WIRELESS COMMUNICATION AND SENSING SYSTEM DESIGN: A LEARNING-BASED APPROACH

By

Shichen Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science—Doctor of Philosophy

2025

ABSTRACT

With the rapid advancement of digital technologies, wireless communication and sensing systems have become increasingly integral to our daily lives. These systems utilize wireless signals not only as data carriers but also as a medium for radio sensing. Model-based approaches have traditionally been a popular choice for addressing existing challenges in communication and sensing. However, model-based approaches struggle to accurately characterize signal propagation, especially at higher frequencies, and optimizing them for communication is even more difficult. Moreover, extracting human motion-related information from these complex signals is often challenging with conventional methods. Recent progress in artificial intelligence (AI) has opened new avenues for addressing these challenges. This thesis explores learning-based approaches to uncover the hidden information embedded within wireless signals. By doing so, it aims to enhance the efficiency of wireless communication systems and enable fine-grained human motion sensing, thereby pushing the boundaries of wireless systems.

The first part of this thesis explores the capability of various RF signals to sense different levels of human motion using learning-based approaches. We begin by proposing AuthIoT, a gesture-based wireless authentication scheme designed for IoT devices. AuthIoT leverages a convolutional neural network (CNN) to learn human gesture features from Wi-Fi channel state information (CSI) and maps them to specific letters for device authentication. To enhance robustness and enable gesture recognition across diverse environments, the system employs a feature fusion approach that integrates location-independent features, ensuring strong transferability. Next, we shift our focus to tiny motions and propose RadSee, a system capable of recognizing fine-grained handwriting. We develop a 6 GHz FMCW radar system along with a tailored deep neural network to identify handwritten letters through walls. The model combines a bidirectional long short-term memory (BiLSTM) network with an attention mechanism to leverage temporal dependencies and capture critical features—such as turning points—in radar phase sequences for accurate recognition. We push the limits of this system further with a novel learning framework and introduce RadEye, a system designed to recognize eye movements. Given the subtle nature of eye motion and the challenge of detecting it in RF signals, we adopt a transformer encoder as the feature extractor to more effectively exploit temporal dependencies in the phase sequences. To further enhance performance, we incorporate a state-of-the-art vision-based method to provide guidance and prior knowledge during the learning process.

The second part of this thesis focuses on leveraging learning-based solutions to improve the efficiency of wireless communication systems, with particular emphasis on enhancing the throughput of mmWave communication systems. We begin by proposing an uplink multi-user MIMO (MU-MIMO) mmWave communication (UMMC) scheme for WLANs. MU-MIMO techniques are well-known for increasing network efficiency and throughput. A key innovation in this work is a learning-based Bayesian optimization (BayOpt) framework for joint beam search across multiple antennas. This approach eliminates the need for complex channel modeling and identifies optimal beamforming directions with only a few search iterations, significantly reducing beamforming overhead. We then further explore the beamforming problem in mmWave communications, shifting our focus to mobile mmWave networks. In such dynamic environments, beamforming overhead becomes more pronounced. To address this challenge, we leverage the temporal correlation of wireless channels to aid in beam selection. Specifically, we propose a Temporal Beam Prediction (TBP) scheme that enables a mobile mmWave device to predict its future beam direction based on its historical beam selection profile. At the core of this scheme is a modified LSTM architecture, complemented by an adversarial learning model to improve the robustness and generalizability of the beam steering process.

This thesis presents efficient communication schemes and novel sensing applications based on learning-driven approaches, paving the way for the design of AI-enabled next-generation wireless communication and sensing systems. It provides detailed descriptions of system implementations, experimental setups, and performance evaluations of the proposed schemes in real-world environments. Furthermore, it offers an in-depth analysis of the limitations of these systems and discusses open challenges in developing future wireless communication and sensing systems using learning-based techniques.

To my parents, Lixia and Yu, and to my wife, Jinghan

ACKNOWLEDGMENTS

I sincerely thank my advisor, Prof. Huacheng Zeng, for his professional guidance, consistent support, and mentorship during my Ph.D. journey. His invaluable insights, high standards, and encouragement have played a crucial role in shaping the direction of my research and in helping me grow as a scholar. I hold deep respect for his conscientious and meticulous approach to research, which has become a model I aspire to follow in my own future work.

I am truly grateful to my committee members—Prof. Li Xiao, Prof. Qiben Yan, Prof. Zhichao Cao, and Prof. Zhaojian Li—for their insightful feedback, constructive input, and thoughtful guidance. Their expertise and perspectives have consistently helped refine my research direction and sharpen my overall approach. Their support has been instrumental in the successful completion of this thesis.

I would like to thank both past and present colleagues from the INSS Lab—Dr. Pedram Kheirkhah Sangdeh, Dr. Hossein Pirayesh, Qijun Wang, Peihao Yan, Bowei Zhang, and Jie Lu—for their collaboration and support throughout my research projects. I am also deeply grateful for the time spent with my lab mates, which brought great joy and meaning to my daily life. Their support has extended beyond research, enriching both my academic journey and personal life.

I would like to extend my sincere thanks to the faculty, staff, and fellow students in the Department of Computer Science and Engineering (CSE) at Michigan State University (MSU) for their support and assistance throughout the completion of my Ph.D. thesis. The intellectually stimulating and supportive environment fostered by the CSE department has played a vital role in advancing my research and contributing to my scholarly growth.

Finally, and most importantly, I would like to thank my mother, Lixia Cui, and my father, Yu Zhang, for their unconditional love and unwavering support. Even when we were separated by distance, their love reached me across oceans. I am deeply grateful to my wife, Jinghan Liu, who crossed oceans to stand by my side. Her unwavering support and countless sacrifices helped me overcome the challenges of my academic journey and turned our love into a lifelong commitment.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

With the prevalence of Internet of Things (IoT) devices, wireless signals have become present in every corner of people's lives. In today's digitized world, there is a growing demand not only for higher data rates in wireless systems but also for these systems to capture and convey information about the physical world. On one hand, wireless systems are advancing into higher frequency bands—such as the mmWave spectrum—enabling new applications like high-quality wireless VR/AR headsets, high-resolution video streaming, and vehicle-to-vehicle (V2V) communications. On the other hand, these systems are evolving into sensors that can capture human motion-related information. These new sensors leverage widely available wireless signals to provide a contactless, privacy-preserving, and resilient sensing approach that functions effectively in low-light and adverse weather conditions. Wireless sensing applications are becoming increasingly common in our daily lives, including measuring vital signs [61, 167], monitoring sleep patterns [195], assisting the elderly with memory-related tasks [46], detecting falls in older adults [164], and even measuring soil moisture [36, 48].

Both evolutionary trends require accurate signal transmission models. In wireless communication systems, such as mmWave communication systems, a reliable transmission model is essential for guiding the beamforming process. Similarly, wireless sensing applications depend on accurate models to describe how human motion affects wireless channels. However, using traditional modeling approaches to achieve these goals is often difficult or inefficient. For example, mmWave communication systems face challenges due to hardware imperfections, such as phase noise, clock jitter, and inaccurate antenna radiation patterns, all of which contribute to imprecise mathematical models. Modeling human motion based on variations in wireless signals is even more challenging. The presence of multipath effects and the often uncorrelated relationship between human motion and signal variation patterns make model-based approaches infeasible.

We will begin by introducing the research background of designing wireless communication and sensing systems, and outline the limitations of existing approaches. Next, we will present the system we developed and explain how we address these limitations using learning-based methods. Finally, we will provide an overview of the organization of this thesis.

1.1: Research Background

1.1.1: Wireless Sensing Systems

Sensing with Communication Signals. Designing wireless sensing systems generally follows two main trends: utilizing existing communication signals or employing dedicated sensing signals. Using existing communication signals for sensing repurposes current infrastructure and spectrum, increases resource utilization, and transforms communication systems into multifunctional platforms. Recently, many applications have emerged in this area. For example, existing works have leveraged Wi-Fi signals for gesture recognition [93, 108, 160, 213] and vital sign detection [167]. Cellular signals have also been used for respiration sensing [47] and even soil moisture measurement [48]. Additionally, low-power signals such as RFID have been applied to touch sensing [122].

These sensing capabilities also show potential for addressing other emerging challenges. One such challenge, which this thesis focuses on, is authentication for IoT devices. Since IoT devices often lack input interfaces, connecting them to Wi-Fi networks can be difficult—users have no straightforward way to input passwords. Existing solutions typically rely on pre-deployed platforms to serve as intermediaries for authentication. However, this thesis explores how existing wireless connections can be leveraged for authentication through gesture recognition, without requiring any additional equipment.

Sensing with Radar Signals. While sensing with communication signals offers advantages by leveraging existing infrastructure, it faces limitations when extended to fine-grained human sensing in complex scenarios. As a result, another design trend focuses on sensing systems based on Frequency-Modulated Continuous Wave (FMCW) radar. Compared to communication signal-based sensing, radar systems are coherent, meaning they do not suffer from hardware-induced phase, frequency, or timing misalignments commonly found in wireless communication systems. Existing work has demonstrated the use of radar to enable human body skeleton sensing [95, 209–

211]. However, these studies primarily target large-scale body motions. The potential of radar systems for detecting fine movements—such as handwriting or eye motion—remains largely unexplored.

In this thesis, we focus on two compelling problems using FMCW radar systems. The first is detecting handwriting through walls, which raises significant security concerns. Through-the-wall detection remains a major challenge for most other sensors, such as cameras or acoustic sensors. Although some RF-based sensing solutions can penetrate walls, achieving the resolution necessary to capture subtle movements like handwriting remains difficult. Furthermore, handwriting detection is particularly susceptible to interference from other moving objects in the environment. The second and more challenging problem we address is the detection of eye motion. Eye tracking is critical in various applications, including human-computer interaction (HCI), virtual reality, and medical diagnostics. While existing camera-based eye-tracking solutions offer high accuracy and usability, they often raise privacy concerns and perform poorly in low-light conditions. Radar-based solutions present a promising alternative, addressing these issues while offering high-resolution, reliable tracking.

1.1.2: Wireless Communication Systems

With the growing demand for data traffic in daily life, wireless communication systems are moving to higher frequency bands—the mmWave spectrum—to support larger bandwidths. This shift is foundational for 5G and beyond, enabling the vision of a smart society and a digitized physical world by delivering ultra-low latency, multi-gigabit per second (Gbps) scalable wireless connections. These capabilities are essential for emerging applications such as virtual reality (VR), cloud-centric real-time AI, and high-resolution video streaming. However, mmWave frequencies suffer from significant path loss, making reliable communication challenging. To address this, mmWave systems heavily rely on analog beamforming to establish and maintain strong links. Beam selection, therefore, becomes a critical challenge in mmWave communications. In this thesis focuses on reducing beamforming overhead across different mmWave communication scenarios, while also proposing efficient communication schemes to enhance performance in mmWave systems.

The combination of mmWave and MU-MIMO technologies has attracted significant interest from both academia and industry due to its potential to deliver data rates in the hundreds of gigabits per second. While extensive research has been conducted on the downlink of mmWave MU-MIMO systems, progress on the uplink remains limited. One of the primary challenges in designing uplink MU-MIMO schemes is the complexity of beamforming across multiple antennas. Existing solutions often rely on accurate antenna models and detailed channel state information (CSI), which are difficult to obtain in real-world deployments. Some model-free approaches have been proposed, but they primarily address single-antenna scenarios. These solutions are not suitable for multi-antenna systems, where joint beamforming decisions must consider spatial correlations and user interference.

Beamforming is not only a challenge in multi-antenna systems, but also in single-antenna scenarios in mobile mmWave networks, where rapid user movement necessitates frequent beam alignment. To address this issue, existing solutions have explored techniques such as out-of-band CSI-assisted beam selection, compressive sensing, and hierarchical beam search. While these methods are effective to some extent, they primarily exploit the spatial characteristics of mmWave channels and often overlook the temporal correlation inherent in beam selection over time. Leveraging this temporal consistency could significantly improve the efficiency and robustness of beamforming in dynamic environments.

1.2: Thesis Contributions

1.2.1: Thesis Overview

This thesis encompasses five of my previously published works, each of which has contributed to the development of the core chapters. The thesis is structured in two parts. The first part focuses on wireless sensing system design: Chapter 2 is based on AuthIoT [200], Chapter 3 builds upon RadSee [201], and Chapter 4 is derived from RadEye [202]. The second part focuses on wireless communication system design, where Chapter 5 is based on UMMC [199] and Chapter 6 is de-

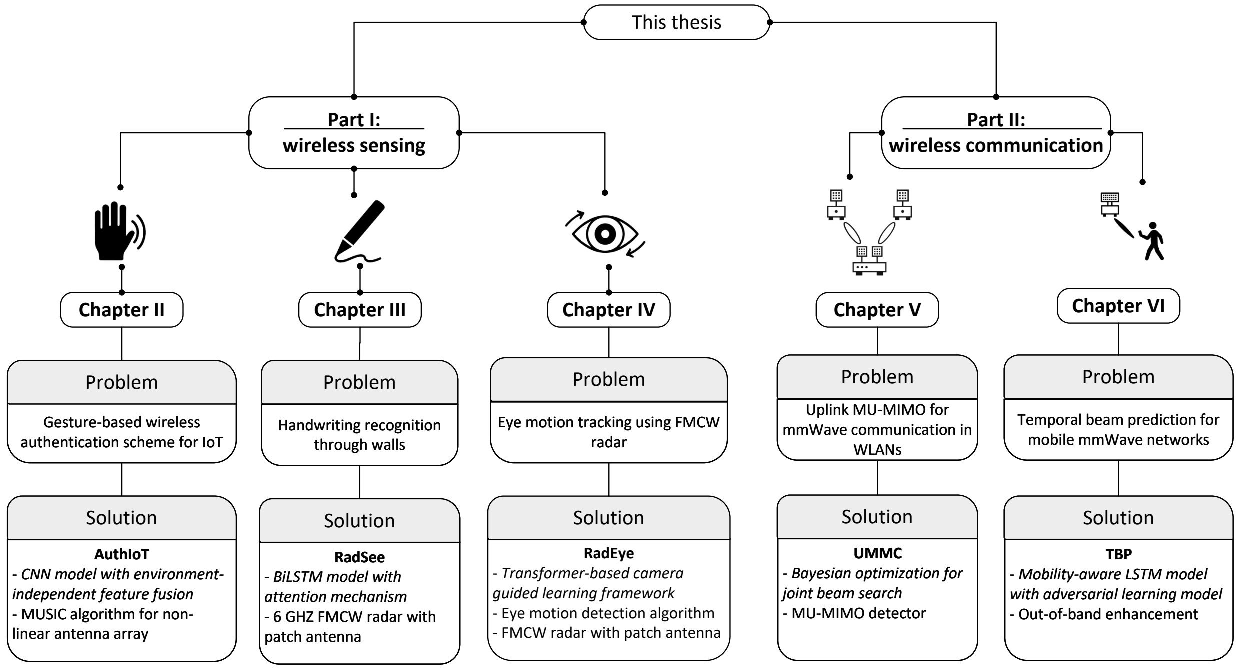

This thesis
Part I:
wireless sensing
Part II:
wireless communication
Chapter II
Chapter III
Chapter IV
Chapter V
Chapter VI
Problem
Gesture-based wireless authentication scheme for IoT
Problem
Handwriting recognition through walls
Problem
Eye motion tracking using FMCW radar
Problem
Uplink MU-MIMO for mmWave communication in WLANs
Problem
Temporal beam prediction for mobile mmWave networks
Solution
AuthIoT
- CNN model with environment-independent feature fusion
- MUSIC algorithm for non-linear antenna array
Solution
RadSee
- BiLSTM model with attention mechanism
- 6 GHZ FMCW radar with patch antenna
Solution
RadEye
- Transformer-based camera guided learning framework
- Eye motion detection algorithm
- FMCW radar with patch antenna
Solution
UMMC
- Bayesian optimization for joint beam search
- MU-MIMO detector
Solution
TBP
- Mobility-aware LSTM model with adversarial learning model
- Out-of-band enhancement

Figure 1.1: Overview of this thesis.

veloped from TBP [203]. More specifically, as illustrated in Fig. 1.1, the first part of this thesis focuses on designing deep learning models to extract human motion features from RF signals. It aims to recognize various types of motion and apply them to the following use cases:

- A gesture-based wireless authentication scheme for IoT devices [200]
- Handwriting recognition through walls using FMCW radar [201]
- Eye motion tracking using FMCW radar [202]

The second part of this thesis focuses on enhancing the efficiency of mmWave communication systems. In particular, we explore learning-based approaches, such as Bayesian optimization and tailored LSTM models, to predict beam directions, and thereby reduce mmWave beamforming overhead across various networking scenarios. The newly designed beamforming schemes are as follows:

- A Bayesian optimization-based beamforming framework for uplink MU-MIMO mmWave communication [199]
- Temporal beam prediction for mobile mmWave networks [203]

In both directions, we explore learning-based approaches to advance wireless systems into a new generation characterized by high throughput and multifunctionality. In the next section, we provide a comprehensive overview of each proposed system, detailing their design components and the specific roles they play in addressing emerging challenges.

1.2.2: Contributions to Wireless Sensing

AuthIoT. We propose AuthIoT, a gesture-based wireless authentication scheme for IoT devices. It directly utilizes channel state information (CSI) from Wi-Fi communications to recognize input passwords, without relying on additional platforms. A novel feature fusion scheme is designed to maintain the system's transferability across different environments. Specifically, we extract an environment-independent feature — the Angle of Arrival (AoA) — and fuse it with channel amplitude to serve as input for the DNN. In addition, we design an extended 2D MUSIC algorithm tailored to this scheme to accurately calculate AoA under various antenna configurations on the access point (AP) side. We have built a prototype of AuthIoT and evaluated its performance in real-world scenarios. Experimental results show that AuthIoT achieves a letter recognition accuracy of 84%.

RadSee. We propose RadSee, a 6 GHz Frequency Modulated Continuous Wave (FMCW) radar system designed to detect handwriting content through walls. The system is developed through a combination of hardware and software design. On the hardware side, RadSee features a 6 GHz FMCW radar equipped with patch antennas. These patch antennas provide a sufficient link power budget, enabling the system to "see" through most walls while operating at low transmission power. On the software side, the system extracts phase features corresponding to the writer's hand movements and utilizes a BiLSTM model with an attention mechanism to classify the letters. The proposed learning framework is specifically designed to identify and extract key features—particularly the turning points in handwritten letters—that are critical for accurate recognition. Extensive experimental results show that RadSee achieves a 75% letter recognition accuracy when subjects write 62 randomly selected letters.

RadEye. We propose RadEye, a radar system capable of detecting fine-grained human eye movements from a distance. RadEye is realized through an integrated hardware and software co-design. It leverages a customized sub-6 GHz FMCW radar and a tailored patch antenna pair to detect millimeter-level eye movements. This hardware combination enables the system to detect subtle motions over an extended range while also minimizing interference from other directions. On the software side, a DNN is employed to enhance detection accuracy, guided by camera-based supervisory training. The DNN incorporates a transformer encoder as the feature extractor, enabling it to effectively capture temporal dependencies between radar sampling points. We have developed a prototype of RadEye, and extensive experimental results demonstrate that it achieves 90% accuracy in detecting human eye rotation directions (up, down, left, and right) across a variety of scenarios.

1.2.3: Contributions to Wireless Communication

UMMC. We propose an efficient Uplink MU-MIMO mmWave Communication (UMMC) scheme for WLANs, which enables multiple stations to simultaneously transmit their data packets to a single access point (AP). A key component of this scheme is a Bayesian optimization (BayOpt) framework, designed to guide the beam search process. BayOpt leverages the posterior probability distribution derived from previously evaluated beam configurations to intelligently explore the beam space. Compared to conventional exhaustive search methods, BayOpt demonstrates remarkable efficiency, often identifying near-optimal beam directions within a constrained airtime budget. In addition to the learning framework, the proposed scheme incorporates a novel MU-MIMO detector capable of decoding asynchronous data packets from multiple user devices. We have developed a prototype of UMMC on a mmWave testbed and evaluated its performance through a combination of over-the-air experiments and extensive simulations. Both experimental and simulation results confirm the effectiveness and efficiency of UMMC in practical network environments.

TBP. We propose the Temporal Beam Prediction (TBP) scheme, which assists mobile mmWave devices in predicting future beam directions based on their historical beam selection profiles. TBP draws inspiration from pedestrian trajectory prediction, employing a Long Short-Term Memory

(LSTM) network to model and predict beam directions in mobile mmWave networks. At the core of TBP is a tailored LSTM module—mobility-aware LSTM (mLSTM)—specifically designed to handle the non-uniform and non-smooth characteristics often observed in mmWave beam angle sequences. An adversarial learning structure is also employed to enhance the system's generalizability across different users. We have implemented a prototype of TBP on a 60 GHz software-defined radio (SDR) mmWave testbed. Experimental results demonstrate that TBP can improve throughput by more than 60% compared to existing beam selection approaches across various scenarios.

1.3: Organization

The rest of the thesis is organized as follows: Chapter 2 presents a gesture-based wireless authentication scheme for IoT devices. Chapter 3 introduces RadSee, a 6 GHz FMCW radar system capable of recognizing handwriting through walls. Chapter 4 presents RadEye, which extends RadSee by incorporating computer vision techniques to recognize eye motions. Chapter 5 describes UMMC, a learning-based beamforming scheme designed to reduce beamforming overhead for uplink MU-MIMO communication in mmWave WLANs. Chapter 6 presents TBP, a deep learning framework that reduces beamforming overhead for mobile mmWave networks. Finally, Chapter 7 summarizes this thesis and outlines future directions from both application and technical perspectives.

CHAPTER 2: A GESTURE-BASED WIRELESS AUTHENTICATION SCHEME FOR IOT DEVICES

2.1: Introduction

The Internet of Things (IoT) has transformed various aspects of our society, playing a vital role in enhancing the way we live and work. According to Statista [145], the number of IoT devices worldwide is projected to reach 40 billion by 2033. In real-world applications, many IoT devices rely on Wi-Fi connections for Internet access and have no input interfaces (e.g., keypad or touchscreen) due to their limits in physical size, power consumption, and/or manufacturing cost. For example, smart home devices such as Gosund Smart Wi-Fi power outlet [11], SYLVANIA Wi-Fi dimmable LED light bulb, and AGSHOME Wi-Fi windows open alert sensors require Wi-Fi network access to be functional, but they have no input interfaces which end users can use to type in Wi-Fi passcode for wireless Internet access. With the proliferation of compact wireless sensors in smart environments, wireless IoT devices lacking input interfaces are expected to become increasingly common.

One widely used approach to authenticate Wi-Fi-enabled IoT devices that lack input interfaces involves leveraging existing platforms such as Google Home Assistant [10] and Amazon Alexa [8]. These platforms facilitate device recognition and authentication via a smartphone or computer connected to the same Wi-Fi network. This method, however, requires end users to have a smartphone/computer with pre-installed proprietary apps such as Google Home and Amazon Alexa. It also requires Internet connection to gain the support of Google or Amazon cloud services. These requirements make this method inapplicable in the scenarios where a smartphone or Internet is not available and where the IoT device owners do not want to get involved in commercial cloud platforms.

In this chapter, we present AuthIoT, a gesture-based wireless authentication scheme for IoT

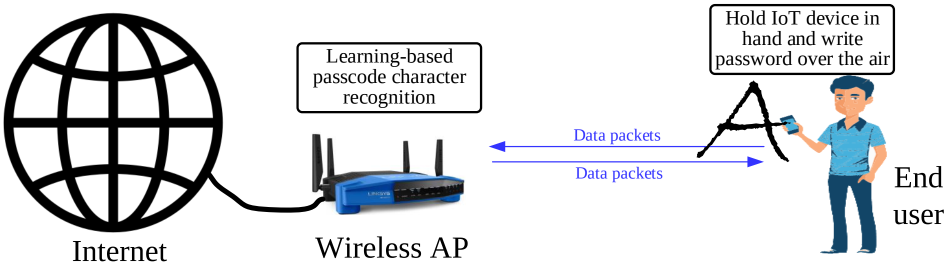

Learning-based passcode character recognition
Hold IoT device in hand and write password over the air
Data packets
Data packets
Internet
Wireless AP
End user

Figure 2.1: A CSI-based authentication scheme for wireless IoT devices without input interfaces.

Table 2.1: Wireless writing and gesture recognition.

Ref.	(Tx,Rx) ant #	Nonlinear antenna array	Dataset	main approach	Learning features	computation complexity	cross-environment transferability	Reported accuracy
WriFi [51]	(2,3)	No	26 capital letters	GMM-HMMs	CSI amplitude	High	No	87%
WiReader [59]	(1,2)	No	26 capital letters	LSTM model	CSI amplitude	Medium	No	90%
LetFi [198]	(1,6)	No	26 capital letters	SOM network	CSI amplitude	Medium	No	95%
WiDraw [147]	(30,3)	No	Any	Trajectory tracking	AoA	Medium	Yes	91%
Wi-Wri [33]	(2,3)	No	26 capital letters	kNN model	CSI amplitude	High	No	82%
AuthIoT	(1,3 or 4)	Yes	48 characters	CNN-based learning	LoS AoA, CSI amplitude	Medium	Yes	84%

devices without input interfaces. AuthIoT requires neither assistance from other devices nor support from an Internet-based software platform. It is a channel state information based (CSI-based) passcode recognition scheme for a Wi-Fi communication system, as shown in Fig. 2.1. It consists of an access point (AP), an IoT device, and an end user. Specifically, AuthIoT works as follows: The end user holds the IoT device in hand and writes the passcode over the air; and the AP leverages recent advances in deep learning to recognize the passcode input from the IoT device based on the spatial and temporal CSI features.

A key challenge in the design of AuthIoT is to maintain its transferability across different environments. As CSI is significantly affected by the multipath effect of a wireless channel, a wireless AP tends to observe different CSI in different environments. Hence, at the wireless AP, using raw CSI for passcode recognition is not a plausible strategy because a deep neural network (DNN) trained with raw CSI in an environment does not work well in another environment (based on our experimental results). To address this challenge, AuthIoT extracts environment-independent features as the input for the training and inference of a DNN. Specifically, AuthIoT computes the angle of arrival (AoA) of the line-of-sight (LoS) signal path by leveraging recent advances in wireless localization [79, 85, 152, 158, 185], and uses the AoA (as well as normalized channel amplitude) as the input for the training and inference of DNN. Since different passcode characters tend to produce distinct AoA patterns, an AP can identify the characters if the DNN is properly trained.

Another challenge in the design of AuthIoT is to compute the LoS AoA of received packets for an AP with a nonlinear antenna configuration. While AoA estimation of wireless packets has been studied in wireless localization (e.g., [79, 85, 152, 158, 183, 185]), most of existing techniques deal with the case where the antenna elements are equally-spaced and linearly installed. However, many Wi-Fi routers and other APs are equipped with antenna elements in a nonlinear shape so as to save space. Existing methods such as MUSIC (MUltiple SIgnal Classification) algorithm cannot be directly used to estimate AoA for a receiving device with nonlinear antenna configuration. To address this challenge, AuthIoT extends two-dimensional MUSIC algorithm to the case where the receiver (wireless AP) is equipped with nonlinear antenna elements. Following the idea from SpotFi [85], AuthIoT jointly considers the AoA and ToF (time of flight) to enhance the AoA resolution of different signal paths.

Based on the environment-independent features (LoS AoA) as well as the normalized amplitude of CSI, AuthIoT employs a DNN to recognize the passcode when an end user continuously writes the passcode characters over the air by holding the IoT device in her hand. Once the AP detects the passcode, it will grant the network access to the IoT device; otherwise, it will wait until the correct passcode is detected or the maximum number of attempts is reached. We have built a prototype of AuthIoT and evaluated its performance on two distinct AP testbeds: i) Intel 5300 Wi-Fi card with three linear antennas, and ii) USRP N310 with four nonlinear (square-positioned) antennas. Experimental results show that AuthIoT achieves 84% successful rate of passcode character recognition on the former testbed and 83% successful rate on the latter testbed, both for cross-environment applications.

The contributions of this work are summarized as follows.

- AuthIoT is, to the best of our knowledge, among the first that explores environment-independent features of CSI for authenticating IoT devices without input interfaces. It is transferable to a new environment for handwriting recognition once its DNN is well trained.

- AuthIoT extends two-dimensional MUSIC algorithm for AoA estimation from linear, equally-spaced antenna configuration to nonlinear antenna configuration.
- We have built a prototype of AuthIoT and demonstrated its performance in real scenarios. Our experimental results show that it can achieve more than 83% passcode recognition accuracy in cross environments for both linear and nonlinear antenna configurations.

2.2: Related Work

We survey the literature in the following category.

Authenticating IoT Devices without Input Interface. As mentioned before, a mainstream authentication method for smart-home IoT devices is to leverage the platforms such as Google Home [10] and Amazon Alex [8]. This method, however, requires users to have a smartphone with pre-installed proprietary apps, to have Internet access, and to share the data with the platforms. In addition to the commercial products, research advances have been made for IoT authentication.

TouchAuth [189] harnesses induced body electric potentials (iBEPs) for IoT authentication by having users wear a wristband to touch an analog-to-digital (ADC) pin of the IoT device. It makes the ADC pin touchable by connecting devices' ADC pins to their conductive exteriors. The authentication is performed by measuring the IBEPs similarities between the wristband and the smart object. P2Auth [96] authenticates IoT devices without input interface by leveraging their inertial measurement unit. It requires users to perform unique petting operations that can be sensed by both an IoT device and a wristband device. It compares the captured data from the two devices and makes a decision for the authentication based on their similarity. SFIRE [55] is a secret-free trust establishment protocol that pairs commercial wireless devices with a hub. It requires a user to move a helping smartphone around the wireless device and measures the similarity of RSS signals for authentication. Move2Auth [197] is another authentication scheme for IoT devices without an input interface. It requires users to hold a smartphone and perform one of two hand-gestures in front of an IoT device.

In contrast to the above works, AuthIoT takes a very different approach to authenticate IoT devices without input interface. It requires neither assistance from smartphones nor hardware/software modifications on IoT devices.

CSI-based Handwriting Recognition. Our work is closely related to the research in this area. Table 2.1 presents a comparison of our work with prior work. WriFi [51] is a CSI-based handwriting system that comprises a Wi-Fi AP, a Wi-Fi client device, and a user writing 26 letters over the air. In this system, CSI amplitude is collected for learning-based recognition. Operations such as principal component analysis (PCA) and fast fourier transform (FFT) have been performed to extract the CSI features for hidden Markov model (HMM) training and inference. The accuracy is reported to be 86%. Similar to WriFi, Wi-Wri [33] is another CSI-based handwriting letter recognition system. It is based on k-nearest neighbors (k-NN) model and uses dynamic time warping (DWT) to calculate the distance between CSI waveform and classified data. It reports 83% recognition accuracy for 26 letters. WiReader [59] is another work in this area. It exploits CSI from commercial Wi-Fi devices to extract activities-related information. It employs long short-term memory (LSTM) model for recognition and adopts PCA and discrete wavelet transform (DWT) for CSI feature extraction. It reports 90% recognition accuracy for 26 letters with intelligence text correction. LetFi [198] is also a CSI-based over-the-air handwriting recognition system in Wi-Fi networks. It employs multi-domain feature extraction method and self-organizing mapping neural networks (with SoftMax regression classifier) to recognize 26 letters. The reported recognition accuracy is 95%. WiDraw [147] is a handwritten recognition system which allows a user to write over the air. It recognizes hand movement trajectory based on the analysis of collected CSI. With the presence of 30 transmitters, it can achieve 91% word recognition accuracy and superior accuracy for hand movement patterns.

As shown in Table 2.1, AuthIoT differs from the above works in several aspects: i) AuthIoT has a larger dataset (48 characters in AuthIoT versus 26 letters in the above-mentioned works); ii) it enables its cross-environment transferability by design; and iii) it works for Wi-Fi AP with nonlinear antenna array.

CSI-based Gesture Recognition. In addition to handwriting recognition, many works have also been done for CSI-based gesture recognition [93, 108, 160, 179, 213], by extracting and recognizing the temporal, spatial, and Doppler features of hand movements. Generally speaking, CSI-based gesture recognition can achieve very high accuracy (over 90%), because it has a small number of dataset (e.g., 6 gestures). In contrast, AuthIoT has 48 characters in its dataset, which is much larger than the above networks. In addition, AuthIoT distinguishes itself from previous works by focusing on cross-environment transferability design.

Wireless Localization. Another research line related to our work is CSI-based wireless localization in Wi-Fi networks [85, 139, 152, 183, 185]. Particularly, SpotFi [85] presents an accurate indoor localization scheme using commercial Wi-Fi devices. It proposes a two-dimensional MUSIC algorithm by leveraging the information in both spectral and spatial domains to enhance the resolution of AoA estimation. It jointly estimates AoA and ToF (time of flight) of incoming Wi-Fi signals using multiple antennas and broadband (40MHz) spectrum. The localization median accuracy is reported to 40cm using the commercial Wi-Fi card. AuthIoT borrows the idea of AoA estimation from the above works, and extends the antenna setting from linear to nonlinear case for IoT authentication applications.

2.3: AuthIoT: Design Overview

2.3.1: System Setting and Operation

AuthIoT is designed for a wireless communication system as shown in Fig. 2.1, which comprises a wireless AP (e.g., Wi-Fi router), an IoT device, and an end user. IoT devices do not have input interfaces such as keypads and touchscreens due to the limits in their physical size, power consumption, and/or manufacturing cost. Examples of such IoT devices include Wi-Fi LED light bulbs, Wi-Fi light switches [9], and window/door open alert sensors [12]. The wireless AP has multiple antennas for data packet reception. This is very common for Wi-Fi routers, most of which are equipped with four or more antennas. In such a system, AuthIoT works as follows.

- **End User:** The end user first triggers wireless AP to exchange packets between itself and the IoT device at a certain rate (e.g., 200 packets/s). She then holds the IoT device in front of the wireless AP with a distance of about 2 meters and ensures that there is a LoS signal path between the IoT device and the wireless AP. After that, the end user writes each of the passcode characters over the air until the IoT device is successfully authenticated.

- **IoT Device:** The IoT device needs no hardware or software modification. It responds to the sounding packets from the wireless AP (e.g., using ACK packets) so that the wireless AP can estimate wireless channel at a desired rate.
- **Wireless AP:** The wireless AP estimates the channel between itself and the IoT device using the packets from the IoT device. It continuously runs a modified MUSIC algorithm to estimate the LoS AoA of the packets from the IoT device and feed the LoS AoA along with normalized CSI amplitude to a DNN for the recognition of each character in the passcode. It authenticates the IoT device once the passcode is detected or the maximum number of attempts is reached.

2.3.2: Challenges and Our Approach

Compared to prior CSI-based recognition work [33, 51, 59, 147, 198], AuthIoT needs to recognize a much larger set of characters, which include upper-case letters, low-case letters, numbers, and special characters. In addition, AuthIoT faces the following challenges in its design and implementation.

Cross-Environment Transferability. A challenge in the design of AuthIoT is to maintain its cross-environment transferability, so that the system can be used in any environment once its DNN has been trained. To address this challenge, AuthIoT uses environment-independent CSI features as its input for passcode recognition. Specifically, it computes the LoS AoA of the received packets from the IoT device based on the estimated CSI by leveraging recent advances in wireless localization [85, 152, 183, 185], and uses the LoS AoA as the main feature for passcode recognition. It should be noted that an end user can always hold the IoT device in front of its wireless AP to ensure the existence of LoS path between the IoT device and its AP.

Nonlinear Antenna Array at AP. Although the LoS AoA estimation techniques have been well studied for wireless localization, most of them consider the case where the receiver is equipped with linearly, equally-spaced antenna array [85, 152, 183, 185]. However, many off-the-shelf wireless APs such as Wi-Fi routers are equipped with nonlinear antenna array (e.g., rectangular-installed) to save space. As expected, the AoA estimation techniques proposed for a device with linear antenna array cannot be directly used for a device with nonlinear antenna array. To address this challenge, AuthIoT revisits the MUSIC algorithm and extends it for the case where the device has nonlinear antenna array. AuthIoT also borrows the idea from SpotFi [85] to jointly estimate AoA and ToF so as to improve the AoA resolution.

Indistinguishable Characters. Another challenge lies in the fact that some character pairs are hard to distinguish in their handwriting format, such as “z” and “Z”, “o” and “O”, “s” and “S”, “v” and “V”, letter “I” and number “1”, etc. Sometimes, these handwritten character pairs even cannot be distinguished by a human. Unfortunately, this challenge is hard to address from a technical perspective. Therefore, AuthIoT resorts to regulation. AuthIoT asks end users to use a passcode that does not include indistinguishable pairs of characters. Excluding some characters will not compromise the passcode security as there are still sufficient characters to be used.

2.3.3: Security of AuthIoT

Essentially, AuthIoT serves as an interface for an AP to receive a passcode from an end user for authenticating a particular IoT device. It does not alter the authentication mechanism and thus has the same authentication safety as existing methods. However, due to the broadcast nature of wireless signals, AuthIoT may face the passcode leakage problem. A malicious user may overhear the signal from IoT device and attempt to infer the passcode for AP access. To address this issue, a substitution cipher [176] can be applied to the passcode at wireless AP, and the substitution rules can be updated regularly to avoid replay attacks.

2.4: AoA Estimation for General Antenna Array

This section first offers a primer on the existing MUSIC algorithm for AoA estimation at a wireless device equipped with uniform linear antenna array, and then extends the MUSIC algorithm to the case where the wireless device is equipped with a general (linear or nonlinear) antenna array.

2.4.1: MUSIC for Uniform Linear Antenna Array (MUSIC-ULAA)

System Modeling. The basic idea of AoA estimation is that different signal propagation paths are likely to have different AoAs at a receiving device. The different AoAs will introduce a corresponding phase shift across the array of antennas. For a uniform linear antenna array, once the antenna space and the phase shift are given, the AoA can be accordingly calculated. To understand AoA estimation, let us consider a receiving device with a uniform linear antenna array as shown in Fig. 2.2, where the number of antennas is M, and the antenna spacing is d. Assume that the number of signal propagation paths is L and let us focus on the lth path shown in the figure. Denote α_l as the complex channel attention experienced by the signal when impinging on the first antenna. Then, the complex channel attention of the signal at the second antenna is the same except for an additional phase shift caused by the additional distance traveled by the signal. Mathematically, the additional phase shift at the mth antenna can be written as $(m-1) \cdot d \cdot \sin(\theta_l) \cdot \frac{2\pi}{\lambda}$, where λ is the wavelength of radio signal. Then, the complex channel attention at the mth antenna can be expressed as $\frac{(m-1) \cdot d \cdot 2\pi}{\lambda} \cdot \sin(\theta_l) \cdot \alpha_l$. Denote $\vec{h}_l$ as the channel coefficient vector for the lth path. Then, $\vec{h}_l = \vec{a}(\theta_l) \cdot \alpha_l$, where

$$\vec{a}(\theta_l) = \begin{bmatrix} 1 & e^{-j\frac{2\pi d \sin(\theta_l)}{\lambda}} & e^{-j\frac{4\pi d \sin(\theta_l)}{\lambda}} & \cdots & e^{-j\frac{2\pi (M-1) d \sin(\theta_l)}{\lambda}} \end{bmatrix}^{\mathsf{T}}. \tag{2.1}$$

At each antenna of the device, the observed CSI is the blend of all paths as well as noise, i.e., $\vec{H} = \sum_l \vec{h}_l = \sum_l \vec{a}(\theta_l)\alpha_l$. Then, the AoA estimation problem can be formulated as follows. Based on the N observations of CSI (i.e., $\vec{H}_n$, $n = 1, 2, \cdots N$, where $\vec{H}_n$ is the nth observation of channel vector), how to estimate θ_l, $l = 1, 2, \cdots, L$.

MUSIC Estimation. MUSIC is a subspace-based algorithm that has been widely used for AoA estimates in wireless localization. The general idea behind MUSIC method is to use all the eigenvectors that span the noise subspace to improve the performance of the Pisarenko estimator. It mainly comprises the following steps.

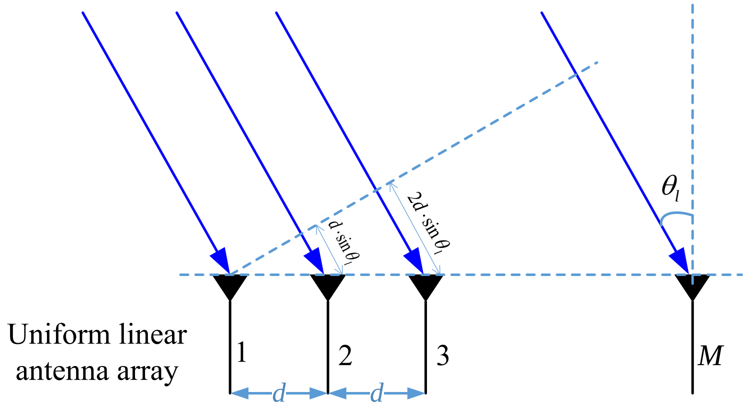


Figure 2.2: Illustration of MUSIC algorithm for AoA estimates at a wireless device with uniform linear antenna array. Only one signal path with AoA θ_l is shown in the figure.

Step 1: Calculate the correlation matrix of CSI observations: $\mathbf{R} = \sum_{n=1}^{N} \vec{H}_n \vec{H}_n^{\mathsf{H}}$, where $(\cdot)^{\mathsf{H}}$ is conjugate transpose operator.

Step 2: Perform eigendecomposition of the correlation matrix: $[\mathbf{E}\ \mathbf{S}] = eig(\mathbf{R})$, where $\mathbf{E}$ is a matrix with its columns being eigenvectors and $\mathbf{S}$ is the diagonal matrix with sorted eigenvalues (in non-decreasing order).

Step 3: Divide $\mathbf{E}$ into two sub-matrices: $\mathbf{E} = [\mathbf{E}_s \mathbf{E}_n]$, where $\mathbf{E}_s$ is the signal subspace and $\mathbf{E}_n$ is noise subspace.

Step 4: Evaluate the following function for all possible θ: $p(\theta) = \frac{1}{\vec{a}(\theta)^{\mathsf{H}} \mathbf{E}_n \mathbf{E}_n^{\mathsf{H}} \vec{a}(\theta)}$, where $\vec{a}(\theta)$ is the steering direction defined in (2.1). The values of θ corresponding to the peaks of $p(\theta)$ are the AoAs of incoming signals.

2.4.2: MUSIC for General Antenna Array (MUSIC-GAA)

The above MUSIC algorithm assumes that the antenna array is equally spaced and linearly installed. However, in practice, most wireless APs are equipped with nonlinear antenna array. For example, many Wi-Fi routers are equipped with four antennas which are installed in a rectangular shape to save the space. In this section, AuthIoT extends the MUSIC algorithm for a wireless device with general antenna array. In addition, it borrows the idea from SpotFi [85] to improve the AoA resolution by jointly estimating AoA and ToF of incoming signals. The rationale behind joint

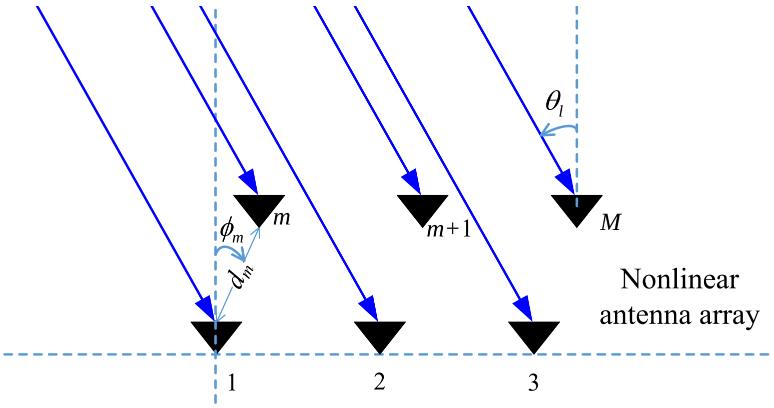


Figure 2.3: Illustration of MUSIC algorithm for AoA estimation at a wireless device with nonlinear (arbitrary) antenna configuration. Only one signal path with AoA θ_l is shown here.

estimation is that, if two incoming signals are indistinguishable in the spatial domain (due to the limited number of antennas), they may be distinguishable in the time domain. Joint estimation makes it possible to distinguish two incoming signals even if they have very similar AoA.

Consider a receiving device with nonlinear antenna array as shown in Fig. 2.3. For notional simplicity, we adopt polar coordinate system for the antennas using the first antenna position as the origin. Denote d_m as the distance between the 1st and mth antennas and ϕ_m as their angle, as illustrated in the figure. Then, the coordinate of the mth antenna can be written as (d_m, ϕ_m). Particularly, the first antenna's coordinate is $(0, 0)$.

Recall that α_l is defined as the complex channel attention of the lth path on the first antenna. The observed channel coefficient (CSI) on the mth antenna over subcarrier k can be modeled as:

$$h_{m,k} = \sum_l \alpha_l \cdot e^{j\frac{2\pi d_m \cos(\phi_m - \theta_l)}{\lambda}} \cdot e^{-j2\pi k f_\delta \tau_l} + n_{m,k}, \tag{2.2}$$

where (d_m, ϕ_m) is the polar coordinate of the mth antenna, f_δ is the subcarrier spacing of OFDM modulation, $(\alpha_l, \theta_l, \tau_l)$ is the complex attention, AoA, and delay of the lth path, respectively. Lastly, $n_{m,k}$ is the CSI observation noise/error at antenna m over subcarrier k.

Collectively, the observed CSI at all antennas and over all subcarriers can be expressed as an $M \times K$ complex matrix, where M is the number of antennas and K is the number of subcarriers. Consider a four-antenna 802.11 Wi-Fi router as an example, which has 52 valid subcarriers in

OFDM modulation. The CSI matrix $\mathbf{H} \in \mathbb{C}^{4\times 52}$ can be written as follows:

$$\mathbf{H} = \left.\overbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} & h_{16} & h_{17} & h_{18} & h_{19} & \cdots \\ h_{21} & h_{22} & h_{23} & h_{24} & h_{25} & h_{26} & h_{27} & h_{28} & h_{29} & \cdots \\ h_{31} & h_{32} & h_{33} & h_{34} & h_{35} & h_{36} & h_{37} & h_{38} & h_{39} & \cdots \\ h_{41} & h_{42} & h_{43} & h_{44} & h_{45} & h_{46} & h_{47} & h_{48} & h_{49} & \cdots \end{bmatrix}}^{\text{Frequency domain over OFDM subcarriers}}\right\} \text{Spatial domain over antennas} \tag{2.3}$$

Solely using spatial degrees of freedom (DoF) provided by antennas for AoA estimate may not be an ideal approach, as it requires the number of antennas is larger than the number of paths. This requirement may not be fulfilled in a real-world indoor environment when the number of antennas on a wireless AP is limited (e.g., four antennas on a Wi-Fi router). To improve the AoA resolution, AuthIoT expands the CSI matrix $\mathbf{H}$ for MUSIC-based AoA estimate by following the idea in [85]. Consider the CSI matrix in (2.3) as an example. AuthIoT can expand the CSI matrix by bonding every three columns as a new column as illustrated below:

$$\mathbf{H}_e = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} & h_{16} & h_{17} & \cdots \\ h_{21} & h_{22} & h_{23} & h_{24} & h_{25} & h_{26} & h_{27} & \cdots \\ h_{31} & h_{32} & h_{33} & h_{34} & h_{35} & h_{36} & h_{37} & \cdots \\ h_{41} & h_{42} & h_{43} & h_{44} & h_{45} & h_{46} & h_{47} & \cdots \\ h_{12} & h_{13} & h_{14} & h_{15} & h_{16} & h_{17} & h_{18} & \cdots \\ h_{22} & h_{23} & h_{24} & h_{25} & h_{26} & h_{27} & h_{28} & \cdots \\ h_{32} & h_{33} & h_{34} & h_{35} & h_{36} & h_{37} & h_{38} & \cdots \\ h_{42} & h_{43} & h_{44} & h_{45} & h_{46} & h_{47} & h_{48} & \cdots \\ h_{13} & h_{14} & h_{15} & h_{16} & h_{17} & h_{18} & h_{19} & \cdots \\ h_{23} & h_{24} & h_{25} & h_{26} & h_{27} & h_{28} & h_{29} & \cdots \\ h_{33} & h_{34} & h_{35} & h_{36} & h_{37} & h_{38} & h_{39} & \cdots \\ h_{43} & h_{44} & h_{45} & h_{46} & h_{47} & h_{48} & h_{49} & \cdots \end{bmatrix}. \tag{2.4}$$

The expanded CSI matrix is of 12 by 50 size, i.e., $\mathbf{H}_e \in \mathbb{C}^{12\times 50}$; and its correlation matrix is of 12 by 12 size, i.e., $\mathbf{H}_e\mathbf{H}_e^{\mathsf{H}} \in \mathbb{C}^{12\times 12}$. This means that, when applying MUSIC to AoA estimate, the expanded matrix renders a larger dimension for noise subspace compared to the original CSI matrix ($12 - L$ versus $4 - L$), thereby tending to offer a better AoA resolution.

In a general case, for CSI matrix $\mathbf{H} \in \mathbb{C}^{M\times K}$, a question is how many columns should be bonded when expanding this matrix for AoA estimate. For this question, we have the following considerations. On one hand, the number of rows of $\mathbf{H}_e$ should be maximized to improve the dimension of noise subspace; on the other hand, the expanded CSI matrix $\mathbf{H}_e$ should be a flat matrix for MUSIC calculation. Denote b as the number of bonding columns in the CSI matrix. Then, these two observations can be formulated as: max(Mb), subject to $Mg \leq K - G + 1$ and $G \in \mathbb{Z}$. Hence, we have $G = \lfloor \frac{K+1}{M+1} \rfloor$. Therefore, the dimension of the expanded CSI matrix is Mg by $K - G + 1$, i.e., $\mathbf{H}_e \in \mathbb{C}^{(Mg)\times(K-G+1)}$. The jth column of $\mathbf{H}_e$ is $[H_j; H_{j+1}; \cdots; H_{j+G-1}]$, where H_j is the jth column of $\mathbf{H}$ and $[;\cdots;]$ is vertical concatenation operator.

For the expanded CSI matrix $\mathbf{H}_e$, we would like to explore its basis for its columns. Based on (2.2), it can be verified that each of its columns is a linear combination of the following L basis vectors:

$$\vec{a}_l = [\underbrace{a_{11}\ \ a_{21}\ \ \cdots\ \ a_{M1}}_{\text{column 1}}\ \ \underbrace{a_{12}\ \ a_{22}\ \ \cdots\ \ a_{M2}}_{\text{column 2}}\ \ \cdots\ \ \underbrace{a_{1G}\ \ a_{2G}\ \ \cdots\ \ a_{MG}}_{\text{column G}}]^{\mathsf{T}} \tag{2.5}$$

for $1 \leq l \leq L$, where $a_{mg} = e^{j\frac{2\pi d_m \cos(\phi_m - \theta_l)}{\lambda}} \cdot e^{-j2\pi g f_\delta \tau_l}$ with $1 \leq m \leq M$ and $1 \leq g \leq G$.

Based on the expanded CSI matrix $\mathbf{H}_e$ and its column basis, the two-dimensional MUSIC algorithm is summarized as follows.

