. Wang, and K. R. Liu, "Wifi-based device-free gesture recognition through-the-wall," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 8017–8021.

[277]   F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello, and M. Rossi, "Environment and person independent activity recognition with a commodity ieee 802.11 ac access point," *arXiv preprint arXiv:2103.09924*, 2021.

[278] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[279] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.

[280] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, IEEE, 2017, pp. 1230–1239.

[281] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "Mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, IEEE, 2020, pp. 2400–2409.

[282] F. Zhang, C. Wu, B. Wang, and K. R. Liu, "Mmeye: Super-resolution millimeter wave imaging," *IEEE Internet of Things Journal*, 2020.

[283] 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 Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[284] J. O. Lacruz, R. Ruiz, and J. Widmer, "A real-time experimentation platform for sub-6 ghz and millimeter-wave MIMO systems," in *ACM MobiSys'21*, 2021.

[285] H. Li, X. He, X. Chen, Y. Fang, and Q. Fang, "Wi-motion: A robust human activity recognition using wifi signals," *IEEE Access*, vol. 7, pp. 153 287–153 299, 2019.

[286] B. Korany, H. Cai, and Y. Mostofi, "Multiple people identification through walls using off-the-shelf wifi," *IEEE Internet of Things Journal*, 2020.

[287] A. Pokkunuru, K. Jakkala, A. Bhuyan, P. Wang, and Z. Sun, "Neuralwave: Gait-based user identification through commodity wifi and deep learning," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2018, pp. 758–765.

[288] W. Jiang, C. Miao, F. Ma, *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 289–304.

[289] Z. Shi, J. A. Zhang, Y. D. R. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, 2020.

[290] G. Lai, X. Lou, and W. Ye, "Radar-based human activity recognition with 1-d dense attention network," *IEEE Geoscience and Remote Sensing Letters*, 2021.

[291] A. Loch, H. Assasa, J. Palacios, J. Widmer, H. Suys, and B. Debaillie, "Zero overhead device tracking in 60 ghz wireless networks using multi-lobe beam patterns," New York, NY, USA: Association for Computing Machinery, 2017.

[292] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. R. Liu, "Mmwrite: Passive handwriting tracking using a single millimeter wave radio," *IEEE Internet of Things Journal*, 2021.

[293] Y. Ren, J. Lu, A. Beletchi, *et al.*, "Hand gesture recognition using 802.11 ad mmwave sensor in the mobile device," in *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, IEEE, 2021, pp. 1–6.

[294] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "Vimo: Multi-person vital sign monitoring using commodity millimeter wave radio," *IEEE Internet of Things Journal*, 2020.

139

[295] Y. Zhu, Y. Zhu, B. Y. Zhao, and H. Zheng, "Reusing 60ghz radios for mobile radar imaging," New York, NY, USA: Association for Computing Machinery, 2015.

[296] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[297] A. Nambiar, A. Bernardino, and J. C. Nascimento, "Gait-based person re-identification: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–34, Apr. 2019.

[298] D. Garcia, J. O. Lacruz, P. Jiménez Mateo, and J. Widmer, "Polar: Passive object localization with ieee 802.11ad using phased antenna arrays," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 1838–1847.

[299] M. I. Ribeiro, "Kalman and extended kalman filters: Concept, derivation and properties," *Institute for Systems and Robotics*, vol. 43, p. 46, 2004.

[300] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *11th international conference on information fusion (FUSION)*, IEEE, Cologne, Germany, Jun. 2008, pp. 1–6.

[301] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, Dec. 2009.

[302] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," in *International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, May 2016.

[303] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.

[304] I. 8. working group, "Ieee draft standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks-specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications-amendment: Enhanced throughput for operation in license-exempt bands above 45 ghz," *IEEE P802.11ay/D3.0*, 2019.

[305] W.-C. Liu, F.-C. Yeh, T.-C. Wei, C.-D. Chan, and S.-J. Jou, "A Digital Golay-MPIC Time Domain Equalizer for SC/OFDM Dual-Modes at 60 GHz Band," en, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 10, p. 10, 2013.

[306] SIVERSIMA, *Evk06002 development kit*, https://www.siversima.com/product/evk-06002-00/, 2020.

[307] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *ICLR: International Conference on Learning Representations*, 2015, pp. 1–15.