Step 1: Measure the CSI matrix $\mathbf{H}$ at M antennas over K subcarriers. Construct the expanded CSI matrix $\mathbf{H}_e$ by letting its jth column be $[H_j; H_{j+1}; \cdots; H_{j+G-1}]$, where H_j is the jth column of $\mathbf{H}$, $[;\cdots;]$ is vertical concatenation operator, and $G = \lfloor \frac{K+1}{M+1} \rfloor$.

Step 2: Calculate the correlation matrix of CSI observations: $\mathbf{R} = \mathbf{H}_e\mathbf{H}_e^{\mathsf{H}}$, where $(\cdot)^{\mathsf{H}}$ is conjugate transpose operator.

Step 3: Perform eigendecomposition of the correlation matrix: $[\mathbf{E}\ \mathbf{S}] = eig(\mathbf{R})$, where $\mathbf{E}$ is a matrix with its columns being eigenvectors and $\mathbf{S}$ is the diagonal matrix with sorted eigenvalues (in non-decreasing order).

Step 4: Divide $\mathbf{E}$ into two sub-matrices: $\mathbf{E} = [\mathbf{E}_s\ \mathbf{E}_n]$, where $\mathbf{E}_s$ is the signal subspace and $\mathbf{E}_n$ is noise subspace.

Table 2.2: Simulation parameters of MUSIC-GAA.

parameter	value	parameter	value
carrier frequency	5 GHz	# of paths	5
bandwidth	40 MHz	path 1: $(\alpha_1, \theta_1, \tau_1)$	$(1.00e^{j1.26}, 15^o, 5ns)$
FFT size	64	path 2: $(\alpha_2, \theta_2, \tau_2)$	$(.40e^{j0.64}, -71^o, 21ns)$
# of valid subcarrier	52	path 3: $(\alpha_3, \theta_3, \tau_3)$	$(.20e^{-j1.86}, 81^o, 38ns)$
# of antennas	4	path 4: $(\alpha_4, \theta_4, \tau_4)$	$(.15e^{j1.64}, -15^o, 65ns)$
antenna configuration	Vertex of 6cm×6cm square	path 5: $(\alpha_5, \theta_5, \tau_5)$	$(.10e^{-j1.51}, 31^o, 89ns)$

Step 5: Evaluate the following function for all possible θ and τ:

$$p(\theta, \tau) = \frac{1}{\vec{a}(\theta, \tau)^{\mathsf{H}} \mathbf{E}_n \mathbf{E}_n^{\mathsf{H}} \vec{a}(\theta, \tau)}. \tag{2.6}$$

Based on (2.5), the steering vector $\vec{a}(\theta, \tau)$ is defined as follows:

$$\vec{a}(\theta, \tau) = \big[\underbrace{a_{11}\ a_{21}\ \cdots\ a_{M1}}_{\text{column 1}}\ \underbrace{a_{12}\ a_{21}\ \cdots\ a_{M2}}_{\text{column 2}} \cdots \underbrace{a_{1G}\ a_{2G}\ \cdots\ a_{MG}}_{\text{column G}}\big]^{\mathsf{T}}, \tag{2.7}$$

where $a_{mg} = e^{j\frac{2\pi d_m \cos(\phi_m - \theta)}{\lambda}} \cdot e^{-j2\pi g f_\delta \tau}$ for $1 \leq m \leq M$ and $1 \leq g \leq G$. The values of (θ, τ) corresponding to the peaks of $p(\theta, \tau)$ are regarded as a path with AoA of θ and delay of τ.

An Example: We use an example to illustrate the performance of MUSIC-GAA. We consider a wireless AP and an IoT device and attempt to estimate the AoA of signal paths at the wireless AP. Table 2.2 lists the parameters that we use for simulation. Particularly, the antennas on the AP are not linear installed; instead, they are installed at the vertex of a 6cm×6cm square. This antenna configuration is more realistic compared to a uniform linear antenna array. In this case, the number of paths is greater than the number of antennas. Fig. 2.4 shows our simulation results when the CSI bears different levels of error. Specifically, Fig. 2.4a depicts the result when the AP has perfect

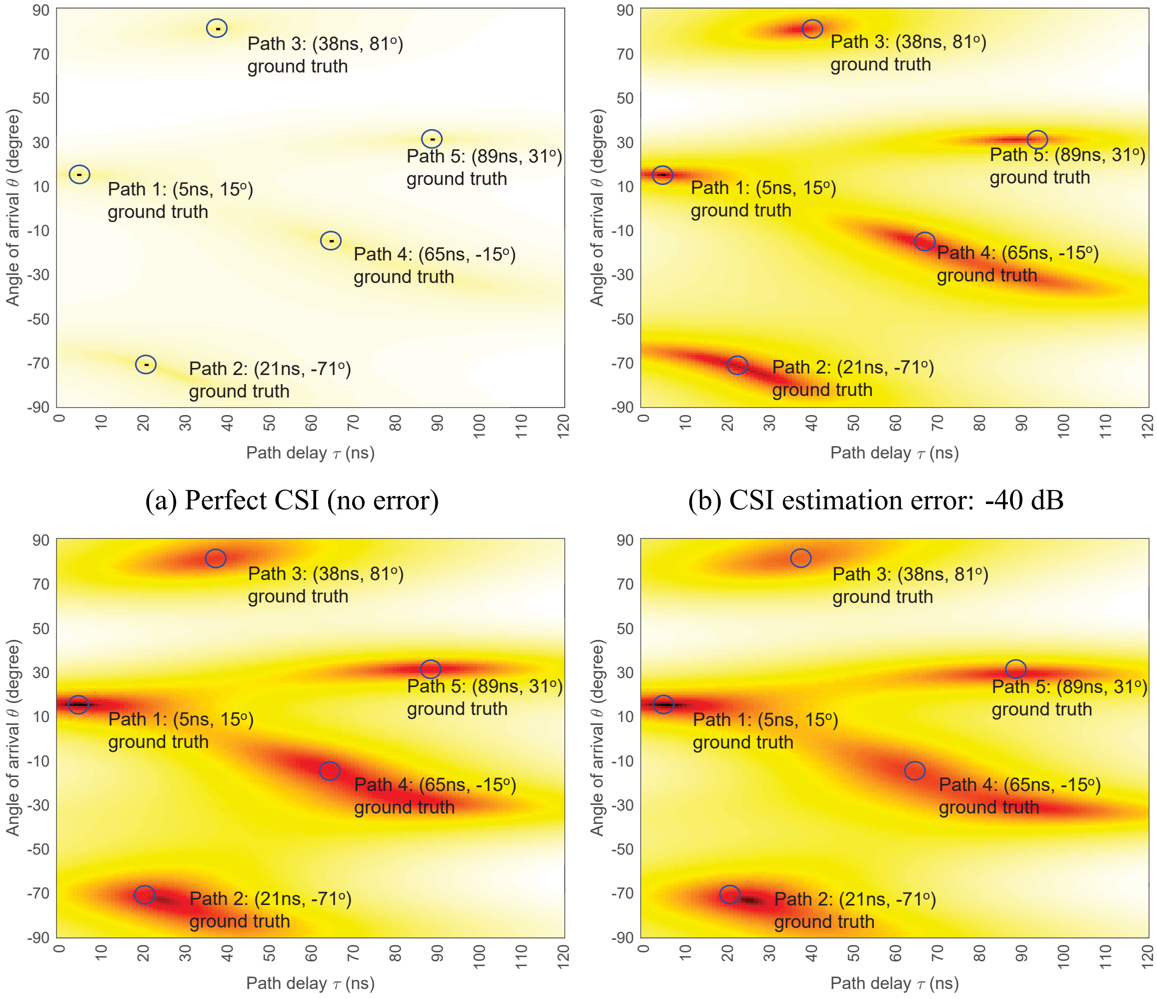
CSI. In this figure, the small circles mark the ground truth, while the black dots in the circles are the results of MUSIC-GAA. The results reveal that MUSIC-GAA finds the exact AoAs and delays of the five paths. Figs. 2.4b-d depict the results when the CSI at the AP has -40 dB, -30 dB, and -20 dB error. It can be seen that the heatmap becomes increasingly blurry when the CSI bears larger error. This indicates that accurate CSI is crucial. Fortunately, AuthIoT has accurate CSI for MUSIC-GAA as the IoT device is physically close to the AP with a LoS path.

Another observation from Figs. 2.4b–d is that the hot spots appear to be horizontally stretched, rendering better accuracy for AoA estimate than for delay estimate. This is because AuthIoT only requires AoA of LoS signal path and does not need the delay information. This phenomenon stems from the CSI expansion operation (see (2.4) for example), where each column of the expanded CSI matrix contains the CSI from all antennas (but the CSI from a subset of subcarriers).

2.4.3: MUSIC-GAA for AuthIoT

Using MUSIC-GAA for AuthIoT to estimate the LoS AoA faces the following two challenges. The first challenge is the very small delay difference of multiple paths indoor environments, especially in a small room with many objects. For example, if the distance difference of two paths is 1m, their delay difference is 3.3ns. To achieve this delay resolution (3.3ns), it requires 300MHz bandwidth. Such a large signal bandwidth is not affordable for most wireless systems. 5GHz Wi-Fi offers 40MHz bandwidth, which is insufficient to distinguish two paths whose distance difference is less than 1m. The second challenge is the CSI quantization error. For example, Atheros Wi-Fi NIC [182] offers 10-bit CSI quantization, rendering a quantization error of $10\log_{10}(1/2^{10}) = -30$ dB; Intel 5300 Wi-Fi NIC [63] offers 8-bit quantization for CSI, and its quantization error is $10\log_{10}(1/2^{8}) = -24$ dB. As shown in Fig. 2.4, the CSI error degrades the performance of MUSIC-GAA.

AuthIoT addresses these two challenges as follows. First, it asks users to keep the IoT device close to the AP ($\sim$2m) so that there is a strong LoS path between the two devices. It also asks users to handwrite the passcode over the air at a large scale (i.e., spanning a 75cm×75cm area for



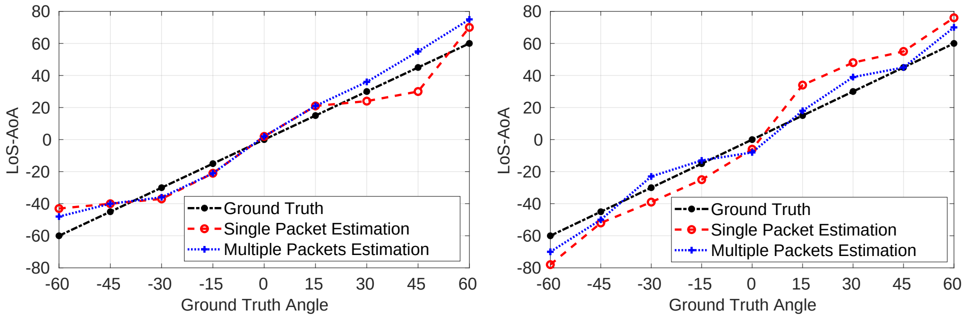
(a) Perfect CSI (no error)

(b) CSI estimation error: -40 dB

(c) CSI estimation error: -30 dB

(d) CSI estimation error: -20 dB

Figure 2.4: Performance of MUSIC-GAA algorithm.



(a) Linear antenna array.

(b) Nonlinear antenna array.

Figure 2.5: Experimental results of MUSIC-GAA in two cases: (a) Intel 5300 card with three linear equal-spaced antennas; (b) USRP N310 with four antennas placed at the vertex of 6cm×6cm square.

each passcode character), so that the AoA change of writing a passcode character is significant. These requirements will be specified on the manual for end users. Second, it combines multiple consecutive packets to improve the LoS AoA estimation through k-means clustering [105]. Details will be given in §2.5.2.

We have evaluated the performance of MUSIC-GAA for AuthIoT via experiments on two cases: i) the AP is an Intel 5300 card with three linear equal-spaced antennas; and ii) the AP is a USRP N310 with four antennas placed at the vertex of 6cm×6cm square. Both testbeds use Wi-Fi signal for data packet transmission, and the packet rate is 1000 per second. It means that the AP can obtain 1000 CSI instances per second. The distance between the AP and the IoT device is 2m, with the presence of LoS path. We conducted the measurement campaign in a two-story apartment with ordinary furniture. Fig. 2.5 shows our experimental results. It can be seen that the estimated LoS AoA increases/decreases as the ground-truth LoS AoA increases/decreases. This observation is consistent for both testbeds. This indicates that the LoS AoA tends to manifest a unique pattern based on the movement of IoT devices.

2.5: Learning-based Passcode Recognition

A passcode is composed of several characters (English alphabets, numbers, and some special characters). AuthIoT recognizes each individual character based on its generated CSI. Fig. 2.6 depicts

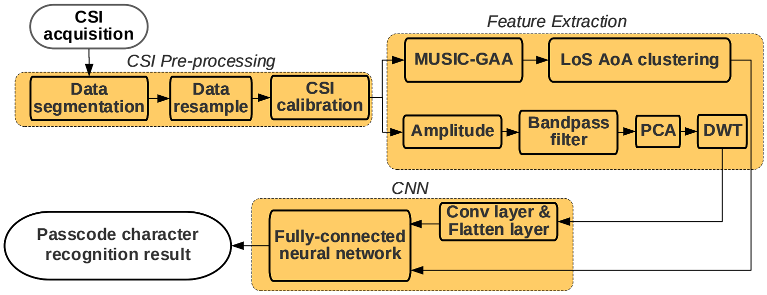


Figure 2.6: Diagram of CSI-based passcode character recognition.

the high-level system diagram of AuthIoT's passcode recognition. As shown in the diagram, AuthIoT uses both LoS AoA and normalized amplitude of CSI as the features for CNN-based character recognition. The reason is that our experiments show, compared to solely using LoS AoA as a feature, adding normalized CSI amplitude as input can considerably improve the recognition accuracy (by 5% on average in our observations). In what follows, we explain each module in Fig. 2.6.

2.5.1: CSI Segmentation, Resampling, and Compensation

CSI Segmentation. When a user continuously writes passcode characters in the air, the AP pings the IoT device at a certain rate (e.g., 200 ping packets per second), so that it can frequently estimate the CSI based on the ACK packets from the IoT device. In practice, an end user may take different amounts of time to write different characters, and different users may take different amounts of time to write the same character. Therefore, it is necessary to separate the collected CSI data in the time domain for each written character. To facilitate the CSI segmentation and improve its accuracy, AuthIoT asks end users to pause (holding IoT device still) one second before they begin to write a character. AuthIoT leverages the pause between two neighboring characters for CSI segmentation. In addition, AuthIoT asks end users to hold IoT device still for two seconds before they start to write passcode and after they complete passcode writing. Since a still IoT device generates unique CSI features, AuthIoT leverages such features to determine the time period of passcode writing.

Fig. 2.7 shows an example of AuthIoT's CSI segmentation, which comprises the following steps.

Step 1: Calculate the following metric: $g(i) = \angle(h_{m,k}(i)h_{n,k}(i)^*) \cdot |h_{m,k}(i)|$, where $h_{m,k}(i)$

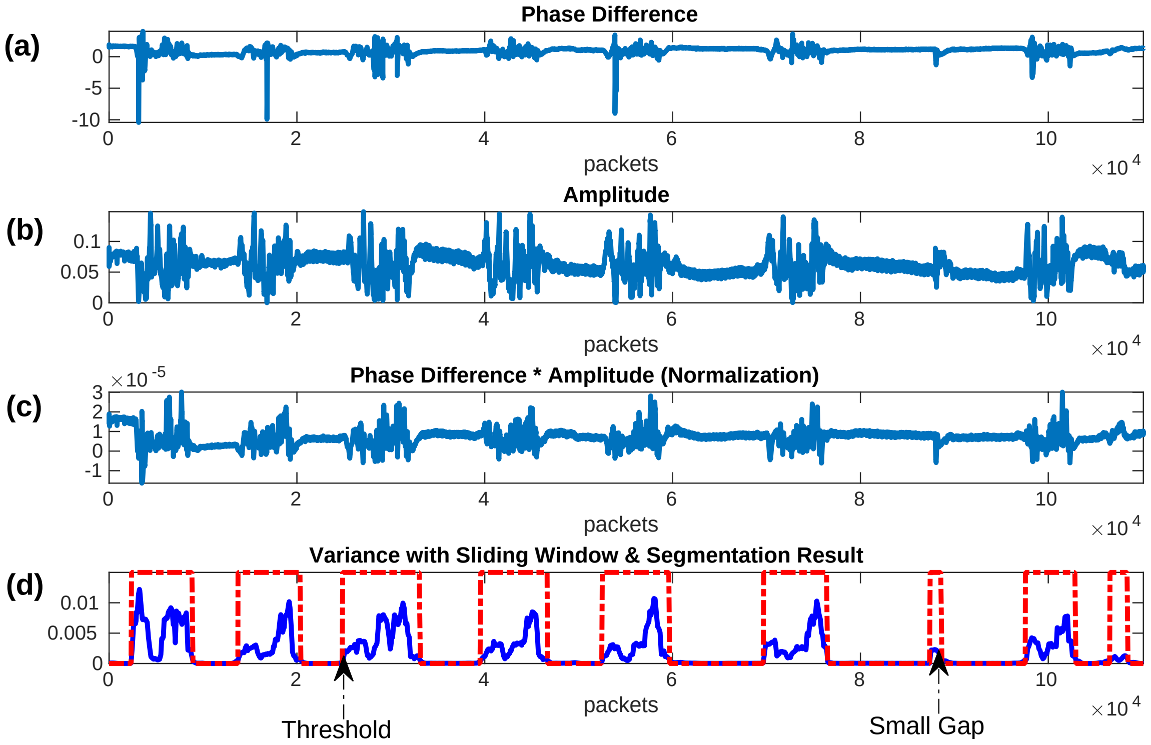


Figure 2.7: An example illustrating CSI segmentation.

is the channel coefficient from antenna m, subcarrier k, and packet i. In our design, m and n are the two antennas that offer strongest CSI, and $k = 1$. Fig. 2.7a shows an instance of phase difference of two channels, i.e., $\angle(h_{m,k}(i)h_{n,k}(i)^*)$. Fig. 2.7b shows an instance of channel amplitude, i.e., $|h_{m,k}(i)|$. Fig. 2.7c shows an instance of $g(i)$.

Step 2: Calculate the window-slided variance as follows: $v(i) = \frac{1}{W}\sum_{j=i}^{i+W-1}|g(j) - \bar{g}|^2$, where $\bar{g} = \frac{1}{W}\sum_{j=i}^{i+W-1} g(j)$. Fig. 2.7d shows an instance of $v(i)$.

Step 3: Compare $v(i)$ with a threshold T_v, where $T_v = 0.03 \times avg\{v(i)\}$. The CSI segment corresponding to $v(i) \geq T_v$ is considered for an individual character. Fig. 2.7d illustrates the windows corresponding to the segments of CSI to be used for character recognition.

Step 4: Check the segmentation length for each letter. If the time duration of a CSI segment is shorter than 1 second or longer than 4 seconds. AuthIoT discards this CSI segment.

CSI Resampling. After CSI segmentation, different CSI segments may have different numbers of CSI samples. The purpose of resampling is to make sure that the number of CSI samples in each CSI segment is identical. Doing so is likely to ease the training and inference of CNN. AuthIoT resamples each CSI segment using linear interpretation and/or decimation on the real and imaginary parts of CSI samples.

CSI Compensation. The CSI data need to be calibrated before feeding to the MUSIC-GAA. Since the receiver and transmitter are not synchronized, the CSI data from a Wi-Fi receiver may suffer from Sampling Time Offset (STO) and Sampling Frequency Offset (SFO). To compensate STO and SFO, a popular method is performing linear regression over multiple consecutive CSI instances in both time and frequency domains [85]. The linear fit of the unwrapped CSI phase for i_{th} packet can be expressed as

$$\tau_{s,i} = arg \min_{\alpha} \sum_{m=1}^{M} \sum_{n=1}^{N} (\phi_i(m,n) + 2\pi f_\delta (n-1)\alpha + \beta) \tag{2.8}$$

The $\tau_{s,i}$ is the STO for i_{th} packet. The f_δ is the frequency spacing between subcarriers. And the $\phi_i(m,n)$ is the wrapped phase at m_{th} antenna and n_{th} subcarrier. After estimating $\tau_{s,i}$ based on (2.8), the compensation is performed by adding $2\pi f_\delta (n-1)\tau_{s,i}$ to subcarrier n, $n = 1, 2, \cdots, N$. The same compensation applies to the CSI from each antenna.

2.5.2: Feature Extraction

LoS AoA Feature Extraction. AuthIoT uses MUSIC-GAA to estimate the AoA-delay profile of the signal paths based on the CSI samples. One observation from our experiments is that the AoA corresponding to the largest profile value is always associated with the minimum delay. This makes sense as there always exists a strong LoS path between the IoT device and the AP. Based on this observation, AuthIoT chooses the AoA corresponding to the largest value as LoS AoA.

As shown in Fig. 2.8, the LoS AoA computed from CSI is noisy due to the imperfection of hardware (e.g., 8-bit quantization) and interference caused by environment changes (e.g., body movement). Sometimes the LoS AoA jumps over 20 degrees for consecutive 10 packets. Obvi-

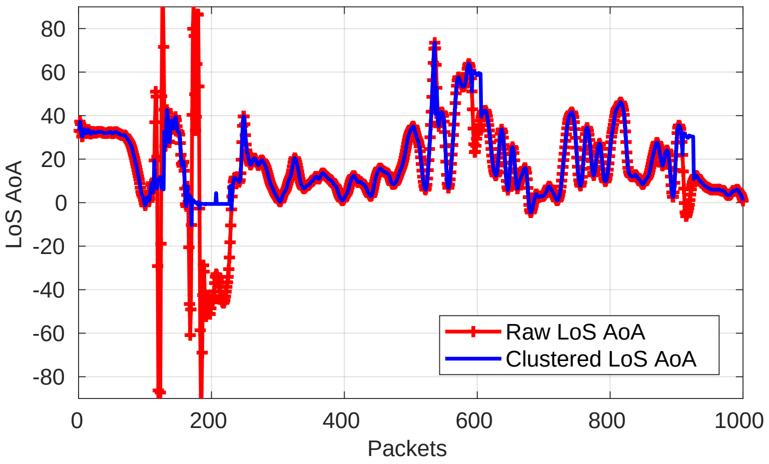


Figure 2.8: Removal of abnormal LoS AoA samples through filtering.

ously, such a big jump is abnormal. To reduce the adverse effect of this phenomenon, AuthIoT employs a clustering algorithm for the elimination of unexpected AoA values. The rationale behind this algorithm is that the AoA should not change over 20 degrees over 10 packets (10 ms). The clustering algorithm works as follows.

Step 1: Slide a window of size 10 to move across the LoS AoA sample sequence using the step size of 5. In each window, the k-means clustering algorithm [105] is employed to divide the 10 LoS AoA samples into 2 groups.

Step 2: Calculate the average values of the samples in the two groups. If the difference is larger than 20 degree and the number of samples in one group is less than 3, then the group of smaller size is regarded as abnormal.

Step 3: Replace every sample in the abnormal group with the average value of the larger group.

Amplitude Feature Extraction. In addition to LoS AoA, AuthIoT uses CSI amplitude as another feature for CNN-based recognition. The raw CSI amplitude is noisy. To enhance the input data quality, AuthIoT employs a Butterworth band-pass filter with frequency band 5Hz–20Hz to eliminate the undesired frequency components and reduce the noise for the CSI amplitude. This is because human's writing movement is in this frequency range [59]. Fig. 2.9 shows an example of the filtering operation.

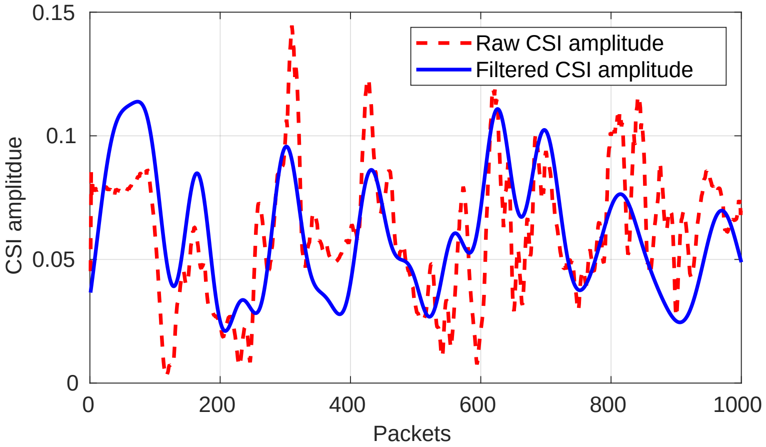


Figure 2.9: CSI amplitude before and after bandpass filtering.

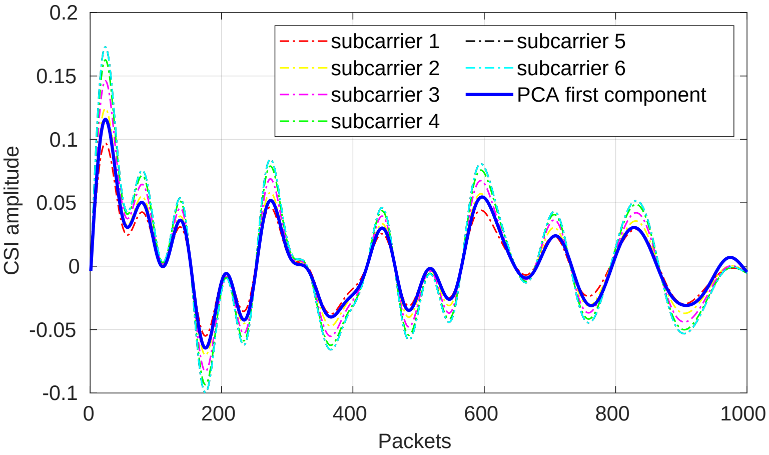


Figure 2.10: Illustration of PCA operation on CSI amplitude.

In indoor environments, wireless channels over neighboring subcarriers are very similar [51]. Hence, AuthIoT applies Principle Component Analysis (PCA) to a group of adjacent subcarriers for data compression. Specifically, AuthIoT groups every 6 subcarriers and applies PCA to each group. The first component of PCA results is kept as the amplitude features, while other components are discarded. Fig. 2.10 shows an example of this operation. As it can be seen, the adjacent 6 subcarriers have similar channel amplitude, and the first component of PCA results maintains the main shape of the channels.

Writing a character over the air mainly comprises a series of strokes. The action of each stroke is the key feature for the CSI-based character recognition. To capture the action of each stroke, AuthIoT performs Discrete Wavelet Transform (DWT) on the CSI amplitude after PCA operation, as shown in Fig. 2.6. Similar to WiReader [59], it performs 8-level discrete wavelet transform on the CSI amplitude samples using `symlet` as the basis function. Fig. 2.11 shows an example of DWT operation on the CSI amplitude, where Fig. 2.11(a) shows the CSI amplitude from PCA and Fig. 2.11(b) shows the DWT results. The DWT results are then sent to the CNN for training and inference.

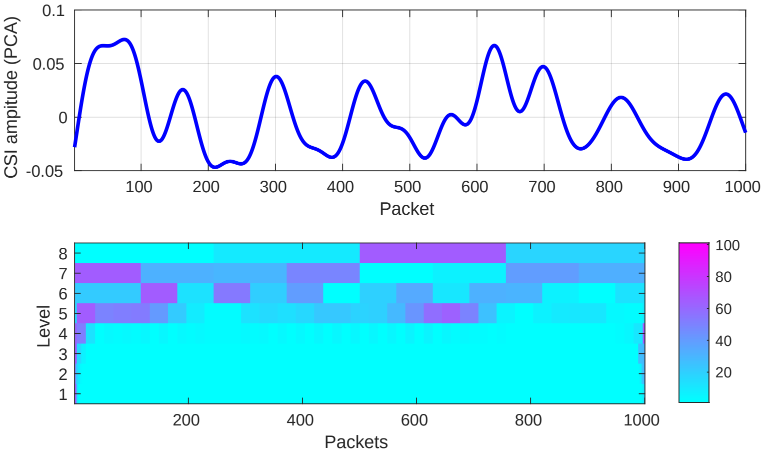


Figure 2.11: Illustration of DWT operation on CSI amplitude.

2.5.3: CNN Settings and Training

CNN Settings. Fig. 2.12 shows the structure of CNN, which is composed of convolution layers, flatten layers and fully-connected layers. Since the CSI amplitude matrix is of high dimension $(1000 \times 40 \times 3)$, AuthIoT employs convolution operations to extract its high-level features and reduce its dimension. Specifically, AuthIoT treats the amplitude DWT spectrum (1000×40) as an image and each of the three antennas as an image channel, similar to the process of RGB channels in colorful image recognition. Two convolution layers are used to compress the amplitude DWT spectrum. The first convolution layer involves 32 kernels of 11×5 size, and the second layer has 16 kernels of 6×4 size. The step size of both kernels is one. The purpose of the convolution layers is to extract the features from amplitude DWT spectrum based on its spatial relationship. It employs kernels moving across the feature matrix and outputs the convolution result with ReLU function. To further reduce the data dimension, AuthIoT employs an averaging pooling layers with

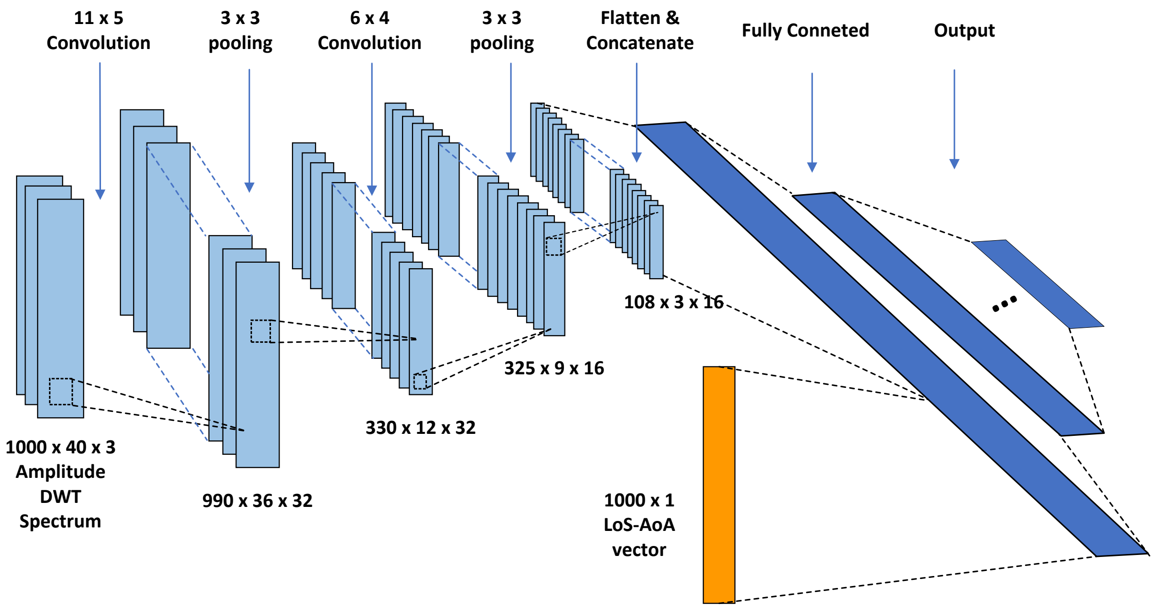


Figure 2.12: CNN Structure.

Table 2.3: Passwords Characters.

Capital Letters	A-Z
Lower-case letters	a,b,d,e,f,g,h,q,r,t
Numbers	3-9
Special Characters	#,$,%,+,=

a size of 3×3 for each of the convolution layers. The pooling layers down-sample the amplitude matrix, thereby reducing the computational complexity. The output of the second pooling layer is flattened for vectorization. AuthIoT then concatenates the resultant amplitude features with the AoA features, and feeds the concatenated data vector to a fully-connected $128 \times 64 \times 32$ neural network. SoftMax activation function is used for the output layer to calculate the probability of each possible passcode character.

CNN Training and Inference. As stated before, some character pairs are not distinguishable in their handwriting format, such as “z” and “Z”, “c” and “C”, “o” and “O”, “s” and “S”, “v” and “V”, letter “I” and number “1”, etc. Unfortunately, this challenge is hard to address from a technical perspective. Therefore, AuthIoT excludes the subset of indistinguishable characters. Table 2.3 lists the 48 characters that can be used for passcode in AuthIoT.

To train the CNN model, CSI data are collected from different locations and diverse users (details given in §2.6). The batch size in our training process is set to 100, and the number of epochs is set to 25. A batch normalization layer is added to the neural network after the activation function. We observed that it could improve the convergence speed in the training process, especially when the CSI is not stable due to the change of environment. In addition, a dropout layer is added after the second (64 neurons) and third (32 neurons) layers to avoid overfitting [143]. It can make the network less sensitive to specific neurons, and in turn make the network better generation. The dropout rate is set to 0.2 for each layer by randomly setting the output to zero. The CNN uses cross-entropy as the loss function and employs Adam optimization algorithm to update the weights.

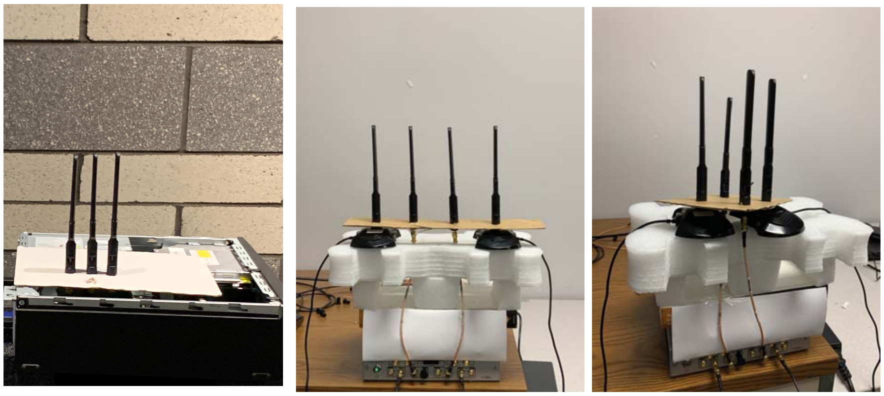
After the CNN is trained, the system is then used for online passcode character recognition in different environments. The CNN model will eventually yield the possibility of the input being each character. The character with the highest probability is regarded as the character being written by the end user.

2.6: Experimental Evaluation

2.6.1: Implementation and Experimental Settings

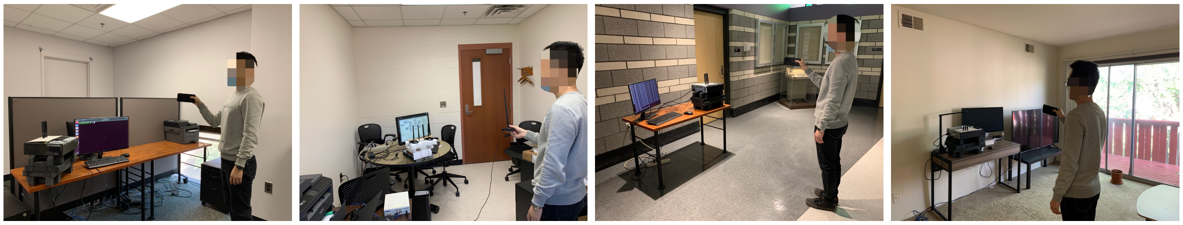
Intel 5300 Testbed. This testbed is implemented using Dell XPS 8940 Desktop with the Intel Wi-Fi NIC 5300 and a Redmi Note 9 Pro cellphone. The desktop serves as AP working in hotspot mode, and the cellphone emulates an IoT device. The desktop is installed with the Ubuntu 14.04 operating system with 802.11 Linux CSI tool [63], which is used to acquire the CSI from the Wi-Fi card. The carrier frequency is 5GHz, and the bandwidth is 40MHz. The packet rate is 600 packets per second. Intel Wi-Fi NIC 5300 is equipped with three antennas, which are linearly placed with equal spacing. The antenna spacing is half wavelength (3cm). Fig. 2.13(a) shows the linear antenna setting of this testbed.

USRP Testbed. This testbed consists of a USRP N310 and a USRP N210. USRP N310 has four antennas. It serves as the AP. USRP N210 has one antenna. It emulates an IoT device by sending data packets to USRP N310. The carrier frequency is 2.4GHz, and the bandwidth is 20MHz. The packet rate is 1000 per second. This testbed has two antenna settings: linear antenna array as



(a) (b) (c)

Figure 2.13: AP antenna settings: (a) Intel 5300 testbed with linear antenna array; (b) USRP testbed with linear antenna array; (c) USRP testbed with nonlinear antenna array.



(a) Lab scenario (b) Office scenario (c) Hallway scenario (d) Home scenario

Figure 2.14: Experimental settings.

shown in Fig. 2.13b and nonlinear antenna array as shown in Fig. 2.13c. For the linear case, the antenna spacing is 6.25cm. For the nonlinear case, the four antennas are positioned at the vertex of a 6cm×6cm square.

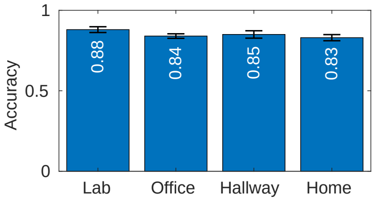


Figure 2.15: Recognition accuracy of AuthIoT on Intel 5300 testbed.

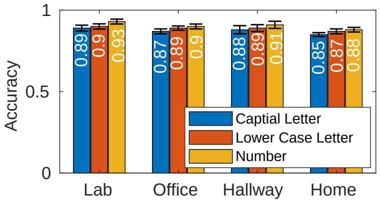


Figure 2.16: Recognition accuracy breakdown of AuthIoT on Intel 5300 testbed.

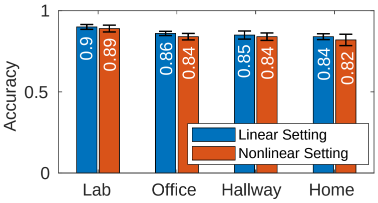


Figure 2.17: Recognition accuracy of AuthIoT on USRP testbed.

Experimental Settings. Four scenarios are considered for the evaluation of AuthIoT: lab, office, hallway, and home, as shown in Fig. 2.14. The AP was placed on a table of 70cm height, and the IoT device was held by the participants. The participants were asked to face the AP and keep an approximate 2m distance. We placed the two testbeds in these four scenarios and collected data to evaluate the performance.

The training data were collected solely from lab, while the evaluation (inference) was performed in four scenarios (lab, office, hallway, and home). The training data were collected from five different participants, while the evaluation was conducted over nine participants (i.e., those five participants for training plus four new participants). In the training phase, each participant was asked to write the 48 characters in Table 2.3, and each character was repeated 12 times. In total, 576 data samples were collected from each participant in the lab scenario for the training purpose.

In the test (inference) phase, each of the nine participants was asked to hold the IoT device and write 500 characters at his/her will at each scenario. The collected data samples were fed into the system for evaluation purpose.

2.6.2: Experimental Results from Intel 5300 Testbed

Intel 5300 is a commercial off-the-shelf Wi-Fi NIC that is widely used for computers and routers. Evaluating AuthIoT on this testbed reveals its performance in real-world Wi-Fi networks.

Overall Accuracy. Fig. 2.15 presents the overall recognition accuracy on this testbed. Literally, AuthIoT reaches 88% recognition accuracy with a standard deviation of 0.018 over the nine participants in the lab scenario; it reaches 85% recognition accuracy with a standard deviation of 0.023 in the hallway scenario; it reaches 84% recognition accuracy with a standard deviation of 0.014 in the office scenario; and it reaches 83% recognition accuracy with a standard deviation of 0.019 in the home scenario. It can be seen that AuthIoT performs best in the lab scenario. This is not surprising, because AuthIoT's CNN model was trained by the dataset collected from the lab scenario.

Fig. 2.18 shows the confusion matrix of passcode character recognition. It can be seen that the accuracy is above 85% for most characters. The majority of errors occur due to the ambiguity of the characters sharing similar hand gestures. For example, AuthIoT is more likely to be confused by letters 'C' and 'O'; it is also hard to distinguish letters 'M' and 'N'.

Accuracy of Individual Category. To obtain more details, we examine the performance of Au-

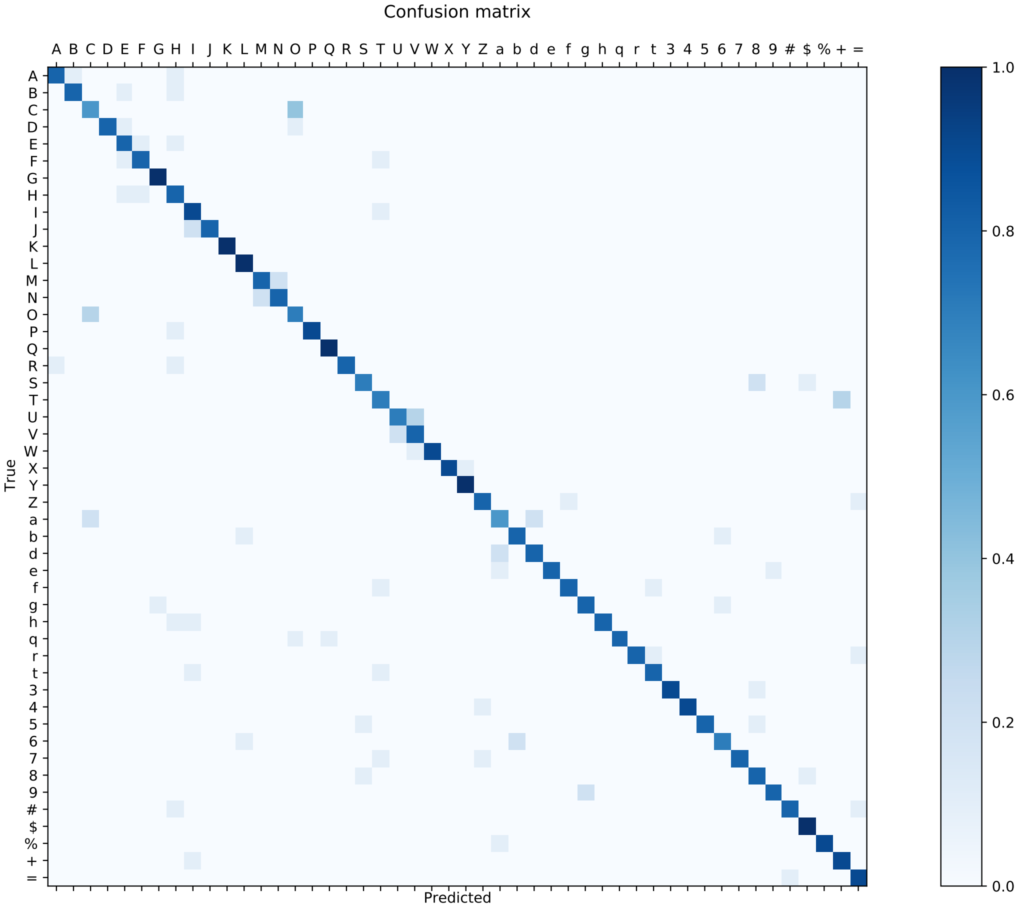


Figure 2.18: Confusion matrix for Intel 5300 testbed (with linear antenna array).

thIoT over three subsets of passcode characters: 26 upper-case letters, 10 lower-case letters, and 10 numbers. Fig. 2.16 shows our test results. It can be seen that the recognition accuracy in all scenarios are beyond 85% for the three subsets of characters.

2.6.3: Experimental Results from USRP Testbed

We further evaluate the performance of AuthIoT on the USRP testbed with linear and nonlinear antenna arrays.

Linear Antenna Array. Fig. 2.17 presents the recognition accuracy on the USRP testbed when it is equipped with four linearly equal-spaced antennas. It can be seen that the recognition accuracy in the lab scenario is better than other scenarios. This is because AuthIoT's CNN model was trained by the dataset collected from the lab scenario. It also can be seen that the recognition accuracy on the USRP testbed is slightly higher than that on Intel 5300 testbed. This can be attributed to the fact that the USRP testbed has one more antenna than the Intel 5300 testbed.

We examine the recognition accuracy for each individual participants. Fig. 2.19 shows our experimental results. The results show that the recognition accuracy is within the range of 81% to 88% for the nine participants. This indicates that AuthIoT is robust against the variation of end users.

Nonlinear Antenna Array. Fig. 2.17 also presents the recognition accuracy when the USRP testbed is equipped with four nonlinear antennas. It can been seen that the two cases (linear antenna array and nonlinear antenna array) have very similar recognition accuracy, with a difference less than 2%. The performance similarity can be traced down to the accuracy of LoS AoA estimation as shown in Fig. 2.5. Since the LoS AoA estimation in the two antenna settings has similar accuracy, it is not surprising that the recognition in the two antenna settings has similar accuracy.

2.6.4: Robustness of AuthIoT

To evaluate the robustness of AuthIoT, we examine its performance when the user is located at different distances and from different directions.

Different Distances. We change the distance between AP and IoT device to examine the perfor-

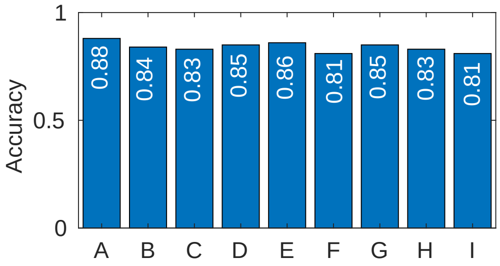


Figure 2.19: Recognition accuracy of each individual participant on USRP testbed.

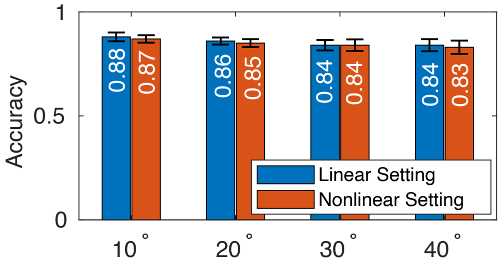


Figure 2.20: Recognition accuracy for user at different angles to the AP.

Table 2.4: Recognition accuracy of AuthIoT when the distance between AP and IoT device changes.

	Lab		Office		Hallway		Home	
distance	Linear	Non-linear	Linear	Non-linear	Linear	Non-linear	Linear	Non-linear
2.0m	89%	89%	86%	84%	85%	84%	84%	82%
2.5m	87%	86%	85%	84%	84%	84%	83%	82%
3.0m	85%	85%	84%	83%	84%	83%	82%	81%
3.5m	85%	84%	83%	82%	83%	83%	82%	81%

mance of AuthIoT. We consider four distances: 2.0m, 2.5m, 3.0m, and 3.5m. We conduct experiments in four scenarios: lab, office, hallway, and home. Table 2.4 presents our experimental results. It can be seen that, in each scenario, AuthIoT has a consistent performance when the distance between AP and IoT device varies from 2.0m to 3.5m. For all cases, the recognition accuracy of AuthIoT is within the range from 81% to 89%, regardless of the experimental scenario, the antenna pattern, and the distance between AP and IoT device. This indicates the robustness of AuthIoT.

Different Directions. To evaluate its robustness to user's facing direction, we let the user keeps the same distance to the AP but moves around with different facing angles ranging from 10 degree to 40 degree. As we can observe from Fig. 2.20, the recognition accuracy of AuthIoT slightly degrades when the angle between the user and AP increases from 0 to 40 degree. This is because the training data are collected at the 0 degree location. However, it can be observed that the accuracy for both linear and nonlinear settings are always above 83%.

Discussions. The overall recognition accuracy of 83% is not perfect but within an acceptable range. In practice, there are some ways to further improve AuthIoT's Quality of Experience (QoE) for end users. For example, an end user can consider using an all-numbers passcode. AuthIoT offers a superior performance when the passcode is all numbers. Meanwhile, an all-number passcode is sufficiently strong in practice. Moreover, a prompt-notification mechanism can be added into AuthIoT to improve the QoE of end users. In essence, AuthIoT is a learning-based classification algorithm. The output of AuthIoT includes not only the corresponding character but also its recognition probability (i.e., the recognition confidence). When AuthIoT has a low confidence for a character recognition, it immediately asks end user to rewrite the previous character. Doing so will offer a better QoE for end users.

2.7: Summary

In this chapter, we studied the communication authentication problem for wireless IoT devices without an input interface. We presented AuthIoT to authenticate such IoT devices in Wi-Fi networks by leveraging the unique CSI pattern generated by the movement of IoT devices. AuthIoT exploits environment-independent CSI features for learning-based character recognition, and therefore is transferable for cross-environment applications. AuthIoT also extends its applications for the case where a Wi-Fi AP is equipped with a nonlinear-installed antenna array by generalizing existing AoA estimation methods. We have built a prototype of AuthIoT and evaluated its performance on the testbeds with linear and nonlinear antenna arrays. Our experimental results confirm that AuthIoT is transferable for cross-environment applications, and show that AuthIoT achieves at least 83% recognition accuracy.

CHAPTER 3: HANDWRITING RECOGNITION THROUGH WALLS USING FMCW RADAR

3.1: Introduction

Despite the rise of digital technologies, handwriting—whether on traditional paper or electronic devices like iPads—continues to be a widely used method for recording and sharing information. Studies show that people still engage in handwriting more frequently than they might expect [18]. In some scenarios, the confidentiality of written content is of paramount importance. A natural question to ask is: *if one is writing important documents on a desk in a private room, is it possible for an attacker outside the room to detect the letters being written through the wall?* Understanding the capability and performance limits of such an attacker would inform the public of not only potential threats but also possible countermeasures, thereby preventing information leakage and enhancing human activity privacy.

Recent years have witnessed significant progress in remote human activity recognition (HAR) using different sensing technologies such as cameras [45, 80, 101], ultrasound [113, 187], Wi-Fi [52, 78, 93, 94, 123, 163], RFID [172, 191], and millimeter-wave (mmWave) radar [19, 72, 98, 99, 161, 174, 188, 206]. In contrast to existing work, the task of detecting handwriting content through a wall is unique yet challenging in the following three aspects.

(a) Through-wall detection. This requirement significantly limits the viable sensing techniques for this task. Camera-based computer vision (CV) can be used for human activity recognition by analyzing video data to identify and classify different human actions and movements [80, 101]. Powered by advanced deep neural network (DNN) techniques, a camera system can easily recognize the handwriting characters from a distance [57, 132]. However, camera-based HAR systems are limited by occlusions and thus not applicable to through-wall detection. Ultrasound

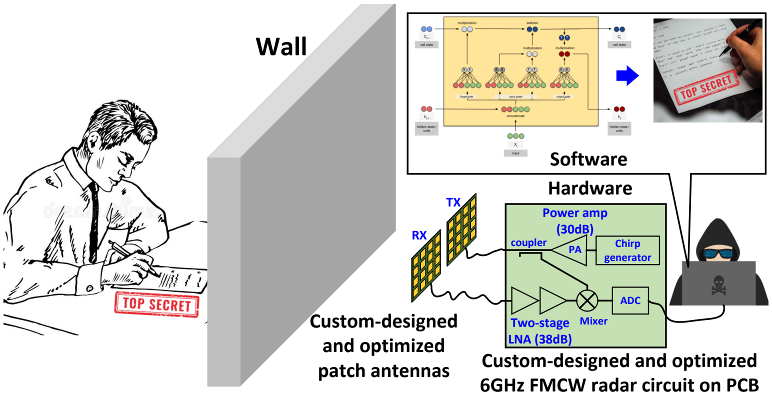


Figure 3.1: RadSee is a joint hardware (radar) and software (deep learning) design to detect the letters being written by a victim behind a wall.

sensors have also been widely used for HAR. They emit high-frequency sound waves that bounce off objects and produce echoes, which can be analyzed to determine the patterns of human activities [50, 113, 187]. Unlike camera sensors, ultrasound can work even in low light conditions and does not require a line-of-sight path. But ultrasound has a very limited ability of passing through walls due to its short wavelength, making it unsuitable for this task. Radio frequency (RF) has emerged as a popular technology for HAR such as gesture recognition [52,66,82,93,123,166,200], keystroke detection [22,190], and vital signal detection [61,167]. Among existing RF technologies, high-frequency signals (e.g., mmWave) have very limited ability to pass through a wall. Therefore, radio signals on sub-10 GHz bands appear to be the plausible carrier for HAR behind walls.

(b) Millimeter-level hand movement for writing. Handwriting features very small movements compared to other human activities. Typically, the movement of a pen-holding hand is smaller than 1 cm for both paper and iPad writings. When using an RF system for handwriting detection, its detection resolution is determined by its signal wavelength. On one hand, high-frequency mmWave (e.g., 60 GHz and 77 GHz) signals are capable of detecting sub-mm movement of an object but cannot pass through a wall. On the other hand, low-frequency (e.g., 915 MHz) microwave signals can easily pass through a wall but cannot detect the mm-level movement of an object. On the middle-frequency spectrum bands such as 2.4 GHz and 5 GHz, channel state information (CSI) in Wi-Fi networks has been extensively used for HAR [52, 78, 93, 94, 163, 171]; and its application on 5 GHz frequency bands seems a possible solution to achieve the desired trade-off between wall penetration and detection resolution. However, Wi-Fi CSI-based HAR is a *non-coherent* detection approach that suffers from phase, frequency and timing misalignments in hardware. As such, it is incapable of detecting mm-level movement in time. Recently, 6 GHz FMCW radar, which is a coherent detection system, has been used for HAR such as human body skeleton construction [16, 95, 209–211]. This approach uses custom-designed hardware and promises high accuracy and stability. However, so far, its applications are limited to the detection of large-scale movements such as people walking and interaction.

(c) Interference resilience. The detection of handwriting may suffer from interference from other moving objects such as a walking person around the writer. It may also suffer from interference from indirect paths between the writer and the detection equipment. Actually, such interference is a notorious issue with RF sensing [86, 119, 184]. This issue is particularly acute in sub-6 GHz RF sensing systems. If not addressed, the interference may appear dominant and place a fundamental limit on the detection performance. Moreover, since different scenarios have different multi-path effects and different moving objects/people, addressing the interference is critical to extract environment-independent features for handwriting recognition and ultimately develop a radio detector that can work in new environments.

In this chapter, we present RadSee, a 6 GHz FMCW radar system for detecting the handwriting activities behind a wall, as shown in Fig. 3.1. RadSee is realized through a joint hardware and software design. In terms of hardware, RadSee builds a 6 GHz FMCW radar with highly optimized patch antennas. In terms of software, RadSee first extracts the phase information of demodulated FMCW signals and employs a deep neural network (DNN) model for letter classification. Combining the hardware and software innovations, RadSee is capable of continuously detecting mm-level handwriting movement over time and recognizing most letters based on their unique phase patterns.

RadSee addresses **Challenge (a)** with its FMCW modulation, its high-gain patch antenna, and its optimized baseband analog filter. RadSee has co-located Tx and Rx RF chains, making it possible to perform coherent signal demodulation for handwriting recognition. In addition, the optimized patch antennas have a total 36 dBi gain for wall penetration. RadSee addresses **Challenge (b)** by using the *phase* information of demodulated FMCW signals to extract the features of handwriting movements. FMCW radar has been widely used for ranging. Its range resolution is $\frac{c}{2B}$, where c is light speed and B is signal bandwidth. Achieving the range resolution of 1 mm requires $B = \frac{3\times10^8}{2\times10^{-3}} = 150$ GHz signal bandwidth, which is impossible in practice. However, the *phase* of demodulated FMCW signals is much more sensitive to the movement of an object. In theory, 1 mm hand movement corresponds to 14° phase change of the demodulated signal, which is easy to detect. Therefore, RadSee uses the *phase* of demodulated FMCW signals as the features of letter recognition. RadSee addresses **Challenge (c)** by demodulating the reflective signals only from the handwriting movement. This is achieved by its FMCW modulation and high-directional patch antennas. The FMCW modulation allows it to focus on the Range-FFT bin that corresponds to the distance of interest; the patch antennas allow it to focus on the reflective signal from the direction of interest. Combining FMCW modulation and antenna directivity, RadSee is capable of detecting a clear phase pattern corresponding to the handwriting movements behind a wall using a small transmission power (20 dBm).

Based on the demodulated FMCW signals, RadSee employs a bidirectional long short-term memory (BiLSTM) model to classify the handwriting characters (a-z, A-Z, and 0-9). Different from other human activities such as keystroke [22, 190], handwriting is a smooth and continuous movement of the pen-holding hand. As such, handwriting tends to generate a unique temporal phase pattern for each letter. That is the reason why RadSee uses BiLSTM to classify a phase data sequence. Of a phase data sequence, some parts may be very important for letter recognition (e.g., those turning points), while some parts may not carry useful information (e.g., horizontal strokes). Therefore, RadSee adds an attention layer to the BiLSTM model so that the model can automatically focus on those important parts of a phase data sequence for letter classification. Powered by the BiLSTM model and its attention mechanism, RadSee is capable of recognizing handwriting letters based on their unique movement patterns.

We have built a prototype of RadSee (through PCB fabrication) and evaluated its performance

Table 3.1: Related work on human activity recognition. $\mathcal{W}$ = "See through wall?", $\mathcal{M}$ = "Mm-level movement detection?", $\mathcal{R}$ = "Resilient to multipath?", $\mathcal{I}$ = "Resilient to interference from other moving objects?", $\mathcal{S}$ = "Classification size".

References	**Objective**	**Technique**	$\mathcal{W}$	$\mathcal{M}$	$\mathcal{R}$	$\mathcal{I}$	$\mathcal{S}$
RF-Capture [16], RF-Avatar [210], RF-Pose [209], RF-Pose3D [211], RF-Action [95]	Human body skeleton	6GHz FMCW radar	✓	✗	✓	✓	N/A
WiSIA [94], WiPose [78], F. Wang [163]	Radio imaging	Wi-Fi	✗	✗	✗	✗	N/A
Tadar [191], RF-HMS [172]	Human tracking	RFID	✓	✗	✗	✗	N/A
mtrack [174]	Hand writing	mmWave	✗	✓	✗	✓	N/A
WiKey [22]	Key stroke	Wi-Fi	✗	✗	✗	✗	37
WiHF [93], WriFi [52], WiSee [123]	Gesture recognition	Wi-Fi	✗	✗	✗	✗	26
Soli [99]		mmWave	✗	✓	✗	✓	4
PhaseBeat [167]	Vital sign	Wi-Fi	✓	✗	✗	✗	N/A
RF-SCG [61]		mmWave	✗	✓	✗	✓	N/A
RadSee (ours)	**Hand writing**	**FMCW radar**	✓	✓	✓	✓	**62**

in various scenarios. Experimental results show that, when placed behind office interior drywalls and external wood/vinyl walls, RadSee achieves 75% letter recognition accuracy when victims randomly write 62 different letters and 87% word recognition accuracy when victims write articles. Notably, RadSee demonstrates its resilience to the interference from walking persons around the victim writer and the interference from other radio devices.

Table 3.1 shows the comparison of RadSee and its related work. It advances the state-of-the-art in the following aspects.

- It designs and implements a 6 GHz FMCW radar device that can detect mm-level movements of an object behind a wall using a small transmission power.
- RadSee is capable of detecting the letters that one is writing behind a wall. Furthermore, it is resilient to the interference from other mobile objects and other radio devices.
- Extensive experimental results show that RadSee can achieve over 75% accuracy when detecting 62 random letters and 87% word recognition accuracy behind walls.

3.2: Related Work

We surveyed the literature in two categories: *through-wall detection* and *fine-grained human activity recognition*. Table 3.1 in Section 3.1 outlined RadSee's uniqueness compared to prior work.

3.2.1: See Through Wall using Radio

See Through Wall using FMCW Radar. Some pioneering works have studied 6 GHz FMCW radar to detect and track human activities behind walls using model-based or learning-based methods [16, 95, 209–211]. For instance, [209–211] focuses on using FMCW radar to generate the heatmap image of human body skeleton through walls. [95] uses FMCW radar to detect the interactions between two people behind walls. However, all these works are based on the ranging detection of FMCW radars. Since the range resolution of an FMCW radar is fundamentally limited by its bandwidth, this method cannot achieve mm-level accuracy for through-wall motion detection. To address this issue, RadSee uses the phase information for through-wall mm-level hand movement detection.

RF-capture [16] is probably the most related work of RadSee. It also uses FMCW radar to recognize the "handwriting" behind a wall. However, the letters that RF-capture aims to recognize are of large size (e.g., 0.5 m×0.5 m). It is actually a gesture recognition rather than normal-sized handwriting detection. Its method is based on range- and angle-based tracking, and thus cannot achieve mm-level accuracy. Therefore, RadSee is fundamentally different from RF-capture.

Through-Wall Detection using Wi-Fi. Wi-Fi signal is ubiquitous and it has a strong ability of passing through a wall. [17] utilizes Wi-Fi signals and multi-antenna techniques to track the movement of people behind a wall. [173] uses Wi-Fi signals to recover the audio sound from a speaker placed behind a soundproof wall. However, due to the no-coherent detection at a Wi-Fi receiver, it is impossible for a Wi-Fi receiver to detect movement at the millimeter level. Therefore, Wi-Fi signals are not suitable for through-wall handwriting detection.

Through-Wall Detection using RFID. Through-wall detection is also possible by using RFID systems. Tadar [191] and RF-HMS [172] demonstrated their capabilities of tracking human moving directions through walls using an array of RFID tags. However, the tracking error in these systems is around 10 cm, indicating their incapability of tracking mm-level hand movements. RFID tag can also be used to measure the vibration pattern of a loudspeaker [162]. But, due to its long wavelength (33 cm), it is not a good candidate for tracking mm-level movements.

3.2.2: Fine-Grained HAR

Handwriting Recognition. Camera-based handwriting recognition is a well-established field [39]. However, the camera cannot see through walls. Recently, RF signals have been studied for handwriting recognition. RF-IDraw [165] attaches an RFID tag to a people's finger and can reconstruct the trajectory of that finger. A multi-resolution positioning technique was designed, yielding a tracing accuracy at the centimeter level. mTrack [174] developed a mmWave (60 GHz) tracking system and achieved mm-level tracking accuracy. It also demonstrated its capability of recognizing handwriting letters. However, mmWave signals are vulnerable to blockage and cannot go through walls. Therefore, it is not suitable for our purpose.

MmWave FMCW Radar Detection. In recent years, mmWave (24 GHz, 60 GHz and 77 GHz) FMCW radars become available on the market for autonomous driving applications. These radars have been widely used for human activity recognition and vital sign detection [19, 61, 72, 98, 99, 161, 174, 188, 206]. Given their large bandwidth and small wavelength, they can easily achieve mm-level accuracy when detecting object movements. However, mmWave signals cannot pass through walls. Therefore, they cannot apply to through-wall handwriting detection.

Gesture and Vital Sign Detection. CSI in Wi-Fi networks has been used for a wide range of sensing applications such as gesture recognition [52, 93, 123], vital sign detection [167], and radio imaging [78, 94, 163]. However, Wi-Fi is a non-coherent system due to the physical separation of its transmitter and receiver. Therefore, its detection accuracy is fundamentally limited by timing, frequency, and phase misalignments. As a result, it is not competent for mm-level handwriting detection.

3.3: Attack Model

Attack Scenarios. We consider a scenario as shown in Fig. 3.1, where one is writing a confidential document on a paper or an electronic device (e.g., iPad and Kindle Scribe) in a private room (e.g., government office, business office, hotel room, and apartment). Inside the room there may be other static objects (e.g., furniture) and people performing daily activities. Outside the room there is a malicious attacker who aims to detect the content (English letters and Arabic numbers) being written by the victim.

Attacker's Assumptions. We assume that the attacker has physical access to the space behind the wall which the victim is facing toward. As such, the radio signals for detecting the victim's handwriting movements would not be blocked by the victim's torso. We also assume that the attacker knows the layout of the room and the approximate location of the victim. However, the accurate location of victim's writing hand will be estimated by the attacker using radar signal. We know that it is not improbable to obtain the knowledge about the location of a desk in a room, as many public spaces such as hotels have standard layouts that are consistent across rooms. Furthermore, we assume that there are no RF-shielding materials inside the wall between victim and the attacker.

Challenges. As we stated before, there are three grand challenges that must be addressed for the design of such an adversarial device, namely, through-wall detection, mm-level recognition, and interference resilience. In addition, handwriting recognition has 62 character candidates (26 low-case letters, 26 upper-case letters, and 10 Arabic numbers) for classification. Such a large character set adds another level of challenge to the task. To address these challenges, it calls for a joint hardware and software design for such an attack device.

3.4: RadSee: Design Analysis

3.4.1: A Primer on FMCW Radar

FMCW radar is an active radio device that uses frequency modulation to generate a continuous wave signal with a linear frequency sweep. This signal is transmitted from the radar antenna toward a target, and the reflection from the target is received by the radar antenna. The frequency difference between the transmitted and received signals, known as the beat frequency, is proportional to the range of the target. By analyzing the beat frequency over time, FMCW radar can determine the distance and velocity of the target.

Fig. 3.2 shows the diagram of an FMCW radar device. It transmits frequency-modulated continuous-wave signals and receives the reflective signals from the surrounding objects. Denote $s_T(t)$ as the transmitting signal and $s_R(t)$ as the received echo from an object. Mathematically, we have

$$s_T(t) = \text{exp}\left(j(2\pi f_0 t + \pi K t^2)\right), \tag{3.1}$$

and

$$s_R(t) = \alpha s_T(t - 2d/c), \tag{3.2}$$

where f_0 is the starting frequency, K is the frequency ramp rate, α is the path attenuation, d is the distance from the radar to the object of interest, and c is light speed.

The received signal and the transmitted signal are mixed together, generating the intermediate frequency (IF) signal. The IF signal can be written as:

$$s_{IF}(t) = s_T(t)s_R(t)^* = \text{exp}\Big(\underbrace{j4\pi K\frac{d}{c}t}_{\text{frequency}} + \underbrace{j4\pi f_0\frac{d}{c}}_{\text{phase}} - \underbrace{j4\pi K\frac{d^2}{c^2}}_{\text{negligible}} \Big). \tag{3.3}$$

As we can see from (3.3), the observed frequency and phase both contain the distance information. Typically, the frequency term in (3.3) is used to estimate the range of an object, while the phase term is used to estimate the velocity of the object. Specifically, the range and velocity of the

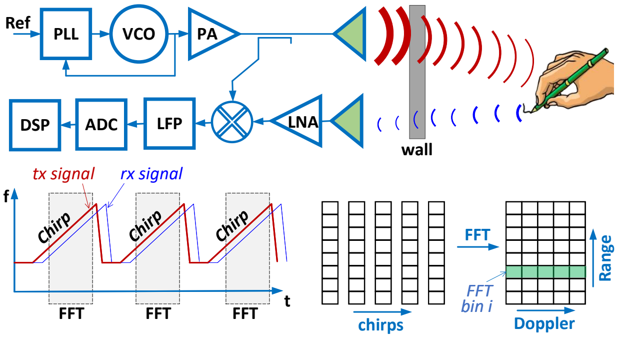


Figure 3.2: Illustration of radar device.

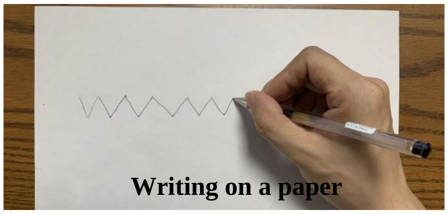


Figure 3.3: Illustration of handwriting pattern.

object are estimated as follows.

- **Range.** As illustrated in Fig. 3.2, the IF signal from each chirp is digitized and converted to the frequency domain through FFT operation. Suppose that the FFT size is N and a peak is identified at the ith FFT bin ($0 \leq i \leq N-1$). Then, the distance of the corresponding object is $d = \frac{c}{2B}i$, where B is the FMCW signal bandwidth. Accordingly, the range resolution is $\Delta d = \frac{c}{2B}$, which is determined solely by the FMCW signal bandwidth.
- **Velocity.** Grouping an array of chirps together, the velocity of the object can be accurately estimated by performing the second FFT operation on the ith Range-FFT bins. Suppose that the time duration of one chirp is T and that the FFT size is M. In this case, a peak is identified at the kth FFT bin, which allows us to calculate the velocity of the object $v = \frac{kc}{2MTf_0}$. Accordingly, the velocity resolution is $\Delta v = \frac{c}{2MTf_0}$, which is determined by three parameters: the initial frequency, the time duration of a chirp, and the number of used chirps.

3.4.2: Feasibility Analysis

To detect fine-grained movements, the first option that came to our mind is mmWave FMCW radar, which is widely available on market at a low price. Particularly, existing work (e.g., [72, 98, 161]) has demonstrated the ability of mmWave radars to "see" through walls made of cotton and glass. A key question to ask is whether a mmWave radar can "see" through typical walls in our daily lives. To answer this question, we conducted experiments using IWR1642BOOST 77 GHz mmWave FMCW radar from Texas Instruments (TI) with a bandwidth of 1.1 GHz. We placed the mmWave radar behind an office drywall to detect the handwriting in a room. Fig. 3.3 shows our writing content. Fig. 3.4(a) shows experimental setting and the corresponding FFT-bin's amplitude and phase over time. We did not observe any amplitude or phase changes over time caused by the handwriting. This indicates that mmWave signals cannot go through the drywall under test.

Another possible approach to this task is to use Wi-Fi-based channel state information (CSI). Since Wi-Fi uses 2.4 GHz and 5 GHz frequency bands, its signal is able to penetrate walls for movement detection. To examine this approach, we conducted experiments in the same scenario as the previous case. Fig. 3.4(b) shows the measured CSI at a receiver when using Wi-Fi channel #3 (2412 MHz–2432 MHz). We observed random CSI changes over time, and did not find any patterns on the CSI's amplitude and phase that are related to the handwriting movement. Similar results are observed for the CSI measured on Wi-Fi channel #36 (5170 MHz–5190 MHz). This can be attributed to the non-coherent detection of a Wi-Fi receiver. Since Wi-Fi transmitter and receiver are driven by different clocks, the measured CSI suffers from carrier frequency and sampling time offsets, making it unreliable to extract the pattern of tiny-scale movements.

In comparison, we replaced the mmWave/Wi-Fi device with RadSee—our custom-designed 6 GHz FMCW radar. Fig. 3.4(c) shows the corresponding FFT-bin's amplitude and phase over time. It can be seen that the phase pattern is significant and that the phase pattern is consistent with the handwriting trajectory on the paper (see Fig. 3.3). This demonstrates the ability of RadSee to "see" through the wall under test.

Why Use 6 GHz FMCW Radar? Some may inquire about the suitability of other frequencies for through-wall and fine-grained movement detection. Low-frequency (0–3 GHz) radio signals have large wavelengths, rendering them incapable of detecting movements at the millimeter level. High-frequency (20–300 GHz) radio signals, on the other hand, have a large path loss and a significant penetration loss; thus they cannot travel through walls with normal transmission power. Radio signals in the range from 3 GHz to 20 GHz, however, should be suitable for this task. We opted for 6 GHz due to the availability and cost-effectiveness of electronic chips for FMCW radar implementation, including phase-locked loop (PLL), voltage-controlled oscillator (VCO), power

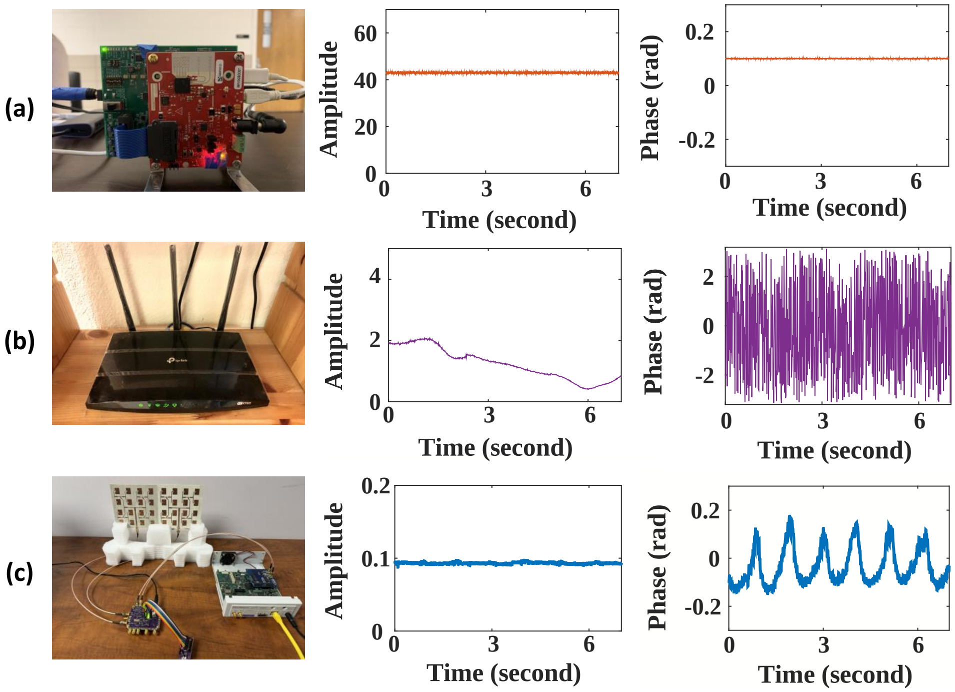


Figure 3.4: (a) The amplitude and phase of the corresponding FFT-bin from IWR1642BOOST mmWave FMCW radar. (b) The amplitude and phase of a subcarrier from a Wi-Fi receiver. (c) The amplitude and phase of the corresponding FFT-bin from RadSee.

amplifier (PA), low-noise amplifier (LNA), etc. On the market, only 6 GHz chips are available for individual customers at a reasonable price, thanks to the widespread production of 5 GHz Wi-Fi industry. The cost of our prototype is approximately \$500.

Millimeter-level Movement Detection. If an FMCW radar wants to achieve 1 mm range resolution, it will need 150 GHz spectrum bandwidth, which is impossible in practice. Therefore, RadSee uses the phase information of its demodulated FMCW signal to infer the movement pattern of handwriting. Based on Eqn. (3.3), when the object moves 1 mm, RadSee will observe $\frac{2df_0}{c}2\pi = 0.25$ radian (about 14°) phase change on the corresponding Range-FFT bin. Typically, handwriting movement is larger than 5 mm, which will generate 70° phase change on the Range-FFT bin. Therefore, the radar will measure the phase pattern over time when a victim is writing, and use the temporal phase pattern to classify the letters being written.

Fig. 3.5 shows the observed phase change of a Range-FFT bin when one is writing back and forth on a paper behind a thick office drywall. The distance between the writing hand and the wall is

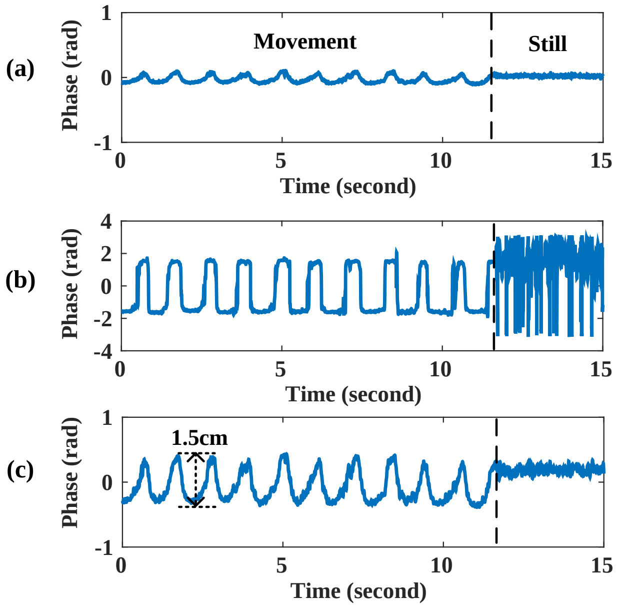


Figure 3.5: Phase observations at a behind-wall radar when one is continuously writing back-and-forth on a paper within 1.5 cm. (a) phase data before DC component removal. (b) phase data after DC component removal. (c) phase data after partial DC component removal.

about 2 m. The radar was placed on the other side of the wall, with a distance of 0.5 m. The person wrote back and forward within a vertical range of 1.5 cm. It can be observed from Fig. 3.5(a) that the phase changes as the pen-holding hand moves. However, the phase dynamic range is small. The small dynamic range is attributed to a DC voltage component of the received signal. The DC component, which can be modeled as a constant complex number, is the reflective signals from static objects (e.g., furniture and human body) of the same distance. Fortunately, the DC component is static over time and thus can be easily removed. Ideally, we should completely remove the DC component to maximize the phase sensitivity. However, when we completely remove the DC component, the time period of no-movement will have an irregular phase pattern as illustrated in Fig. 3.5(b), making it hard for RadSee to detect the gap between two consecutive letters. Therefore, we partially remove the DC component to strike a balance between movement detection sensitivity and the phase stability of non-movement periods. Fig. 3.5(c) shows the observed phase of a Range-FFT bin over time.

Interference Resilience to Other Mobile Objects. In the proximity of a target writer, there may be many static objects such as desks, chairs, books, and lamps. Fortunately, the static objects

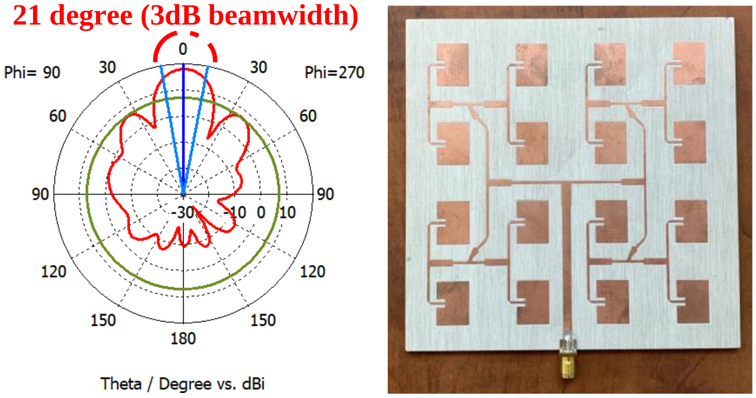


Figure 3.6: The gain pattern of the patch antenna (left). The custom-designed directional antenna (right).

will not generate interference for the detection of RadSee as their reflective signals appear to be a constant complex number (DC component) over time. Such a constant can be easily removed or adjusted to extract the useful phase information. As stated before, RadSee may suffer from interference from two sources: *(i) channel multi-path*, and *(ii) movement of other objects* (e.g., a walking person). Actually, RadSee is resilient to the interference from these two sources, thanks to its FMCW modulation and antenna directivity. We explain the reasons below.

- *FMCW Modulation (Distance Filter).* If two moving objects have different distances to the radar and their range difference is larger than the radar's range resolution, their phase-change patterns will appear on different Range-FFT bins and will not interfere with each other. Therefore, increasing the range resolution of RadSee is critical for reducing the interference from mobile objects. RadSee uses 1.1 GHz (5.4-6.5 GHz) bandwidth and thus has a range resolution of 14 cm. This means that, if separated by 14 cm, a mobile object (e.g., writer's chest movement of breathing) will not generate interference to RadSee's handwriting detection.

- *Patch-array Antenna (Directional Filter).* In addition to offering high link gain, the patch-array antenna also serves as a directional filter to suppress the interference from undesired azimuth/elevation angles. We designed and optimized the patch-array antenna using CST Studio Suite [42] and fabricated the path-array antenna as shown in Fig. 3.6. The main

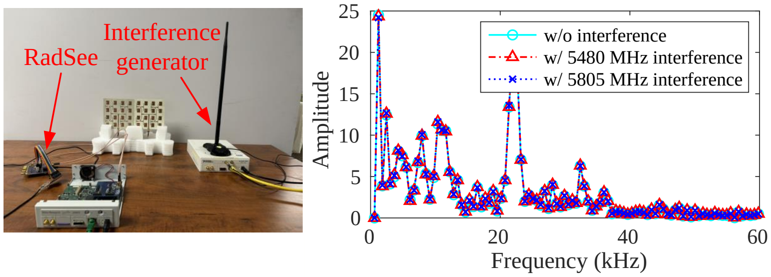


Figure 3.7: Study an FMCW radar's resilience to radio interference from Wi-Fi devices: experimental setup (left) and experimental results (right).

lobe of the antenna has an angular width of 21° (3 dB), which means that this antenna can effectively mitigate interference from mobile objects when they are positioned 21° or more away from the writer.

Combining its FMCW modulation and patch-array antenna, RadSee is capable of extracting the phase information corresponding to the movement within a small spot of interest, while being resilient to interference from other moving objects.

Interference Resilience to In-band Wi-Fi Devices. Although RadSee operates on a frequency band that overlaps with 5 GHz Wi-Fi, it differs significantly from Wi-Fi in two key aspects. First, RadSee has a bandwidth of 1.1 GHz, while Wi-Fi devices typically operate within a bandwidth of 20 or 40 MHz. Second, RadSee utilizes an FMCW waveform, whereas Wi-Fi devices use an Orthogonal Frequency-Division Multiplexing (OFDM) waveform. OFDM waveforms are characterized by pseudo-noise-like signals. When an OFDM signal is correlated with an FMCW signal over time, the correlation result is nearly zero. Therefore, in theory, RadSee is resilient to radio interference from the Wi-Fi devices in its proximity.

To validate the above theory, we conducted experiments by observing RadSee's IF signals in two cases: with and without radio interference from a Wi-Fi device, as shown in Fig. 3.7. To better control the experiments, we use a Universal Software Radio Peripheral (USRP) device for continuous Wi-Fi signal generation at two frequencies: 5.480 GHz and 5.805 GHz. The bandwidth of Wi-Fi signals is 20 MHz. The scene is static during the experiments. Fig. 3.7 presents RadSee's IF signals (i.e., the input of DNN) in three cases: i) no radio interference from the Wi-Fi device, ii)

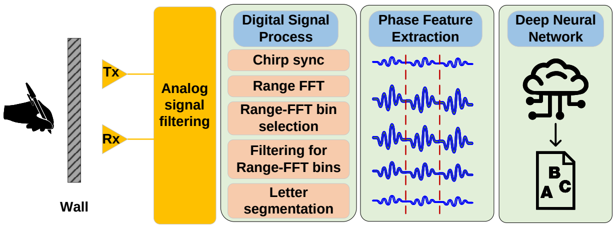


Figure 3.8: RadSee process overview.

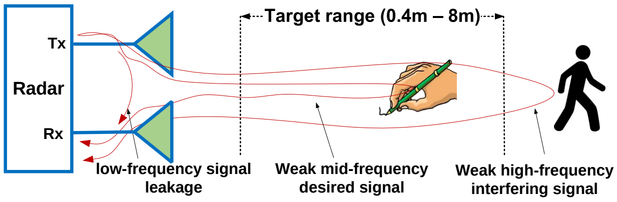

Target range (0.4m – 8m)
Tx
Radar
Rx
low-frequency signal leakage
Weak mid-frequency desired signal
Weak high-frequency interfering signal

Figure 3.9: Illustration of the received signals at the radar.

radio interference from 5.480 GHz Wi-Fi device, and iii) radio interference from 5.805 GHz Wi-Fi device. It can be seen that the IF signals generated by RadSee are almost the same in these three cases. This indicates that RadSee is resilient to radio interference from Wi-Fi devices.

3.5: RadSee: Data Processing

In this section, we present the signal processing pipeline of RadSee, as outlined in Fig. 3.8. We first elaborate on the signal processing modules for phase feature extraction and then use k-nearest neighbor (kNN) to validate the extracted features.

3.5.1: Signal Processing

Analog Signal Filtering. The received signal at RadSee may have different components, including RF leakage on PCB, desired echo from handwriting, and undesired echo from other moving objects, as shown in Fig. 3.9. Since the RF leakage signal is very close to zero frequency, RadSee uses a high-pass filter with 5 kHz cutoff frequency to suppress the RF signal leakage. Meanwhile, the undesired high-frequency signal from other moving objects may generate interference to the desired signal if not suppressed in the analog domain. To do so, RadSee employs a first-order low-pass filter with a bandwidth of 100 kHz for the suppression of high-frequency echoes from undesired moving objects. Combining the high-pass and low-pass filters, RadSee has a band-pass filter from 5 kHz to 100 kHz, corresponding to a target range from 0.4 m to 8 m for handwriting detection.

Range-FFT. RadSee sets its chirp cycle time to 1 ms. For each chirp cycle, RadSee sets its transmission time to 0.6 ms and idle/delay time to 0.4 ms as shown in Fig. 3.10(a). As the PLL

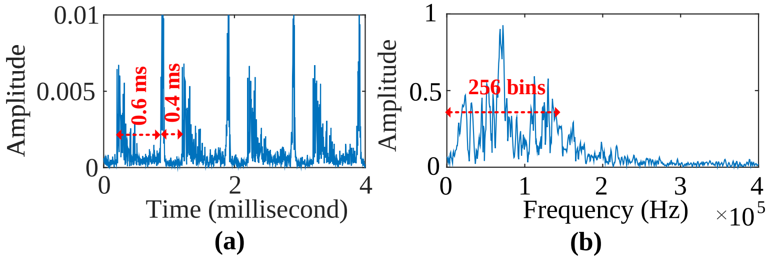


Figure 3.10: Illustration of the IF signal. (a) the IF signal in time domain. (b) the IF signal after FFT operation.

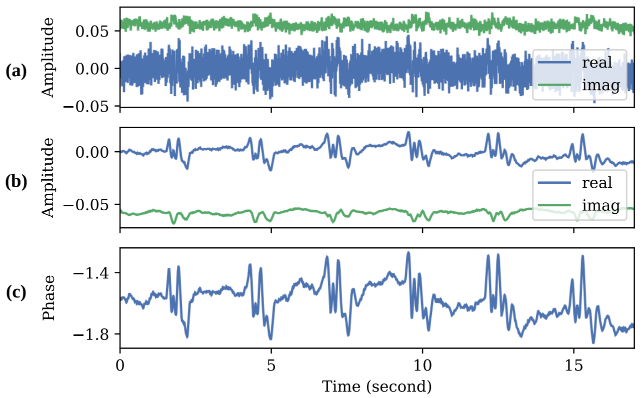


Figure 3.11: (a) The original signal of one Range-FFT bin (one sample per chirp cycle); (b) the Range-FFT bin after DC adjustment and low-pass filter; (c) phase of the signal in (b).

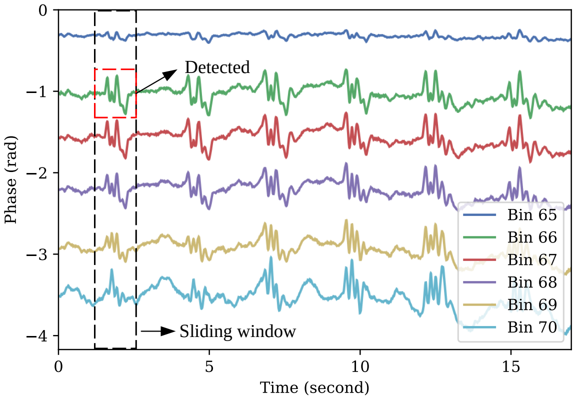


Figure 3.12: Phase sequence of six Range-FFT bins.

and VCO are typically not very stable at the beginning and end of their frequency ramping, RadSee discards 0.05 ms at the beginning and at the end of its transmission period, resulting in only 0.5 ms for useful signal reception. To best observe this useful signal in the digital domain, RadSee samples its received signal at 5 MSps. As a result, it obtains 2,500 complex samples from each chirp cycle. To further improve the range resolution, RadSee adds zeros behind the 2,500 samples to perform 8,192-point Range-FFT operation. The resultant Range-FFT bins are shown in Fig. 3.10(b). Of the resulting 8,192 bins, only the first 256 are under examination.

Filtering for Range-FFT Bins. For each Range-FFT bin of interest, RadSee first adjusts its DC component to the dynamic range of its real and imaginary parts, and then applies a low-pass filter to remove the high-frequency component. As per [154], RadSee sets the low-pass filter's bandwidth to 5 Hz. Fig. 3.11 compares the data sequences of one Range-FFT bin *before* and *after* the DC adjustment and low-pass filter. It can be observed that the process can manifest the phase pattern of handwriting effectively.

FFT Bin Selection. Experiments show that handwriting will cause multiple bins to fluctuate. This can be attributed to the high range resolution and the multi-path effect within antenna's aperture. Instead of using a single Range-FFT bin, RadSee uses *multiple consecutive* Range-FFT bins to extract their phase patterns. The questions need to be answered: (i) how many Range-FFT bins should be selected, and (ii) which Range-FFT bins should be used. For the first question, RadSee empirically selects *five* consecutive Range-FFT bins and uses their phase information for letter classification. For the second question, RadSee selects the Range-FFT bins of the *smallest* index but with its phase variance larger than a predefined threshold. RadSee's bin selection algorithm is provided in Alg. 1. Its core idea is to identify five consecutive FFT-Range bins based on their phase variances, so that the handwriting movement pattern can be captured along the line-of-sight (shortest) through-wall path. These five bins are then fed into our DNN for letter recognition. Fig. 3.12 shows a sample of our observed Range-FFT bins in handwriting detection. In this case, RadSee selects bins 66 to 70 as the input of its DNN model for letter classification.

Algorithm 1 RadSee's bin selection algorithm.

Require: Range-FFT phase matrix $[S(i,t) \in \mathbb{R}]_{N \times T}$, where i is bin index ($0 \leq i < N$), t is time index ($0 \leq t < T$), window size W, predefined lower bound of variance θ_{lw}, predefined upper bound of variance θ_{up}. ▷ *In our experiments,* $W = 500$, $N = 256$, $T = 5000$, $\theta_{lw} = 0.03$, $\theta_{up} = 0.18$.

Ensure: The smallest bin index i where the phase variance exceeds θ_{lw} but is lower than θ_{up}.

```
for $t = 0$ to $T - W$ do
    for $i = 0$ to $N$ do
        Calculate window-slided variance as follows:
        $v(i,t) = \frac{1}{W}\Sigma_{j=t}^{t+W-1}|S(i,j) - \mu|^2$,
        where $\mu = \frac{1}{W}\Sigma_{j=t}^{t+W-1}S(i,j)$.
        if $v(i,t) > \theta_{lw}$ & $v(i,t) < \theta_{up}$ then
            return $i$
        end if
    end for
end for
return $-1$                ▷ Indicate no writing activity is detected.
```

Data Segmentation. RadSee performs data segmentation on the phase stream of the selected

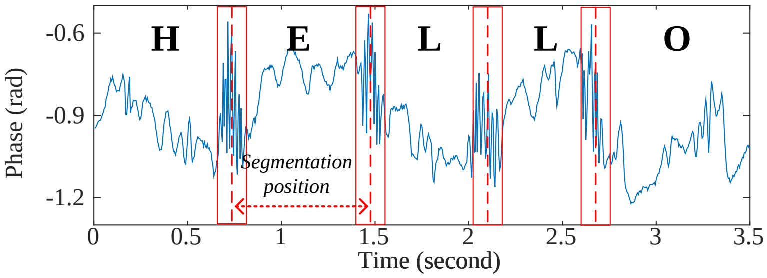


Figure 3.13: Illustrating the rapid phase change of a target Range-FFT bin during the transition of writing letters.

Range-FFT bins to extract the meaningful features that correspond to individual letters. RadSee employs different methods for phase data segmentation at the training and test phases. We elaborate them as follows.

(i) During Training Phase: Since we have full control of the training data collection, we ask every participant to stop and be still for one second after writing each letter. By doing so, RadSee can easily segment phase sequence and extract meaningful phase data for individual letters. *(ii) During Test Phase:* In this phase, RadSee has no control over the writing style of a victim. Likely, the victim writes in a continuous manner without a stop in the middle. Interestingly, we always observed a rapid phase change during the transition from writing one letter to another. Fig. 3.13 shows an example of our observations. This is caused by the pen-holding hand's quick movement during the transition period. RadSee leverages this signature to segment the phase data streams. Since the time duration of writing different letters may be different, the data sequences corresponding to different letters are of heterogeneous length.

Extracted Phase Features. Based on the above process, RadSee will obtain the phase data segments corresponding to individual letters being written. Fig. 3.14 shows some samples of its obtained phase segments from different users. From the figure we have the following observations. First, for the same user, the phase patterns of different letters are different. This is an encouraging observation as the uniqueness of phase patterns is the foundation of letter classification. Second, for the same letter (e.g., letter 'A' in Fig. 3.14), the phase patterns from different users look different. So far, it is not clear if those phase patterns will be classified to the same letter through an advanced

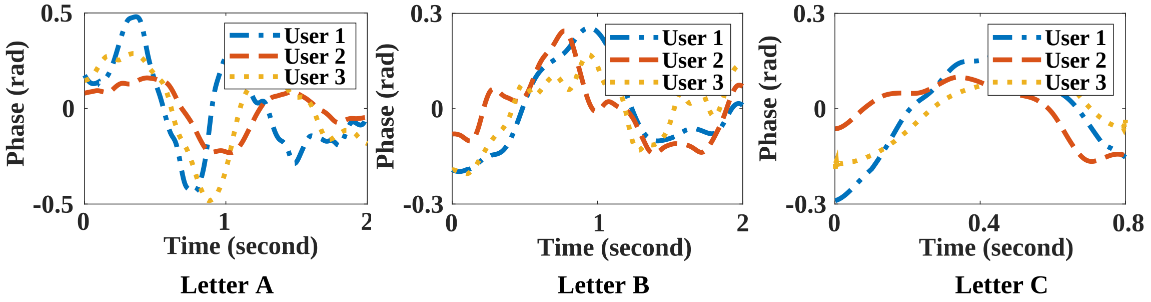


Figure 3.14: The observed phase sequences when three users are writing letters ‘A’, ‘B’, and ‘C’.

transformation. To better understand this question, we conduct feature validation using kNN.

3.5.2: kNN-based Feature Validation

We use the kNN model [41] to validate the effectiveness of the extracted features. kNN is a simple data classification method that estimates the belonging of a new data sample based on a set of labeled data samples. When a new data sample comes, the distance between this new sample and all labeled samples is calculated. Then, the k closest neighbors are selected. The selected k closest neighbors cast weighted votes (using their distance) to make the final classification decision for the new data sample. One issue with kNN in this case is that the length of data samples (phase sequences) is not fixed, i.e., different phase sequences have different lengths. To address this issue, we employ Dynamic Time Warping (DTW), which has been widely used in speech recognition [43] and data mining [83]. DTW can find an optimal alignment between the two sequences by warping the time axis non-linearly.

Data Set. We collected the phase data samples for 62 letters (a-z, A-Z, and 0-9) from 12 users. Each user was asked to write in print writing style on a desk that is one meter away from the wall. Our radar was placed just behind the wall to collect the phase data. Each letter has 10 samples from a user and a total of 120 samples from those 12 users. In total, 7,440 samples were collected for all 62 letters, all of which were labeled during the data collection. The data samples are divided into two groups: those from the first 6 users are used for training, while those from the second 6 users are used for test.

Validation Results. We perform kNN on the collected data set. As an example, Fig. 3.15 shows

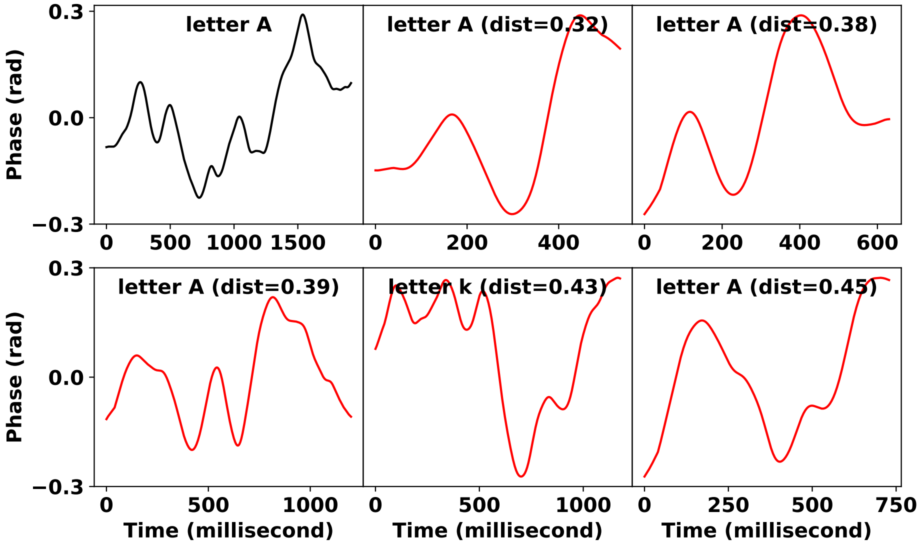


Figure 3.15: Results of using kNN to search 5 closest neighbors for a new data sample. The top-left figure shows the phase sequence of the new data sample. The remaining 5 figures show the found 5 closest data samples (and their corresponding letters) in our training data set.

the search results of kNN when the new data sample is the phase sequence of letter ‘A’. It can be seen that, of the five closest data samples in the training data set, four are correct (labeled with ‘A’) and one is incorrect (labeled with ‘k’). The five closest data samples cast votes to make the final decision. The weighted vote for ‘A’ is 10.54, while the weighted vote for ‘k’ is 2.32. Based on the voting result, this new data sample is classified to letter ‘A’, which is correct.

Fig. 3.16 shows kNN’s classification accuracy when the test data samples are from 6 different users. We note that the test data samples and the training data samples are from different users. As we can observe, the classification accuracy is from 53% (user 4) to 77% (user 3). This could be attributed to two factors: i) most of training data are from Asian participants; and ii) User 4 is an American participant while other five users are Asian participants.

We then evaluate kNN’s classification accuracy using the data samples from User 6 when the radar was placed at different distances (1 m, 2 m, and 3 m). The training data samples were collected from six different users when the radar was placed at 1 m distance. Fig. 3.17 presents the classification results. It shows that the classification accuracy is 68% when the test was conducted at the same distance. However, when RadSee has a different distance from the victim, its detection

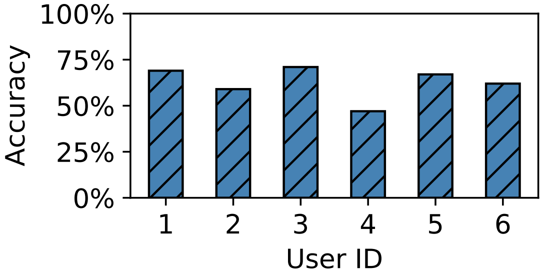


Figure 3.16: kNN's classification accuracy when test and training data are from different users.

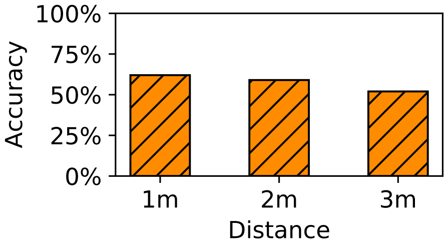


Figure 3.17: kNN's classification accuracy when radar is at different distances.

accuracy decreases to 58%.

Limitations of kNN. The kNN-based classification results indeed manifest the effectiveness of phase features in handwriting letter classification. But this approach has two limitations. First, it has a very high computational complexity and thus limits the size of the labeled (training) data set. Second, it uses the phase sequence from only one Range-FFT bin for classification. Using those five Range-FFT bins together may improve the classification accuracy. In what follows, we design a DNN-based approach for handwriting recognition, with the aim of overcoming the above limitations and improving the classification accuracy.

3.6: RadSee: DNN-based Recognition

In this section, we focus on designing a DNN model for through-wall handwriting recognition using the phase features extracted in the previous section. Compared to kNN, DNN is much more efficient in computation and is more appealing for practical use.

3.6.1: DNN Model

In essence, this letter recognition problem is a classification problem with its input being multi-dimensional phase sequences and its output being the probability of each letter in the candidate set (a-z, A-Z, and 0-9). We found that this task is similar to many classification tasks in natural language processing (NLP), such as information status classification [70] and stress detection [178]. Following the state-of-the-art classification techniques in NLP, we employ an attention-based Bidi-

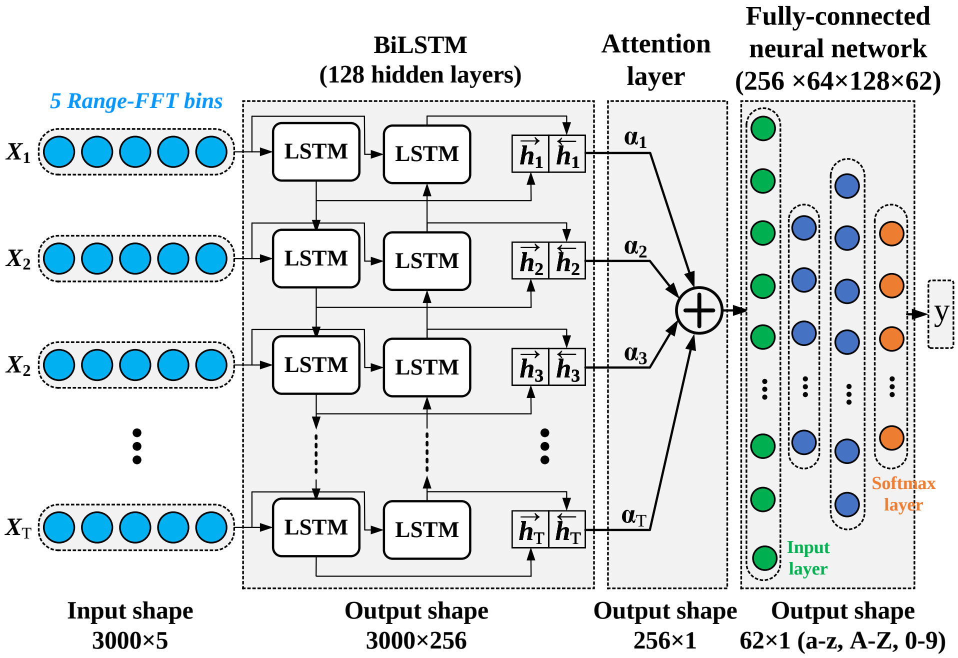


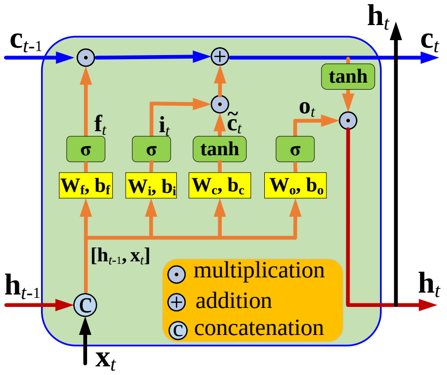
Figure 3.18: The structure of an attention-based BiLSTM model for letter recognition. The input is the phases of selected 5 Range-FFT bins over $T = 3000$ ms, and the output is the classified one from the 62 characters.

rectional LSTM (BiLSTM) model for RadSee's letter classification.