[308] A. Shastri, M. Canil, J. Pegoraro, P. Casari, and M. Rossi, "Mmscale: Self-calibration of mmwave radar networks from human movement trajectories," in *Proc. of IEEE RadarConf22*, New York, USA, 2022, pp. 1–6.

[309] D. Gubelli, O. A. Krasnov, and O. Yarovyi, "Ray-tracing simulator for radar signals propagation in radar networks," in *Proc. of EuRAD*, Nuremberg, Germany, 2013, pp. 73–76.

[310] S. Sundar Ram and H. Ling, "Simulation of human microdopplers using computer animation data," in *Proc. of IEEE RadarConf*, Rome, Italy, 2008, pp. 1–6.

[311] A. D. Singh, S. S. Ram, and S. Vishwakarma, "Simulation of the radar cross-section of dynamic human motions using virtual reality data and ray tracing," in *Proc. of IEEE RadarConf*, Oklahoma City, USA, 2018, pp. 1555–1560.

[312] Y. Deep, P. Held, S. S. Ram, *et al.*, "Radar cross-sections of pedestrians at automotive radar frequencies using ray tracing and point scatterer modelling," *IET Radar, Sonar Navig.*, vol. 14, 833–844(11), 6 Jun. 2020.

[313] A. Berson, *Client/Server Architecture (2nd Ed.)* USA: McGraw-Hill, Inc., 1996.

[314] B. A. Forouzan and S. C. Fegan, *TCP/IP Protocol Suite*, 2nd. McGraw-Hill Higher Education, 2002.

[315] C.-Y. Chong, S. Mori, W. Barker, and K.-C. Chang, "Architectures and algorithms for track association and fusion," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 15, no. 1, pp. 5–13, 2000.

[316] K. Chang, R. Saha, and Y. Bar-Shalom, "On optimal track-to-track fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 4, pp. 1271–1276, 1997.

[317] K. Bernardin, A. Elbs, and R. Stiefelhagen, "Multiple object tracking performance metrics and evaluation in a smart room environment," in *Proc. of IEEE Int. Wkshp. on Vis. Surveillance*, vol. 90, 2006.

List of publications

Journals

[1] A. Shastri, N. Valecha, E. Bashirov, *et al.*, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Communications Surveys & Tutorials*, 2022.
Main contribution of the author is the Section 2.6 where the Device-free radar-enabled mmWave localization and sensing techniques are discussed

[4] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, "Rapid: Retrofitting ieee 802.11 ay access points for indoor human detection and sensing," *arXiv preprint arXiv:2109.04819*, 2021.
Main contributions of the author are the management of measurement campaigns (Sec. 3.5), the preprocessing of the signal and modelling of the NN (Sec. 3.4.1)

Conferences

[5] J. Pegoraro, D. Solimini, F. Matteo, E. Bashirov, F. Meneghello, and M. Rossi, "Deep learning for accurate indoor human tracking with a mm-wave radar," in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, 2020, pp. 1–6.
Main contributions of the author are HAR section of the RAPID design (Sec. 4.3.5)

[6] E. Bashirov, M. Canil, and M. Rossi, "Radnet: A testbed for mmwave radar networks," in *Proceedings of the 1st International Workshop on Emerging Topics in Wireless*, 2022, pp. 1–5.

Acknowledgments

I would like to express my gratitude and appreciation to several individuals and groups who have supported me throughout my research journey.

First and foremost, I would like to thank my supervisor, Michele Rossi, for his guidance, support, and encouragement throughout my Ph.D. His expertise, insight, and patience have been invaluable to me.

I am also deeply grateful to my co-supervisor Joerg Widmer from IMDEA Networks (Spain, Madrid), for his mentorship and valuable inputs from his expertise and my co-supervisors Mikko A. Uusitalo and Dani Korpi (Finland, Espoo) for their guidance and support from Nokia Bell Labs.

In addition, I would like to acknowledge the ESRs, supervisors, and research group from the ITN MINTS project, for their support, collaboration, and encouragement throughout my research journey.

I would also like to express my appreciation to everyone in the UNIPD DEI department, especially the SIGNET research group, for their support, advice, and assistance.

Finally, I would like to thank my family for their unwavering support, love, and encouragement throughout my studies.