Fig. 3.18 shows the high-level structure of our attention-based BiLSTM model. The BiLSTM component is used to extract the temporal features in the time-series phase sequence. The attention layer is used to capture the key movement information of handwriting. This is critical as the key information of handwriting movement likely lies in some turning points. This attention layer will allow the model to focus on specific parts (e.g., those turning points) of the phase sequences, thereby improving the accuracy and efficiency of classification.

3.6.2: BiLSTM

BiLSTM is a variant of the LSTM network [69] and has demonstrated its effectiveness for a wide range of NLP tasks such as machine translation [153], part-of-speech tagging [97], and sentiment analysis [169, 215]. In a BiLSTM, the input sequence is processed in both forward and backward directions using two separate LSTM layers. This allows the model to capture both past and future context for each input element. This is crucial for handwriting recognition, because the turning



$$
\begin{aligned}
\mathbf{f}_t &= \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f)\\
\mathbf{i}_t &= \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i)\\
\mathbf{o}_t &= \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o)\\
\tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c)\\
\mathbf{c}_t &= \mathbf{f}_j \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t\\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t),
\end{aligned}
$$

Figure 3.19: The structure and operation of an LSTM cell ($\mathbf{h}_t \in \mathbb{R}^{128\times 1}$, $\mathbf{c}_t \in \mathbb{R}^{128\times 1}$, and $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_c, \mathbf{W}_o \in \mathbb{R}^{128\times 133}$).

points of handwriting movement carry the key information for letter classification but the turning points may appear at the beginning, in the middle, and at the end of a phase sequence. The use of BiLSTM allows the model to capture those turning points at any pace of the input phase sequence.

Input Data. We set the input data shape to be 3000×5, where 3,000 is the number of chirps and 5 is the number of selected Range-FFT bins. Recall that each chirp is 1 ms. This means that the maximum time of writing a letter is 3 seconds. In most cases, one can finish the writing of a letter less than 3 seconds. If the phase sequence is less than 3,000 points, we simply pad zero behind the phase sequence as the input of BiLSTM. If the phase sequence is greater than 3,000, we trim the head and tail of the phase sequence, retaining only 3,000 points in the middle as input for the BiLSTM.

LSTM Cell. LSTM has been used in a wide range of learning tasks. It is the key component of the BiLSTM model as shown in Fig. 3.18. It allows the model to selectively retain or forget information at each time step. The cell structure includes three gates: an input gate, a forget gate, and an output gate. The input gate determines which information should be stored in the cell, the forget gate determines which information should be discarded, and the output gate determines which information should be used for the current output. Fig. 3.19 shows the structure and parameters of each LSTM cell.

BiLSTM Structure. As shown in Fig. 3.18, BiLSTM has two LSTM cells: one is for forward information flow, and the other is for backward information flow. In each iteration t, it combines the hidden states of forward and backward LSTMs through concatenation: $\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t]$, where $\overrightarrow{\mathbf{h}}_t$ is the hidden state from the forward LSTM, $\overleftarrow{\mathbf{h}}_t$ is the hidden state from the backward LSTM, and $\mathbf{h}_t$ is the hidden state of the BiLSTM. Since each LSTM has 128 hidden layers, we have $\mathbf{h}_t \in \mathbb{R}^{256\times 1}$, with $t = 1, 2, \ldots, 3000$. Then, the combined hidden states are fed to the attention layer for further processing.

3.6.3: Attention Layer

The attention mechanism is probably one of the most important inventions for deep learning and it has been used for many applications such as GPT [32, 125, 169, 193]. With the attention layer, the model learns to focus on some key parts of the data sequence. During the handwriting of a letter, some turning points may carry critical information for letter classification. The attention layer attempts to learn the importance of each part of the phase sequence and then assigns them with proper weights. To calculate the corresponding weights, it first feeds $\mathbf{h}_t$ to a one-layer Multilayer Perceptron (MLP) to learn a hidden representation u_t, and then normalizes the weights to generate α_t. Mathematically, it can be written as follows:

$$u_t = \tanh(\mathbf{W}_h^\top \mathbf{h}_t + b_h), \tag{3.5a}$$

$$\alpha_t = \frac{\exp(u_t)}{\sum_{k=1}^{T} \exp(u_k)}, \tag{3.5b}$$

$$\mathbf{s} = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t, \tag{3.5c}$$

where $\mathbf{W}_h \in \mathbb{R}^{256\times 1}$ is the training weights, $b_h \in \mathbb{R}$ is a training bias, and $\mathbf{s} \in \mathbb{R}^{256\times 1}$ is the weighted vector for the fully-connected neural network in Fig. 3.18. The fully-connected network is of $256 \times 64 \times 128 \times 62$ size. The last layer is a SoftMax layer to calculate the possibility of each letter candidate (a-z, A-Z, and 0-9). The letter of the highest possibility is selected as the output y.

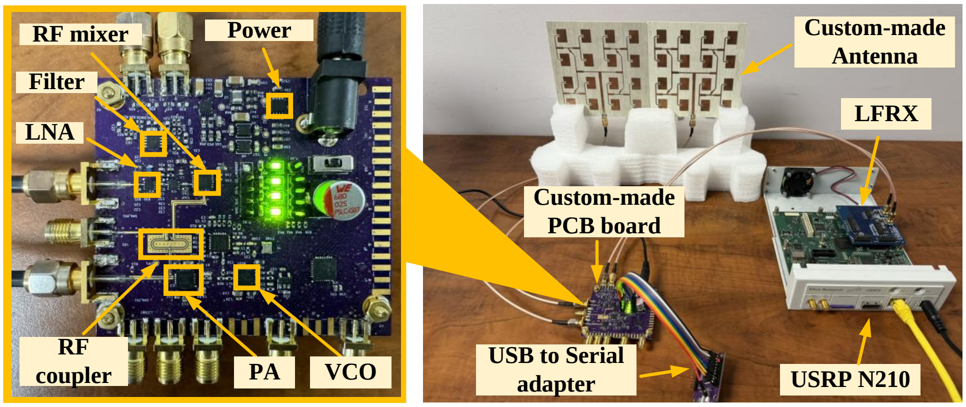


Figure 3.20: Radar PCB (left) and a picture of RadSee (right).

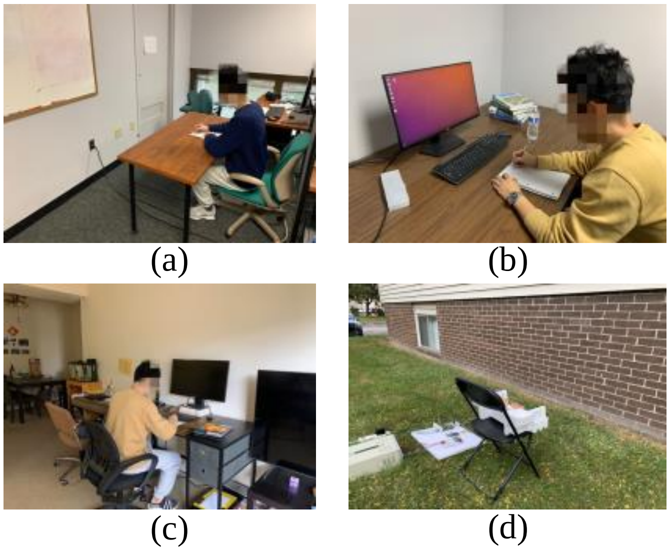
3.7: Implementation

3.7.1: Hardware

Fig. 3.20 shows the hardware components of RadSee. We fabricated a radar PCB board as shown in this figure. The electronic components of this board include VCO, LNA, PA, Tx/Rx 16 dB RF coupler, RF quadrature mixer, and baseband filter. This PCB was made by OSH Park using FR408 substrate. We designed, simulated, and optimized 4×4 patch-array antennas using HFSS for radio signal transmission and reception. These antennas offer 18 dBi antenna gain for both transmission and reception. In total, it offers 36 dBi gain for the link path, making it possible to compensate the signal penetration loss of a wall. The total cost of RadSee is approximately $500, including $50 for PCB fabrication, $50 for antennas, and $400 for chips. We use USRP N210 with LFRX daughterboard to convert the analog signal to digital I/Q samples, which were then sent to a computer for data process. Transmission power is set to 20 dBm. The FMCW radar sweeps from 5.4 GHz to 6.5 GHz. The time duration of one chirp period is 1 ms, including 600 μs for frequency sweeping and 400 μs for idle.

3.7.2: Algorithms

Digital Signal Processing. We implemented the data processing algorithms on a laptop in C++ using GNU Radio Out-of-Tree (OOT) module. The laptop receives a continuous data stream from



(a) (b) (c) (d)

Figure 3.21: Evaluation setting: (a) Laboratory scenario. (b) Office scenario. (c) Apartment scenario. (d) RadSee attacks from outside of the apartment.

the radar. It needs to synchronize the chirp signal and extract the useful data samples of each chirp. Fortunately, due to the presence of 400 μs idle period of each chirp, it is easy to identify the useful data samples from the data stream. Specifically, we use the high peaks as shown in Fig. 3.10 to extract the useful data samples. One fundamental issue with the current hardware design is the lack of clock synchronization between ADC and FMCW chirps. To address this issue, we use a high sampling rate 5 MSps and perform fine-grained synchronization to identify the first data sample corresponding to the starting moment of each chirp.

Data Collection for DNN Training.[1] We collected training data in a laboratory. The radar was placed behind an interior drywall at a distance of 0.5 m. A writing desk was placed in front of the wall at a distance of 1 m, as shown in Fig. 3.21(a). Eighteen participants (4 American, 3 Indian, 4 Middle East, 7 Chinese) were asked to write 62 characters (a-z, A-Z, and 0-9) on the desk. Each participant wrote every character 60 times. In total, we collected $18 \times 62 \times 60 = 66,960$ data samples. Of the eighteen participants, twelve were asked to write in the *print* style, while six were asked to write in the *cursive* style. Regarding handedness, two of them were left-handed writers while the rest were right-handed writers. The handedness and writing styles of the participants are

[1] The experiments were conducted under FCC experimental spectrum license with Call Sign # WM2XWQ and File # 0954-EX-CN-2022.

Table 3.2: Participants for training and test data collection.

	Participants for training				Participants for test			
Handedness	Right-handed		Left-handed		Right-handed		Left-handed	
Writing style	Print	Cursive	Print	Cursive	Print	Cursive	Print	Cursive
# of participants	11	5	1	1	7	3	1	1

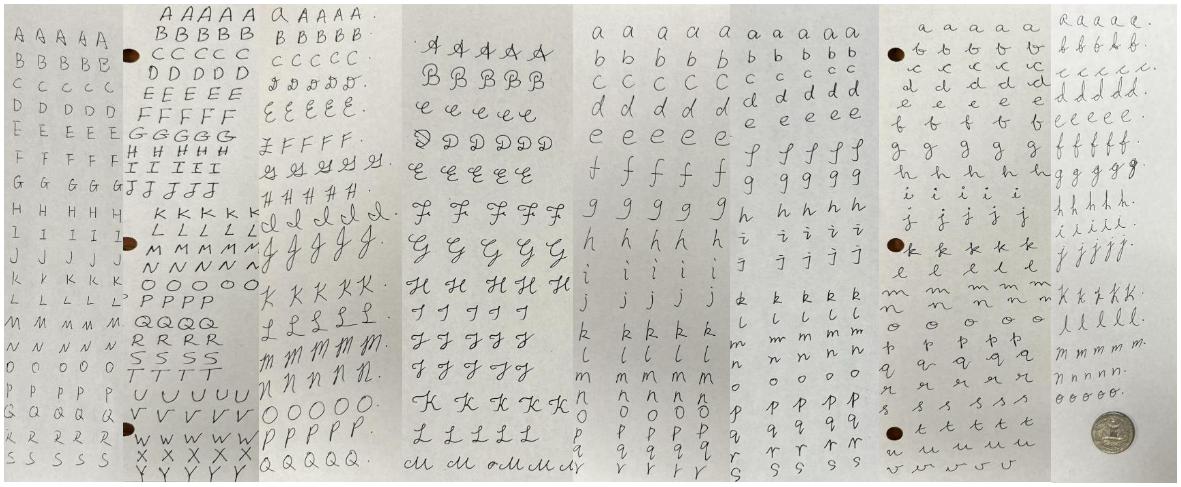


Figure 3.22: Writing samples from participants for training.

summarized in Table 3.2. Some writing samples from the participants are provided in Fig. 3.22.

DNN Training. The DNN model was implemented using TensorFlow's Keras library. We used cross entropy as loss function. During the training process, we set the batch size to 2,000 and trained the model for 500 epochs. We used Adam optimizer with a learning rate of 7e^{-4} to train the model.

3.8: Experimental Evaluation

3.8.1: Letter Recognition Accuracy

Write on A4 Papers. Recall that our training data was collected in a laboratory from eighteen participants. To evaluate the recognition accuracy of RadSee, we completely separate the training and test datasets. We invited twelve new participants (4 American, 4 Chinese, 2 Indian, 2 Middle East) to write letters in the same setting (i.e., sitting 1 m away from the wall and facing to the radar). None of these twelve people participated in the training data collection. Each of them wrote 300

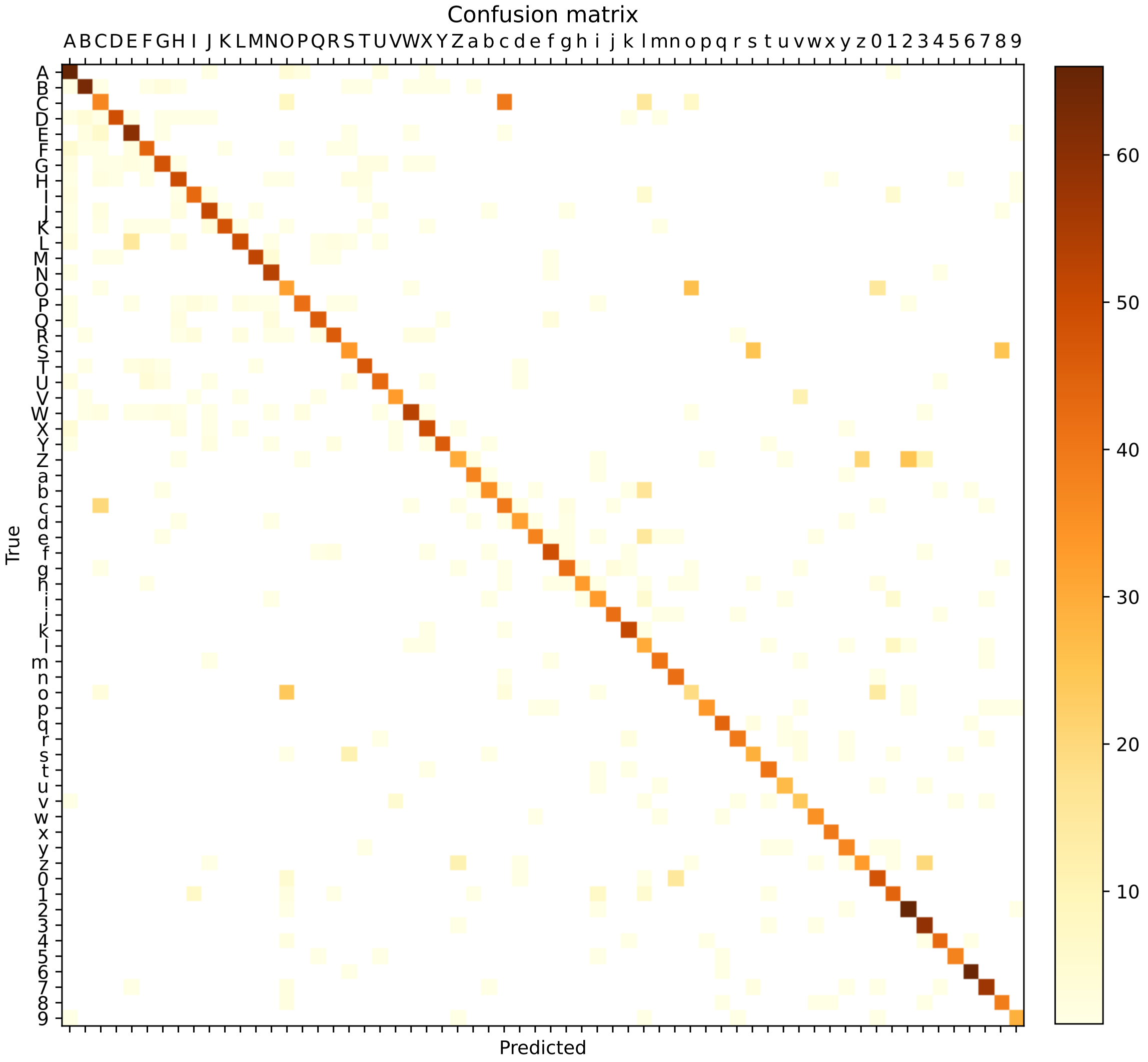


Figure 3.23: Confusion matrix of RadSee's letter recognition results.

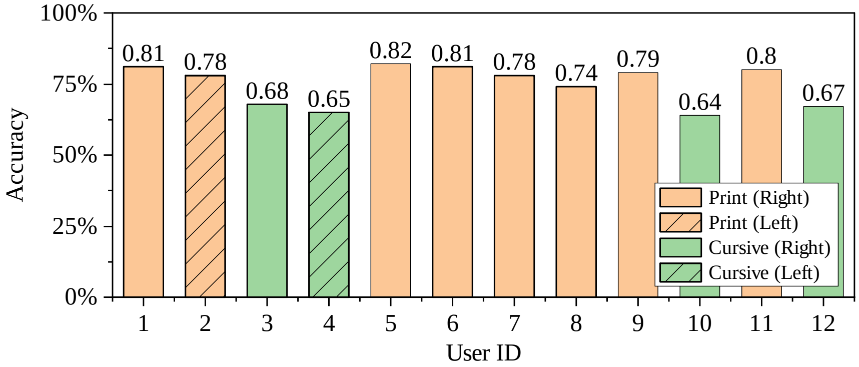


Figure 3.24: RadSee's letter recognition accuracy when participants wrote on A4 papers. Users 1-4 are Americans, users 5-8 are Chinese, users 9-10 are Indians, and users 11-12 are from Middle East.

random letters on A4 papers. During the test, eight participants were asked to write in the *print* style, and four were asked to write in the *cursive* style. Both print and cursive writing letters are within the size of 5 mm to 10 mm. Regarding handedness, ten participants were right-handed writers, while two were left-handed writers. The handedness and writing style are summarized in Table 3.2.

Fig. 3.23 shows the confusion matrix of RadSee's letter recognition results. It is evident that RadSee can recognize most of the letters. RadSee is prone to making mistakes for some letters. For instance, it can easily confuse 'O' with 'o', 'C' with 'O', and 'I' with '1'. Other errors can arise from cursive writing, such as confusing 'S' with '8' and 'Z' with '3'. This is understandable, as their handwriting patterns are similar to each other. Overall, RadSee achieves 75% letter recognition accuracy.

Print vs. Cursive. Fig. 3.24 presents RadSee's letter recognition accuracy for the 12 individual participants. As observed, RadSee has a lower recognition accuracy for the participants who wrote in cursive style compared to those who wrote in print style. This observation can be attributed to two factors. First, cursive writing is more individualized and diverse, making it challenging for the model to extract consistent features across different participants, despite having cursive-style data in the training dataset. Second, our segmentation method relies on detecting signal transitions between letters, which becomes more difficult when people write in cursive style.

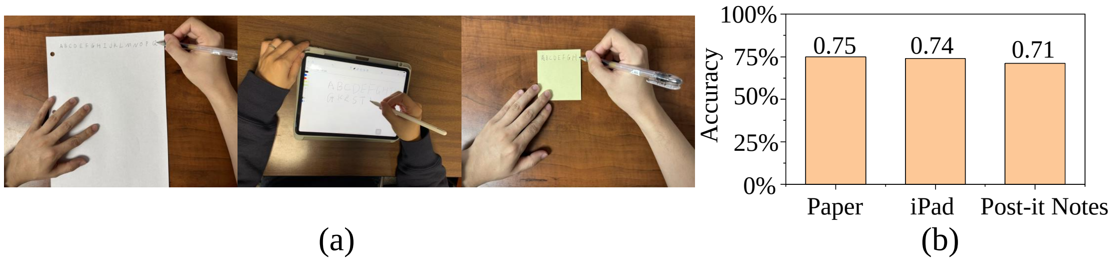


Figure 3.25: Writing on different media. (a) Writing on papers, iPad, and Post-it notes. (b) The recognition accuracy of RadSee when writing on different media.

Writing Handedness. Besides writing style, handedness is another factor that may affect RadSee's letter recognition accuracy. However, experimental results show that handedness affects RadSee very slightly. As shown in Fig. 3.24, RadSee has a very similar performance for both left-handed and right-handed users. This can be attributed to the fact that most left-handed individuals have the same writing movement pattern as right-handed individuals, i.e., write from left to right and from top to bottom.

Write on iPad and Post-it Notes. Tablets, such as Apple iPad, have become increasingly popular for writing activities, with many individuals opting to use them for important documents instead of traditional pen and paper. To evaluate the performance of writing on an iPad, we repeated our measurements by asking twelve participants to write 300 random letters using an Apple Pencil. The experimental results are shown in Fig. 3.25(b). RadSee achieves 74% letter recognition accuracy. In the same setting, RadSee achieves 75% letter recognition accuracy when participants write on A4 papers. This indicates that RadSee has almost the same performance for A4 paper and iPad writing recognition. Another commonly used medium for writing is Post-it notes. Given their smaller size, we asked participants to write 20 random letters on Post-it notes. RadSee's letter recognition accuracy for Post-it notes is 71%, as presented in Fig. 3.25(b). As shown in Fig. 3.25(a), these three writing media have different horizontal writing ranges. Since RadSee has similar performance for them, it suggests that RadSee effectively accommodates the horizontal range for writing on A4 papers, iPad, or Post-it notes.

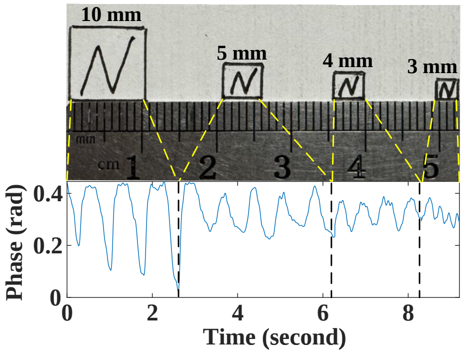


Figure 3.26: RadSee's phase signal for different letter size.

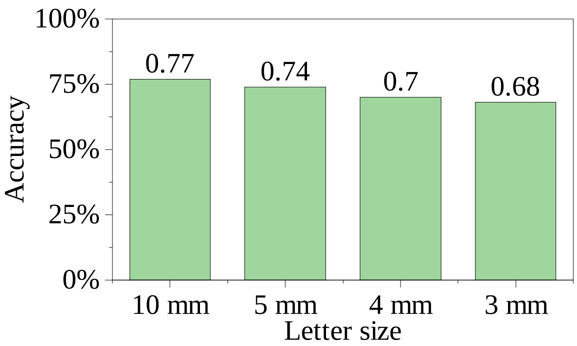


Figure 3.27: RadSee's accuracy for letters of different sizes.

3.8.2: Impact of Letter Size

We conducted experiments to better understand RadSee's ability of detecting small-size letters. Fig. 3.26 presents RadSee's signal changes when a participant wrote letter 'N' of different sizes. Evidently, RadSee is capable of detecting as small as 3 mm handwriting movement. We further asked one participant to write on A4 papers with grid boxes of different sizes: 3 mm × 3 mm, 4 mm × 4 mm, 5 mm × 5 mm, and 10 mm × 10 mm. The participant was instructed to write letters within the boundaries of the grid boxes. However, for the 3 mm × 3 mm grids, since the boxes were too small, a considerable portion of the written letters exceeded the boundaries. Fig. 3.27 presents RadSee's letter recognition accuracy in these four cases. It is evident that RadSee's accuracy decreases with the letter size. But notably, RadSee achieves 68% recognition accuracy even in the case where the letter size is confined within 3 mm.

3.8.3: Impacts of Distance and Angle

When an attacker attempts to detect the handwriting behind a wall, it may not know the distance from itself to the victim and the angular direction of the victim. The attacker may use RadSee to do an exhaustive search to find the best pointing direction for the radar's antennas, but the radar-antenna-pointing direction may not be accurate. To evaluate RadSee's robustness, we examine its accuracy in different settings: (i) the writers are 1 m, 2 m, and 3 m behind the wall; and (ii) RadSee's antenna is pointing to different angles (0°, 10°, 20°, and 30°). The combination constitutes 12

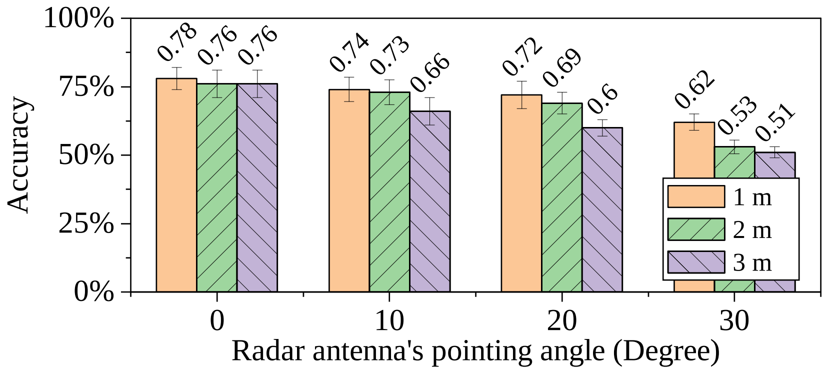
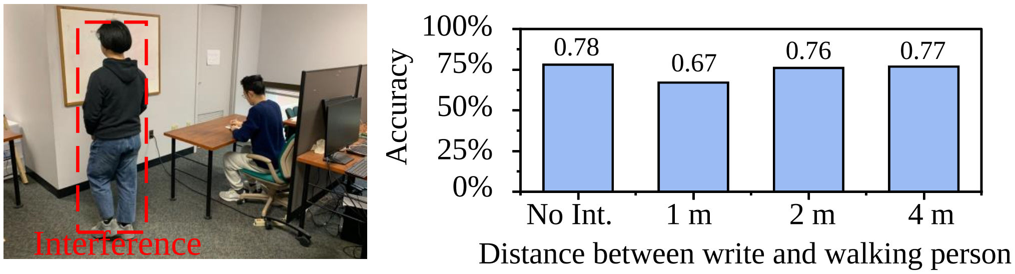


Figure 3.28: Letter recognition accuracy of RadSee when writers are at different distances and different angles from the wall.

different cases. In each case, we instructed eight participants to write 300 letters using their normal handwriting habits.

Fig. 3.28 presents our measured accuracy and deviation. It can be seen that RadSee is robust to the distance change. This can be explained by its design. In nature, FMCW radar is capable of precisely capturing the movement features at different distances. When the distance between the writer and the wall changes from 1 m to 3 m, RadSee will identify another 5 Range-FFT bins for phase feature extraction. Since the handwriting movement patterns are not related to the wall distance, the extracted features will remain unchanged. Therefore, RadSee is robust to distance changes.

Fig. 3.28 also presents our measurement results when RadSee's antennas was pointing to different angles. Evidently, RadSee's accuracy decreases when its directional error increases from 0° to 30°. Specifically, when RadSee was pointing to 0°, it achieved 77% recognition accuracy. When RadSee was pointing to 30°, it achieved 55% recognition accuracy. In all cases, the standard deviation is almost the same, i.e., 4%. This degradation can be attributed to the directivity of the patch-array antennas, as shown in Fig. 3.6. When the writer deviates from its central direction, the patch antenna's effective radiation power decreases, making noise and other imperfections more significant and thus leading to a decreased accuracy.



(a) (b)

Figure 3.29: Interference test. (a) Interferer is 2 meters from writer. (b) RadSee's resilience to interference from a walking person.

3.8.4: Impact of Interference from Other Moving Objects

Experimental results in Fig. 3.7 have confirmed that RadSee is immune to radio interference from in-band (5 GHz) Wi-Fi devices. All experiments in this work were conducted in office and laboratory environments, which are rich with interference from multiple Wi-Fi sources. Therefore, the experimental results presented have already taken into account the radio interference from multiple Wi-Fi sources. Additionally, RadSee is not affected by static objects (e.g., desks and chairs) around a writer as they appear to be a constant in the received signal, which can be easily mitigated. Therefore, we focus on studying RadSee's performance in the presence of moving objects (e.g., walking persons) in the proximity of the writer. We emulated this scenario by asking another person to walk around the writer as shown in Fig. 3.29(a). We measure the recognition accuracy of RadSee in three cases, i.e., the distance between a writer and a walking person is 1 m, 2 m, and 4 m. We asked eight participants to write 300 random letters in each case and measured RadSee's letter recognition accuracy.

Fig. 3.29(b) depicts our measured results. We can see that the performance degradation depends on the distance between the writer and the interferer. The closer the interferer is, the larger performance degradation RadSee has. For the case where interferer is 1 m away, RadSee demonstrates 67% letter recognition accuracy, with 11% accuracy degradation compared to the case without interference. When the interferer is 2 m away, RadSee rapidly increases its accuracy to 76%, which is close to its accuracy in the case without interference. We note that the participants in all experiments

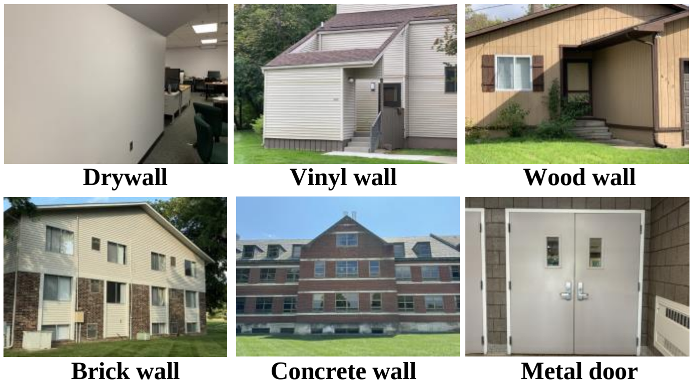


Figure 3.30: Illustration of six different types of wall materials.

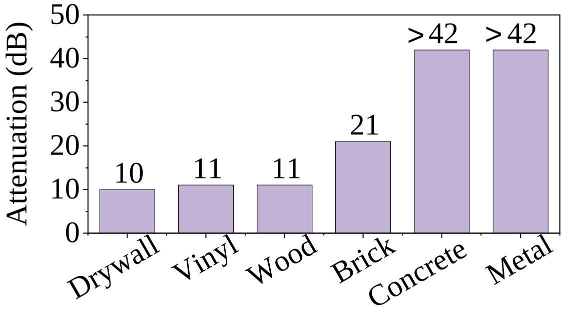


Figure 3.31: RF signal's power attenuation for penetrating a wall of different materials.

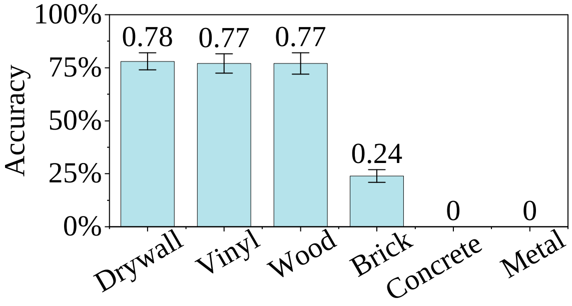


Figure 3.32: RadSee's recognition accuracy when placed behind six wall materials.

maintained normal physiological activities, such as breathing and respiration. The experimental results reported above have already taken into account those normal physiological activities of the writers.

3.8.5: Impact of Different Wall Materials

RF signals have varying penetration abilities depending on the type of wall. We conducted experiments to evaluate the performance of RadSee in detecting letters through different wall materials. Specifically, we considered six wall materials as shown in Fig. 3.30: drywall (12 cm), vinyl wall (20 cm), wood wall (19 cm), brick wall (22 cm), concrete wall (23 cm), and metal door (4 cm). We first measured their penetration loss, which refers to the power attenuation of radio signals as they pass through a wall. Fig. 3.31 presents our measurement results. It is evident that drywall, vinyl and wood walls have similar penetration loss for radio signal, which is about 10 dB. Brick wall is more lossy for radio signal compared to wood wall. Its penetration loss is about 21 dB. However,

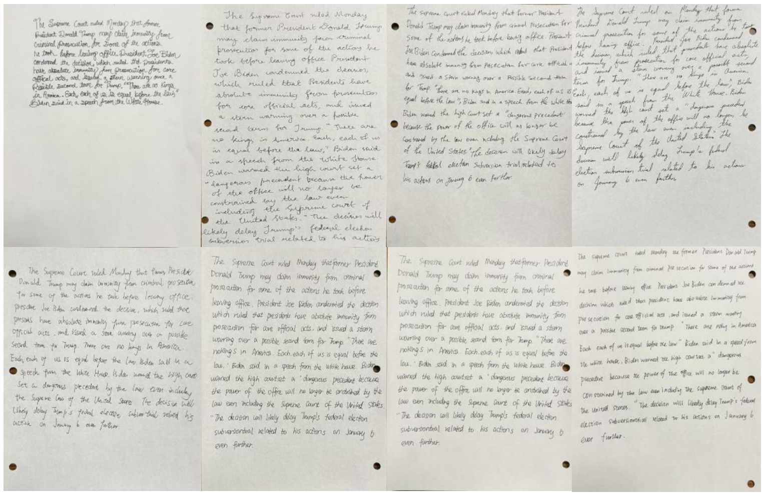


Figure 3.33: Writing samples from participants as they transcribed CNN news articles in both print and cursive styles.

concrete walls and metal doors completely block radio signals. Their attenuation loss is greater than 42 dB.

We then conducted experiments to measure RadSee's letter recognition accuracy. Eight participants took part in the experiments. They were seated 1 meter away from the wall, while RadSee was positioned 0.5 meters away on the other side of the wall as shown in Fig. 3.21. Each of the eight participants wrote 300 random letters using his/her own writing style. Fig. 3.32 presents the experimental results. It shows that RadSee achieves similar performance when participants wrote behind drywall, vinyl, and wood walls. This similarity is due to the comparable electromagnetic properties of these materials. In contrast, a brick wall significantly reduces recognition accuracy, with RadSee achieving only 24% letter recognition accuracy in this scenario. Furthermore, concrete walls and metal doors completely obstruct letter detection.

3.8.6: Word Recognition Accuracy in Content

In addition to detecting individual letters, we evaluate RadSee's performance of recovering entire sentences. This is important because an attacker's interest may lies in the content that a victim is writing, rather than individual letters. We asked twelve participants to reproduce an CNN News article, which is about 300 words. Some writing samples from the participants are provided in

Table 3.3: A case study of RadSee detecting the sentences written by a person behind a lab drywall.

Ground truth	**Letters recognized by RadSee**	**Segmented by** `Wordsegment` **[75]**	**Corrected by** `TextBlob` **[106]**
`football is popular in the united states'	`ecctbollispo pulaintheuni tedstate'	`ecc', `t', `boll', `is', `popula', `in', `the', `united', `state'	`etc', `t', `ball', `is', `popular', `in', `the', `united', `state'
`Bill is a hardworking student'	`Billiislhar dworkimg studena'	`bill', `i', `isl', `hard', `work', `img', `studena'	`bill', `is', `hard', `work', `ing', `student'
`My favourite fruit is apple'	`mgfavouri teffruitl qapple'	`mg', `favourite', `f', `fruit', `lq', `apple'	`my', `favourite', `fruit', `is', `apple'

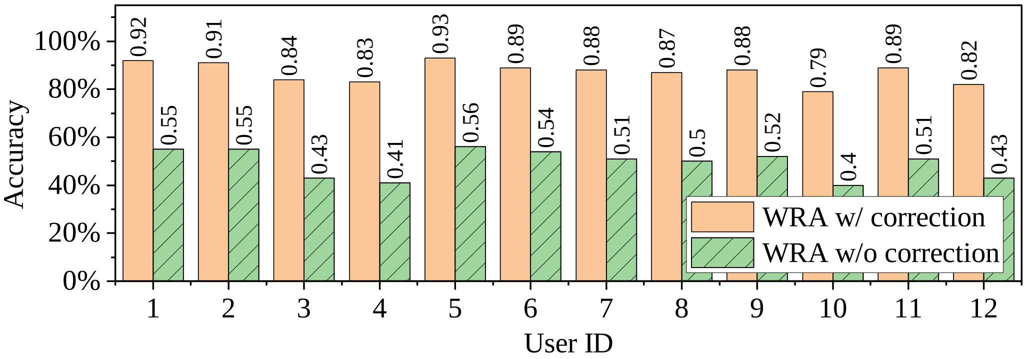


Figure 3.34: Word recognition accuracy of RadSee with and without correction for different users.

Fig. 3.33. The experimental setting is the same as described above.

RadSee employs two open-source software tools to translate its detected letters into word sentences: `Wordsegment` [75] and `TextBlob` [106]. It first sends the detected letters to `Wordsegment` for word segmentation. Then, it sends the segmented text to `TextBlob` for automatic spelling correction. Table 3.3 presents samples of the sentence recognition results. Leveraging these two open-source tools, RadSee demonstrates impressive performance in word and sentence recognition. It nearly recognized the first sentence in the table and accurately recovered both the second and third sentences.

We then use *word recognition accuracy* as the metric to evaluate the performance of RadSee. According to [112], word recognition accuracy is defined as $WRA = \frac{N-S-D-I}{N}$, where N is the number of words in the ground-truth text, S is the number of word substitutions, D is the number of word deletions, and I is the number of word insertions. Fig. 3.34 shows RadSee's WRA with and without using `TextBlob` for automatic spelling correction. It can be seen that without automatic spelling correction, RadSee's WRA ranges from 40% to 56% across the twelve participants. In contrast, when automatic spelling correction is applied, RadSee's WRA significantly improves, ranging from 79% to 93%. On average, RadSee's WRA hovers around 87% with automatic spelling correction. This level of word recognition accuracy is sufficient for an attacker to comprehend the content written by a victim.

3.9: Countermeasures and Other Applications

3.9.1: Countermeasures

Handwriting Safety Tips. RadSee demonstrated a serious threat to handwriting privacy. Based on the study, we have the following tips for those who have concerns about their handwriting information leakage. **Tip 1:** Do not write important documents in a room with drywall or vinyl wall. Instead, write them in a room with thick concrete or any metal walls. These walls can largely reduce the radio signal and thus reduce the probability of information leakage. **Tip 2:** Do not face yourself to a wall behind which a radar may be placed. Instead, face against that wall. Your body/torso will significantly reduce the radio signal strength and thus reduce the probability of your content being detected by an attacker. **Tip 3:** If possible, write important documents on a desk far from all walls rather than a desk against a wall. This will increase the distance between yourself and a radar, thereby reducing its recognition accuracy.

Protection Strategies. One natural approach to protecting handwriting content is to install multi-layer RF shielding materials inside the walls of your room [90]. Common materials used for RF shielding include metals such as aluminum, copper, and steel, as well as conductive coatings or paints. Another approach is to take advantage of recent advances in reconfigurable intelligent surface (RIS), which has also been studied under other names such as electromagnetic metasurface or radio relay. RIS can be used to create virtual multipath from radar's Tx to its Rx. By manipulating its phase shifting and beam steering, RIS is capable of generating fake phase patterns for the radar, preventing it from recovering the handwriting content. Unfortunately, neither of the above approaches is easy or economical to deploy.

3.9.2: Other Applications

While RadSee was designed to better understand the radio attacks related to handwriting privacy, it can also be used for many other applications. For instance, RadSee can be installed on a laptop as an input method. When an end user physically writes something on paper in front of his/her laptop, the content is automatically recognized by RadSee and digitally recorded on his/her laptop. In this case, RadSee does not need to use a 4×4 patch-array antennas since there is no need to penetrate through walls. Rather, a small patch antenna should be sufficient. RadSee can also be used as a human-computer interface for smart TVs. End users can write using their bare hands, and a TV equipped with RadSee can recognize the letters being written.

3.10: Summary

While mmWave FMCW radar has been extensively studied for autonomous driving and HAR, sub-10GHz FMCW radar has not received as much attention. This is of particular interest due to its *see-through-wall* capability, which may pose significant threats to the privacy of human activities. In this chapter, we presented RadSee, a 6 GHz FMCW radar system designed for detecting handwriting content behind walls. Through a combined hardware and software design, RadSee is capable of detecting mm-level handwriting movements and recognizing most letters based on their unique phase patterns. Additionally, it is resilient to the interference from other moving objects and coexisting radio sources. Extensive experimental results show that RadSee achieves 75% letter recognition accuracy when victims write 62 different letters and 87% word recognition accuracy when they write articles. In light of these realistic threats, we offered handwriting safety tips and defense strategies to help the public protect their handwriting information.

CHAPTER 4: EYE MOTION TRACKING USING FMCW RADAR

4.1: Introduction

Eye motion tracking has a wide range of applications across various fields. According to the Amyotrophic Lateral Sclerosis (ALS) Association, more than 5,000 people in the U.S. are diagnosed with ALS each year [26]. Individuals with ALS progressively lose control of their muscles, which affects their ability to move, speak, eat, and breathe [136]. Eye movements often become the only means of communication for individuals with ALS [28]. A non-intrusive, privacy-preserving eye-tracking system can help better interpret their intended messages and enable more efficient communication. In addition to assisting individuals with disabilities, eye motion tracking can serve as an effective human-computer interaction (HCI) tool in various scenarios. For example, it can help immobile patients communicate, as illustrated in Fig. 4.1. It can also be used to remotely control devices such as smart TVs, home appliances, elevators, and virtual reality systems. Furthermore, a reliable eye-tracking system has broader applications in healthcare, including psychology research [126], marketing analysis [196], and early disease detection [151].

Existing contactless eye tracking solutions employ various sensors, including cameras, acoustic, and radar. While cameras have demonstrated high accuracy in eye motion detection [53, 156], their application may pose privacy concerns in some scenarios. Additionally, cameras do not perform well in poor lighting conditions. Recently, acoustic signals have been studied for eye tracking on smartphones [38, 103]. However, due to the propagation nature of sounds, acoustic-based eye tracking systems are limited to eye blink detection within a small distance. Millimeter-wave (mmWave) radio frequency (RF) radar has also been studied for facial recognition and eye blink detection (e.g., [71,204]). While mmWave radar can achieve mm-level motion detection, its detection range is very limited due to its small wavelength. Additionally, existing mmWave-based sensing

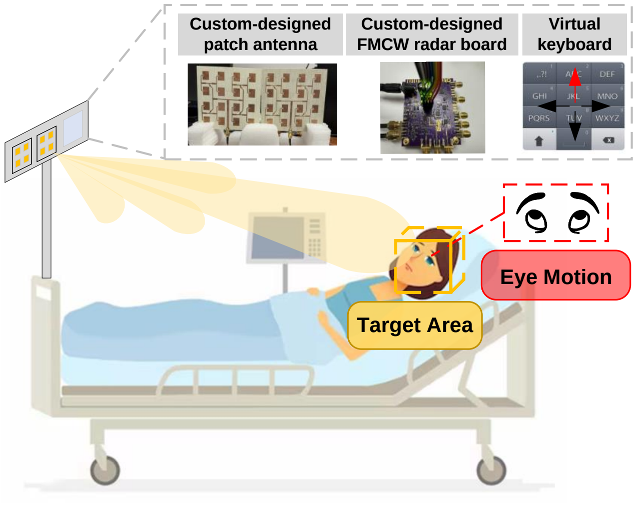

Custom-designed patch antenna
Custom-designed FMCW radar board
Virtual keyboard
Eye Motion
Target Area

Figure 4.1: Illustration of RadEye.

work focuses mainly on eye blink detection rather than eye motion tracking. Low-frequency radio signals have been widely leveraged for fine-grained human activity recognition (HAR), such as Wi-Fi sensing [52, 78, 93, 94, 163, 167], RFID sensing [162, 191], and 4G/5G sensing [47, 175]. However, due to their large wavelength as well as their non-coherent sensing approaches, those systems may not be able to detect such subtle eye motions. So far, there is no RF-based system that can track human eye motions from a distance.

In this chapter, we present RadEye, an RF sensing system that can track eye motions from a distance. Compared to camera-based eye tracking, RadEye not only mitigates privacy concerns but also performs reliably in poor lighting conditions. The privacy-preserving nature of RF signals stems from their inherent characteristics. Unlike camera images, RF signals are not visually interpretable by humans and inherently possess low spatial resolution. As a result, they are unlikely to reveal detailed, identifiable features of individuals. Extracting personal information from RF signals involves complex processes that require advanced signal processing and AI models, making misuse significantly less accessible compared to camera systems. Furthermore, RF sensing systems are typically designed to capture only coarse-grained human activities—such as presence, movement, or positioning—rather than detailed personal characteristics like facial features or voice. This makes RF-based sensing systems inherently more privacy-preserving than cameras, ensuring better protection of individual privacy.

Table 4.1: Comparison of RadEye and existing eye detection works.

Reference	Technique	Max. distance	Track eye motion	Work in low light
Blink Listener [103]	Acoustic	0.8 m	✗	✓
TwinkleTwinkle [38]	Acoustic	0.6 m	✗	✓
BlinkRadar [71]	IR-UWB	0.8 m	✗	✓
X. Zhang [207]	mmWave	1.2 m	✗	✓
C. Ryan [134]	Event Camera	0.6 m	✓	✗
GazeRecorder [53]	Web Camera	0.7 m	✓	✗
RadEye	Sub-6GHz FMCW Radar	5 m	✓	✓

Compared to acoustic- and mmWave-based eye detection approaches, RadEye extends both the RF sensing capability (from eye blink to eye motion) and detection range (from less than 1 m to more than 5 m). We note that the eye motion tracking task is very different from eye blink detection. The former is a regression problem, while the latter is a binary classification problem. Such an extension will significantly enlarge RadEye's application landscape in real life.

In the design of RadEye, we face two challenges. **Challenge #1: subtle eye movement and long detection range.** On one hand, eye rotation, encompassing eyelid and eye muscle displacement, involves movement around 1 mm [91]. Such a subtle motion makes it hard to detect for an RF system. On the other hand, an eye-tracking sensor may be used in indoor or outdoor scenarios. The eye detection distance varies significantly, ranging from 0.5 m (e.g., from smartphone to eyes) to 5 m (e.g., from smart TV to eyes). Devising an RF system that can detect subtle (mm-level) eyeball rotation movement from a distance is not a trivial task. **Challenge #2: interference mitigation by design.** An eye-tracking system may suffer from interference from three sources: i) multipath from the target person to the RF sensor, ii) the movement of the target person's other body parts such as chest breathing, arm waving, and leg shaking, and iii) other moving objects/people in the area. For instance, when the RF sensor detects eye movement, another person may walk around in the same room, generating interference to the received signals at the RF sensor. In general, interference is a notorious problem for RF sensing. Given the subtlety of eyeball movement, the interference must be mitigated by design so as to accurately detect eye's movement.

RadEye addresses the above two challenges through a joint hardware and software design. To achieve the required detection resolution and range (i.e., Challenge #1), we design and optimize a 5 GHz FMCW radar for eye movement tracking. We choose 5 GHz radar for two reasons: i) high-frequency radio wave (e.g., mmWave) is suited for detecting tiny motions, but has a small detection range; and ii) low-frequency radio wave is suited for long detection, but not suited for detecting subtle movement. A tradeoff between detection resolution and range leads to our selection of 5 GHz frequency band. Additionally, the market has rich electronic devices (e.g., power amplifiers, mixers, and low noise amplifiers) at 5 GHz frequency bands due to the maturity of Wi-Fi industry. Thus, it is cost-friendly to build 5 GHz radars. To mitigate interference from multipath and other moving objects (i.e., Challenge #2), we combine four techniques: i) FMCW modulation for the 5 GHz radar, ii) a sophisticated signal processing pipeline for eye-related feature extraction, iii) a transformer-based deep neural network (DNN) for eye motion detection, and iv) a camera-guided supervisory training method for the DNN model. Together, these four techniques make RadEye capable of separating the eye motion features from the interference from multipath and other objects. More importantly, these four techniques make RadEye transferable to unseen scenarios, enhancing its generalizability in practice.

We have built a prototype of RadEye and evaluated its performance in multiple scenarios. Experimental results show that, for a person at a 5 m distance, the average estimation error of eye rotation is 24 degrees in azimuth and 21 degrees in elevation. By formulating the eye rotation problem to a classification (up, down, left, and right) problem, RadEye achieves 90% accuracy. Extensive results confirm the generalizability of RadEye in unseen scenarios as well as its resilience to interference.

Table 4.1 shows how RadEye advances the state-of-the-art (SOTA) RF sensing technology. The main contributions of RadEye are summarized as follows:

- To the best of our knowledge, RadEye is the first-of-its-kind system that utilizes RF signals to estimate eye rotation angles from a distance.
- RadEye presents a joint hardware and software scheme for subtle eye motion detection in

the presence of interference.

- Extensive experimental results validate the performance, robustness, and generalizability of RadEye.

4.2: Related Work

4.2.1: Eye Motion Recognition

Acoustic-based Detection. As speakers and microphones are now commonplace on mobile devices, acoustic signals have become widely utilized for recognizing human daily activities. BlinkListener [103] can detect eye blink motions using acoustic signals, modeling variations caused by eye blinks and interference. By leveraging interference, they identify an optimal position to maximize the variation induced by eye blinks. TwinkleTwinkle [38] addresses a similar objective using a different approach. They employ a phase difference-based method to detect potential blink motions, followed by a model-based approach to distinguish subtle motions. Additionally, they establish a language input system based on ASCII code and Morse code. While RadEye is capable of recognizing eye motions, the mentioned works focus solely on detecting eye blinks at limited distances.

RF-based Detection. In the study by Zhang et al. [207], an off-the-shelf mmWave FMCW radar is employed to detect eye blinks. They introduce an Adaptive Variational Mode Decomposition (AVMD) algorithm to extract the blink signal, achieving an effective detection distance of up to 1.2 meters. Several other studies [34, 107, 170] have taken a similar approach with mmWave FMCW radar. In addition to the mmWave signal, BlinkRadar [71] employs UWB radar to detect eye blinks in a driving scenario. They implemented a customized impulse-radio ultra-wideband (IR-UWB) radar. By analyzing signal features in the complex domain, the system can isolate eye blinks without interference from other motions.

Camera-based Detection. Due to the ubiquity of web cameras and smartphone cameras, significant progress has been made in using cameras to track eye motion. In the computer vision research domain, deep learning networks have been effectively employed to predict gaze direction [87, 146, 208, 216]. Furthermore, in the security domain, studies have shown that the camera on a mobile phone can even track the gaze trace, raising concerns about potential password leakage [37, 168]. Additionally, there are commercial eye-trackers available on the market [53, 156] that support high-accuracy eye-tracking at an affordable price. However, it's worth noting that all camera-based eye-tracking solutions may raise privacy concerns and may not function effectively in low-light scenarios.

Wearable-based Detection. With the prevalence of VR devices, smart glasses offer another solution for eye motion detection. Google Glass [74] and Jins MEME pair of eyeglasses [44] demonstrated the potential to detect eye motions several years ago. Building on existing approaches, Liu et al. [104] attached a copper electrode to the glass frame to sense eye blinks by utilizing the capacitance variation between the electrode and eyelid. However, wearable devices like these require users to keep the glasses on their heads, which may be inconvenient for daily use.

4.2.2: Fine-grained HAR

MmWave-based Recognition. MmWave FMCW radar has reached great performance these years due to its fine-grained detection ability and affordable cost. They are extensively used in human activity recognition and vital sign detection [61, 98, 181, 188, 205, 206]. Thanks to their substantial bandwidth and diminutive wavelength, they attain millimeter-level accuracy in detecting object movements.

Wi-Fi-based Recognition. Channel State Information (CSI) in Wi-Fi networks has been applied across various sensing applications, including gesture recognition [52, 93, 200], vital sign detection [167], and radio imaging [78, 94, 142, 163]. Nevertheless, Wi-Fi, characterized as a non-coherent system due to the physical separation of its transmitter and receiver, faces limitations in detection accuracy stemming from timing, frequency, and phase misalignments.

4.3: RadEye: Design Analysis

4.3.1: Background

RadEye leverages the FMCW signal to detect the eye motions. The signal is transmitted from the radar's TX antenna towards the eyes; and the reflective signal from the eyes is received by the radar's RX antenna. The difference between the transmitted and received signal is used to extract the eye motion features. As shown in Fig. 4.2, the FMCW signal starts from frequency f_0 and ramps up linearly over time T. The transmitted signal can be written as:

$$S_T(t) = e^{-j2\pi(f_0 t + \frac{B}{2T}t^2)}. \tag{4.1}$$

The received signal reflected from the target can be written as:

$$S_R(t) = \alpha e^{-j2\pi(f_0(t-\tau) + \frac{B}{2T}(t-\tau)^2)}. \tag{4.2}$$

The transmitted and received signals are mixed together, leading to an immediate frequency (IF) signal as follows:

$$S_M(t) = S_T(t)S_R(t)^* = \alpha e^{-j2\pi(f_0\tau + \frac{B}{T}\tau t - \frac{B}{2T}\tau^2)}. \tag{4.3}$$

RadEye uses the IF signal to infer the eye motions. As we can see from the Eqn. 4.3, both the frequency and phase of the IF signal are proportional to the delay of the signal. The frequency of the IF signal $f_m = \frac{B}{T}\tau$. The time delay can be calculated as $\tau = \frac{f_m T}{B}$ and, as a result, the distance can be calculated by $d = \frac{c\tau}{2} = \frac{c f_m T}{2B}$.

To separate the signal reflected from different objects, we do range-FFT on each chirp of the IF signal, as illustrated in Fig. 4.2. Each range bin represents the signal coming from different distances. The range resolution $\Delta d = \frac{c}{2B}$ is determined only by the bandwidth of the signal. To identify the FFT bin corresponding to eye motions, a user will be asked to blink his/her eye as a

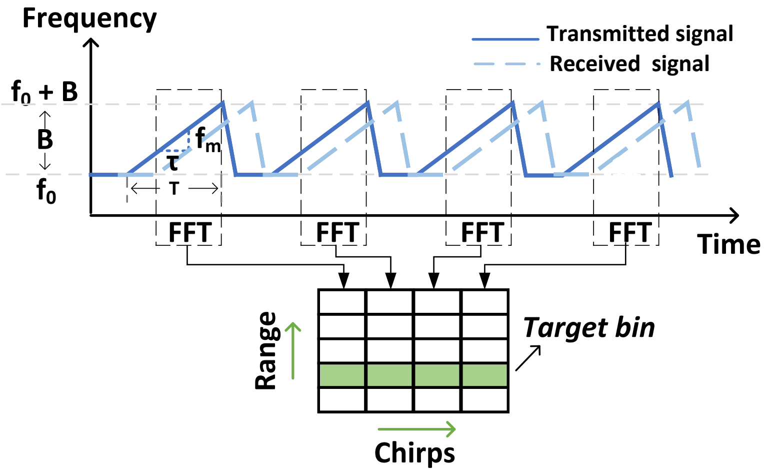


Figure 4.2: Illustration of FMCW signal.

reference. The algorithm will be presented in §4.4.

4.3.2: Detectability of Human Eye Rotation

The kinematics of eye rotation is a complex process involving the stretching and contraction of six extraocular muscles. The combined movement of these muscles alters the shape of the reflection surface, affecting the length of the signal reflection path and influencing the phase shift of the FMCW signal. Additionally, these muscle movements impact signal attenuation, as muscles and surrounding tissues absorb and scatter FMCW signals to varying degrees based on their density, composition, and position. When eye muscles move, the spatial distribution of these tissues changes, altering the amount of signal absorbed or scattered and leading to variations in attenuation.

In addition to muscle movements, eyelid motion also affects reflected FMCW signals. During eye rotations, the eyelids fold or stretch, and this change in thickness modifies the distance of the signal reflection path. Furthermore, eyelid movement alters the size of the exposed area of the eyeball, further influencing the attenuation of the reflected signal.

4.3.3: Feasibility Analysis

We conducted experiments to compare the performance of RadEye with a 60 GHz FMCW mmWave radar (i.e., AWR6843 [155]), both of which have 1.1 GHz bandwidth. Specifically, a participant performed eye blinks at distances of 3 m, 4 m, and 5 m. Fig. 4.3 presents our experimental mea-

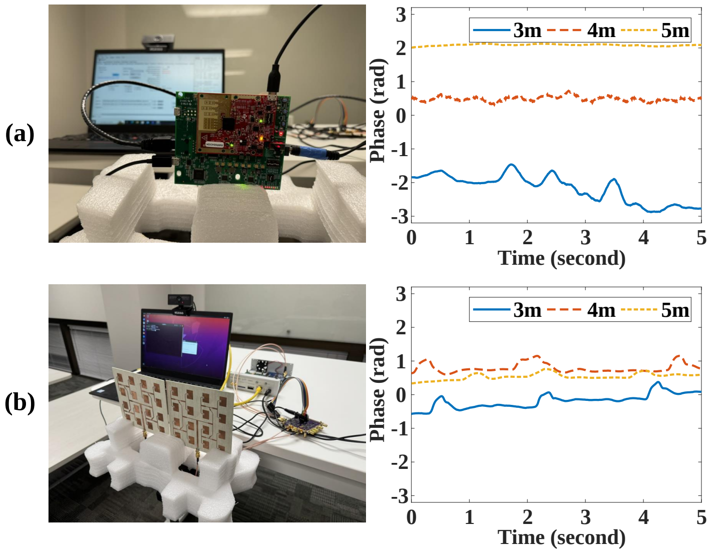


Figure 4.3: (a) The phase change of the corresponding FFT-bin from the mmWave radar. (b) The phase change of the corresponding FFT-bin from the RadEye.

surements. The experimental results show that mmWave radar can detect human eye blinks within a range of 3 meters. However, the detectability decreases rapidly as the distance increases. In contrast, RadEye exhibits a consistent capability of detecting human eye blinks at those three distances.

In some cases, the line-of-sight path from human eyes to the radar device might be blocked. Thus, we conducted comparative tests to evaluate the ability of two types of radars to track eye movements under obstructed conditions. To simulate such cases, we repeated the same test at a distance of 3 m but placed a wooden door between the radar and the participant. As shown in Fig. 4.4, RadEye is capable of detecting eye blinks even behind the door, whereas the mmWave radar fails to do so. This limitation of the mmWave radar can likely be attributed to the high attenuation of mmWave signals. It is worth noting that these experiments were conducted in the same environment and used an identical signal processing pipeline. The detailed parameters of the two systems are provided in Table 4.2.

Millimeter-Level Motion Detection. As we mentioned earlier, the ranging resolution of an FMCW radar is not enough to detect the tiny eye motion. However, the phase of the demodulated FMCW signal can reflect the eye rotation motion. Eye rotation involves eyelid and eye muscles displacement, which moves at millimeter level [91]. Based on Eqn. (4.3), one-millimeter movement

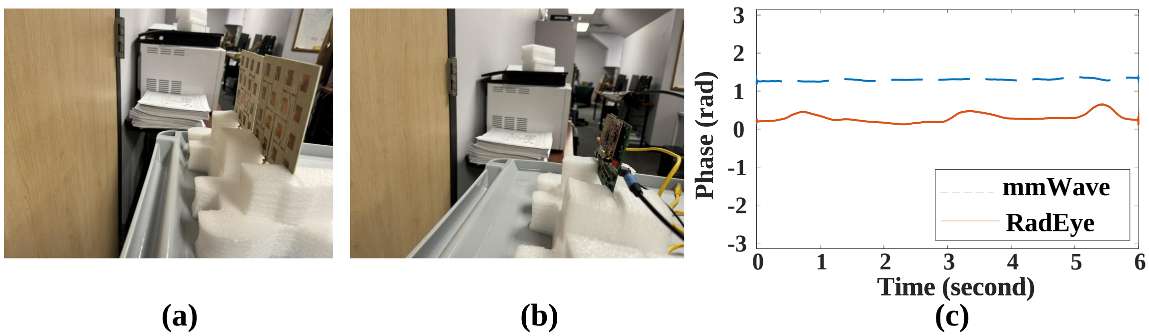


Figure 4.4: (a) RadEye tracking behind a wooden door. (b) mmWave radar tracking behind a wooden door. (c) Phase change of the corresponding FFT-bin from both systems.

Table 4.2: Detailed Parameters for the RadEye and mmWave Radar.

Systems Parameters	RadEye	AWR6843
Tx / Rx antenna gain	15 dBi	7 dBi
Transmission power	15 dBm	12 dBm
Chirp duration	600 μs	600 μs
Idle time	400 μs	400 μs
Bandwidth	1.1 GHz	1.1 GHz
Gain figure (Rx chain)	–	48 dB
Noise figure (Rx chain)	7 dB	12 dB
Gain from baseband amplifier	8 dB	–

of eyeball can cause a phase change of FMCW signal by: $2\pi f_0 \frac{2d}{c} = 0.25$ radian (i.e., 14°), which is easy to detect and measure on the corresponding Range-FFT bin.

Resilience to Interference. An eye-tracking system may suffer from interference from three sources: i) multipath from the target person to the RF sensor, ii) the movement of the target person's other body parts such as chest breathing, arm waving, and leg shaking, and iii) other moving objects/people in the area. To mitigate such interference, RadEye employs wideband FMCW modulation and high-optimized directional antennas. The FMCW modulation with 1.1 GHz offers a ranging resolution of 14 cm, allowing RadEye to distinguish objects separated by 14 cm. The FMCW modulation can effectively filter out the interference from the target person's other body motions such as chest breathing. A custom-designed patch antenna is used for signal transmission and reception, serving as an angular filter for suppressing interference from other directions. As shown in Fig. 4.5, the patch antenna has a 3 dB beamwidth of 21 degrees. In addition to the hard-

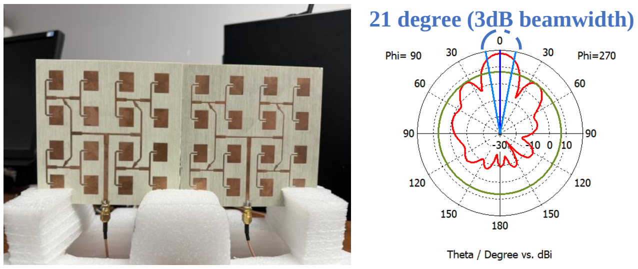


Figure 4.5: The custom-designed patch antennas (left) and their gain pattern (right).

ware design and optimization, a transformer-based DNN, trained through a video-guided pipeline, will be useful to focus on the desired features while eliminating the interfering features through a self-attention mechanism.

4.3.4: Feature Validation

We conducted a preliminary study of the sub-6GHz radar's capability for detecting eye motions. A participant was seated 3 m in front of the radar and instructed to rotate his eyeballs in four different directions (up, down, left, and right). Fig. 4.6(a) depicts the signal amplitude from the corresponding Range-FFT bin during these eye rotations. The "Ground truth" points in the figure were captured by a camera with the SOTA computer vision-based eye detection algorithm [89], marking the moment of real eye rotations. It is evident that the eyeball rotations towards the four directions indeed induce the amplitude change of the radar's IF signal. This indicates that the 5 GHz FMCW radar is capable of capturing the eye motions. When delving into eye rotation signals in different directions, we observe that up/down movements exhibit a more substantial change compared to left/right movements. This is not surprising, because the eyelid of vertical actions has larger movements. Additionally, the up-rolling of eyes results in an amplitude increase. This is because the eyelid's movement during the upward gaze involves more parts of the eyeball in reflection. The water-textured nature of the eyeball, in contrast to the skin, enhances signal reflection. Given that the different eye rotations cause different amplitude changes, we use the amplitude variation ratio as one of the features for inference.

Features in Complex Domain. In addition to the observation in the temporal domain, we further observe the IF signal in the complex domain as shown in Fig. 4.6(b). For the eye motion

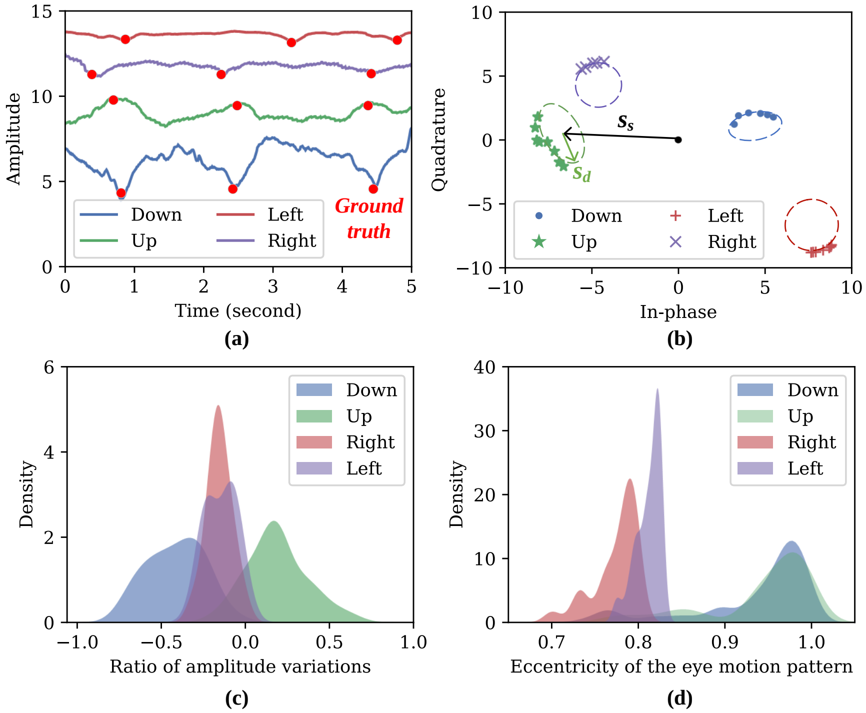


Figure 4.6: The feature for eye rotations. (a) The signal amplitude when eye rotates toward different directions. (b) The signal in the complex domain when eye is rotating. (c) The kernel density estimation of the signal amplitude variation ratio. (d) The kernel density estimation of the motion pattern eccentricity.

signal S_e from the corresponding Range-FFT Bin, it can be decomposed into a static component S_s and a dynamic component S_d. They can be written as:

$$S_e = S_d + S_s = \alpha_d e^{\phi_d} + \alpha_s e^{\phi_s}, \tag{4.4}$$

where α_s and ϕ_s are the amplitude and phase of the static component. α_d and ϕ_d are the amplitude and phase of the dynamic component. As shown in Fig. 4.6(b), the curve shape of the dynamic component in the complex domain is determined by both amplitude and phase.

Different Eye Rotation Directions. For different directional movements of the eyeball, the folding of the eyelid and the rotation of the eyeball create a unique relative relationship, manifesting through the changes in IF signal amplitude and phase. We characterize this feature by utilizing the curvature of the curve. Specifically, we perform regression on the curve, identify an elliptical equation [49], and use the eccentricity of the ellipse to characterize this feature. As shown in

Fig. 4.6(b), the ellipse generated by the up/down eye motions exhibits a more elongated shape, while the right/left motions lead to a more circular shape. This can be attributed to the fact that up/down eye motions involve more eyelid movements, leading to changes in reflective surface, while left/right eye rotations mainly cause changes in the length of the reflection path.

Experimental Validation. To verify the robustness of the amplitude and eccentricity features, we conduct experiments involving five participants. We repeat the experiments in the same setting as described above. Each participant performs his/her eye rotations in each direction 50 times. Then, we perform the kernel density estimation for all the participants. Fig. 4.6(c) presents the density estimation results of amplitude changing ratio. The up and down motions are centered on the 0.3 and -0.4, respectively. The left and right motions have a relatively smaller variation; and they are centered around zero. Fig. 4.6(d) shows the eccentricity density estimation results. The up/down motions are centered close to 1; and the left/right motions are close to 0.8. The statistical results across different individuals align with the findings described earlier for a single person. This consistency suggests the potential of segregating these motions based on the radar's signal. To enhance the recognition of eye motions, we employ a DNN model, which will be described in §4.5.

4.4: RadEye: Signal Processing

In this section, we describe the signal processing of the radar's IF signal for eye rotation detection. Fig. 4.7 shows the overall structure of the system. In what follows, we introduce the signal processing techniques for RadEye, which include the selection of the range bin and the extraction of eye motions.

Range-FFT. RadEye sets the chirp duration to 1 ms for detecting eye motions. In each cycle, the chirp takes 0.6 ms, and the delay takes 0.4 ms. RadEye employs a sample rate of 2.5 MSps to observe the signal in the digital domain. Consequently, in each cycle, 1500 complex numbers are acquired. Subsequently, RadEye appends zeros to the end of the samples and performs a 4096-point range-FFT operation to obtain the signal at different ranges. In the 4096 bins, only the first

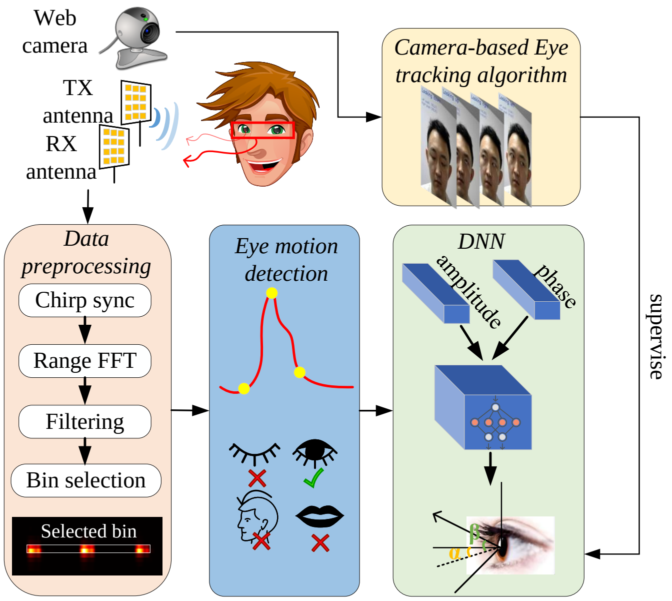


Figure 4.7: The system overview of RadEye.

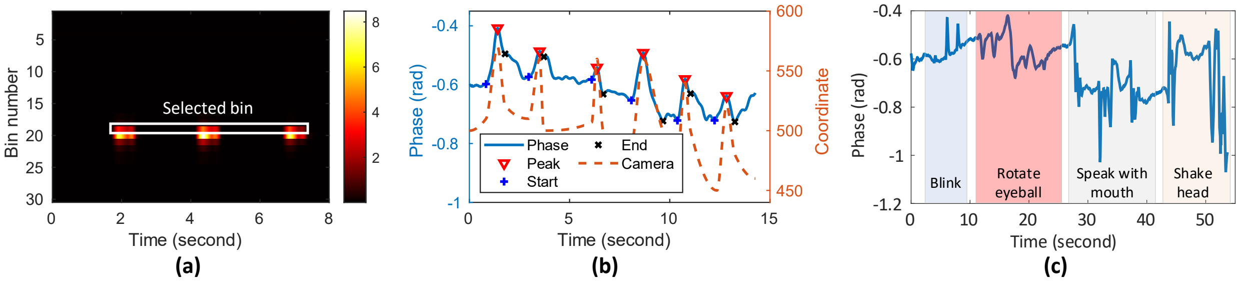


Figure 4.8: (a) The signal amplitude variance for different Range-FFT bins during eye blinks. (b) Eye motion detected based on signal phase (with the camera-based ground truth marked). (c) Comparison of Eye motions and interfering motions from the target person's head.

256 will be used since it already covers the range to 13 m.

Filtering for Range-FFT. RadEye applies a second-order Butterworth bandpass filter to suppress noise and out-of-band interference. It sets the filter's pass band to 1 Hz∼5 Hz [204, 205]. We note that, although the eye blink frequency may overlap with the chest breathing frequency (0.1 Hz∼0.5 Hz), the input eye motion command for RadEye has a higher frequency and will not be affected by the filter. One may ask whether the heartbeat will cause the slight head shaking and thus pollute the eye motion signals, our answer is no. Based on our experimental results, the heartbeat is too weak to cause the head shaking that can be captured by RadEye.

FFT Bin Selection. Eye rotation motions have a very small dynamic range. Therefore, it is nontrivial to find the FFT bin that carries the eye motion features. To do so, RadEye requires users to blink their eyes three times with an interval of 2 seconds as a 'start button' to initiate the control process. The user only needs to provide the initialization command once. In this period, the users should keep their head still (no movement more than 14 cm). If the user's head position moves beyond this range, the initialization command must be re-entered. Since this is a human input device, it is reasonable to require the user to remain relatively stationary within a short time period. RadEye utilizes the amplitude dynamic range as an indicator to find the candidate bin. The reason for using the amplitude is that, when the eyes switch between opening and closing, the blink motion causes the reflective surface to switch from the water-textured eyeball to the skin-textured eyelid [103], causing the amplitude change of the radar's IF signal.

We describe the bin selection algorithm as follows. RadEye calculates the window-slides variance for each Range-FFT bin i by processing the signal as: $v_i(j) = \frac{1}{W}\sum_{m=j}^{j+W-1}(|y_i(m)| - \bar{y}_i)^2$, where $\bar{y}_i = \frac{1}{W}\sum_{m=j}^{j+W-1}|y_i(m)|$, and w is the window size which is set to 200 to fit the duration of the eye blink. If $v_i(j)$ is larger than a predefined threshold T, the timestamp j will be recorded as t_n. Here, T is empirically set to 0.05. Only when the next detected timestamp j satisfies the $2000 < j - t_n < 3000$ (fit in the interval of blink), it will be counted as the continuous blink. Once the three continuous blinks have been detected, we mark the Range-FFT bin i as the candidate bin. Multiple Range-FFT bins might satisfy this condition as shown in Fig. 4.8(a). In this case, RadEye chooses the bin that has the smallest index. The smallest-index Range-FFT bin represents the shortest path of signal travel, which best reflects the eye motion pattern.

Eye Motion Detection. After RadEye identifies the Range-FFT bin, it will continue to monitor this bin and estimate eye motions based on it. Each eye motion can be separated into three phases: start moving the eyeball, the eyeball reaches the edge, and the eyeball backs to the start position. RadEye tries to detect these three positions for each eye motion. Although both amplitude and phase contain information about eye motions, we found that phase exhibits a more significant pattern when detecting eye rotations.

Eyeball rotations involve repetitive movement. It rolls back to its central point when reaching the edge. This motion occurs swiftly, resulting in a repetitive phase change pattern. Hence, the positions of local phase extremums correspond to where eye movement reaches the edge. Additionally, we noticed that there are inflection points in the signal phase at the beginning and end of eye movements. RadEye utilizes these features to extract eye motions, it first searches for the local maximum/minimum on phase with an interval of 1 second. After identifying the local peaks, RadEye will search along the gradients of samples before/after the peak. The position in which the gradient is equal to zero will be defined as the start/end position. Fig. 4.8(b) presents the phase of the signal when a participant repeats the look-up motion, and the detected start, peak, and end positions are marked on the figure.

To mitigate interference from the target person's other body parts, we utilize both phase shift and time duration to refine the detection results. Fig. 4.8(c) shows eye blinks, eye rotations, head motions, and mouth motions. It can be seen that head and mouth motions induce significant changes in the signal phase. Consequently, if the phase shift surpasses a specified threshold, the signal is discarded. For motions falling below the threshold, the time duration is considered. Only signals within the duration range of 200 ms$\sim$600 ms are deemed as valid eye motion signals. Doing so will effectively filter out eye blinks, which typically last for less than 100 ms.

4.5: RadEye: DNN-based Eye Movement Detection

In this section, we present a DNN model for eye rotation recognition by using the amplitude and phase of the radar's IF signal. RadEye utilizes a transformer encoder to extract features and feeds these features into a fully connected layer to output the azimuth and elevation angles of a target person's eyeball. The DNN is trained using a camera-guided method, transferring the knowledge from computer vision to radio sensing.

4.5.1: Sequential Signal

The input signal to our DNN model is a time-series signal with a high sampling rate. As the subject's eyeballs rotate toward different angles, the swift motions of the eyeballs and eyelids cause fluctuations in the amplitude and phase of the input signal over time. Therefore, to accurately

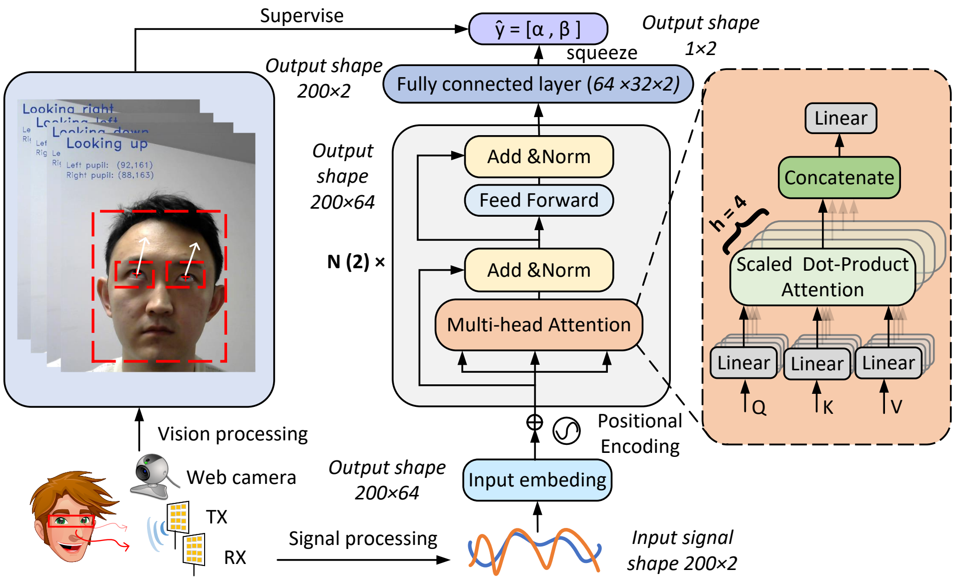


Figure 4.9: A camera-guided DNN structure for RadEye.

model the temporal dependencies between various sampling points, we employ a DNN model that is capable of efficiently encoding information over the temporal domain.

Traditional time-series models, such as Recurrent Neural Network (RNN) [133] and LSTM [69], are capable of temporal sequence modeling. However, they struggle with gradient vanishing and exploding problems when dealing with long sequence inputs, limiting their capabilities of capturing dependencies over a long distance. In contrast, Transformer [159], employing a self-attention mechanism, can effectively overcome these issues. The self-attention mechanism allocates a weight to the output of each position in a time series, reflecting the degree of attention that the position pays to other positions within the sequence. This method allows for the computation of correlations between any two positions in the sequence without being constrained by their physical separation, thus better capturing long-range dependencies. Furthermore, the multi-head attention mechanism within Transformers can project eye movement signals into various subspaces, including different frequency spaces. Frequency analysis can better distinguish certain angle information, as the movement of the eyeball at different angles will have different speeds.

4.5.2: Camera-guided DNN Framework

The overall structure of our designed model, as shown in Fig. 4.9, primarily utilizes Transformer to predict an angle vector based on the input time-series data, which includes both the amplitude and phase derived from the corresponding Range-FFT bin. We introduce its key elements as follows.

- **Input Data.** The input data for the eye motion signal is captured by RadEye from the corresponding Range-FFT bin. Each input data sample consists of 200×2 dimensions, where 200 is the time length of the data sample, and 2 is the number of features: amplitude and phase. The data have been normalized to ensure that they share the same dynamic range.
- **Signal Encoder.** Before sending to Transformer, the signal data are first passed through a projection layer to up-sample them to a higher-dimensional representation. The output size of this projection layer is 200×64. Additionally, to enable the model to discern the temporal relationships within the input sequence embeddings, each signal representation is augmented with positional embeddings.
- **Backbone.** The transformer encoder can extract features from the embedded data. RadEye uses two transformer encoders; and each encoder has four heads. The self-attention mechanism in the transformer encoder can build connections across different time steps in the signal and also attend to different parts of the signal. These connections enable the model to easily derive information about the eye rotation angle.
- **Prediction Head.** The extracted features finally feed into the fully connected layer, which has a size of $64 \times 32 \times 2$. It combines features from previous layers and flattens the output into the appropriate shape. The output of the model is a direction vector $y = [\alpha, \beta]$, where α is the eyeball's azimuth angle and the β is the eyeball's elevation angle.
- **Camera-guided Training.** The vision processing module initially tracks the user's face and subsequently localizes the position of the eyes. It then calculates the eye rotation angle based on the relative position of the pupil within the eye region. Guided by these vision-based techniques, the DNN model endeavors to create a feature extractor similar to those used in vision processing, but specifically designed to handle RF signals. The vision processing module

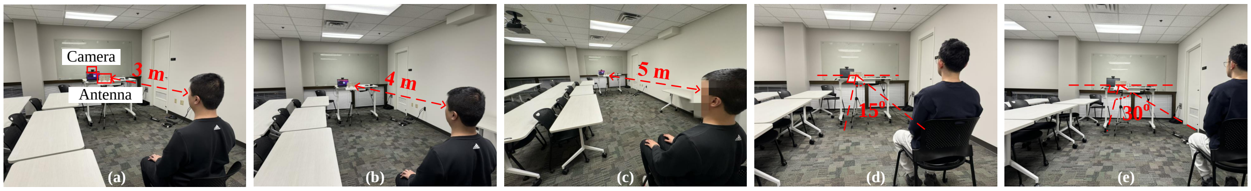


Figure 4.10: Experiment settings in a lab for different distances and angles. (a) 3 meters. (b) 4 meters (c) 5 meters. (d) $15°$. (e) $30°$.

ultimately provides the ground truth angle to the DNN, which then utilizes Mean Squared Error (MSE) to calculate the loss for angle estimation. The loss function is as follows:

$$\mathcal{L}_{angle} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{4.5}$$

where y_i is the predicted angle vector and the $\hat{y}_i$ is the ground truth angle vector.

4.5.3: Data Collection

We gathered training data exclusively in a controlled laboratory setting, where participants engaged with the radar and camera setup positioned on a table before them. Participants were seated 3 m away from the radar, facing it directly as shown in Fig. 4.10(a). A total of 8 participants took part in the data collection. In each data collection session, participants were instructed to rotate their eyes in up, down, right, and left directions. Recognizing the potential fatigue associated with eye rotation, each test session had a limited duration of 3 minutes. Each participant will repeat 15 sessions, generating a total of 27,000 data samples.

We note that during the test phase, the eye motion signals are directly captured using RadEye (no camera presents). In this scenario, the length of the signal vector may differ from the training data. To ensure consistent input dimensions, downsampling or interpolation techniques are employed to normalize the data dimension.

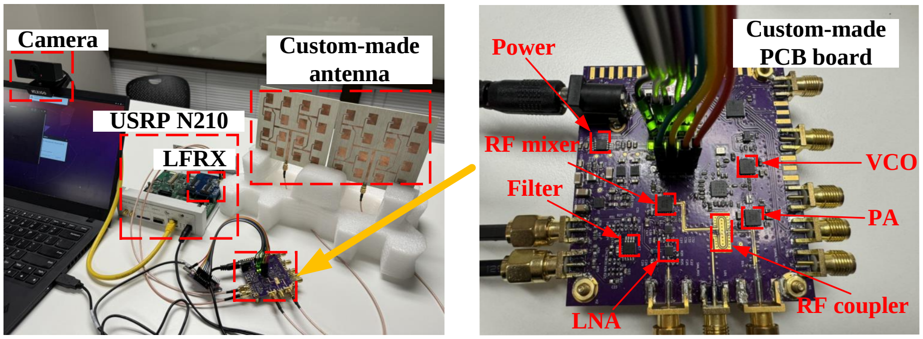


Figure 4.11: The system setting for RadEye.

4.6: Experimental Evaluation

In this section, we conduct experiments to evaluate the performance of RadEye. Particularly, we aim to answer the following questions.

- **Q1 (§4.6.4)**: What is RadEye's detection rate of eye motions?
- **Q2 (§4.6.5)**: What is RadEye's accuracy in estimating eye rotation angles?
- **Q3 (§4.6.6)**: What is RadEye's resilience to environmental changes and interference?
- **Q4 (§4.6.7)**: What is RadEye's zero-shot performance (for unseen users and in unseen scenarios)?

4.6.1: Implementation

Hardware. Fig. 4.11 shows the hardware of RadEye. We have a fabricated a PCB board capable of transmitting and receiving FMCW signals at 5 GHz, using Wi-Fi's electronic components. The received signal undergoes amplification with a power amplifier (PA), and then it is mixed with the transmitted signal. The electronic components of this board also include Tx/Rx 16 dB coupling, RF I/Q mixer, and baseband filtering. Additionally, we have custom-designed and optimized a 4×4 patch antenna using HFSS for signal transmission and reception. A single patch antenna provides an 15 dBi gain, resulting in a total gain of 30 dBi. The patch antenna design maintains the signal beam within a narrow range while providing significant gain, enabling RadEye to detect eye motions from a distance. The mixed signal is subsequently fed into a USRP N210 with an

LFRX daughterboard to convert the analog signal into baseband I/Q samples. The FMCW signal generated by RadEye sweeps from 5.4 GHz to 6.5 GHz. Each chirp has a time duration of 1 ms, with 600 μs for frequency ramping and 400 μs idle.

Software. We implemented our data preprocessing module in C++ using the GNURadio out-of-tree module. A crucial function of this module is to synchronize the chirps, facilitating the extraction of useful samples. However, owing to the absence of clock synchronization between the USRP ADC and FMCW chirps, only the software-based method can be employed for synchronization. To address this issue, we utilize a high-sampling rate of 2.5 MSps and the idle period for synchronization. Initially, we detect the idle period based on the smooth amplitude during this interval, followed by fine-grained detection to identify the first sample of the chirp. The DNN model was implemented using PyTorch with the Adam optimizer. Throughout the training process, a batch size of 200 and 50 epochs were set.

4.6.2: Experimental Setting

During the experiments, participants were seated on a chair facing the antennas of RadEye. The antennas were positioned 1.1 m above the ground on a tripod. A varifocal camera was placed on top of the laptop to capture the participants' face video for their eye motion detection using the SOTA gaze tracking tool [89]. Our experimental studies show that this camera-based eye detection tool achieves about 98% accuracy as shown in Fig. 4.13. While it is not perfect, we use the detection results from the camera-based tool as the ground-truth labels to supervise the training of RadEye's DNN model. During the inference, we also use the camera-based detection results as the ground-truth to evaluate the estimation accuracy of RadEye.

RadEye and camera operated concurrently, synchronized with the PC clock, to estimate participants' eye rotations. RadEye, with a higher sample rate than the camera's frame rate (10 frames/s), resulted in each camera-captured direction being mapped to 200 continuous chirps. Each training sample in our dataset is a 200×2 matrix, where 200 is the time dimension, and 2 is the feature dimensions (azimuth and elevation).

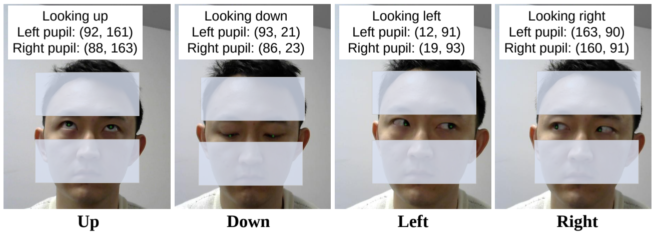


Figure 4.12: The eye rotation directions captured by a camera. It is used as ground truth for DNN training and evaluation.

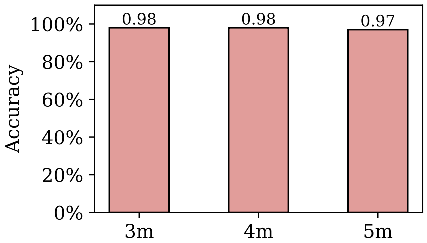


Figure 4.13: Tracking accuracy of eye rotation with camera at varying distances.

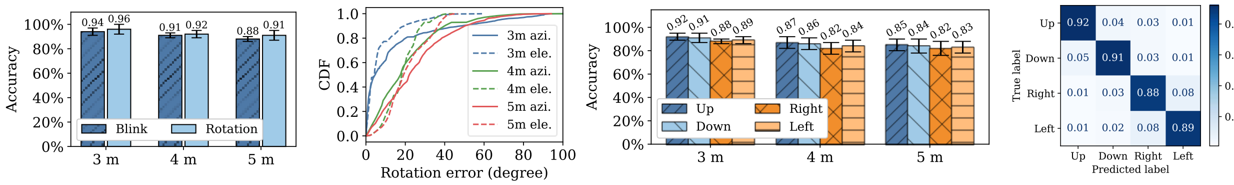
4.6.3: Performance Metrics

We consider the following three performance metrics.

- **Eye Motion Detection Rate (EMR).** The eye motion here is defined as the eye blink and eye rotation. The signal processing module extracts the eye motion signals from RF data before sending them to the DNN model. We define $\textit{Detection rate} = \frac{\textit{Number of eye motions detected}}{\textit{Total eye motions performed}}$.
- **Estimation Error of Eye Rotation Angle (ERA).** The eye rotation angles are illustrated in the bottom right corner of Fig. 4.7. The estimation error of azimuth angle is defined as $e_\alpha = |\alpha - \hat{\alpha}|$, where $\hat{\alpha}$ is the estimated eye rotation azimuth angle and α is the eye rotation angle ground truth provided by the camera. Similarly, the estimation error of elevation angle is defined as: $e_\beta = |\beta - \hat{\beta}|$.
- **Estimation Accuracy of Eye Rotation Direction (ERD).** Some applications of RadEye (e.g., remote TV control) may not require precise angle measurements for functionality but instead focus on eye rotation direction. This metric evaluates the accuracy of classifying eye movement directions into four categories: up, down, left, and right. Specifically, we define $Accuracy = \frac{\textit{Number of correct direction estimations}}{\textit{Total eye rotations performed}}$.

4.6.4: Eye Motion Detection Rate

RadEye's eye motion detection ability, including eye rotation and eye blink detection, serves as the foundation of many eye tracking applications. While RadEye focuses on estimating eye rotation angles, eye blink detection is also one of its key components. This feature not only enhances the



(a) (b) (c) (d)

Figure 4.14: (a) RadEye's eye blink/rotation detection rate. (b) RadEye's eye rotation estimation error at different distances. (c) RadEye's accuracy of estimating eye rotation directions. (d) The confusion matrix of RadEye's eye rotation direction estimation at 3 m distance.

input of RadEye but also enriches its functionalities. Therefore, we evaluate RadEye's success rate of detecting eye rotation and blink.

We instructed eight participants to perform eye rotations and blinks from three different distances: 3 m, 4 m, and 5 m, as shown in Fig. 4.10(a)-(c). A total of 5 minutes of data were collected at each distance for each participant. Fig. 4.14(a) shows RadEye's average eye motion detection rate for 8 individuals at various distances. The highest detection rate for eye blink and eye rotation is 94% and 96%, respectively. This was observed at the distance of 3 m. Overall, the detection rate is consistent. Even at a distance of 5 m, RadEye achieves 88% detection rate for eye blinks and 91% detection rate for eye rotation. This confirms the robustness of RadEye in eye motion detection in different environmental settings. Numerically, the standard deviation of eye blink detection across the eight individuals is about 2%. For the eye rotation detection, the standard deviation is about 4%. This slight difference can be attributed to the simplicity and similarity of eye blink motion across different individuals. Additionally, the detection rate of eye rotations is consistently higher than that of eye blinks in all cases. This is not surprising, as eye rotations involve more significant facial muscle movements compared to eye blinks.

4.6.5: Eye Rotation Angle/Direction Estimation

Eye Rotation Angle Estimation. We conducted the experiments in the same way as described in §4.6.4. Fig. 4.14(b) presents the cumulative distribution function (CDF) of the angle estimation errors of RadEye for all participants at three different distances. The mean azimuth/elevation errors at 3 m, 4 m, and 5 m are approximately 14°/7°, 20°/18°, and 24°/21°, respectively. Evidently, the eye rotation angle estimation error increases as the distance increases. This is not surprising, as the radio signal has a larger attenuation over a longer distance. Additionally, we observed that the elevation angle estimation error is consistently smaller than the azimuth angle estimation error. This observation agrees with our previous observation in §4.3, i.e., eye's vertical movements (up and down) generate more pronounced changes in radar signal's amplitude and phase compared to eye's horizontal movements (right and left).

Eye Rotation Direction Estimation. As some applications of RadEye need only the eye rotation direction information, we first classify the estimated eye rotation angle into four directions (up, down, left, and right) and then evaluate the estimation accuracy. Since the slight eye motions are always accompanied by humans, we consider the eye rotation action effective only when the azimuth $\alpha < 50°$ or $\alpha > 130°$, or the elevation $\beta < 50°$ or $\beta > 130°$, as exemplified by Fig. 4.12. Using the effective input captured by the camera as ground truth, we can measure RadEye's estimation accuracy. Fig. 4.14(c) plots the average estimation accuracy for eight people at three different distances. The average estimation accuracy at 3 m, 4 m, and 5 m is 90.0%, 84.7%, and 83.5%, respectively. These accuracy levels are suited for most daily applications requiring human input. The standard deviations at 3 m, 4 m and 5 m are 3%, 5% and 5.5%. This indicates RadEye's robustness when detecting eye rotation of different users. Additionally, Fig. 4.14(d) presents the confusion matrix in the 3-meter case. It is evident that distinguishing between right and left eye rotations is more challenging compared to up and down eye rotations. This suggests that developing an application with binary input for up and down movements could enhance RadEye's robustness.

4.6.6: RadEye's Robustness

RadEye's Field of View. RadEye has two patch antennas for signal transmission and reception. Ideally, the target person should be perpendicularly facing RadEye's antenna. In practice, the target person may not be ideally positioned. Therefore, we conducted experiments to evaluate RadEye's field of view by examining its estimation accuracy when the target person was located in different directions, as illustrated in Fig. 4.10(d)-(e). Specifically, five participants performed eye rotations

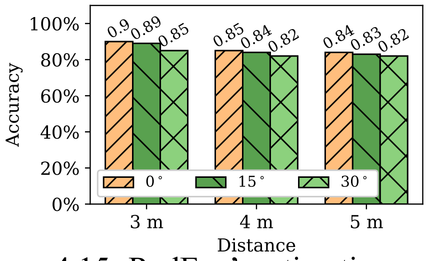


Figure 4.15: RadEye's estimation accuracy of eye rotation directions when the target person is located at different distances and angles.

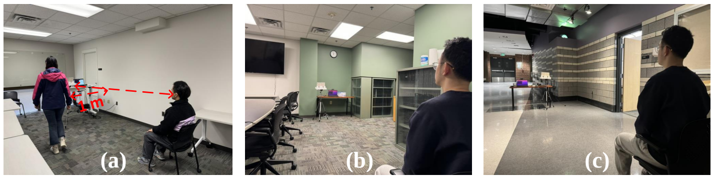


Figure 4.16: (a) The interference test in a lab. (b) Test at 5 m in the conference room. (c) Test at 5 m in the hallway.

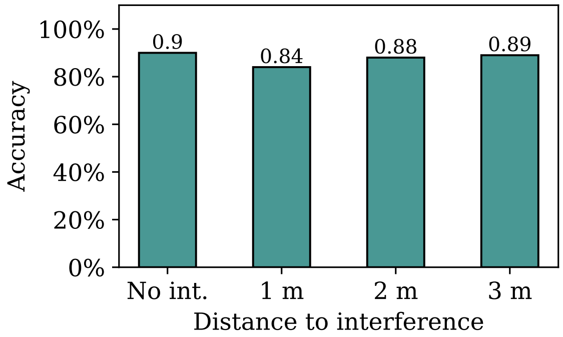


Figure 4.17: RadEye's estimation accuracy of eye rotation directions when experiencing interference from walking people.

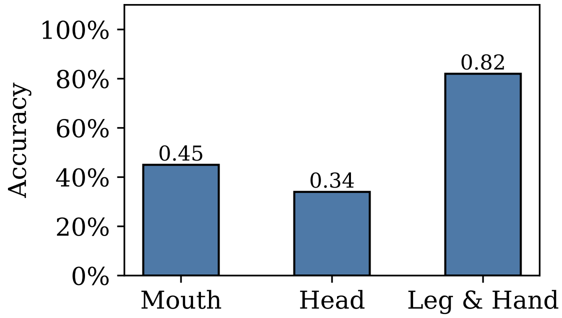


Figure 4.18: RadEye's estimation accuracy of eye rotation directions during self-body motions.

at angles of 15° and 30° from three different distances. In total, six scenarios were studied. In each scenario, participants performed eye rotations for 5 minutes. Fig. 4.15 shows our experimental results. It can be seen that RadEye achieves a high estimation accuracy when the target person is located at 0°, 15°, and 30°. Overall, the estimation accuracy remains above 82% in all cases. This indicates that RadEye has at least 60° field of view.

Impact of Moving Objects. Since static objects can easily be filtered out in the received signal, we further evaluated RadEye's resilience to interference caused by nearby walking individuals. In the experiments, another person was asked to walk around the user in close proximity, as depicted in Fig. 4.16(a). We measured the accuracy in three scenarios where the distance between the user and the walking person was 1 m, 2 m, and 3 m. In each scenario, five participants performed eye rotations for 5 minutes. Fig. 4.17 shows the results. The presence of a walking person causes a slight decrease in RadEye's estimation accuracy. Overall, RadEye achieves accuracies of 84%, 88%,

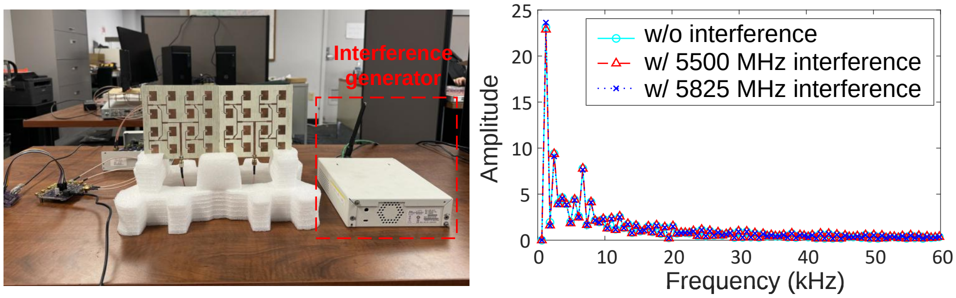


Figure 4.19: Wi-Fi interference test: experimental setup (left) and experimental results (right).

and 89% when the walking person is 1 m, 2 m, and 3 m away from the participant, respectively. We note that all these experiments were conducted in normal scenarios. The participants were only instructed to keep their heads still when performing eye rotations. No other restrictions were made to avoid interference from multipath effects or other normal physiological activities of the participants.

Impact of Self-Body Motions. Besides nearby moving objects, the participant's own body movements may also affect RadEye's performance. To evaluate RadEye's usability in practical scenarios, we studied the cases where the participant was speaking, shaking head, or engaging in leg or hand motions. A participant was asked to perform these three activities separately while executing eye rotations. In each scenario, the participant performed eye rotations for five minutes at a distance of 3 m. Fig. 4.18 presents our measurement results. It can be seen that RadEye's accuracy remains at 82% when the participant performed leg or hand motions. However, RadEye's accuracy decreases to 34% when he was shaking his head and to 45% when he was speaking. This reduction could be attributed to the limited resolution of RadEye. Since the legs and hands are more than 14 cm away from the eyes, their movements have minimal impact on RadEye's detection accuracy. However, head and mouth movements interfere with the eye rotation signal, leading to a lower detection accuracy.

Impact of Wi-Fi signals. As RadEye operates at 5 GHz, overlapping with part of the Wi-Fi spectrum, we conducted experiments to evaluate the impact of Wi-Fi signal interference on RadEye. A Wi-Fi interferer was set up using the USRP and placed next to RadEye, as shown in Fig. 4.19, continuously transmitting Wi-Fi packets. The Wi-Fi signals were generated at two frequencies, 5.5 GHz and 5.825 GHz, with a bandwidth of 20 MHz. The experiments were conducted in a static environment. We compared the IF signals from RadEye with and without Wi-Fi interference, and the results are shown in Fig. 4.19. We observed that the IF signals remain nearly identical, regardless of the presence of Wi-Fi signals.

RadEye appears to be resilient to Wi-Fi interference due to two key factors. First, RadEye operates with a broad bandwidth of 1.1 GHz, whereas Wi-Fi signals occupy only 20 MHz. Second, RadEye employs FMCW modulation, which contrasts with Wi-Fi's OFDM modulation. OFDM signals exhibit pseudo-noise characteristics, and when they are correlated with FMCW signals over time, the resulting correlation is close to zero. This theoretical outcome explains why RadEye can effectively resist interference from nearby Wi-Fi devices.

4.6.7: Zero-Shot Performance

Since RadEye has a DNN component for eye rotation detection, it is critical to evaluate its zero-shot performance against new users and new scenarios.

New Users. The generality of the trained model is crucial as it allows for easy extension to new users with minimal effort. To evaluate this, we conducted a cross-user test for RadEye. Specifically, four new users who were not involved in the training data collection were invited to perform eye rotations at a distance of 5 m. Among these users, participants B and C wore glasses. Each participant contributed 5 minutes of data. Fig. 4.20 reports RadEye's estimation accuracy for these four different users. The highest accuracy is 86% for user A, and the lowest accuracy is 77% for user D. Notably, wearing glasses does not seem to affect the results much. RadEye achieves an average accuracy higher than 80% for new users. This demonstrates the generalizability of RadEye to new users.

New Environments. In addition to evaluating RadEye's generalizability to new users, we also assessed its zero-shot performance in unseen scenarios. Four new participants performed eye rotations at a distance of 5 m in both a conference room and a hallway. Fig. 4.16(b)-(c) illustrates our experimental settings, where participants faced RadEye at an angle of 0°. Fig. 4.21 presents

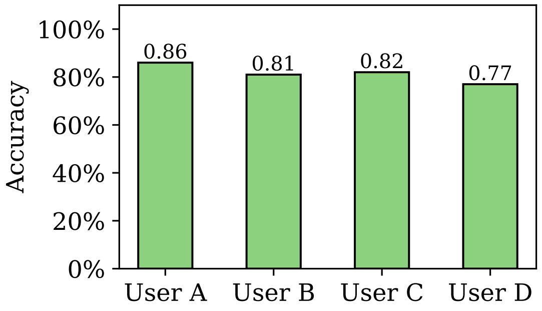


Figure 4.20: The accuracy for users not in the training set.

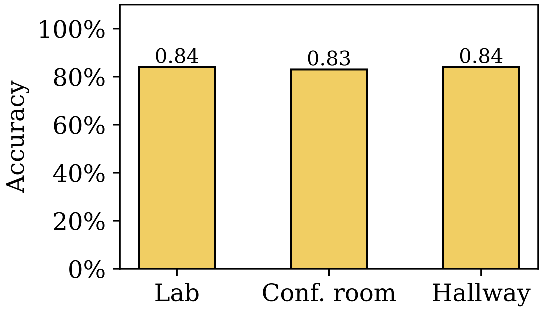


Figure 4.21: The accuracy in the different environments.

the measurement results. RadEye achieves accuracies of 83% and 84% in the conference room and hallway, respectively. These results demonstrate RadEye's ability to generalize to unseen environments as well.

4.7: Limitations and Discussions

In this section, we point out the limitations of RadEye and discuss potential solutions to address them.

- **Interference Caused by Head and Mouth Movements.** While RadEye is resilient to interference from surrounding environments, it requires the users to keep their heads still during use. Movements such as head shaking, speaking, or other facial expressions can obscure the eye rotation signals, resulting in unsuccessful detection. To address this issue, one approach is to increase the bandwidth of RadEye. When the bandwidth is sufficiently large, RadEye can differentiate eyes from mouth in the frequency domain, thereby eliminating the interference from head and mouth movements for eye motion detection.
- **RadEye versus mmWave Radar.** MmWave radar is capable of detecting subtle movements, such as eye motion, and is commercially available on the market. However, its detection range is relatively short due to the rapid signal attenuation of mmWave propagation. In contrast, RadEye offers a significantly larger range for eye motion detection but requires a wide spectrum bandwidth at lower frequencies. Therefore, both mmWave radar and RadEye have distinct advantages and limitations. MmWave radar is better suited for short-range eye

tracking, while RadEye is more appropriate for long-range use cases.

- **Physical Size of RadEye.** Our current prototype of RadEye is not compact enough for certain applications, such as installation on wheelchairs. This limitation arises because our prototype has not yet been optimized. In fact, the current design has significant potential for size reduction through various optimizations, including using smaller packages (e.g., SMD) for electronic components, more efficient power management chips, improved patch antenna designs, and shorter cables. Moreover, integrating the patch antennas directly into the PCB could significantly reduce the system's physical size, making it more suitable for space-constrained applications.
- **User Fatigue.** RadEye currently recognizes eye rotations in only four directions, requiring users to rotate their eyes multiple times to input a word. This tends to cause eye fatigue. Future work will focus on developing a system capable of continuously tracking eye rotation directions with high accuracy, rather than limiting recognition to four discrete directions. This improvement would enhance input efficiency and significantly reduce user fatigue.

4.8: Summary

Remote eye tracking has many potential applications ranging from HCI-based input to eye disease detection. While camera has been widely studied for eye tracking, its application in practice may raise privacy concerns in some scenarios. In this chapter, we presented RadEye, an RF sensing system capable of recognizing fine-grained human eye movement from a long distance. The challenge in the design of RadEye is to detect tiny eyeball movements in the presence of interference from other moving objects. RadEye addresses this challenge through a joint hardware and software design. For hardware, RadEye custom-designed a sub-6GHz FMCW radar for feature extraction and interference mitigation. For software, a camera-guided DNN model has been crafted to improve RadEye's detection accuracy. Extensive experiments show that RadEye achieves 90% accuracy when detecting people's eye rotation directions (up, down, left, and right) in various scenarios.

CHAPTER 5: UPLINK MU-MIMO COMMUNICATION IN MMWAVE WLANS

5.1: Introduction

Recently, the integration of millimeter-wave (mmWave) and multi-user multiple-input multiple-output (MU-MIMO) technologies has attracted significant research and development attention in wireless local area networks (WLANs), due to their potential to deliver data rates of hundreds of Gbps through the simultaneous transmission of multiple independent data streams [54]. As a concrete step towards its real-life applications, downlink mmWave MU-MIMO has been standardized by IEEE 802.11ay [5], and its theoretical data rate can reach 176 Gbps.

However, the advancement of mmWave MU-MIMO is mainly limited to its downlink. Very limited progress has been made so far for its uplink. While both 802.11ac (sub-6GHz) and 802.11ay (60GHz) support downlink MU-MIMO, neither of them supports uplink MU-MIMO. This stagnation underscores the grand challenges in the design of practical yet efficient uplink mmWave MU-MIMO communication schemes. In addition, the demand of uplink data rate is dramatically increasing in emerging applications such as autonomous driving and video streaming. Ericsson predicts that the amount of global uplink traffic will reach 70 EB per month in 2027 [2]. Therefore, there is a critical need to fill this gap.

In this chapter, we present a practical yet efficient uplink MU-MIMO mmWave communication scheme (UMMC) for a wireless local area network (WLAN). UMMC allows multiple stations to *simultaneously* send their data packets to an access point (AP) while not requiring fine-grained inter-station synchronization. We address two challenges in the design of UMMC. *The first challenge lies in the analog beamforming for a multi-antenna AP.* While the literature has a wealth of analog beamforming work, existing approaches can be generally classified into two categories: model-based optimization (e.g., [111], [88, Table V]) and model-free beam search (e.g., [62, 65, 118, 148, 150]). While model-based approaches offer the optimal antenna weight vectors (AWVs) for analog beamforming, they require accurate antenna model and channel knowledge, which are hard to obtain. Therefore, these approaches are not amenable to practical use. Model-free approaches do not require the above knowledge as they aim to find the best beam in a predefined beambook. However, most of them focus on maximizing the signal strength for a single-antenna mmWave device while minimizing their beam search overhead. While maximizing signal strength is equivalent to maximizing data rate in single-antenna systems, it is not the case in MU-MIMO systems. This is because the capacity of an MU-MIMO channel is dependent upon not just the signal strength but also the correlation of MIMO channels. When two stations have highly-correlated channels, the AP may not be capable of decoding their packets even if the signals are strong. In addition, exhaustive search is notorious for its large airtime overhead and thus not suitable for practical use.

To address this challenge, we design a Bayesian optimization (BayOpt) framework for joint beam search at the AP. This framework is inspired by two facts: i) the relation between a selected beam and its achievable data rate in MU-MIMO communications is complex and unknown in real systems; and ii) BayOpt has been proved to be an effective technique for finding an optimal or near-optimal solution to an optimization problem whose objective function and constraints are unknown and costly to evaluate. The key idea of the BayOpt framework is to guide beam search using the posterior probability derived from those beams that have already been evaluated. The more beams we evaluate, the more accurate information we have for the remaining beams. Compared to exhaustive search, BayOpt appears to be surprisingly efficient in finding a near-optimal beam within a given airtime budget.

Another challenge in the design of UMMC is the synchronization among stations. Actually, the signal detection in uplink MU-MIMO transmission has been well studied in sub-6GHz wireless networks, and some signal detection methods such as zero-forcing (ZF) and minimum mean square error (MMSE) have been widely used in practice. However, existing signal detectors are based on an important assumption — the data packets from different stations are synchronized in time when impinging on the AP. Particularly, in OFDM systems, the time misalignment of the packets when arriving at the AP must be less than the time duration of an OFDM symbol's cyclic prefix (CP). While this requirement can be achieved in narrow-band (20 MHz) sub-6GHz systems (e.g., using timing advance protocols), it is extremely challenging to achieve in ultra wideband mmWave systems. For instance, using conventional MU-MIMO detectors, the time misalignment of packets in 802.11ay must be less than 36ns, which is hard to maintain in practice. Due to this stringent requirement, uplink MU-MIMO has not yet been supported by 802.11ay standard [5].

To address this challenge, we argue that it is more desirable living with the packet misalignment at the AP instead of employing an onerous protocol to synchronize stations. Towards this goal, we observed that existing MU-MIMO detectors work in the *spatial* domain while the packet misalignment is an imperfection in the *temporal* domain. Since these two domains are orthogonal, spatial MU-MIMO detectors should be immune to temporal misalignment of data packets. In fact, the real problem is that the construction of existing MU-MIMO detectors requires the knowledge of channel, which relies on orthogonal pilots (reference signals) in data packets to estimate. However, misaligned packets cannot maintain the orthogonality of their pilots, making it hard to estimate channels. To solve this problem, we design an asynchronous MU-MIMO detector through a transformation of existing MMSE MU-MIMO detector. This new detector is capable of decoding asynchronous packets from multiple stations without the need of explicit channel knowledge. The key idea behind our design is to use the interfered pilots within each packet to train its detection filter. Doing so eliminates the need of channel matrix in the construction of the detection filter while achieving a surprisingly good performance. The new detector fundamentally relaxes the synchronization requirement for stations in uplink MU-MIMO transmission, making UMMC amenable to practical implementation.

We have evaluated UMMC through a blend of over-the-air experiments and extensive simulations. We implemented UMMC on a two-user MIMO mmWave (60GHz) testbed and demonstrated that it enables real-time uplink packet transmission in the absence of inter-user synchronization (see our demo video via the anonymous link in [29]). Experimental results show that, compared to exhaustive beam search, BayOpt achieves 92% throughput while reducing the overhead by 98%. In addition, simulation results from a 100-user mmWave network show that, compared to exhaustive beam search, BayOpt achieves more than 80% of its throughput while entailing less than 5% of its overhead in all two-user, three-user, and four-user MIMO cases.

The contributions of this work are summarized as follows.

- We design a practical uplink MU-MIMO mmWave communication scheme for WLANs. We demonstrate that it works in realistic scenarios via over-the-air experiments.
- We introduce the first-of-its-kind BayOpt framework for beam search in mmWave MU-MIMO systems, and show its efficiency through both experimental and simulation results.
- We propose a new MU-MIMO detector that can decode the asynchronous data packets from multiple user devices. For the first time, it demonstrates via theory and experiments that fine-grained inter-user synchronization is not needed for uplink MU-MIMO mmWave transmission.

5.2: Related Work

This work is relevant to mmWave MIMO communications, beam search, MU-MIMO detection, and system prototyping.

802.11ad/ay and Cellular Networks. In 2012, the IEEE 802.11ad amendment standardized communication in 60 GHz unlicensed band to offer up to 6.75 Gbps data rate in a short range [1]. While 802.11ad devices may have multiple antennas, they do not support MU-MIMO transmission. As a follow-up, 802.11ay was standardized in 2020 [5], which supports new features including channel bonding, higher-order modulation, and downlink MU-MIMO. However, it does not support uplink MU-MIMO yet.

The 3GPP specification for 5G cellular networks has already supported MU-MIMO, hybrid beamforming, and mmWave communications in the 24–53 GHz band [13]. While abundant literature has studied the beam design and MU-MIMO for mmWave, most of them are limited to signal processing and numerical analysis [67, 116]. Very likely, future cellular networks will em-

Table 5.1: Representative work on beam search in literature.

Beam search technique	**Approach**
Learning-based search: [23, 30, 68, 84]	Train the machine learning model to predict the current best beam direction.
Out-of-Band assistance: [21, 62, 64, 118, 135, 148]	Utilize the out-of-band information (sub-6GHz, light, camera) to align the beam.
Compressive sensing: [65, 127]	Finding the best alignment beam direction with sparse measurement.
Hierarchical search: [76, 92, 120, 141, 157, 194]	Design a beamforming codebooks and training in a hierarchical way.

ploy sophisticated synchronization protocols (e.g., timing advance) and flexible numerology (e.g., long CP) to support uplink MU-MIMO. However, this approach is not suitable for WLANs as they target low-cost applications.

Beam Search. There is a large body of work on beam training for mmWave communications. Table 5.1 lists some of representative work and their basic ideas. Of existing work, most focuses on finding the best beam in a predefined beambook to maximize *signal strength* while minimizing the associated cost. As we explained before, maximizing the signal strength is not a good strategy for MU-MIMO.

In addition to beam selection, there is a considerable amount of work that studies analog beamforming from a signal processing perspective by formulating the AWV design problem to an optimization [92, 157, 180, 194]. However, this kind of work is not amenable to practical implementation for several reasons: i) they assume that the phased-array antenna has an ideal radiation pattern; ii) they require the over-the-air channel knowledge for the design of AWV; and iii) they suffer from high computation in solving an optimization problem.

Uplink MU-MIMO in Sub-6GHz Networks. Uplink MU-MIMO has been supported in 4G cellular networks and will be supported by 5G and beyond [3]. In contrast, the way of uplink MU-MIMO to 802.11 standards was rocky. Thus far, no on-market Wi-Fi devices support uplink MU-MIMO. Similar to 802.11ay, 802.11ac supports downlink MU-MIMO but does not support uplink MU-MIMO [4]. This can be attributed to the fact that WLANs are distributed, contention-based systems and lack inter-user coordination. Although 802.11ax will support uplink MU-MIMO, the symbol-level synchronization remains an outstanding challenge [140].

Inter-User Synchronization for Uplink MU-MIMO. Timing advance (TA) is the main mechanism used in wireless networks to compensate inter-user time misalignment and offset the signal propagation delays for uplink MU-MIMO and other multi-access technologies. Per [14] and [110], the timing error achieved by TA in cellular networks cannot meet the requirement of mmWave MU-MIMO based on the 802.11ay numerology. [212] validated the throughput gain of MU-MIMO via offline experiments but did not address the timing problem.

5.3: Problem Description

We consider the uplink MU-MIMO communication in a WLAN as shown in Fig. 5.1, where an AP wishes to decode concurrent data packets from multiple stations. Our objective is to maximize the uplink throughput through the design of analog and digital beamforming for the AP. In the pursuit of this objective, we assume that a beam has already been selected for each station using an existing beam search scheme such as sector-level sweep (SLS) and beam refinement protocol (BRP) [1]. We focus on the analog and digital beamforming at the AP for uplink MU-MIMO transmission.

5.3.1: Problem Formulation

Analog Beamforming. Denote M as the number of phased-array antennas (RF chains) on the AP and N as the number of stations involved in the uplink MU-MIMO transmission (assuming $N \leq M$). We assume that all the phased-array antennas on the AP are identical. Suppose that a linear phased-array antenna intends to steer its beam energy to the direction of θ. Then, its antenna weight vector (AWV) can be modeled as: $G_{\text{ap}}(\theta) = [e^{j\frac{d_{\text{ap}}}{\lambda} i \sin(\theta)}]_{0 \leq i \leq N_{\text{ap}}-1}$, where d_{ap} is the patch element spacing, λ is the wavelength, and N_{ap} is the number of patch elements. Similarly, for the phased-array antenna on a station, suppose it intends to steer its beam energy to the direction of ϕ. Then, its AWV can be modeled as: $G_{\text{sta}}(\phi) = [e^{j\frac{d_{\text{sta}}}{\lambda} i \sin(\phi)}]_{0 \leq i \leq N_{\text{sta}}-1}$, where d_{sta} is the patch element spacing and N_{sta} is the number of patch elements. Then, the signal received by the AP's mth RF

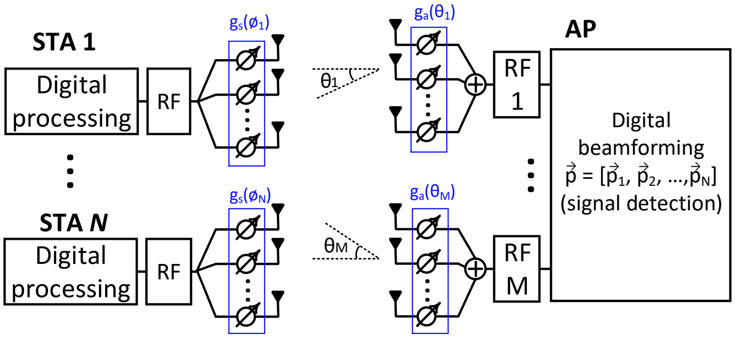


Figure 5.1: Uplink MU-MIMO transmission in WLAN.

chain can be written as:

$$y_m = \sum_{n=1}^{N} G_{\text{ap}}(\theta_m)\, \mathbf{H}_{mn}\, G_{\text{sta}}(\phi_n)^\top x_n + w_m, \tag{5.1}$$

where x_n is the signal transmitted by the nth station, w_m is the received noise, $\mathbf{H}_{mn} \in \mathbb{C}^{N_{\text{ap}} \times N_{\text{sta}}}$ is the over-the-air channel between the AP's mth antenna and the nth station's antenna.

Digital Beamforming. At the AP (receiver), digital beamforming serves for the purpose of MU-MIMO Detection. Denote $\vec{y} = [y_1, y_2, \cdots, y_M]^\top$ as the received signals and $\vec{p}_n$ as the AP's spatial filter for decoding the data packets from station n. Then, the decoded version of the signal from station n can be written as: $\hat{x}_n = \vec{p}_n^{\,\mathsf{H}}\, \vec{y}$, for $1 \leq n \leq N$, where $(\cdot)^{\mathsf{H}}$ is the conjugate transpose operator.

Design Objective. At the AP, denote $\vec{\theta} = [\theta_1, \theta_2, \cdots, \theta_M]$ as the beam angle vector, which can be directly used to calculate the AWV for analog beamforming. Denote $\vec{p} = [\vec{p}_1, \vec{p}_2, \cdots, \vec{p}_N]$ as the detection vector. Denote EVM$_n$ as the error vector magnitude (EVM) of the decoded signals from station n, i.e., $\text{EVM}_n \equiv \frac{\mathbb{E}[|x_n - \hat{x}_n|^2]}{\mathbb{E}[|x_n|^2]}$. Without loss of generality, we assume that the transmit power at stations are normalized, i.e., $\mathbb{E}[|x_n|^2] = 1$. Then, we have

$$\text{EVM}_n = \mathbb{E}[|x_n - \hat{x}_n|^2]. \tag{5.2}$$

The link capacity (spectral efficiency) between station n and the AP can be written as: $c_n = \log_2(1 + \frac{1}{\text{EVM}_n})$.

In uplink MU-MIMO, it is important not only to maximize the data rate but also ensure the fairness among users. Thus, our objective is to pursue the best analog and digital beams so that the bottleneck link data rate can be maximized. Mathematically, it can be formulated as:

$$[\vec{\theta}^{*},\ \vec{p}^{*}] = \underset{\vec{\theta}\in\mathcal{B},\ \vec{p}}{\arg\max}\left(\min_{n}\log_2(1+\frac{1}{\text{EVM}_n})\right), \tag{5.3}$$

where $\mathcal{B}$ is the predefined beambook that includes all possible beam angle vectors.

The optimization problem in (5.3) can be divided into two subproblems: i) analog beam selection (determining $\vec{\theta}$), and ii) MU-MIMO detector construction (determining $\vec{p}$). These two subproblems are tightly coupled with each other. Given the complex nature of this problem, it is intractable to pursue a global optimal solution in real systems. Therefore, we develop a practical yet efficient scheme to solve the two subproblems.

5.3.2: Key Challenges

Inaccurate Models. Solving the above optimization is nontrivial as the gradients of the objective function are unknown, so first-order methods like gradient descent cannot be applied. In addition, we used $G_{\text{ap}}(\theta)$ and $G_{\text{sta}}(\phi)$ to model the response of ideal linear phased-array antennas. In practice, phased-array antennas have many imperfections in their radiation patterns. Their actual mathematical models are unknown. The discrepancy between the ideal and real antenna model significantly affects the beamforming design.

Channel Correlation. The capacity of MU-MIMO transmission is determined by not only *signal strength* but also *MIMO channel correlation*. Existing approaches based on signal strength only are not suitable for beam search in MU-MIMO. Therefore, it calls for a beam search scheme that can *jointly* identify the best beams for *all* antennas. One straightforward approach is exhaustive search. However, it will entail a large airtime overhead and thus compromise the throughput gain of MU-MIMO. *Therefore, an efficient joint beam search scheme is needed.*

Inter-Station Timing Synchronization. Uplink MU-MIMO detection has been well studied. However, existing schemes require fine-grained inter-user timing synchronization for signal detection. That is, the time misalignment of data packets from different stations must be less than OFDM CP length. In 802.11ay [5], the normal guard interval duration (CP) is 36.36ns. Maintaining the inter-user synchronization within 36.36ns not only entails a large overhead but also complicates the network design and operation. For this reason, neither 802.11ac (sub-6GHz) nor 802.11ay (60GHz mmWave) supports uplink MU-MIMO.

5.4: Overview of UMMC

In this section, we first highlight our approaches to overcoming the above challenges and then present the overall system diagram of UMMC. In what follows, we denote $f(\vec{\theta}) = \max_n\{\text{EVM}_n\}$. When $\vec{p}$ is given, the optimization in (5.3) is equivalent to minimizing $f(\vec{\theta})$.

5.4.1: Our Approaches

Analog Beam Search. To address the beam search challenge, we design a BayOpt scheme for joint beam search. BayOpt has been proved to be an effective technique for solving sequential optimization problems where the objective function is complex (treated as a black-box), the (sub-)gradient is unknown, and the evaluation is expensive [114]. To illustrate the idea behind BayOpt, let us consider the beams in a beambook $[\vec{\theta}_1, \vec{\theta}_2, \cdots, \vec{\theta}_{3600}]$. Suppose that we have measured two beams, say $\vec{\theta}_{10}$ and $\vec{\theta}_{1000}$, and found that $f(\vec{\theta}_{10}) = 5$ and $f(\vec{\theta}_{1000}) = 0.1$. Then, in the next iteration we should select a beam in the neighborhood of $\vec{\theta}_{1000}$ to evaluate, because the global minimum is more likely sitting in the neighborhood of $\vec{\theta}_{1000}$ compared to $\vec{\theta}_{10}$. BayOpt is a principled strategy to guide the process of joint beam search based on posterior probability.

MU-MIMO Detection. Inter-station synchronization is a fundamental problem for uplink MU-MIMO. Achieving the required timing alignment for packet transmission among distributed stations is extremely hard. In light of this, we live with the timing misalignment among the stations and focus on enabling asynchronous MU-MIMO detection. To this end, we revisit the conventional (synchronous) MMSE detector and find that a transformation can make it applicable to decoding asynchronous data packets from independent stations.

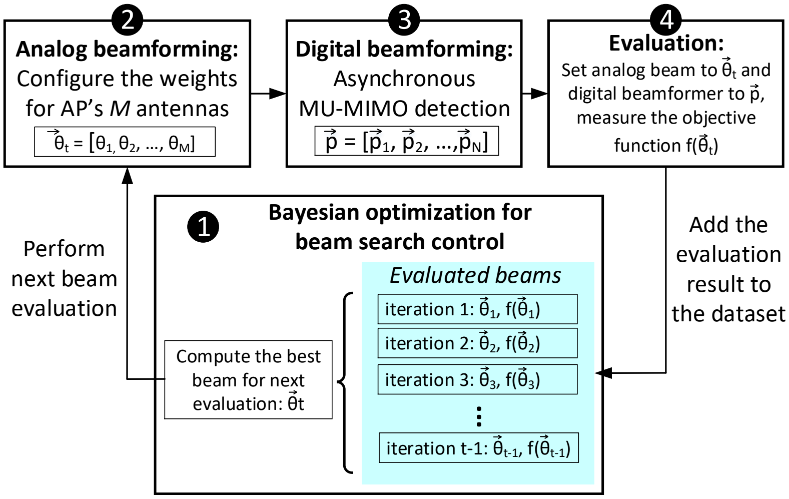

2
Analog beamforming:
Configure the weights for AP's M antennas
θ⃗t = [θ1, θ2, ..., θM]
3
Digital beamforming:
Asynchronous MU-MIMO detection
p⃗ = [p⃗1, p⃗2, ...,p⃗N]
4
Evaluation:
Set analog beam to θ⃗t and digital beamformer to p⃗, measure the objective function f(θ⃗t)
1
Bayesian optimization for beam search control
Perform next beam evaluation
Add the evaluation result to the dataset
Evaluated beams
iteration 1: θ⃗1, f(θ⃗1)
iteration 2: θ⃗2, f(θ⃗2)
iteration 3: θ⃗3, f(θ⃗3)
iteration t-1: θ⃗t-1, f(θ⃗t-1)
Compute the best beam for next evaluation: θ⃗t

Figure 5.2: The high-level system diagram of UMMC.

5.4.2: System Diagram

Fig. 5.2 shows the system diagram of UMMC. The AP measures the performance of a sequence of analog beams $[\vec{\theta}_1, \vec{\theta}_2, \cdots, \vec{\theta}_t, \cdots, \vec{\theta}_T]$, where t is the evaluation/iteration index and T is the predefined maximum number of evaluations/iterations allowed (e.g., $T = 30$). In the end of T iterations, UMMC chooses the beam that yields the best performance. In each iteration t, the operations of UMMC include the following four steps:

- Step 1: The AP selects a beam $\vec{\theta}_t$ for evaluation in the current iteration based on the posterior probability derived from the past evaluations, i.e., $(\vec{\theta}_{t\prime}, f(\vec{\theta}_{t\prime}))$ for $1 \leq t\prime < t$. Details are presented in Section 5.5.
- Step 2: The AP reconfigures its phased-array antennas by setting their beam patterns to $\vec{\theta}_t$.
- Step 3: The AP first calculates its digital beamformers (a.k.a. MU-MIMO detector) $\vec{p} = [\vec{p}_1, \vec{p}_2, \cdots, \vec{p}_N]$, and then uses them to decode asynchronous signal frames from the N stations. Details are presented in Section 5.6.
- Step 4: The AP measures the EVM of the decoded signals from each station. By doing so, it obtains $f(\vec{\theta}_t)$. Then, $(\vec{\theta}_t, f(\vec{\theta}_t))$ is added to the dataset and will be used to guide the future beam search.

5.5: Bayesian Optimization for Beam Search

In this section, we assume that the algorithm for determining $\vec{p}$ is given and focus on the BayOpt design to find a near-optimal beam $\vec{\theta}$ for the AP. The design of $\vec{p}$ will be presented in the next section.

5.5.1: Why Bayesian Optimization?

Recall that the objective function is $f(\vec{\theta}) = \max_n\{\text{EVM}_n\}$. It has the following salient features.

- *$f(\vec{\theta})$ has a complex structure:* Fig. 5.3 shows an example of $f(\vec{\theta})$ obtained through exhaustive beam search on our two-user MIMO 60GHz mmWave testbed[2]. It is evident that $f(\vec{\theta})$ is hard to optimize due to its non-convexity.
- *$f(\vec{\theta})$ is unknown:* Practical mmWave communication systems typically suffer from hardware imperfections such as phase noise and clock jitters [192], which are hard to characterize and model. As such, the beam pattern may largely deviate from its ideal model $G_{\text{ap}}(\theta)$. The accurate objective function $f(\vec{\theta})$ is unknown and can only be obtained via exhaustive experimental measurements.
- *Evaluating $f(\vec{\theta})$ is costly:* To evaluate $f(\vec{\theta})$ for a given $\vec{\theta}$, the AP needs to physically set up the beam pattern and measure the resultant signal quality. This process incurs a fixed airtime overhead. For example, in 802.11ay, measuring the value of $f(\vec{\theta})$ for a given $\vec{\theta}$ may take the time of one Control PHY Preamble (about 3.7μs), let alone other airtime overhead incurred in this process. Therefore, there is a tradeoff between the quality of $\vec{\theta}$ and the number of evaluations of $f(\vec{\theta})$.

Fortunately, BayOpt is an effective technique to optimize such a function that is unknown yet expensive to evaluate [114]. It makes use of the laws of probability to combine prior belief with observed data to compute posterior distribution of the objective function. Therefore, we will design a BayOpt framework for analog beam search.

[2] The detailed experimental setup is presented in Section 5.7.1.

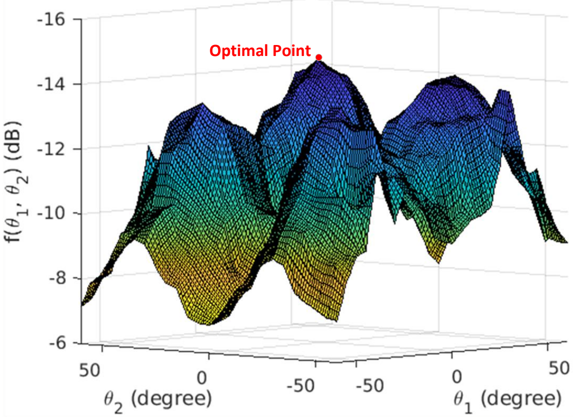


Figure 5.3: An instance of $f(\vec{\theta})$ obtained from experimental measurements on a two-user MIMO 60GHz testbed, where $\vec{\theta} = [\theta_1, \theta_2]$ and $f(\vec{\theta}) = \max(\text{EVM}_1, \text{EVM}_2)$ in dB.

5.5.2: A Bayesian Optimization Framework

To perform BayOpt, one needs to address two problems: i) finding a statistical process to model the function being optimized, and ii) selecting an acquisition function as a surrogate approximation to guide the search in each iteration. In what follows, we address these two problems in order.

Gaussian Process Regression. We model the iterative beam search problem as a Gaussian process. In the tth iteration, the AP has observed $t-1$ beams. Denote $\mathbf{\Theta} = \{\vec{\theta}_i\}_{i=1}^{t-1}$ as the set of beams that the AP has already observed. Denote $f(\mathbf{\Theta}) = \{f(\vec{\theta}_i)\}_{i=1}^{t-1}$ as the objective function values of those observed beams. We treat $f(\mathbf{\Theta})$ as a multi-variate Gaussian distribution, with $\mu(\mathbf{\Theta})$ as its mean and $k(\mathbf{\Theta}, \mathbf{\Theta})$ as its covariance kernel. Here, $\mu(\mathbf{\Theta})$ is a $(t-1)\times 1$ vector, while $k(\mathbf{\Theta}, \mathbf{\Theta})$ is a $(t-1)\times(t-1)$ matrix. Let $\vec{\theta}$ be an arbitrary beam in the beambook. Then, per the definition of Gaussian process, the joint distribution of the function values corresponding to $\vec{\theta}$ and $\mathbf{\Theta}$ should satisfy:

$$\begin{bmatrix} f(\mathbf{\Theta}) \\ f(\vec{\theta}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{\Theta}) \\ \mu(\vec{\theta}) \end{bmatrix}, \begin{bmatrix} k(\mathbf{\Theta}, \mathbf{\Theta}) & k(\mathbf{\Theta}, \vec{\theta}) \\ k(\vec{\theta}, \mathbf{\Theta}) & k(\vec{\theta}, \vec{\theta}) \end{bmatrix} \right), \tag{5.4}$$

where $\mu(\cdot)$ and $k(\cdot, \cdot)$ should be understood as an element-wise operational function. There are various definition for Gaussian kernel, such as *Matérn* kernel, *exponentiated quadratic* kernel, and *radial basis function* kernel [177]. In our experiments, we choose radial basis function kernel,

$k(\vec{\theta}_i, \vec{\theta}_j) = \exp(-\frac{1}{2\sigma^2}||\vec{\theta}_i - \vec{\theta}_j||^2)$, where σ is a hyper-parameter that governs the kernel width. In our experiments, we let $\sigma = 1$.

The posterior distribution on the arbitrary beam $\vec{\theta}$ can be calculated through standard Bayesian rules. Specifically, the distribution of $f(\vec{\theta})$ can be modeled as:

$$f(\vec{\theta}) \sim p\big(f(\vec{\theta})|\vec{\theta}, \mathbf{\Theta}, f(\mathbf{\Theta})\big) = \mathcal{N}\big(\mu(\vec{\theta}), \Sigma(\vec{\theta})\big), \tag{5.5}$$

where

$$\mu(\vec{\theta}) = k(\vec{\theta}, \mathbf{\Theta})k(\mathbf{\Theta}, \mathbf{\Theta})^{-1}f(\mathbf{\Theta}), \tag{5.6}$$

$$\Sigma(\vec{\theta}) = k(\vec{\theta}, \vec{\theta}) - k(\vec{\theta}, \mathbf{\Theta})k(\mathbf{\Theta}, \mathbf{\Theta})^{-1}k(\mathbf{\Theta}, \vec{\theta}). \tag{5.7}$$

Acquisition Function. There are different acquisition functions available for BayOpt problems such as Probability of Improvement (PoI), Expected Improvement (EI), and Gaussian process Upper Confidence Bound (GP-UCB) [177]. We choose EI for two reasons: i) compared to PoI, it has been shown to be better-behaved; and ii) unlike GP-UCB, it does not involve tuning parameters [138]. The acquisition function can be written as:

$$\text{EI}(\vec{\theta}) = \mathbb{E}\big[\max(f(\vec{\theta}) - f(\vec{\theta}^{+}), 0)\big], \tag{5.8}$$

where $\vec{\theta}^{+}$ is the best beam found so far. Under the Gaussian process model, it can be analytically written as follows:

$$\text{EI}(\vec{\theta}) = \big(\mu(\vec{\theta}) - f(\vec{\theta}^{+}) - \xi\big)\,\text{CDF}(Z) + \Sigma(\vec{\theta})\,\text{pdf}(Z), \tag{5.9}$$

where $Z = \frac{\mu(\vec{\theta}) - f(\vec{\theta}^{+}) - \xi}{\Sigma(\vec{\theta})}$, $\text{CDF}(\cdot)$ and $\text{pdf}(\cdot)$ are the cumulative distribution function and the probability density function of standard normal distribution, respectively, and ξ is a parameter that determines the amount of exploration during the optimization. A large value of ξ leads to more exploration, while a small value leads to more exploitation. In our experiments, we empirically set ξ to 0.1.

Beam Selection. Then, in the tth iteration, the beam selected for evaluation is obtained by solving the following problem:

$$\vec{\theta}_t = \underset{\vec{\theta} \in \mathcal{B} \setminus \mathbf{\Theta}}{\arg\max} \; \mathrm{EI}(\vec{\theta}), \tag{5.10}$$

where $\mathcal{B}$ is the set of all predefined beams and $\mathbf{\Theta}$ is the set of beams that has been evaluated so far. It is worth noting that (5.10) is easy to solve because (5.9) is a simple, disciplined function.

5.5.3: Practical Considerations

There are two challenges associated with the above BayOpt framework when it is applied to beam search. In the following, we first point out the challenges and then present our solutions.

Limited number of evaluations. MmWave systems have a fixed airtime budget for beam search/-training, which determines the maximum number of evaluations/iterations that can be performed before data transmission. In practice, given the limited airtime budget for beam search, it is unlikely to find the optimal beam for data transmission. Therefore, the beam search problem is further constrained by the number of evaluations. To address this challenge, we propose a *recenter-and-shrink* (RaS) scheme for the Gaussian process regression. This scheme was inspired by [144]. The basic idea is that, when approaching the evaluation budget, we *recenter* the search space to the current optimal beam and *shrink* the search space. Doing so increases the probability of finding a better beam when we reach the evaluation budget. Following this idea, we modify the acquisition function in (5.10) to:

$$\begin{aligned} \vec{\theta}_t &= \underset{\vec{\theta} \in \mathcal{B} \setminus \mathbf{\Theta}}{\arg\max} \; \mathrm{EI}(\vec{\theta}) \\ \text{s.t. } \theta_m &\in \begin{cases} [-\frac{\pi}{2}, \; \frac{\pi}{2}] & \text{if } 1 \leq t < T/2 \\ [\theta_m^+ - \frac{\phi_t}{2}, \; \theta_m^+ + \frac{\phi_t}{2}] & \text{if } T/2 \leq t \leq T. \end{cases} \end{aligned} \tag{5.11}$$

where t is the iteration/evaluation index, T is the maximum number of evaluations, $\vec{\theta}^{+} = [\theta_m^{+}]_{m=1}^{M}$ is the best beam found so far, and ϕ_t is the reduced search range. Empirically, we set $\phi_t = (\frac{3}{2} - \frac{t}{T})\pi$ in our experiments.

Cubic Computational Complexity. The computational complexity of Gaussian process regression is cubic to the number of data samples, i.e., $\mathcal{O}(t^3)$, where t is the number of evaluations that have been performed [177]. Clearly, the computation rapidly increases as the evaluation procedure evolves. To overcome the computation challenge of Gaussian process, a wealth of sparse approximations have been recently suggested, such as the subset of data (SoD) approximation, the subset of regressors (SoR) approximation, the deterministic training conditional (DTC) approximation, and partially independent training conditional (PITC) approximation [124]. *In these methods, a subset of the latent variables are treated exactly while the remaining variables are treated approximately to reduce the computation.* Here, we employ the SoR approximation for the beam search as it demonstrates a good tradeoff between performance and computation (see Tables 8.1 & 8.2 in [124]).

Denote $\mathbf{\Phi}$ as the subset of training data samples that are selected for exact regression, where $\mathbf{\Phi} \subset \mathbf{\Theta}$. Per [124], the Gaussian process regression can be characterized by the approximate mean and covariance as follows:

$$\mu(\vec{\theta}) = \sigma^{-2} k(\vec{\theta}, \mathbf{\Phi}) \boldsymbol{Q}^{-1} k(\mathbf{\Phi}, \mathbf{\Theta}) f(\mathbf{\Theta}), \tag{5.12}$$

$$\Sigma(\vec{\theta}) = k(\vec{\theta}, \mathbf{\Phi}) \boldsymbol{Q}^{-1} k(\mathbf{\Phi}, \vec{\theta}), \tag{5.13}$$

where $\boldsymbol{Q} = \sigma^{-2} k(\mathbf{\Phi}, \mathbf{\Theta}) k(\mathbf{\Theta}, \mathbf{\Phi}) + k(\mathbf{\Phi}, \mathbf{\Phi})$.

A question to ask is how to select the active data samples for $\mathbf{\Phi}$. Empirically, we define an integer number $\tau \in \mathbb{Z}$ which is smaller than t. We choose the τ beams in $\mathbf{\Theta}$ that are closest to $\vec{\theta}^{+}$ as the active samples for $\mathbf{\Phi}$. Denote $g(\vec{\theta}) \triangleq ||\vec{\theta}^{+} - \vec{\theta}||^2$ as the metric for $\vec{\theta}$. Based on this metric, we sort the elements in $\mathbf{\Theta}$ in a non-decreasing order and denote the resulting vector as

Algorithm 2 Bayesian optimization for analog beam search.

```
**Required:** $T$: the budgeted number of evaluations.
**Output:** A beam $\vec{\theta}^*$ in the predefined beambook $\mathcal{B}$ for data packet reception at the AP
Initialization $\mathbf{\Theta} = [\vec{0}]$.
**for** $t = 1, 2, \cdots, T$ **do**
    Calculate $\mathbf{\Phi}$ using (5.14)
    Calculate $\mu(\vec{\theta})$ using (5.12) and $\Sigma(\vec{\theta})$ using (5.13)
    Construct the surrogate function EI($\vec{\theta}$) using (5.9)
    Find the next beam direction $\vec{\theta}_t$ by solving (5.11)
    Add $\vec{\theta}_t$ to $\mathbf{\Theta}$
**end for**
**return** $\vec{\theta}^* = \arg\min_{\vec{\theta}\in\mathbf{\Theta}} f(\vec{\theta})$.
```

$\mathbf{\Theta}_{srt} = [\vec{\theta}_{s_1}, \vec{\theta}_{s_2}, \cdots, \vec{\theta}_{s_t}]$. Then, we let:

$$\mathbf{\Phi} = [\vec{\theta}_{s_1}, \vec{\theta}_{s_2}, \cdots, \vec{\theta}_{s_\tau}]. \tag{5.14}$$

With the approximation in (5.12)-(5.14), the computational complexity of Gaussian process regression in the tth iteration decreases to $\mathcal{O}(\tau^2 t)$ [177]. More importantly, the complexity scales linearly (rather than cubically) with the number of iterations.

We present the proposed BayOpt algorithm in Alg. 2. In a nutshell, it is a non-parametric online learning algorithm that guides the beam search using the posterior probability of those data samples that have been evaluated so far.

5.6: Asynchronous MU-MIMO Detection

In this section, we first review the MMSE MU-MIMO detector, and then present a transformation for MMSE MU-MIMO detector so that it can decode asynchronous data packets. The resulting detector fundamentally relaxes the inter-user synchronization for uplink MU-MIMO, and thus is particularly suited for mmWave communications. Finally, we conduct performance analysis of the proposed detector in mmWave networks.

5.6.1: Conventional (Synchronous) MMSE MU-MIMO Detector

Consider the uplink MU-MIMO transmission from N stations to an M-antenna AP as shown in Fig. 5.1. Suppose that data packets from the N stations are perfectly aligned in time when impinging on the AP. Then, the signal transfer model in the digital domain can be written as:

$$\vec{y} = \mathbf{H}\vec{x} + \vec{w}, \tag{5.15}$$

where $\vec{y} \in \mathbb{C}^{M\times 1}$ is the received digital baseband signal vector at the AP, $\vec{x} = [x_1, x_2, \cdots, x_N]^\top$ is the transmit signal vector, where x_n is the signal at the nth station, $\vec{w} \in \mathcal{C}^{M\times 1}$ is the noise vector, and $\mathbf{H} = [H_{mn}]_{1\leq m\leq M, 1\leq n\leq N} \in \mathbb{C}^{M\times N}$ is the compound channel between the N stations and the AP.

To decode the N data packets, the AP can first estimate the compound channel using the orthogonal pilots (a.k.a. reference signals) in the N data packets and then construct the MMSE MIMO detector as follows:

$$\mathbf{P} = \mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\sigma_w^2}{\sigma_x^2}\mathbf{I})^{-1}, \tag{5.16}$$

where **I** is an identity matrix of proper dimension, σ_x^2 is signal power, and σ_w^2 is noise power. After constructing the MMSE detector, the AP can perform MU-MIMO detection as follows: $\hat{\vec{x}} = \mathbf{P}\vec{y}$, where $\hat{\vec{x}}$ is an estimated copy of $\vec{x}$.

Conventional MU-MIMO detectors can work only when the data packets are well aligned in time. Roughly speaking, the time misalignment of data packets must be less than the time duration of an OFDM symbol's cyclic prefix [27]. For example, in 802.11ay, the time misalignment must be less than 36.36ns [5]. In real systems, this requirement is extremely hard to satisfy as many factors (e.g., propagation delays, digital processing delays, and clock jitters) contribute to the time misalignment. For this reason, uplink MU-MIMO is not standardized in IEEE 802.11ac (sub-6GHz) [4] and 802.11ay (60GHz) [5].[3]

[3] Note that 802.11ax is the only WLAN standard that supports uplink MU-MIMO. Yet, there is still no 802.11ax product that supports this feature. In addition, 802.11ax is more of a centralized rather than distributed network.

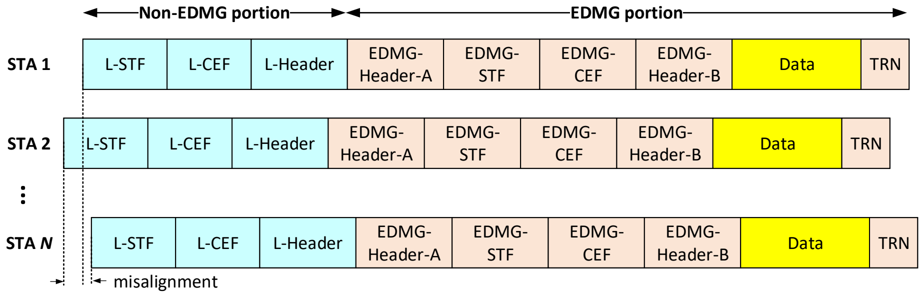


Figure 5.4: An illustration of the received asynchronous packets from multiple stations at the AP in an 802.11ay WLAN.

5.6.2: A Transformation of MMSE MU-MIMO Detector

Since it is hard to maintain the time alignment of the data packets for the AP, we wish to design an MIMO detector for the AP so that it can decode the misaligned data packets as shown in Fig. 5.4. In this case, if the AP knows the MMSE MIMO detector **P** in (5.16), then it should still be able to decode those asynchronous data packets. This is because **P** is a *spatial* filter and its effectiveness is not affected by the *temporal* imperfections (i.e., time misalignment of data packets). In other words, the spatial and temporal properties of data packets are orthogonal to each other. The key question here is how to obtain **P** when the AP receives asynchronous data packets. In synchronous MU-MIMO, the data packets from different stations carry orthogonal pilots for the AP to estimate the channel matrix **H**, based on which the AP can calculate **P** using (5.16). In asynchronous MU-MIMO, the data packets from different stations cannot maintain the orthogonality of their pilots. As a result, the AP cannot estimate the channel **H** and thus (5.16) does not work for this case.

To overcome this challenge, we show that a transformation of the MMSE detector in (5.16) can eliminate the need of channel knowledge **H** and obtain an approximation of **P**, which allows the AP to decode those asynchronous data packets separately. Denote $\mathcal{R}_n\{\cdot\}$ as the nth row of a matrix or a vector. Per the conventional MMSE detection, we have

$$\hat{x}_n = \mathcal{R}_n\{\hat{\vec{x}}\} = \mathcal{R}\{\mathbf{P}\vec{y}\} = \mathcal{R}_n\{\mathbf{P}\}\vec{y}\,. \tag{5.17}$$

Denote $\mathbf{R}_x$ as the correlation matrix of $\vec{x}$, i.e., $\mathbf{R}_x = \mathbb{E}[\vec{x}\vec{x}^{\mathsf{H}}]$. Denote $\mathbf{R}_w$ as the correlation matrix of $\vec{w}$, i.e., $\mathbf{R}_w = \mathbb{E}[\vec{w}\vec{w}^{\mathsf{H}}]$. In practice, signal and noise are always independent. Then, we have $\mathbf{R}_x = \sigma_x^2\mathbf{I}$ and $\mathbf{R}_w = \sigma_w^2\mathbf{I}$. Per (5.16), we have

$$\begin{aligned}
\mathcal{R}_n\{\mathbf{P}\} &= \mathcal{R}_n\big\{\mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\sigma_w^2}{\sigma_x^2}\mathbf{I})^{-1}\big\} \\
&\overset{(a)}{=} \mathcal{R}_n\big\{\mathbf{R}_x\mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{R}_x\mathbf{H}^{\mathsf{H}} + \mathbf{R}_w)^{-1}\big\} \\
&= \mathcal{R}_n\big\{\mathbb{E}[\vec{x}\vec{x}^{\mathsf{H}}]\mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbb{E}[\vec{x}\vec{x}^{\mathsf{H}}]\mathbf{H}^{\mathsf{H}} + \mathbb{E}[\vec{w}\vec{w}^{\mathsf{H}}])^{-1}\big\} \\
&= \mathbb{E}[\mathcal{R}_n\{\vec{x}\vec{x}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\}]\mathbb{E}[\mathbf{H}\vec{x}\vec{x}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + \vec{w}\vec{w}^{\mathsf{H}}]^{-1} \\
&= \mathbb{E}[\mathcal{R}_n\{\vec{x}\}\vec{x}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}]\mathbb{E}[\mathbf{H}\vec{x}\vec{x}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + \vec{w}\vec{w}^{\mathsf{H}}]^{-1} \\
&= \mathbb{E}[x_n(\mathbf{H}\vec{x})^{\mathsf{H}}]\mathbb{E}[(\mathbf{H}\vec{x} + \vec{w})(\mathbf{H}\vec{x} + \vec{w})^{\mathsf{H}}]^{-1} \\
&\overset{(b)}{=} \mathbb{E}[x_n(\mathbf{H}\vec{x} + \vec{w})^{\mathsf{H}}]\mathbb{E}[(\mathbf{H}\vec{x} + \vec{w})(\mathbf{H}\vec{x} + \vec{w})^{\mathsf{H}}]^{-1} \\
&= \mathbb{E}[x_n\vec{y}^{\mathsf{H}}]\mathbb{E}[\vec{y}\vec{y}^{\mathsf{H}}]^{-1}, \qquad (5.18)
\end{aligned}$$

where (a) and (b) follow from the assumptions that $\mathbf{R}_x$ is of full rank and $\mathbb{E}[x_n\vec{w}] = 0$, respectively. Both assumptions are always valid in practice.

Eq. (5.18) shows that the MMSE detector can be computed without channel knowledge, but using $\mathbb{E}[x_n\vec{y}^{\mathsf{H}}]$ and $\mathbb{E}[\vec{y}\vec{y}^{\mathsf{H}}]$. Now a question to ask is how to compute these two terms. In UMMC, we use the sample averaging operation to approach statistic expectation based on the fact that every packet in practical systems carries reference signals (a.k.a., pilots or preamble) for signal detection. Consider the 802.11ay frame shown in Fig. 5.4 for example. The reference signals include L-STF, L-CEF, EDMG-STF, and EDMG-CEF, which are pre-defined and known to all stations and APs. These reference signals will be used to compute $\mathbb{E}[x_n\vec{y}^{\mathsf{H}}]$ and $\mathbb{E}[\vec{y}\vec{y}^{\mathsf{H}}]$ in (5.18).

In the following, we slightly abuse the notation by introducing l as the index of OFDM symbol and k as the index of OFDM subcarrier. Denote $\mathcal{A}_n(k)$ as the set of reference symbols (pilots) in the data packet transmitted by station n on OFDM subcarrier k. Then, we have

$$\begin{aligned} \vec{p}_n(k) &\triangleq \mathcal{R}_n\big\{\mathbf{P}(k)\big\} \overset{(5.18)}{=} \mathbb{E}[x_n\vec{y}^{\mathsf{H}}]\mathbb{E}[\vec{y}\vec{y}^{\mathsf{H}}]^{-1} \\ &\triangleq \Big[\sum_{(l,k\prime)\in\mathcal{A}_n(k)} x_n(l,k\prime)\vec{y}(l,k\prime)^{\mathsf{H}}\Big]\Big[\sum_{(l,k\prime)\in\mathcal{A}_n(k)} \vec{y}(l,k\prime)\vec{y}(l,k\prime)^{\mathsf{H}}\Big]^{\dagger}, \end{aligned} \tag{5.19}$$

where $(\cdot)^{\dagger}$ is the pseudo-inverse operator, and $x_n(l,k\prime)$ and $\vec{y}(l,k\prime)$ represent the transmitted and received reference signal on OFDM symbol l and subcarrier $k\prime$, respectively.

With the MU-MIMO detector in (5.19), the AP decodes the data packet from station n as follows: $\hat{x}_n(l,k) = \vec{p}_n(k)^{\top}\vec{y}(l,k)$, where $\vec{y}(l,k)$ is the received payload signal vector at the AP and $\hat{x}_n(l,k)$ is its decoded payload signal from station n, $1 \le n \le N$.

5.6.3: Performance Analysis and Discussions

Performance Analysis. Since analyzing the performance of the proposed detector in general settings is extremely hard, we focus on an ideal case. Suppose that the number of reference signals (e.g., pilots in L-STF, L-CEF, EDMG-STF, and EDMG-CEF in Fig. 5.4) is greater than or equal to the number of stations, i.e., $|\mathcal{A}_n(k)| \ge N$. Then, we have the following lemma.

Lemma 1 *: If $M \ge N$ and $\sigma_w = 0$, the MU-MIMO detector in* (5.19) *can perfectly recover the misaligned signals from the asynchronous stations, i.e., $\hat{x}_n(l,k) = x_n(l,k)$ for $1 \le n \le N$, $1 \le k \le K$, and $1 \le l \le L$.*

Proof Sketch. We omit the subcarrier index k to simplify the notation. Given that $M \ge N$, **H** is a square or tall/thin matrix. Then, based on (5.19), we have:

$$\begin{aligned} \vec{p}_n &\overset{(a)}{=} \Big[\sum_{l\in\mathcal{A}_n} x_n(l)\vec{y}(l)^{\mathsf{H}}\Big]\Big[\sum_{l\in\mathcal{A}_n} \vec{y}(l)\vec{y}(l)^{\mathsf{H}}\Big]^{\dagger} \\ &\overset{(b)}{=} \Big[\sum_{l\in\mathcal{A}_n} x_n(l)\vec{x}(l)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\Big]\Big[\sum_{l\in\mathcal{A}_n} \mathbf{H}\vec{x}(l)\vec{x}(l)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\Big]^{\dagger} \\ &\overset{(c)}{=} \Big[\mathcal{R}_n\{\hat{\mathbf{R}}_{\mathbf{x}}\}\mathbf{H}^{\mathsf{H}}\Big]\Big[\mathbf{H}\hat{\mathbf{R}}_{\mathbf{x}}\mathbf{H}^{\mathsf{H}}\Big]^{\dagger} \\ &\overset{(d)}{=} \mathcal{R}_n\big\{\mathbf{H}^{\dagger}\big\}, \end{aligned} \tag{5.20}$$

where (a) follows from (5.19) by omitting the subcarrier index k; (b) follows from the fact that $\vec{y} = \mathbf{H}\vec{x}$ when $\sigma_w = 0$; (c) follows from our definition that $\hat{\mathbf{R}}_x = \sum_{l \in \mathcal{A}_n} \vec{x}(l)\vec{x}(l)^{\mathsf{H}}$; (d) follows from the fact that $\mathbf{H}$ is a square or tall matrix since $M \geq N$ and that $\hat{\mathbf{R}}_x$ is a square matrix of full rank since $|\mathcal{A}_n(k)| \geq N$. Based on (5.20), we have $\hat{x}_n(l,k) = \vec{p}_n(k)\vec{y}(l,k) = \mathcal{R}_n\{\mathbf{H}(k)^{\dagger}\}\vec{y}(l,k) = x_n(l,k)$. ■

In practice, the assumption of $|\mathcal{A}_n(k)| \geq N$ and $M \geq N$ are typically valid, but $\sigma_w \neq 0$. For the realistic case, we will evaluate this detector through experiments in Section 5.7.1.

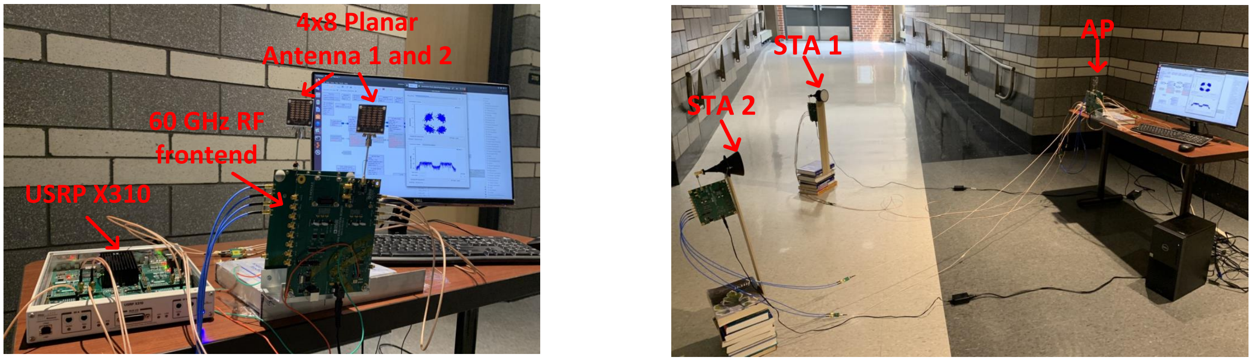
Explicit Channel Knowledge is Not Needed. It is evident that the MU-MIMO detector in (5.19) does not require explicit channel knowledge $\mathbf{H}$ for packet detection. Instead, it uses the reference signals in data packets to compute the detectors for each individual data stream. As such, this MU-MIMO detector is particularly suitable for an AP to decode asynchronous data packets, while the conventional MMSE detector is incapable of doing so.

Unique Features of mmWave MU-MIMO. MmWave communication systems are typically equipped with directional antennas (e.g., phased-array antenna), which significantly reduce the multipath effect of channels. As a result, the mmWave channels are more frequency-flat compared to sub-6GHz systems. In addition, compared to SISO mmWave WLANs (e.g., 802.11ad), MU-MIMO mmWave networks (e.g., 802.11ay) have pilots in both legacy preamble (L-STF and L-CEF) and enhanced preamble (EDMG-STF and EDMG-CEF); see Fig. 5.4. Lemma 1 shows that these two properties make the proposed asynchronous MMSE detector particularly suitable for 802.11ay networks.

5.7: Performance Evaluation

5.7.1: Experimental Results (Two-User MIMO Case)

Implementation. We built a 60 GHz mmWave MU-MIMO testbed that comprises an AP and two stations as shown in Fig. 5.5. The AP was built using two HMC6300 Boards (60 GHz RF Frontend) and one USRP X310. We modified the clock circuits of the two HMC6300 boards to synchronize their clock for MU-MIMO applications. The AP was equipped with two planar antennas, each of which has 4×8 patch elements. Each station was built using one HMC6300 Board and one



(a) Two-antenna AP.

(b) Experimental setup.

Figure 5.5: Illustration of our prototype and experimental setup.

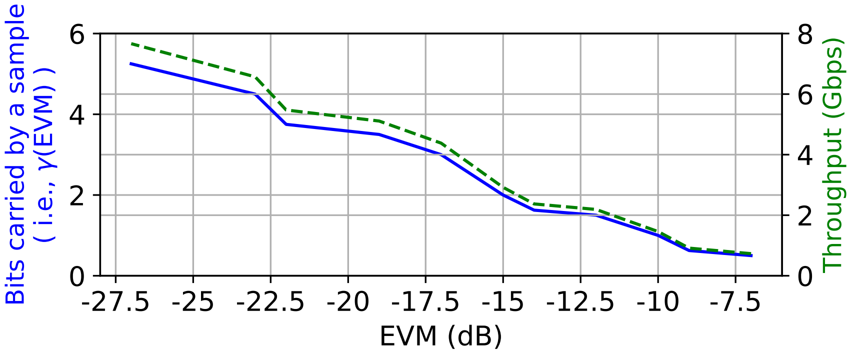
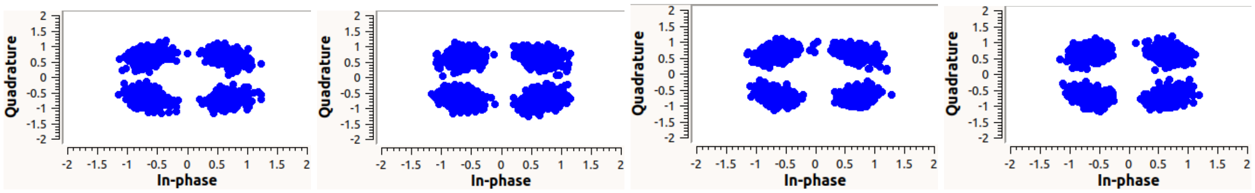


Figure 5.6: EVM specified in IEEE 802.11ay standard [5].

USRP X310, and connected with a horn antenna. The two stations worked independently, and there is no external clock to synchronize their packet transmissions. The instantaneous bandwidth of this MU-MIMO testbed is 100 MHz. We used GNURadio OTT (in C++) to implement the signal processing modules of a simplified 802.11ay PHY layer (512 FFT for OFDM modulation, QPSK, without LDPC codes) for the uplink MU-MIMO transmission. A demo video can be found in [29].

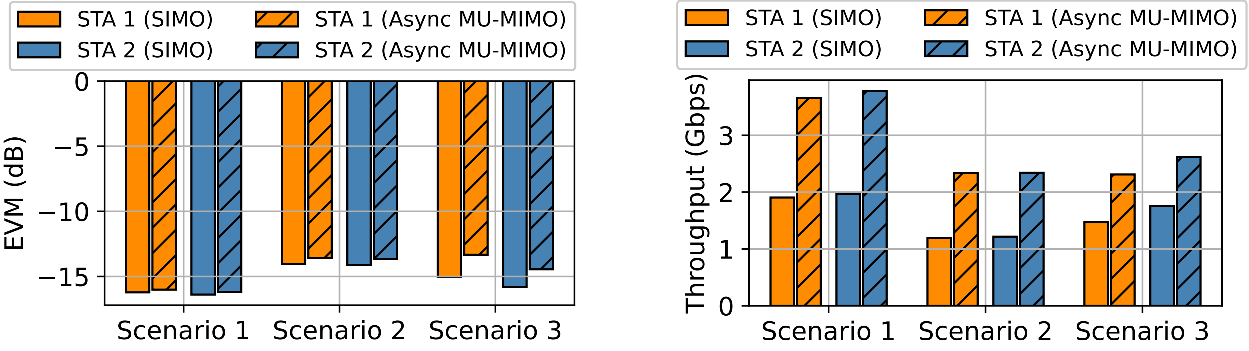
Experimental Setting. We consider three indoor scenarios for our experiments. Scenario 1: short distance (2.5m) for both stations. Scenario 2: long distance (5m) for both stations. Scenario 3: short distance (2.5m) for station 1 and long distance (5m) for station 2.

Performance Metrics. We use EVM and throughput as the performance metrics. EVM is widely used for the performance measurement of wireless receivers in industry. It was defined in (5.2). Based on the measured EVM, we calculate the throughput of 802.11ay networks as follows: $r_n = B \cdot \frac{\tau_{ofdm}}{\tau_{gi}+\tau_{ofdm}} \cdot \frac{N_{data}}{N_{fft}} \cdot \gamma(\text{EVM}_n)$, where $B = 2.64$GHz is the sampling rate, $\tau_{gi} = 36.36$ns is the



(a) STA 1's signal in MU-MIMO (b) STA 2's signal in MU-MIMO (c) STA 1's signal in SIMO (d) STA 2's signal in SIMO

Figure 5.7: Constellation diagram of the decoded signals at the AP.



(a) EVM (b) Throughput

Figure 5.8: Comparison of our proposed asynchronous MU-MIMO technique and conventional SIMO technique.

normal guard interval, $\tau_{ofdm} = 194.56$ns is the OFDM symbol duration, $N_{data} = 336$ is the number of subcarriers for data, $N_{fft} = 512$ is the FFT size, and $\gamma(\text{EVM}_n)$ is the adaptive rate specified by [5] and shown in Fig. 5.6. Recall that our objective is to maximize the minimum of user's throughput. Therefore, we denote $\text{EVM} = \max(\text{EVM}_1, \text{EVM}_2)$ and $\text{Throughput} = \min(r_1, r_2)$.

Asynchronous MU-MIMO Detection. We first validate the feasibility of the proposed asynchronous MU-MIMO detector on the testbed, where the two stations are continuously transmitting data packets but have no synchronization mechanism. For both AP and stations, we perform exhaustive search to find their best analog beams. Fig. 5.7(a-b) shows the two constellation diagrams observed at the AP. It is clear that the proposed detector is able to decode the data packets in the absence of inter-station synchronization.

As a comparison baseline, we also implemented the single-input and multiple-output (SIMO) transmission scheme on the testbed in the same settings. In this case, each station uses a half of the airtime for packet transmission in turn (i.e., TDMA mode). When serving each station, the

Table 5.2: Comparison of exhaustive *separate* beam search and exhaustive *joint* beam search.

	Scenario 1		Scenario 2		Scenario 3	
Search approach	Joint	Separate	Joint	Separate	Joint	Separate
Best angle for ant 1 (θ_1^*)	-30°	-45°	-30°	-13°	30°	-47°
Best angle for ant 2 (θ_2^*)	15°	-23°	-30°	-24°	30°	25°
EVM (dB)	-16.2	-13.0	-13.7	-10.0	-14.5	-10.1
Throughput (Gbps)	3.65	2.28	2.28	0.91	2.37	1.46

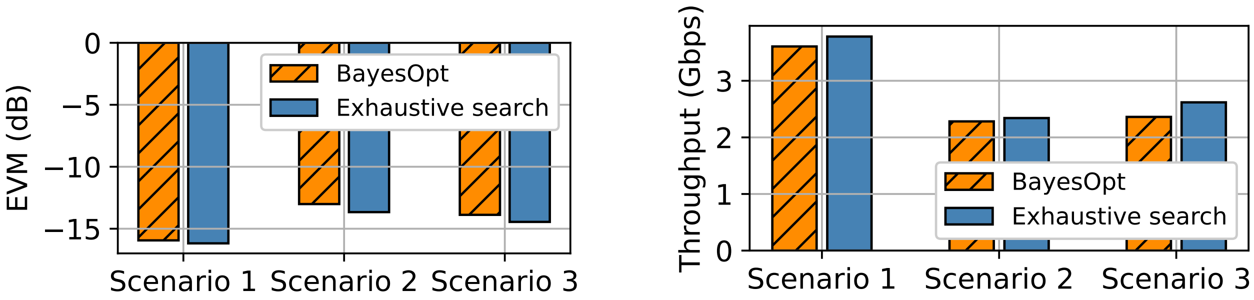
AP selects its best antenna to decode its data packets. Fig. 5.7(c-d) shows the two constellation diagrams observed at the AP. It can be seen that the AP observes similar constellation diagrams in the two cases. This reveals the effectiveness of our proposed MU-MIMO detector in decoding asynchronous data packets.

We repeated the above tests in all three scenarios to quantify the EVM and throughput of the two techniques (Async MU-MIMO and SIMO). Fig. 5.8(a) shows the EVM comparison. It shows that the two techniques have a similar EVM. This is a bit surprising, because in theory SIMO should offer a better (e.g., 3 dB) EVM performance than Async MU-MIMO. We conjecture that it was caused by the non-negligible phase noise of 60GHz mmWave RF devices. Phase noise increases linearly with carrier frequency in communication systems. When phase noise is strong, it dictates the communication performance and marginalizes the difference caused by other factors.

Fig. 5.8(b) shows the throughput comparison. It can be seen that Async MU-MIMO almost doubles the throughput of SIMO. This is because the AP can only serve the stations in turn in SIMO, while Async MU-MIMO allows the AP to serve both stations simultaneously.

Impact of MU-MIMO Channel Correlation. For the two antennas at the AP, we consider two approaches for their beam search: i) exhaustive *separate* search, and ii) exhaustive *joint* search. In the separate search, each individual antenna finds the beam angle that maximizes its signal strength. In the joint search, the two antennas try all possible beam combinations to find the one that maximizes the bottleneck of user data rates.

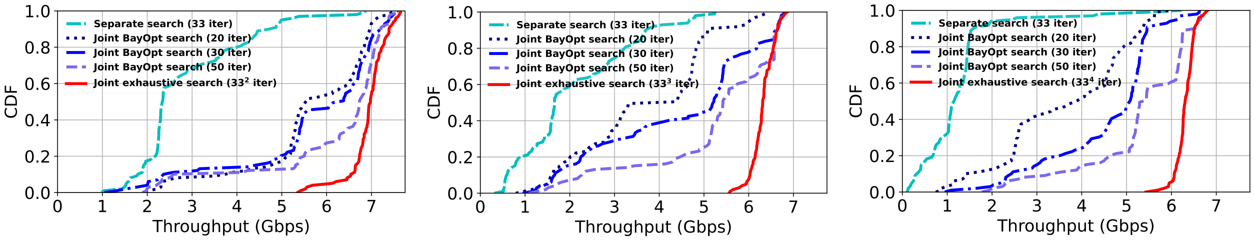
Table 5.2 shows our experimental results. It is clear that joint and separate search approaches lead to different beam results. Consider scenario 1 for example. When using separate beam search, the optimal angle is -45° for antenna 1 and -23° for antenna 2. This combination is optimal in terms



(a) EVM

(b) Throughput

Figure 5.9: Comparison of BayOpt and exhaustive search.



(a) Two-user MIMO case.

(b) Three-user MIMO case.

(c) Four-user MIMO case.

Figure 5.10: Throughput comparison of Separate beam search, BayOpt beam search, and exhaustive search.

of the signal strength at each individual antenna, but it is not optimal in terms of user throughput. For the joint search approach, the combination (-30°, 15°) yields the best EVM and thus the best user throughput. Similar phenomena can also be observed in scenarios 2 and 3. This confirms that signal strength is not a good criterion for beam search in MU-MIMO mmWave systems.

BayOpt Search versus Exhaustive Search. Using the proposed MU-MIMO detector, we compare two joint beam search approaches: *BayOpt search* and *exhaustive search*. For exhaustive search, we search the beams for each antenna every 5 degrees, and the search range is from -60° to 60°. So the total number of beam combinations for search is $(120/5+1)^2 = 625$. Fig. 5.3 shows an instance of exhaustive search results. For BayOpt search, we fix the number of search iterations (evaluations) to 20. Therefore, the overhead of BayOpt search is only 3.2% of the exhaustive search.

Fig. 5.9 shows the comparison of these two joint beam search approaches in three scenarios. It can be seen that BayOpt can achieve a similar EVM and throughput performance of exhaustive search. More accurately, BayOpt achieves 94.3% throughput of exhaustive search. It is important to point out that the throughput in Fig. 5.9(b) does not take into account the airtime overhead of beam training. If the beam training overhead is taken into consideration, BayOpt would easily outperform exhaustive search.

5.7.2: Simulation Results (More-User MIMO Case)

Due to the hardware limitation, we resort to simulations for the evaluation of BayOpt in more-user MIMO cases. We consider a 400ft^2 conference room where the AP is deployed on a wall and 100 users are uniformly and randomly distributed over the whole room. We use the model in [109] to calculate the path loss based on the distance between a user and the AP, and use the model in [129] to generate the gain of phased-array antennas for a given direction. In each time slot, the AP randomly selects N users for uplink MU-MIMO transmission, where $N \in \{2, 3, 4\}$ as defined in 802.11ay. In the simulations, we focus on the comparison of three different beam search approaches (separate exhaustive search, joint exhaustive search, and BayOpt) without considering packet misalignment issue. An ideal MMSE detector is used to decode concurrent packets and calculate their EVM and throughput.

We present the simulation results in Fig. 5.10. Compared to separate exhaustive search, BayOpt-30 (BayOpt with 30 iterations) has a similar airtime overhead (30 vs. 33 iterations), but it improves the throughput by 95.8%, 109.8%, and 267.2% in the two-user, three-user, and four-user cases, respectively. Compared to joint exhaustive search, BayOpt-50 achieves 88.6% throughput with 4.6% overhead in two-user MIMO case, 82.1% throughput with 0.1% overhead in three-user MIMO case, and 83.5% throughput with 0.004% overhead in four-user MIMO case. Note that the throughput presented in Fig. 5.10 does not take into account beam search overhead.

5.8: Summary

In this chapter, we presented a practical yet efficient uplink MU-MIMO communication (UMMC) scheme for mmWave networks. This scheme has two key components: BayOpt for beam search and asynchronous MU-MIMO detection. UMMC provides the first BayOpt framework for beam search in mmWave MU-MIMO systems, and introduced a new MU-MIMO detector that can decode asynchronous data packets from multiple users. It has demonstrated through both theory and experiments that fine-grained inter-user synchronization is not needed for uplink MU-MIMO transmission. We have evaluated the performance of UMMC through a blend of experiments and simulations. Experimental and simulation results confirm the practicality and efficiency of UMMC.

CHAPTER 6: TEMPORAL BEAM PREDICTION FOR MOBILE MMWAVE NETWORKS

6.1: Introduction

Millimeter wave (mmWave) technology is expected to serve as the foundation for 5G and future wireless networks, enabling the vision of a smart society and a digitized physical world. It offers ultra-low latency, multi-gigabit-per-second (Gbps), and scalable wireless connectivity for emerging applications such as virtual reality (VR), cloud-based real-time artificial intelligence (AI), and high-resolution video streaming [131]. In mmWave networks, devices commonly rely on analog beamforming to mitigate the effects of high path loss. In practice, a set of beam angles are predefined, and the analog beamforming operation for a mmWave device is equivalent to the selection of the best beam index that can maximize the signal strength at a receiver. When candidate beams are probed sequentially and exhaustively, the beam search procedure incurs significant airtime overhead. Unfortunately, most existing mmWave devices perform beam search in this exhaustive manner, which leads to poor scalability in environments with high user density [25].

To reduce the airtime overhead of beam selection, different approaches have been studied for the management of beam search, such as out-of-band CSI-assisted beam selection [21,64,118,148], compressive sensing [65, 127, 128], hierarchy beam search [24, 92, 157, 194], and learning-based beam search [23, 40, 58, 100, 121, 130]. These approaches have demonstrated a great success in the acceleration of beam selection and the reduction of its airtime overhead. However, most of prior efforts focus on the beam search optimization in a snapshot of networks by exploiting *spatial* features of mmWave channels. The exploitation of *temporal correlation* of mmWave channels for beam selection remains limited.

In this chapter, we exploit temporal channel correlation of a mobile mmWave device to predict its future beam direction based on its history beam selection profile, with the aim of reducing the airtime overhead of beam search in mobile mmWave networks. Specifically, we present a Temporal Beam Prediction (TBP) scheme for a mobile mmWave device to estimate its future beam direction based on its history beam selection profile. TBP was motivated by the recent success of pedestrian trajectory prediction [20, 60, 186], for which recurrent neural network (RNN) models have demonstrated a great potential for accurate prediction. TBP borrows the idea from pedestrian trajectory prediction by using LSTM [69] as the model for beam prediction in mobile mmWave networks. LSTM has proved its efficiency and effectiveness in capturing the dynamic pattern of beam directions by retaining information about past directions. The internal memory can encapsulate details pertaining to the environment and user mobility, thereby achieving accurate prediction of the future beam direction. Specifically, TBP asks each mmWave device to record its beam angles adopted by its past packets, and uses its past beam angle profile to predict the best beam angles for its current or next packet transmission. As expected, the predicted beam angle might not be the best. So a beam refinement algorithm is employed to find the best one through a local search.

Compared to trajectory prediction, beam prediction has two new challenges. First, the past data samples (history beam selection results) are non-uniform over time due to the bursty nature of data traffic. For example, the time interval of VoIP traffic ranges from 5 ms to 40 ms [137], depending on the voice intensity. The non-uniformity of data samples calls for a new LSTM model that can take into account data timestamp for the beam angle prediction. Second, unlike movement trajectory, the best-beam angle trajectory may have sharp changes due to the multipath effect of wireless channels and the imperfect radiation pattern of phased-array antennas. Consequentially, the best-beam angle trajectory is typically non-smooth over time, making it challenging to perform an accurate prediction.

To address these two challenges, TBP proposes a mobility-aware LSTM (mLSTM) model for the beam angle prediction. The novelty of mLSTM lies in a new structure of its cells, which takes both data samples (history beam angles) and their timestamps as the input to predict future beam angles. This is in sharp contrast to traditional LSTM, which does not consider the timestamps.

The inclusion of data timestamps makes it possible for the mLSTM model to extract the time-dependent features, which is critical for improving the beam prediction accuracy. In addition, TBP employs an adversarial learning structure to extract the user-independent features for the beam prediction. The combination of CNN-based feature extractor, mLSTM-based beam angle predictor, and adversarial-learning discriminator appears to be an efficient model for the temporal beam prediction in mobile mmWave networks.

In addition, a wireless device may be equipped with both mmWave and sub-6GHz radios for its communications. For such a case, we enhance the design of TBP by leveraging the out-of-band channel state information (CSI) from co-located sub-6GHz radio to improve the accuracy of mmWave beam prediction. The key challenge here stems from the heterogeneity of data samples from the two radios. Specifically, mmWave radio generates beam indices (i.e., beam angles), while sub-6GHz radio generates channel coefficients (or channel matrix). Per our experiments, simple concatenation of data samples from the two radios as the input for beam prediction yields an inferior performance. To address this challenge, TBP converts sub-6GHz CSI to the corresponding beam angles by exploiting their inherent spatial relations. The converted beam angles will then be combined with mmWave beam angles for training and inference. Such a CSI-assisted learning model is particularly useful for cases where a mmWave radio does not have sufficient data samples for prediction (e.g., when a mmWave radio just wakes up from sleep mode).

We have built a prototype of TBP on a software-defined radio (SDR) 60GHz mmWave testbed. The mmWave device is equipped with a planar antenna with 4×8 patch antenna elements, and installed on a building's ceiling. We evaluated the performance of TBP in four scenarios: lab, conference room, hallway and apartment. Experimental measurement shows that the average prediction error of TBP is less than 7 degrees for both beam azimuth and elevation angles in most of our studied cases. Experimental measurement also shows that the utilization of out-of-band CSI can further improve the beam prediction accuracy for a mmWave device. Based on the measurement results, we simulate the throughput gain of TBP in a mobile mmWave network with representative traffic settings. The simulation results show that TBP can improve the throughput by more than

60% compared to existing approaches in all four scenarios.

The contributions for this paper are summarized as follows.

- To the best of our knowledge, TBP is the first system-focused beam prediction scheme that exploits the *temporal* correlation of mmWave channels along a device's movement trajectory to reduce the airtime overhead of beam search.
- TBP proposes a deep-learning network structure with new LSTM cells, which is capable of accommodating non-uniform, non-smooth data samples for accurate prediction. It also leverages out-of-band CSI to improve the beam prediction accuracy.
- TBP has been evaluated on a 60GHz mmWave testbed. Extensive experimental measurement confirms its effectiveness of beam prediction.

6.2: Related Work

We review prior works in the following categories.

Learning-based Beam Search. Pioneering works have been done to predict the beam direction using deep learning. [23,58] studied the beamforming problem in highly-mobile mmWave systems. Their primary emphasis is on vehicular communications within outdoor scenarios featuring multiple base stations. They demonstrate the capability to predict the optimal beam by jointly analyzing the uplink signal received at these stations. However, it is noteworthy that the network traffic pattern for vehicular communications significantly differs from that of indoor scenarios. Vehicular environments involve high-speed mobility, but the trajectory dynamics are less pronounced compared to indoor scenarios. Consequently, these studies exclusively employ traditional deep neural networks without considering the time interval.

In the communication scenarios beyond vehicular contexts, [130] specifically targets indoor environments. This study takes into account the orientation of the user's device at each location, employing spatial features in 3D space to train the neural network and align the beam. DeepIA [40] trains a deep learning model to predict the best beam based on received signal strength (RSS) obtained from a subset of beams. However, this approach still necessitates scanning a subset of beams each time, resulting in a significant overhead for mobile millimeter networks. [100] is the most relevant work to our work. It proposed a deep learning-based beam tracking scheme for mobile mmWave devices. Unlike our work, which primarily relies on the user's moving trajectory to predict the optimal beam direction, this study focuses on estimating dynamic channels resulting from the user's minor motion. Additionally, it incorporates Inertial Measurement Unit (IMU) sensor measurements as additional input data. The works mentioned above do not consider practical issues such as non-uniform and non-smooth data samples, and they are evaluated through simulation. In contrast, TBP is a practical design and is evaluated via realistic experiments.

Deepbeam [121] is the only work that has been implemented on a testbed. It listens to the data transmission between the AP and other users, learning unique patterns from the in-phase-quadrature (I/Q) representation of the waveform. This allows it to predict the beam utilized by the transmitter and AoA on the receiver side. However, it relies heavily on other devices in the environment and primarily focuses on leveraging spatial features, a distinction from TBP. On the other hand, [73] improves this method by proposing a deep regularized waveform learning (DRWL) strategy. This approach demonstrates the ability to predict beams even with limited samples, showcasing advancements beyond its predecessor.

The work by [81] focuses on beam tracking in UAV-mmWave communication, modeling the problem as a multi-variable Gaussian process and using the Gaussian Process Machine Learning (GPML) method to address it. However, it is noteworthy that UAVs typically follow a pre-defined path with an existing LoS path for communication, distinguishing it from our approach.

Out-of-Band Assistance for Beam Search. Our work is also related to the research in this area. Table 6.1 compares our work with prior works. The works [21,64,118,148] harness the Wi-Fi band to infer the beam directions for mmWave communications. Most of them use linear antenna arrays for beam search and are limited to the beam search in a 2D plane. MUST [148] studied the beam search in 3D space using an array of three antennas. It achieves 71.2% prediction accuracy and 10° beam tracking error. [35] proposes a self-supervised deep learning approach, directly mapping the CSI from sub-6GHz to mmWave beams. TBP was inspired by these works but uses the out-of-band

Table 6.1: Out-of-Band beamforming for mmWave device.

Ref.	Out-of-Band	sensor or antenna array	2D/3D	Error range
BBS [118]	Wi-Fi band	8 antennas	2D	4°
[64]	sub-6GHz	4 antennas	2D	10°
MUST [148]	Wi-Fi band	3 antennas	3D	10°
[21]	sub-6GHz	8 antennas	2D	N/A
Listeer [62]	Light	light sensor array	3D	3.5°
[135]	Images	Camera	3D	N/A
TBP	Wi-Fi band	5 antennas	3D	7°

CSI in a different way. It converts the sub-6GHz CSI information to the best-beam azimuth and elevation angles based on history sub-6GHz CSI and mmWave beam angles.

Besides leveraging out-of-band information from the radio side, LiSteer [62] determines the mmWave beamforming direction by tracking indicator LEDs on APs. Since it relies on light resources, LoS is required for this solution. Taking a further step, [135] inputs images captured by the transmitter and receiver into a deep neural network, identifying beam directions based on these images. However, both of these methods require an additional sensor installation on the AP, which may not be suited for practical mmWave systems.

Compressive Sensing. In addition to out-of-band beamforming, compressive sensing technique has been studied for mmWave beamforming in order to reduce the beam search overhead [65, 127]. In [127], compressive sensing is directly applied for beamforming. But it relies on the accurate phase measurement. In practice, accurate phase information may not be available due to the existence of carrier frequency offset. Agile-Link [65] hashes the beam directions and utilizes a voting mechanism to recover the directions. It can identify the best path by tracking the change of energy across different bins in a logarithmic number of measurements. The works in this area focus on exploiting the spatial correlation of wireless channels to facilitate the beam search. In contrast, TBP exploits the temporal channel correlation for beam prediction.

Hierarchical Beam Search. To accelerate the beam search process, some works have formulated the beam search problem as an optimization problem [92, 157, 194]. Sophisticated algorithms have been developed to search the possible beams in a hierarchical manner so as to minimize the

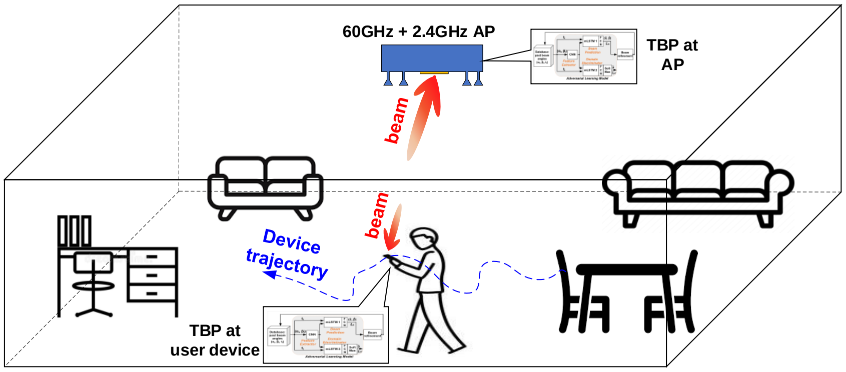


Figure 6.1: Illustration of temporal beam prediction (TBP) for both both mmWave AP and station (user device).

search time. In practice, the signals from different paths may cancel each other, leading to an inaccurate estimation of the power in the beam search process. However, these works are all based on simulation and without considering practical issues such as non-ideal antenna radiation and multipath effect. Clearly, TBP differs from this research line.

Model-based Beam Forecast. [214] studied model-based beam forecast by utilizing the spatial correlation of 60GHz near-field channels to predict the future channel when Tx/Rx moves. It relies on the anchor point channel profile to reconstruct the channel profiles for nearby points. However, when a user moves too far from the anchor point, the beam scanning will still be triggered.

6.3: Problem Description

We consider a mmWave communication network as shown in Fig. 6.1, where an access point (AP) is installed on the ceiling of a building to serve a set of stationary or mobile stations (user devices). To combat the high path loss, analog beamforming is adopted at both AP and station sides for signal energy steering. In practice, a set of beam angles are typically predefined for selection. As such, analog beamforming is equivalent to the selection of the best azimuth and elevation beam angles for a mobile device's phased-array antenna. In what follows, we first briefly introduce the beam search approaches in 802.11ad and 5G NR, and then present our design objective.

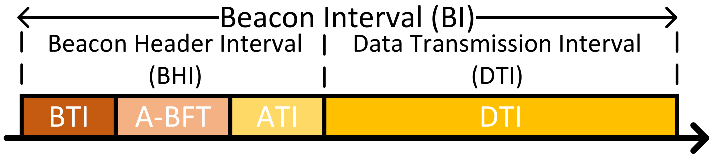

Beacon Interval (BI)
Beacon Header Interval (BHI)
Data Transmission Interval (DTI)
BTI
A-BFT
ATI
DTI

Figure 6.2: The structure of beacon interval (BI) in 802.11ad.

6.3.1: Preliminaries

Beam Search in IEEE 802.11ad/ay. IEEE 802.11ad [6] is a 60GHz mmWave communication standard. The beamforming training in 802.11ad comprises two phases: i) Sector Level Sweep (SLS), and ii) Beam Refinement Protocol (BRP). Fig. 6.2 shows the beacon interval (BI) structure in 802.11ad. The SLS takes place in the beacon header interval (BHI), while the BRP takes place in the data transmission interval (DTI). In the SLS phase, the user device configures its antenna to an omnidirectional radiation pattern, while the AP sweeps its beam over all possible directions. At the end of this process, the user device identifies the AP's best beam index and reports it to the AP. After identifying the AP's best beam index, a similar operation is performed on the user side to find its best beam index. More specifically, the AP uses its identified best beam index, and the user device sweeps its beam over all possible directions. In the end, the AP will find the user device's best beam index and send it to the user. SLS is mandatory in IEEE 802.11ad, and its beam training process takes about 1.54 ms for 7-degree beamwidth [118]. While the SLS phase is mandatory, the BRP phase is optional in 802.11ad. In this phase, the beam selected from SLS is refined through an iterative procedure. While the goal of SLS is to establish the connection between two devices at the control mode rate, the goal of BRP is to optimize devices' antenna settings by making use of the TRN field in a frame. It allows for multiple measurements in the same packet and thus enables coherent measurements, leading to a significant performance improvement compared to SLS.

Compared to 802.11ad, 802.11ay introduces various enhancements and new concepts that improve beam training and extend its support for new applications. Some of the training-related enhancements are highlighted as follows: i) a new beamforming procedure, called BRP Transmit Sector Sweep (TXSS), was introduced to improve the efficiency of beamforming; ii) first path beamforming training is defined to support positioning applications; iii) group beamforming is specified to reduce overhead by enabling the training for multiple stations simultaneously; and iv) a new BRP packet called EDMG BRP-RX/TX packet is defined to enable concurrent Tx and Rx beam training.

Beamforming in 5G mmWave Networks. MmWave communication (on 24–47 GHz) is a key component of 5G New Radio (NR) in order to increase the network throughput. The beam training procedure in 5G is similar to that in 802.11ad. The base station initiates the beam search process by sweeping its beam over all possible directions. At the end of this period, the user equipment identifies the best beam index for the base station and reports it to the base station. This process repeats with a smaller beamwidth at the base station, until the base station obtains its best beam index. After that, the base station fixes the beam direction and the user equipment sweeps over its beam angles to find the best one.

6.3.2: Design Objective

The airtime overhead of beam training in both 802.11ad/ay and 5G NR is $\mathcal{O}(N)$, where N is the number of beam candidates. While many schemes (e.g., [62,65,148,149,214]) have been proposed to reduce the overhead, most of them are limited to the exploitation of spatial-domain channel features in a snapshot of the network. Inspired by the pedestrian trajectory prediction [20, 60, 186], this paper focuses on the system-perspective design of efficient beam search strategies for a mobile mmWave device by exploiting the *temporal channel correlation* over its movement trajectory, aiming to reduce the airtime overhead of its beam training procedure.

6.4: TBP: Design

6.4.1: Overview

Fig. 6.3 shows the system architecture of TBP for a mobile mmWave device to predict its future beam angles based on its past beam selection results. The system has a database to record the past beam information, i.e., $(\alpha_i, \beta_i, t_i, s_i)$, where i is the data sample (beam) index ($i = 1, 2, \cdots, N$),

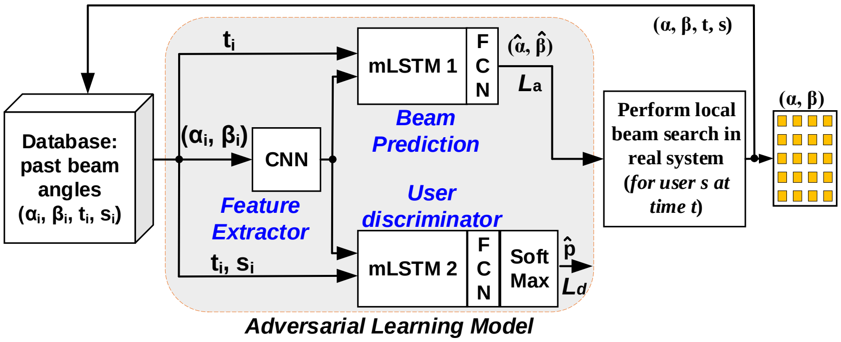

(α, β, t, s)
t_i
Database: past beam angles (α_i, β_i, t_i, s_i)
(α_i, β_i)
CNN
Feature Extractor
mLSTM 1
FCN
($\hat{\alpha}$, $\hat{\beta}$)
L_a
Beam Prediction
User discriminator
t_i, s_i
mLSTM 2
FCN
Soft Max
$\hat{p}$
L_d
Adversarial Learning Model
Perform local beam search in real system (for user s at time t)
(α, β)

Figure 6.3: The overview of TBP.

α_i is the beam azimuth beam angle, β_i is the elevation beam angle, t_i is the time moment when this data sample is generated, and s_i is the ID of the user device for whose data sample is generated. A CNN is used to extract the features of beam azimuth and elevation angles. Two mLSTM branches are adopted in the system. One is used for prediction, and the other is used as the user device discriminator for adversarial learning. The rationale behind this design is that we wish to extract the beam features that are independent of individual user devices, so that this design is generally applicable. The user device discriminator (i.e., mLSTM 2 in Fig. 6.3) is used for this purpose.

The predicted beam angles go through the *local* beam search procedure in real mmWave systems to find the optimal beam angles. The final selected beam is used by the antenna for signal steering and sent to the database for future use. The key components of TBP are highlighted as follows.

- *CNN for Feature Extraction:* As shown in Fig. 6.3, a CNN is used to extract features before sending data to the mLSTM modules. The CNN has 12 kernels with a size of 4, and it uses ReLU as the activation function.

  The kernel size and the number of kernels are adjusted based on the model's performance. Smaller kernel sizes excel at capturing fine-grained features, while larger ones are adept at capturing broader patterns. Given that TBP aims to capture dynamic patterns, the beam prediction is oriented towards the current moment; hence, larger kernel sizes might overlook the optimal beam direction at the present time. While a higher number of kernels enhances the feature space, it amplifies the model complexity. Based on our experimental observations,

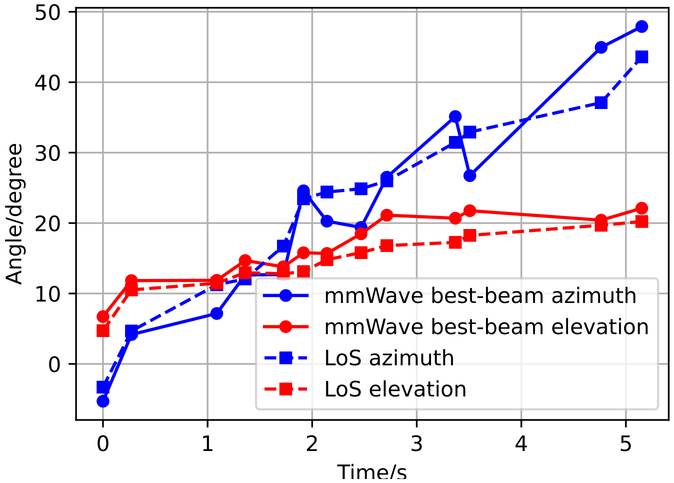


Figure 6.4: Experimental data that show the comparison between a mmWave AP's best-beam direction and its LoS direction when communicating with a station.

we employ 12 kernels to extract the features of the data. Besides, the causal padding is employed to avoid the time length changes.

- *mLSTM and Adversarial Learning:* Referring to Fig. 6.3, two identical mLSTM networks are used in TBP. These two mLSTM networks are structured for adversarial learning, following the architecture in [77]. The first mLSTM is for predicting the beam angles based on the history information, while the second mLSTM is used to predict the user device. We note that the training purpose is to enable the CNN to deceive the second mLSTM, thereby allowing the CNN to extract user-independent features. Both mLSTMs are connected to the fully connected layers. In addition, mLSTM 2 is added with a SoftMax layer after a fully-connected layer. The output of SoftMax is the possibilities of the devices by which the data samples are generated.
- *Local Beam Search in Real System:* Since the predicted beam may not be the best one, a local beam search module is employed to perform beam refinement in real systems. It follows the specified protocol, with the aim of finding the optimal beam for data transmission and therefore improving the efficiency of transmission. Details are given in Section 6.4.4.

6.4.2: mLSTM: Mobility-Aware LSTM

The prediction of mmWave beam azimuth and elevation angles has two unique challenges: *non-uniform* and *non-smooth* data samples over time. Fig. 6.4 shows our experimental results of comparison between the best-beam direction and the LoS direction in a lab scenario. It can be seen that the time intervals of consecutive data samples are not identical. This is because data traffic is bursty in nature. It can also be seen that the best beam angles may differ from and less smooth than LoS angles over time. This is caused by the multipath effect and non-ideal antenna radiation. To address these challenges, we propose an LSTM model for a mobile device to predict its beam angles based on its past beam selection profile.

RNN has been widely used for processing time-series sequential data. Connections between hidden units form a cycle which can send the past memory to the current cell. In this way, RNNs are particularly useful for dealing with the problem where the past memory has an strong effect to the current status. However, RNNs is incapable of capturing long-term dependence because of the vanishing and exploding gradient problems [69]. LSTM, a special case class in RNN, can handle the long-term memory without the problem mentioned above.

A traditional LSTM cell has four parts: forget gates, input gates, output gates, and a cell state. The input time-series data for the traditional LSTM cell is assumed uniformly distributed, disregarding the time intervals between samples. This means that the time gaps between consecutive samples are not taken into consideration during the data processing. However, the time series data is not always uniformly distributed especially when we consider communication requests. The communication frequency depends on the users' needs and also the type of data the user requested. The data from the mmWave band only obey the Poisson distribution.

Given the non-uniformity and non-smoothness of the beam angles over the device's movement trajectory, traditional LSTM models may not work well for the prediction (e.g., [102, 115]). This coincides with our experimental observations. To solve this problem, we propose a new mobility-aware LSTM (mLSTM) model for the beam prediction over time, as shown in Fig. 6.5. The input includes both data samples (past beam angles) and their timestamp. The output are the predicted

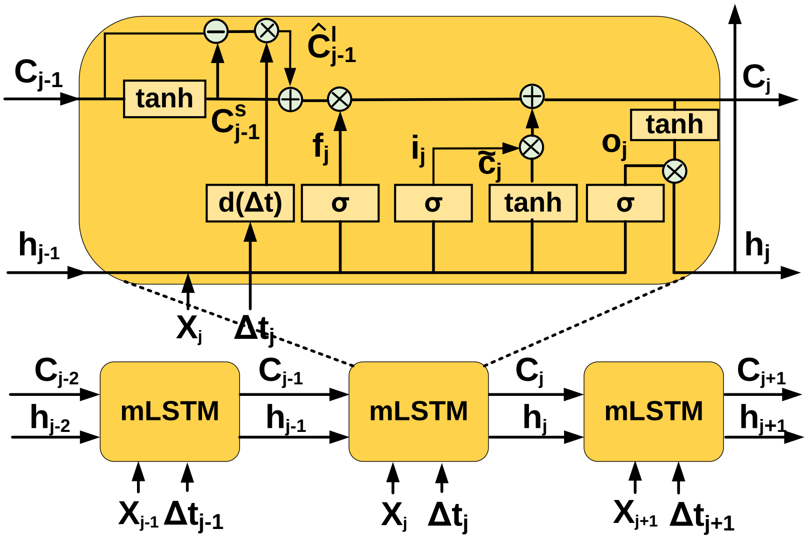


Figure 6.5: The structure of the mLSTM.

beam angles at a given time moment. In this mLSTM structure, the memory from previous time slot is first decomposed into short-term memory through a data driven method. Unlike the time-aware LSTM in [31], which discounts on the short-term memory, our mLSTM extracts the long-term memory and discounts on it. The intuition behind this operation is that, for the beam prediction problem, the short-term memory should carry higher weights for the prediction. In other words, the near-past beam samples should play a more important role in the prediction than the far-past beam samples. Still, the long-term memory is kept to capture the general moving tendency. Comparing with the traditional LSTM, the memory is adjusted with a larger amount of short-term memory and a fewer long-term memory. To discount the long-term memory, mLSTM utilizes a non-increasing function on the time interval and multiplies it with the long-term memory.

Fig. 6.5 shows the diagram of our proposed mLSTM, where each of its cells can be mathematically expressed as follows:

$$\mathbf{c}^{\mathrm{s}}_{j-1} = \mathrm{tanh}(\mathbf{W}_s \mathbf{c}_{j-1} + \mathbf{b}_s) \tag{6.1a}$$

$$\mathbf{c}^{\mathrm{l}}_{j-1} = \mathbf{c}_{j-1} - \mathbf{c}^{\mathrm{s}}_{j-1} \tag{6.1b}$$

$$\Delta t_j = t_j - t_{j-1} \tag{6.1c}$$

$$\hat{\mathbf{c}}^{\mathrm{l}}_{j-1} = \mathbf{c}^{\mathrm{l}}_{j-1} \odot d(\Delta t_j) \tag{6.1d}$$

$$\hat{\mathbf{c}}_{j-1} = \hat{\mathbf{c}}^{\mathrm{l}}_{j-1} + \mathbf{c}^{\mathrm{s}}_{j-1} \tag{6.1e}$$

$$\mathbf{i}_j = \sigma(\mathbf{W}_{xi} \mathbf{x}_j + \mathbf{W}_{hi} \mathbf{h}_{j-1} + \mathbf{b}_i) \tag{6.1f}$$

$$\mathbf{f}_j = \sigma(\mathbf{W}_{xf} \mathbf{x}_j + \mathbf{W}_{hf} \mathbf{h}_{j-1} + \mathbf{b}_f) \tag{6.1g}$$

$$\mathbf{o}_j = \sigma(\mathbf{W}_{xo} \mathbf{x}_j + \mathbf{W}_{ho} \mathbf{h}_{j-1} + \mathbf{b}_o) \tag{6.1h}$$

$$\tilde{\mathbf{c}}_j = \mathrm{tanh}(\mathbf{W}_{xc} \mathbf{x}_j + \mathbf{W}_{hc} \mathbf{h}_{j-1} + \mathbf{b}_c) \tag{6.1i}$$

$$\mathbf{c}_j = \mathbf{f}_j \odot \hat{c}_{j-1} + \mathbf{i}_j \odot \tilde{\mathbf{c}}_j \tag{6.1j}$$

$$\mathbf{h}_j = \mathbf{o}_j \odot \mathrm{tanh}(\mathbf{c}_j), \tag{6.1k}$$

where $\mathbf{x}_j$ is the input data, $\mathbf{h}_{j-1}$ is the previous hidden state, and $\mathbf{c}_{j-1}$ is the previous memory. $\mathbf{W}_{xi}, \mathbf{W}_{hi}, \mathbf{b}_i$ are the parameters for input gate. $\mathbf{W}_{xf}, \mathbf{W}_{hf}, \mathbf{b}_f$ are the parameters for forget gate. $\mathbf{W}_{xo}, \mathbf{W}_{ho}, \mathbf{b}_o$ are the parameters for output gate. $\mathbf{W}_{xc}, \mathbf{W}_{hc}, \mathbf{b}_c$ are the parameters for candidate memory. $\mathbf{c}^{\mathrm{s}}_{j-1}$ is the short-term memory, which is decomposed from previous memory cell $\mathbf{c}_{j-1}$; $\mathbf{W}_s$ and $\mathbf{b}_s$ are new weight matrix and bias vector defined for this operation, respectively. $\mathbf{c}^{\mathrm{l}}_{j-1}$ is the long-term memory, and $\hat{\mathbf{c}}^{\mathrm{l}}_{j-1}$ is the discounted long-term memory. Here, the non-increasing function being used is $d(\Delta t) = 1/\Delta t$. $\hat{\mathbf{c}}_{j-1}$ is the final adjusted memory which includes the full short-term memory and the discounted long-term memory. It is used to update the memory. This mLSTM is the fundamental building block for the design of TBP.

6.4.3: Automated Training Process

Data Collection Automation. A mmWave device first works in the traditional beam training mode (e.g., using 802.11ad or 5G NR beam training protocol [7,56]) to collect data samples for the training of TBP. A data sample can be denoted as $(\alpha_i, \beta_i, t_i, s_i)$, where i is the sample index. After sufficient data samples have been collected, TBP starts to train its models. As shown in Fig. 6.3, after TBP completes its model training and enters its inference phase, it will still add data samples to its database, which can be further used to train its model if necessary. It should be noted that the training process will not disrupt the normal communications of a mmWave device, as the data samples are side information from standard-compatible mmWave communication.

Training of TBP. Following the structure in Fig. 6.3, mLSTM 1 is trained to minimize the following loss function: $\mathcal{L}_{\mathrm{a}} = \frac{1}{N}\sum_{i=1}^{N}[(\hat{\alpha}_i - \alpha_i)^2 + (\hat{\beta}_i - \beta_i)^2]$, where (α_i, β_i) are the beam angles of a data sample, $(\hat{\alpha}_i, \hat{\beta}_i)$ are the prediction results (see Fig. 6.3), and N is the total number of training samples in this mini-batch. Denote $\mathbf{W}_{\mathrm{a}}$ as the weights of mLSTM 1. Then, they are updated as follows: $\mathbf{W}_{\mathrm{a}} \leftarrow \mathbf{W}_{\mathrm{a}} - \mu_{\mathrm{a}}\frac{\partial \mathcal{L}_{\mathrm{a}}}{\partial \mathbf{W}_{\mathrm{a}}}$, where μ_{a} is the update step size ($\mu_{\mathrm{a}} = 0.01$ in our experiments).

mLSTM 2 serves as a user device discriminator. It shares the identical structure of mLSTM 1. Denote $\vec{p_i}$ as the output probability vector of the SoftMax layer when the input data sample is $(\alpha_i, \beta_i, t_i, s_i)$. We define its loss function as: $\mathcal{L}_{\mathrm{d}} = -\frac{1}{N}\sum_{i=1}^{N}\log(g(\vec{p_i}, s_i))$, where $g(\vec{p_i}, s_i)$ returns the element of $\vec{p_i}$ that corresponds to user device s_i. Based on this loss function, the weights of mLSTM 2 are updated by: $\mathbf{W}_{\mathrm{d}} \leftarrow \mathbf{W}_{\mathrm{d}} - \mu_{\mathrm{d}}\frac{\partial \mathcal{L}_{\mathrm{d}}}{\partial \mathbf{W}_{\mathrm{d}}}$, where μ_{d} is the update step size ($\mu_{\mathrm{d}} = 0.01$ in our experiments).

The training of CNN has two purposes: i) minimizing the prediction loss $\mathcal{L}_{\mathrm{a}}$, and ii) maximizing the domain discrimination loss $\mathcal{L}_{\mathrm{d}}$. We define the combined loss function as follows: $\mathcal{L}_{\mathrm{e}} = \gamma\mathcal{L}_{\mathrm{a}} - \mathcal{L}_{\mathrm{d}}$, where γ is a tuning parameter ($\gamma = 0.2$ in our experiments). Based on this loss function, the weights of CNN are updated by: $\mathbf{W}_{\mathrm{e}} \leftarrow \mathbf{W}_{\mathrm{e}} - \mu_{\mathrm{e}}\frac{\partial \mathcal{L}_{\mathrm{e}}}{\partial \mathbf{W}_{\mathrm{e}}}$, where μ_{e} is the update step size ($\mu_{\mathrm{e}} = 0.01$ in our experiments). As a feature extractor, the CNN tries to cheat the user discriminator by maximizing its loss function $\mathcal{L}_{\mathrm{d}}$ while improving the performance of beam prediction by minimizing the loss function $\mathcal{L}_{\mathrm{a}}$. With this adversarial learning structure, TBP tends to extract the user-independent features for the beam prediction.

6.4.4: Prediction with Local Beam Search

After the model is trained, it then can be used for beam prediction based on the past beam samples in the database. In the inference phase, the predicted beam angles (the output of mLSTM 1) may or may not be accurate enough for packet transmission. Hence, TBP performs a *local* beam search, with the aim of finding the best beam angle for signal steering. Suppose that a mmWave device has a set of pre-defined beam azimuth angles $\{\alpha_p : 1 \leq p \leq N_p\}$ and a set of pre-defined beam elevation angles $\{\beta_q : 1 \leq q \leq N_q\}$. Also, recall that $(\hat{\alpha}, \hat{\beta})$ are prediction results of our model, i.e., the output of mLSTM 1 in Fig. 6.3. Then, the task of beam refinement module can be formulated as: $(p^*, q^*) = \text{arg max}_{p,q} f(\alpha_p, \beta_q)$, subject to $|\alpha_p - \hat{\alpha}| \leq \tau_\alpha$ and $|\beta_q - \hat{\beta}| \leq \tau_\beta$, where τ_α and τ_β are the thresholds for azimuth and elevation angles, respectively. $f(\alpha_p, \beta_q)$ is the resulting signal strength at receiver when the transmitter uses (α_p, β_q) as the azimuth/elevation beam angles. To find the optimal beam direction, we can perform the beam probing protocols in Section 6.3.1. Since TBP only needs to perform a local search, its airtime overhead of beam training is much less than that of existing beam training schemes.

6.5: TBP: Out-of-band Enhancement

MmWave communication systems feature high bandwidth, small coverage, and susceptibility to blockage. Thus, it is expected that mmWave communication systems will coexist with sub-6GHz Wi-Fi systems as they complement each other. In this section, we consider an indoor wireless communication network where each device is equipped with both mmWave and sub-6GHz Wi-Fi radios. We aim to take advantage of widely-available sub-6GHz Wi-Fi CSI to enhance the beam prediction for mmWave radio. This design is particularly useful for the case where mmWave radio is in the sparsity of past data samples for beam prediction (e.g., mmWave radio just wakes up from a sleep mode, mmWave radio is inactive for a long time). Although the literature has many works on out-of-band beamforming [21, 62, 64, 118, 135, 148], their focus is mainly on simplifying beam search in the spatial domain. Here, TBP focuses on the temporal prediction of beam angles.

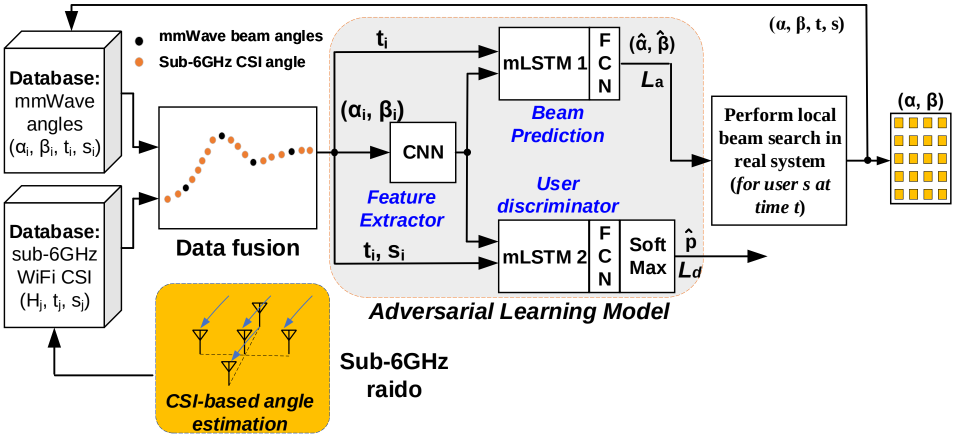

mmWave beam angles
Sub-6GHz CSI angle
Database: mmWave angles (αi, βi, ti, si)
Database: sub-6GHz WiFi CSI (Hj, tj, sj)
Data fusion
CSI-based angle estimation
Sub-6GHz raido
ti
(αi, βi)
CNN
Feature Extractor
ti, si
mLSTM 1
F C N
(α̂, β̂)
La
Beam Prediction
User discriminator
mLSTM 2
F C N
Soft Max
p̂
Ld
Adversarial Learning Model
(α, β, t, s)
Perform local beam search in real system (for user s at time t)
(α, β)

Figure 6.6: The structure for TBP when taking into account sub-6GHz CSI for beam angle prediction.

6.5.1: Design

Fig. 6.6 shows the overall structure of TBP for the case when sub-6GHz CSI is available for beam prediction. Compared to the learning model in Fig. 6.3, the only difference is that it adds data from sub-6GHz radio for its training and inference. As shown in Fig. 6.6, the data sample from mmWave radio is denoted as $(\alpha_i, \beta_i, t_i, s_i)$, while the data sample from sub-6GHz radio is denoted as $(\mathbf{H}_j, t_j, s_i)$. Apparently, the data samples from the two radios are in very different format.

An important question is how to combine the data from the two radios for training and inference. For this question, a straightforward method is concatenation, i.e., feeding all raw data to the CNN and letting the CNN to extract the useful features. However, this method did not perform well in our experiments. We guess the reason is that the CNN is incapable of extracting meaningful features from the heterogeneous data samples. To address this problem, we could unify the data format from the two sources by converting the sub-6GHz CSI to the azimuth and elevation angles of the LoS path (see Fig. 6.7 for example), and then combine the best beam azimuth/elevation angles and the LoS azimuth/elevation angles as the input of CNN. While this method performs better than the previous method, its performance still remains unsatisfactory in our experiments.

It could be attributed to two reasons. The first one lies in the fact that the best-beam direction may differ from the LoS direction. Through a careful study of the CNN model's input and out-

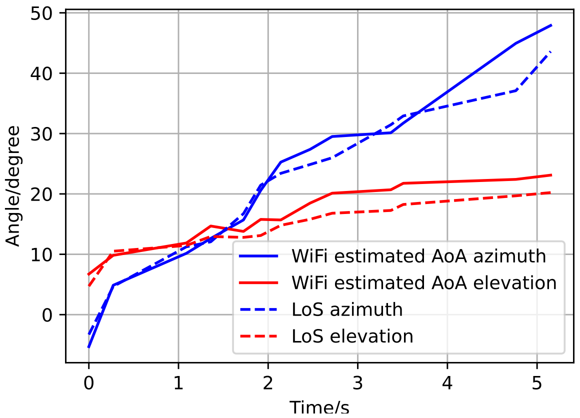


Figure 6.7: The continuous angle-of-arrival (AoA) estimation from Wi-Fi band and ground truth measured by laser meter.

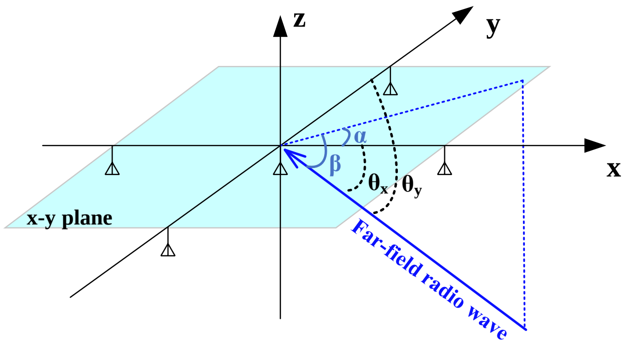


Figure 6.8: Illustration of signal angles (α, β, θ_x, and θ_y in 3D space.

put data, we found that in a large portion of input data, there is a discrepancy between the LoS direction (calculated from sub-6GHz CSI) and the best-beam direction (obtained from mmWave search). This is not surprising. In practice, the best-beam direction deviates from the LoS direction due to the imperfect radiation pattern of patch antennas in mmWave systems and the presence of strong Non-LoS paths. The second reason is that CNN may not be capable of differentiating the best-beam direction from the LoS direction during training and inference. Due to the discrepancy between LoS and best-beam directions within the training data, the model faces challenges in distinguishing the best-beam direction from the LoS direction in the inference phase. This is because the direct-merging method depends solely on time alignment, without utilizing the best-beam direction obtained through mmWave search to correct the corresponding sub-6GHz LoS direction. Furthermore, the larger amount of sub-6GHz CSI data significantly diminishes the impact of the best-beam data generated from the mmWave search.

Based on the above observations, we propose a new method to convert sub-6GHz CSI to its azimuth and elevation angles. It comprises two steps: i) convert CSI to LoS azimuth/elevation angles, and ii) estimate the corresponding best-beam azimuth/elevation angles based on the LoS azimuth/elevation angles. The details are presented in the next subsection. After merging the data from the two radios, the system model is trained and operated in the same way as presented in Section 6.4.

6.5.2: Data Fusion

We consider the case where the sub-6GHz radio is co-located with mmWave radio. We assume that sub-6GHz radio is equipped with multiple equally-spaced antenna elements along both x and y axes as illustrated in Fig. 6.8. To unify the data format for training and inference, we convert the sub-6GHz CSI data (i.e., $(\mathbf{H}_i, t_i)$) to corresponding azimuth/elevation angles (i.e., (α_i, β_i, t_i)). Fig. 6.9 shows the diagram of our data conversion method. It comprises two steps:

Step 1: *Identifying anchor data samples.* In this step, we find those sub-6GHz CSI data samples that coincide with a mmWave data sample in time. In Fig. 6.9, CSI data $(\mathbf{H}_i, t_i)$ and $(\mathbf{H}_{i'}, t_{i'})$ are two examples showing the coincidence. Then, the corresponding azimuth/elevation angles of these two samples can be found, as shown in the figure. These CSI data samples will be used as anchors to calculate the azimuth/elevation angles for the rest of CSI data samples. In practice, as long as the time gap between the CSI data sample and mmWave sample is less than 1 ms, we consider them in coincidence.

Step 2: *Converting the resting CSI data samples.* The data conversion process is illustrated through the example of $(\mathbf{H}_j, t_j)$ in Fig. 6.9. We explain this process by the following three steps.

① **Calculate $(\theta_{\text{x},i}, \theta_{\text{y},i})$ for Reference Data Sample:** Consider an incoming signal in the 3D space as shown in Fig. 6.8. $\theta_{\text{x},i}$ is the angle-of-arrival (AoA) of incoming signal for the linear antenna array on x-axis, while $\theta_{\text{y},i}$ is the AoA of incoming signal for the linear antenna array on y-axis. Then, the relation between azimuth/elevation angles (α_i, β_i) and signal AoA $(\theta_{\text{x},i}, \theta_{\text{y},i})$ can be expressed as:

$$\theta_{\text{x},i} = \cos^{-1}(\cos(\alpha_i)\cos(\beta_i)), \tag{6.2}$$

$$\theta_{\text{y},i} = \cos^{-1}(\sin(\alpha_i)\cos(\beta_i)). \tag{6.3}$$

② **Calculate $(\theta_{\text{x},j}, \theta_{\text{y},j})$ for Data Sample j:** We first focus on the calculation of $\theta_{\text{x},j}$ using the

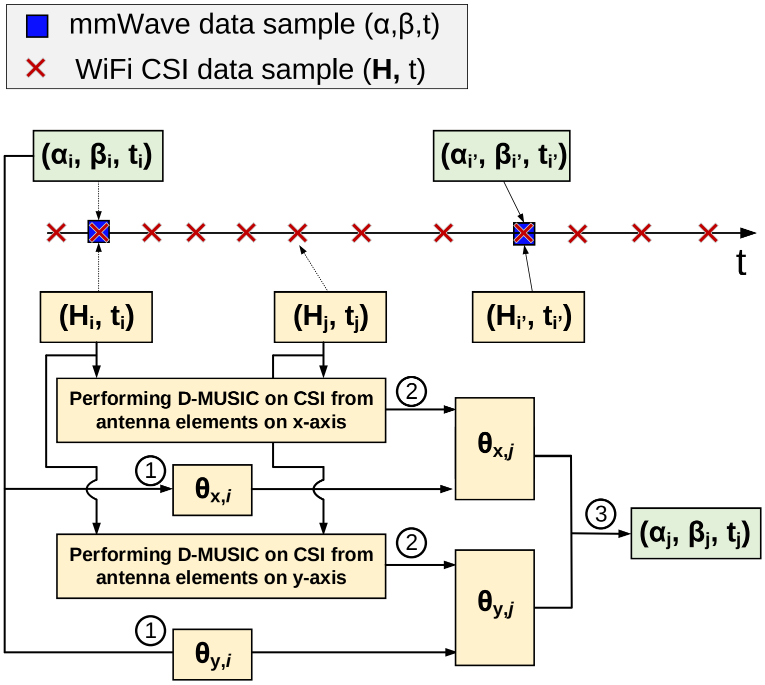

■ mmWave data sample (α,β,t)
× WiFi CSI data sample (**H**, t)
(α_i, β_i, t_i)
$(\alpha_{i'}, \beta_{i'}, t_{i'})$
t
(H_i, t_i)
(H_j, t_j)
$(H_{i'}, t_{i'})$
Performing D-MUSIC on CSI from antenna elements on x-axis
2
$\theta_{x,j}$
1
$\theta_{x,i}$
3
(α_j, β_j, t_j)
Performing D-MUSIC on CSI from antenna elements on y-axis
2
$\theta_{y,j}$
1
$\theta_{y,i}$

Figure 6.9: The diagram of data fusion.

antenna elements on x axis as shown in Fig. 6.8. The calculation of $\theta_{\mathrm{y},j}$ will follow the same token using the antenna elements on y axis. Consider antenna k on x axis. The additional phase shift of its received signal with respect to the first antenna can be written as:

$$\phi_{\mathrm{x},k} = 2\pi\frac{(k-1)d}{\lambda}\cos(\theta_{\mathrm{x}}), \tag{6.4}$$

where λ is the wavelength, d is the antenna spacing, θ_{x} is the AoA of incoming signal.

Denote $\phi_{\mathrm{x},k,i}$ and $\phi_{\mathrm{x},k,j}$ are the additional phase shift on antenna k at time t_i and t_j, respectively. Then, we have

$$\phi_{\mathrm{x},k,j} - \phi_{\mathrm{x},k,i} = 2\pi\frac{(k-1)d}{\lambda}\left(\cos(\theta_{\mathrm{x},j}) - \cos(\theta_{\mathrm{x},i})\right). \tag{6.5}$$

Based on (6.5), we further have

$$\cos(\theta_{\mathrm{x},j}) - \cos(\theta_{\mathrm{x},i}) = \frac{\lambda}{2\pi(k-1)d}\left(\phi_{\mathrm{x},k,j} - \phi_{\mathrm{x},k,i}\right). \tag{6.6}$$

Denote $\mathbf{H}_{\mathrm{x},j} = [H_{\mathrm{x},1,j},\ H_{\mathrm{x},2,j},\ \cdots, H_{\mathrm{x},K,j}]$ as the channel vector measured on the antenna elements on x axis at time t_j. In (6.6), $\phi_{\mathrm{x},k,j}$ is a component of the phase of channel coefficient $H_{\mathrm{x},k,j}$; and $\phi_{\mathrm{x},k,i}$ is a component of the phase of channel coefficient $H_{\mathrm{x},k,i}$. It can be seen that the estimation problem in (6.6) is similar to the classic AoA estimation problem. Therefore, the left-hand side of (6.6) can be estimated using MUSIC algorithm with the input of $\mathbf{H}_{\mathrm{x},j} \odot (\mathbf{H}_{\mathrm{x},i})^*$, where $\odot$ is element-wise product and $(\cdot)^*$ is conjugate operator. Mathematically, we have

$$\cos(\theta_{\mathrm{x},j}) - \cos(\theta_{\mathrm{x},i}) = \text{MUSIC}\left(\mathbf{H}_{\mathrm{x},j} \odot (\mathbf{H}_{\mathrm{x},i})^*\right), \tag{6.7}$$

where $\mathbf{H}_{\mathrm{x},j}$ and $\mathbf{H}_{\mathrm{x},i}$ are measured channels from the antenna elements on x axis at time t_j and t_i, respectively.

Based on (6.7), we have

$$\cos(\theta_{\mathrm{x},j}) = \cos(\theta_{\mathrm{x},i}) + \text{MUSIC}\left(\mathbf{H}_{\mathrm{x},j}, \mathbf{H}_{\mathrm{x},i}\right), \tag{6.8}$$

By the same token, we have

$$\cos(\theta_{\mathrm{y},j}) = \cos(\theta_{\mathrm{y},i}) + \text{MUSIC}\left(\mathbf{H}_{\mathrm{y},j}, \mathbf{H}_{\mathrm{y},i}\right). \tag{6.9}$$

③ **Calculate** (α_j, β_j, t_j) **for Data Sample** j**:** Given $\cos(\theta_{\mathrm{x},j})$ in (6.8) and $\cos(\theta_{\mathrm{y},j})$ in (6.9), we can calculate the desired (α_j, β_j) as follows:

$$\alpha_j = \tan^{-1}\left(\frac{\cos(\theta_{\mathrm{y},j})}{\cos(\theta_{\mathrm{x},j})}\right), \tag{6.10}$$

$$\beta_j = \cos^{-1}\left(\frac{\cos(\theta_{\mathrm{x},j})}{\cos(\alpha_j)}\right). \tag{6.11}$$

This completes the conversion from $(\mathbf{H}_j, t_j)$ to (α_j, β_j, t_j).

6.5.3: Training and Inference

The only purpose of out-of-band CSI from sub-6GHz radio is to enrich the dataset for the temporal mmWave beam prediction. It does not alter the training and inference procedure of TBP. That said, the training and inference operations in this case are the same as those in Section 6.4.

6.6: Performance Evaluation

6.6.1: Implementation

60 GHz mmWave Testbed. We built a mmWave testbed for the evaluation of TBP using EK1HMC6350 RF front-ends from Analog Devices. Fig. 6.10 shows the overall diagram of our testbed; and Fig. 6.11 shows a picture of the mmWave board. HMC6300 supports carrier frequency from 57GHz to 64GHz, and the bandwidth of each channel is 1.8GHz. Two planar antennas, each with 4×8 patch elements, are used for this testbed. One is for transmitter, and the other is for receiver. Since the planar antennas cannot steer its beam electronically; two stepper motors are installed to control the beam direction in the 3D space. One stepper motor controls the beam's azimuth angle; the other controls the beam's elevation angle. The angle resolution of stepper motors is 1.8 degree. The two stepper motors are controlled by the host computer via its USB interface. The mmWave radio RF front-end is connected to USRP X310 through baseband I/Q differential interface, and USRP X310 is then connected to a high-performance computer through 10Gbps SFP+ cable. In our experiments, USRP X310 is installed with BasicTx and BasicRx daughterboards to generate baseband signals. All signal processing modules were implemented in the host computer to measure the signal strength at receiver. Overall, the mmWave testbed can support 100MHz instantaneous bandwidth for real-time communication and 2-dimensional beam steering.

Sub-6GHz Radio for CSI Acquisition. As shown in Fig. 6.10, a 5-antenna SDR receiver was built using an array of synchronized USRP N210 devices to obtain the CSI for the evaluation of TBP when CSI is taken into account. The 5 omnidirectional antennas are deployed in a cross shape with 3 antennas on x axis and 3 on y axis. The sub-6GHz system implements commodity 802.11

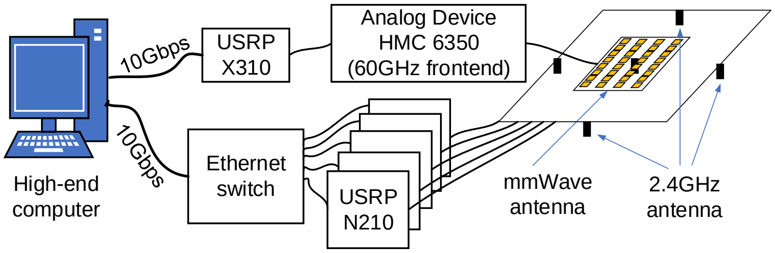


Figure 6.10: The diagram of our testbed installed on the ceiling of an office.

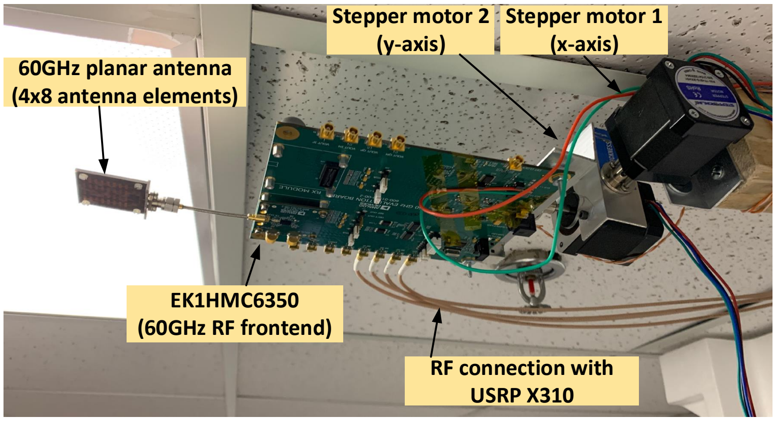

Stepper motor 2 (y-axis)
Stepper motor 1 (x-axis)
60GHz planar antenna (4x8 antenna elements)
EK1HMC6350 (60GHz RF frontend)
RF connection with USRP X310

Figure 6.11: The testbed installed on the ceiling of a lab.

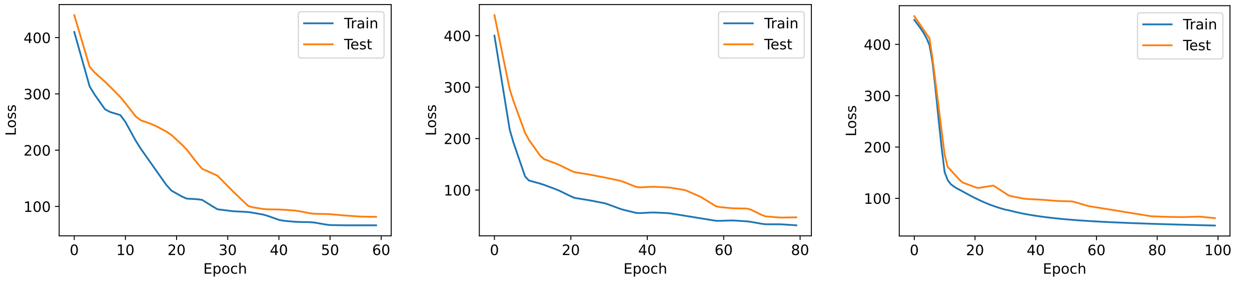
protocol with a bandwidth of 20MHz.

TBP Implementation. We implement TBP in the host computer as shown in Fig. 6.10 using TensorFlow [15]. The database records the past beam selection information, which will be used for the beam prediction over time. The data collection is automated using the beam angles from the local beam search module in Fig. 6.3.

6.6.2: Experimental Settings

Our experiments were conducted in four scenarios: a lab, a conference room, a hallway and an apartment. The lab is 220 ft^2, with typical cubicles and furniture. The conference room is about 170 ft^2, with a big table and multiple chairs. The hallway is relatively large and empty with a few display cases near the wall. The apartment is about 260 ft^2, furnished with common items such as a tea table, chairs, sofa, and TV. In each scenario, the mmWave radio was installed on the ceiling, communicating with a mmWave device carried by six different persons on the floor.

In each scenario, six persons walked along their routing paths sequentially, and 1,920 trace



(a) Single-user case.

(b) Single-user with out-of-band CSI.

(c) Multi-user case.

Figure 6.12: Training and test loss for TBP in lab scenario.

samples were collected in total. 30% of the trace data are randomly selected and used for testing purposes. Along the routing path, the beam angle samples are recorded in irregular time intervals varying from tens to hundreds of milliseconds, and sub-6GHz Wi-Fi CSI was measured once per millisecond.

6.6.3: Training Process

The model was trained in each individual scenario. Fig. 6.12 presents the training and test loss in the lab scenario across various cases. Examining the training loss for a single-user case in Fig. 6.12a, we observed that the model converges after approximately 60 epochs. When combining mmWave data with out-of-band CSI, we observed that the convergence time extends to 80 epochs, as depicted in Fig. 6.12b. This can be attributed to the increased complexity of the data. Comparing the cases with and without out-of-band CSI, we observed that the loss decreases more rapidly at the beginning of training when incorporating the out-of-band CSI. This suggests that the beam direction pattern becomes more discernible with the additional information. Fig. 6.12c shows the training loss for the multi-user case, taking around 100 epochs for the model to converge due to the high complexity of multi-user data.

6.6.4: Performance Metrics and Comparison Baseline

Metrics. We use the prediction error as the performance metric. Specifically, referring to Fig. 6.3, the prediction error of azimuth angle is $e_\alpha = |\alpha - \hat{\alpha}|$, where $\hat{\alpha}$ is the predicted beam azimuth angle while α is the beam angle after beam refinement. Similarly, the prediction error of elevation angle is $e_\beta = |\beta - \hat{\beta}|$.

Comparison Baselines. Two schemes are used as the comparison baselines for the evaluation of TBP.

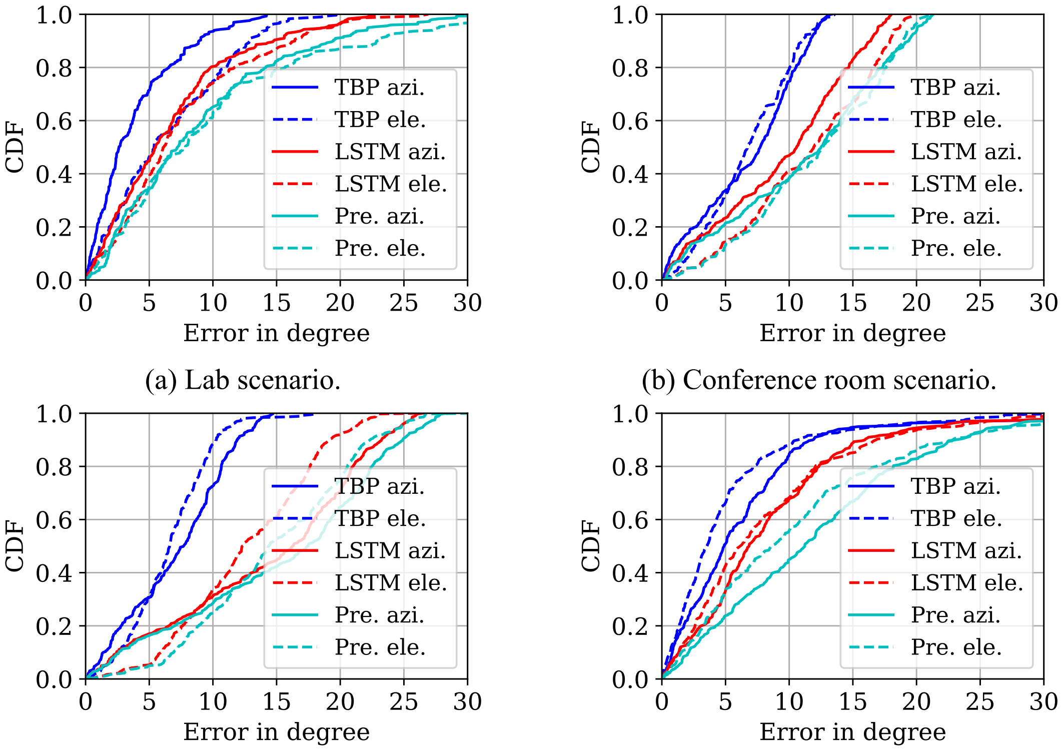
- *Previous Azimuth/Elevation Angle:* For this scheme, we simply use the previous beam's azimuth/elevation angle as the beam direction for the current packet transmission. Apparently, the performance of this scheme is highly dependent on the mmWave data sampling rate and the movement speed of the target mmWave device as well as the dynamics of the environment.
- *LSTM model:* This scheme uses traditional LSTM as the model to predict the beam angles based on the history beam angle information. Specifically, we replace the mLSTM in Fig. 6.3 with a traditional LSTM for the beam prediction and remove the adversarial learning components.

6.6.5: Experimental Results: Beam Angle Errors

In this subsection, we measure the performance of TBP. In addition, we explore the answers to the following questions: For a mmWave AP, is it necessary to create and train a model for each individual user device? If a mmWave AP maintains a separately trained model for each individual user device, would it offer a better performance than the case where the mmWave AP uses a single model for all user devices? To seek the answers, we conduct experiments in two cases: single-user case and multi-user case, as detailed below.

Single-User Case. We first evaluate the performance of TBP based on the history beam selection profile (without CSI from sub-6GHz radio).

Fig. 6.13 shows the CDF of the prediction errors in four scenarios, and Table 6.2 summarizes their average and 95-percentile prediction errors. It can be seen that TBP performs better than the other two schemes. In most cases, the average prediction error of TBP is less than 7 degrees, and the 95-percentile of its prediction error is less than 16 degrees. Both of them are smaller than the other two schemes. Particularly, TBP significantly outperforms the LSTM-based scheme. This



(a) Lab scenario.

(b) Conference room scenario.

(c) Hallway scenario.

(d) Apartment scenario.

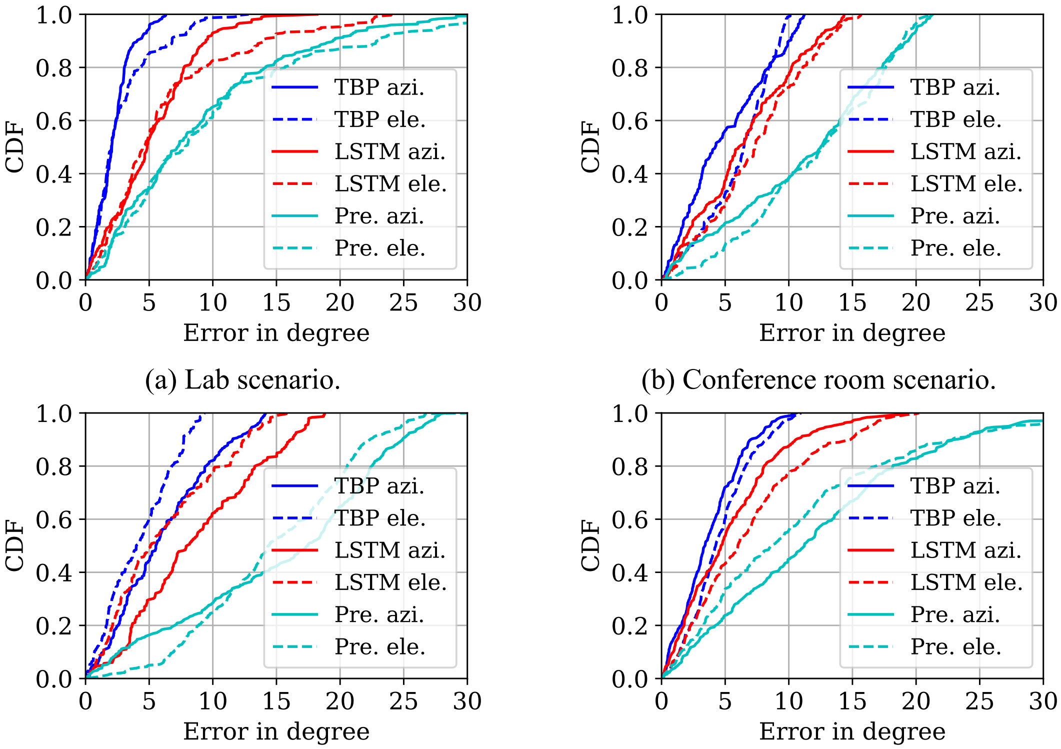
Figure 6.13: Prediction error for TBP: single-user case.

indicates that our proposed mLSTM structure is much more efficient for beam angle prediction than the traditional LSTM structure.

Table 6.2: Prediction errors of TBP: single-user case.

	TBP's prediction error (degree)								LSTM's prediction error (degree)								Previous Point's prediction error (degree)			
	w/o CSI				w/ CSI				w/o CSI				w/ CSI				–			
	Avg		95%		Avg		95%		Avg		95%		Avg		95%		Avg		95%	
	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele
lab	3.8	6.3	11.2	14.7	2.1	2.8	5.0	8.2	6.7	7.5	18.8	19.3	5.2	6.2	11.6	18.6	8.9	9.8	22.2	27.8
conference	6.9	6.8	12.2	12.1	4.9	6.1	10.7	9.6	9.6	11.4	17.0	18.7	6.5	7.3	13.6	14.0	11.3	12.1	20.1	19.8
hallway	7.1	6.6	13.3	11.4	6.0	4.1	12.8	8.7	14.3	12.7	24.8	21.3	8.7	6.1	17.3	13.9	15.3	14.8	26.5	24.8
apartment	6.5	5.0	15.6	16.8	3.8	4.5	8.5	9.3	8.6	8.1	21.1	23.1	5.3	6.8	13.3	16.2	12.4	10.6	27.3	27.5

We now report the experimental results of TBP when it takes advantage of available CSI data from co-located sub-6GHz radio for its training and inference. Fig. 6.14 shows the CDF of the measured prediction errors, and Table 6.2 shows the comparison between the cases with and without CSI data from sub-6GHz radio. It can be seen that, with the utilization of CSI from sub-6GHz radio, the average prediction error of TBP is less than 3 degrees in the lab scenario. The average prediction



(a) Lab scenario.

(b) Conference room scenario.

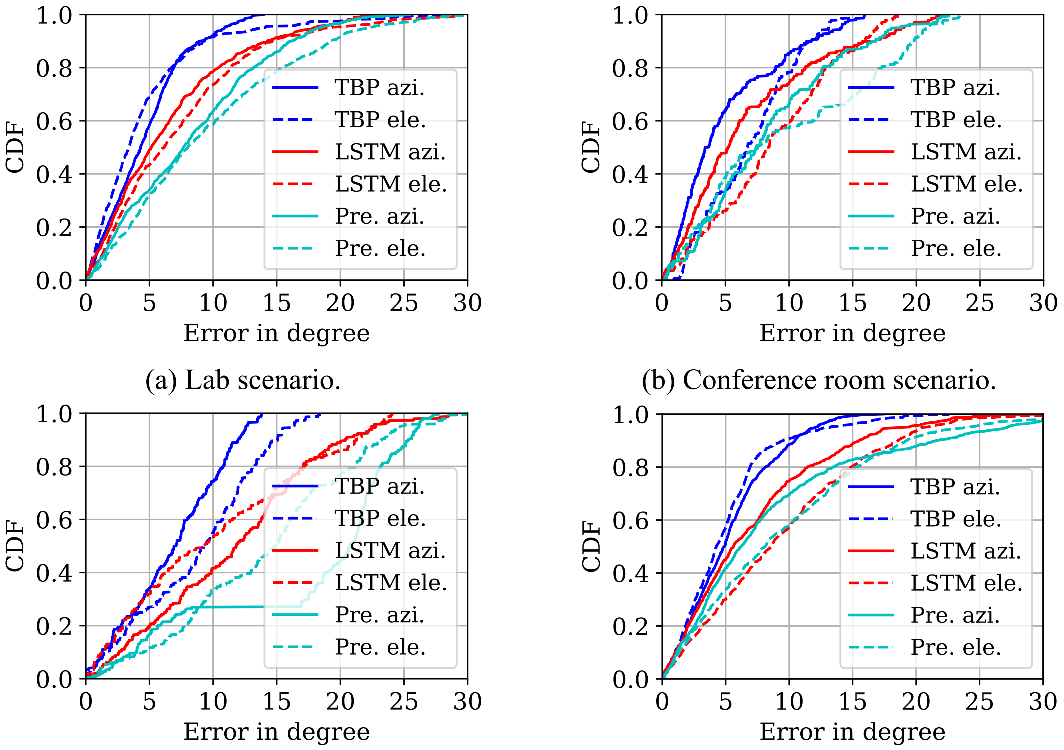
(c) Hallway scenario.

(d) Apartment scenario.

Figure 6.14: Prediction error for TBP: Single-user case with out-of-band CSI enhancement.

error is around 5 degrees in the conference room, hallway, and apartment scenarios. It is larger than that in the lab scenario mainly because of their large size. It can also be seen that the use of CSI data can notably improve the prediction accuracy in the lab scenario and slightly improve the prediction accuracy in the conference room, hallway and apartment scenarios. This is because the lab has many furniture and equipment and thus is more reflective than the other three scenarios.

Multi-User Case. We conduct experiments in the four scenarios by creating and training a single TBP model for six user devices, and measure the prediction errors to evaluate its performance. Fig. 6.15 presents the CDF of our measured prediction errors in the four scenarios, and Table 6.3 summarizes the average and 95th percentile of the measured prediction errors. It can be seen that, for most cases, the average prediction error of TBP is less than 7 degrees. In addition, TBP significantly outperforms its counterparts (LSTM-based scheme and previous beam scheme). Compared to the results presented in Table 6.2 and Table 6.3, we found that TBP has similar performance in the single-user and multi-user cases. This indicates that a mmWave AP does not need to maintain



(a) Lab scenario.

(b) Conference room scenario.

(c) Hallway scenario.

(d) Apartment scenario.

Figure 6.15: Prediction error for TBP: multi-user case.

(create, train, and re-train) different TBP models for different user devices. In other words, it only needs to maintain a single TBP model for all user devices.

Table 6.3: Prediction errors for TBP: Multi-User Case.

scenario	TBP's prediction error (degree)				LSTM's prediction error (degree)				Previous Point's prediction error (degree)			
	Avg		95%		Avg		95%		Avg		95%	
	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele	azi	ele
lab	4.9	4.8	10.9	13.8	6.5	7.3	18.0	19.1	8.3	9.5	18.3	22.4
conference	5.1	6.8	14.1	13.1	6.8	8.6	18.9	16.9	8.3	9.6	18.5	20.8
hallway	6.9	8.7	12.6	16.2	11.6	10.1	22.5	22.7	17.1	14.1	26.2	24.7
apartment	5.3	5.0	12.1	13.2	7.2	9.3	19.1	20.6	8.7	9.6	27.1	24.2

6.6.6: Throughput Gain

Based on the measured beam angle prediction errors, we now assess the throughput gain of TBP in some representative scenarios.

Comparison Baseline. We use the beam search approach in IEEE 802.11ad (see Section 6.3.1)

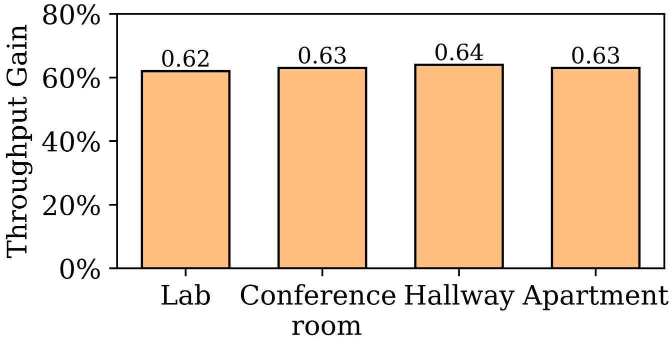


Figure 6.16: Throughput gain of TBP over the 802.11ad beam search approach based on the measured beam prediction errors.

as our comparison baseline.[4] Following the beam search parameters in [118], we assume that the beam search range is from $0°$ to $180°$, with the step size being $7°$. Also following the setting in [118], we assume that the measurement of one beam direction takes $60\ \mu$s. Therefore, to search for the best azimuth and elevation angles in the 3D space, it takes about $\frac{180}{7} \times \frac{180}{7} \times 60\mu\text{s} = 39.6$ms. This is the airtime overhead of beam search during a beacon interval in conventional mmWave networks.

To evaluate the throughput, another important factor is the time duration of a beacon interval (see Fig. 6.2). Theoretically, a longer beacon interval will improve the throughput by amortizing the beam search overhead. However, in practice, the maximum time duration of a beacon interval is constrained by channel coherence time. Here, we follow the parameters in [117] by setting the beacon interval to 100 ms.

Throughput Gain. When a mmWave device uses TBP, it does not need to search for the whole angle range (i.e., $0°$ to $180°$). It only needs to refine the beam direction within its prediction error range. Consider TBP in the lab scenario for example. Its 99-percentile prediction error is 13.0 degrees for azimuth angle and 24.4 degrees for elevation angle. Therefore, the beam search time can be estimated by $\frac{13\times 2}{7} \times \frac{24.4\times 2}{7} \times 60\mu\text{s} = 1.55$ms, where 7 is the angle search step size and 60μs is the airtime of one search. Recall that we assume the beacon interval is 100ms. Then, TBP has $100 - 1.55 = 98.45$ms for data transmission. In contrast, the conventional beam search approach

[4] One may wonder why we use IEEE 802.11ad beam search (rather than the more advanced beam search schemes in the literature) as our comparison baseline. We argue that, while there are many efficient beam search schemes in the literature [23, 24, 65, 118, 148], most of them are limited to the spatial domain. TBP is the first temporal beam prediction approach and complementary to those schemes. Simply put, TBP can be used on top of those beam search schemes to further improve the throughput of mmWave networks.

(comparison baseline) has $100 - 39.6 = 60.4$ms for data transmission. This means that TBP can improve the throughput by 62%. Following the same approach, we also calculate the throughput gain of TBP in other three scenarios. Fig. 6.16 presents the projected throughput gain of TBP. It can be observed that TBP improves the throughput by more than 60% in all four scenarios. This shows the efficiency and robustness of TBP in different environments.

6.7: Summary

Analog beamforming is a fundamental problem in mmWave communication systems. One key problem related to analog beamforming is how to reduce the airtime overhead of beam selection as it is critical for improving the efficiency of mmWave communications. In this chapter, we presented TBP for beam prediction in a 3D space by leveraging the temporal correlation of mmWave channels. The innovation of TBP lies in the design of a new LSTM model, which is capable of performing accurate beam prediction by taking non-uniform, non-smooth history data samples. We further enhanced TBP by taking out-of-band CSI from sub-6GHz radio as its input for training and inference. A novel data fusion method was developed to unify the format of the data samples from the mmWave and sub-6GHz radios. We have evaluated the performance of TBP on a 60GHz mmWave testbed. Experimental results show that the average prediction error of TBP is less than 7 degrees in most of our tested cases and that TBP can improve the throughput by more than 60% in representative mmWave networks.

Our future work will focus on three directions. First, we will develop a high-fidelity testbed for the evaluation of TBP. We will evaluate it on both 28GHz and 60GHz mmWave testbeds following the 5G/Wi-Fi standards. Second, we are currently using the LSTM model for the beam prediction. More advanced deep-learning models have been developed for computer vision and natural language processing. We will enhance TBP by employing new models and evaluating their performance in real-world systems. Third, we will design tools and protocols to enable the automation of data collection for the model training of TBP. Such tools will significantly improve the practicality and generalizability of TBP.

CHAPTER 7: CONCLUSION AND FUTURE WORK

7.1: Conclusion

Wireless communication systems serve as the backbone of today's digitized world, supporting a wide range of applications such as AR/VR, autonomous vehicles, V2X communication, industrial automation, and smart cities. These systems are no longer limited to data communication alone; they also function as sensing platforms, capturing information about the physical world. This transformation has enabled a variety of emerging applications, including elderly care, security and intrusion detection, gesture and activity recognition, and sleep and vital sign monitoring. In this thesis, we designed wireless communication and sensing systems using learning-based approaches, with the goal of improving communication efficiency and enabling fine-grained human motion sensing. We developed learning models in conjunction with signal processing algorithms and customized hardware designs to advance the capabilities of wireless communication and sensing systems, thereby enabling new applications.

The first part of the thesis presented the use of various learning frameworks to design wireless sensing systems for fine-grained human motion sensing. We began by introducing a gesture-based wireless authentication scheme for IoT devices, which employed a convolutional neural network with a feature fusion strategy. Specifically, it combined Wi-Fi CSI amplitude and AoA to enable location-independent gesture recognition. This scheme served as a novel authentication method for widely deployed IoT devices that lack traditional input interfaces. Next, we explored a more challenging topic—handwriting detection through walls. We addressed this challenge through a joint hardware and software design. On the hardware side, we developed a 6 GHz FMCW radar with patch antennas. These components enabled the system to detect motion behind walls while minimizing interference. On the software side, we designed a tailored deep neural network to recognize handwritten letters through obstructions. The model integrated a BiLSTM with an attention mechanism to capture temporal dependencies and extract critical features—such as turning points—from radar phase sequences for accurate recognition. We further extended this system to support eye motion recognition by incorporating a camera-guided deep neural network. This framework used a Transformer encoder as the feature extractor and integrated a state-of-the-art vision-based approach to guide the learning process, enabling the extraction of subtle eye motion features from RF signals. We prototyped all these wireless sensing applications and evaluated them in real-world scenarios, with the hope that they will serve as a foundation for future research.

The second part of the thesis proposed two learning-based solutions aimed at reducing beamforming overhead to enhance the throughput of mmWave communication systems under different network settings. First, we introduced an uplink MU-MIMO mmWave communication (UMMC) scheme for WLANs, which utilized a Bayesian optimization framework for joint beam search across multiple antennas. By significantly reducing beamforming overhead, UMMC improved the network throughput achieved by MU-MIMO in WLANs. Second, we proposed TBP, a Temporal Beam Prediction framework, which focused on reducing beamforming overhead in mobile mmWave networks. TBP was a learning-based beam prediction scheme equipped with a mobility-aware LSTM model. This customized module incorporated timestamp information during training, enabling it to predict beam directions across variable time intervals. By effectively minimizing beamforming overhead, TBP enhanced the throughput performance of mobile mmWave networks.

Investigating wireless communication and sensing together aims to advance the development of future Integrated Sensing and Communication (ISAC) systems. ISAC systems can improve the efficiency of spectrum and hardware utilization, transforming communication platforms into multifunctional systems. This aligns with the vision for 6G networks, which are expected to provide users with a more diverse experience—ranging from data transmission to the delivery of sensing information. This thesis serves as a foundation for future research on the integration of sensing and communication.

7.2: Future Research

Our future research focuses on integrating AI into the next generation of wireless communication and sensing systems. We outline the future direction along two dimensions. First, from an application perspective, we will explore potential use cases for wireless sensing and communication. Then, we will discuss several challenges that must be addressed for practical deployment and for expanding the applicability of our proposed solutions.

7.2.1: Applications

Sensing with Terahertz Communication Signals. Future communication systems are expected to move into the Terahertz (THz) frequency band to support significantly larger bandwidths. The use of high-frequency signals implies shorter wavelengths, which opens up a wide range of novel applications focused on micro-scale information. THz signals can be used to detect fine particles in the air, making them suitable for air quality monitoring. For the same reason, they are promising for applications such as food quality assessment, as shown in Fig. 7.1. Additionally, due to their ability to penetrate the surface of the skin, THz signals offer the potential for non-invasive detection of certain skin diseases in daily life.

Wireless Sensing with LLMs. Wireless sensing leverages the pervasive presence of RF signals, enabling sensing capabilities to extend into every aspect of our daily lives. The continuous stream of data collected by these systems can be effectively processed using Large Language Models (LLMs), significantly enhancing user experience through intelligent interpretation and interaction. By integrating input from wireless sensing and tracking, LLMs gain an additional layer of perception—functioning as a new type of sensor that connects the digital and physical worlds. In the future, LLM-based agents may generate context-aware responses informed by users' activities, locations, and even health conditions, as shown in Fig. 7.2. Synthesizing these technologies to create advanced cyber-physical systems (CPS) will be essential for enabling intelligent, adaptive, and human-centric applications in the next generation of smart environments.

Ubiquitous Vital Sign Measurement. Vital sign measurement is a well-established topic in

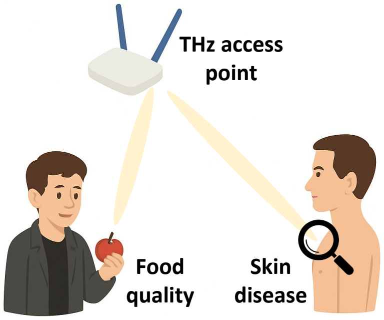


Figure 7.1: THz communication signals are used to assess food quality and detect skin diseases.

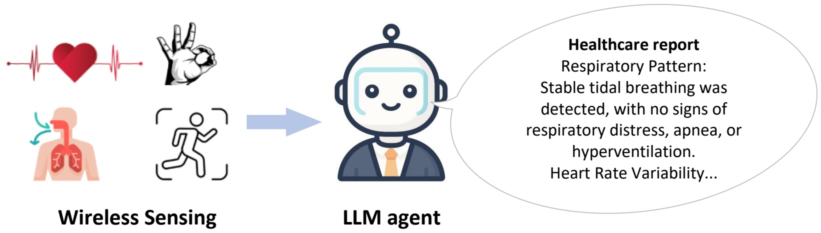

Wireless Sensing
LLM agent
Healthcare report
Respiratory Pattern:
Stable tidal breathing was detected, with no signs of respiratory distress, apnea, or hyperventilation.
Heart Rate Variability...

Figure 7.2: Healthcare reports are generated by an LLM agent based on wireless sensing data.

both academia and industry. However, achieving accurate, contactless monitoring over a large area remains an unsolved challenge. This capability is particularly important for elderly individuals, whose health status should ideally be monitored continuously and unobtrusively throughout the day. RF signals have emerged as a promising candidate for this application, as they can already detect vital signs such as respiration and heartbeat without physical contact. Nonetheless, effectively covering an entire room or large area with consistent accuracy continues to be a significant hurdle that must be addressed for widespread deployment.

RIS-Aided Sensing and Communication. As communication signals move into higher frequency bands, they experience greater attenuation and are more easily blocked by obstacles. Reconfigurable intelligent surfaces (RIS), which can manipulate electromagnetic waves in a controlled manner, offer a promising solution to this challenge. By intelligently redirecting signals, RIS can help avoid blockages, extend coverage areas, and focus signals in specific directions—enhancing both communication performance and sensing capabilities. Designing and integrating such surfaces will be a key component of future 6G networks.

7.2.2: Towards the Advancement of Techniques

Generalization to Different Environments. Generalizability of learning models is a key challenge across many domains, and it becomes particularly critical when dealing with RF data, which is highly sensitive to environmental changes. Although AuthIoT has explored environment-independent gesture recognition and has been evaluated in multiple settings, its performance in more complex or dynamic environments remains uncertain. A similar issue arises in TBP, where the current system requires fine-tuning with data collected from the target environment for accurate beam prediction. Training learning models with RF data that can generalize well across diverse environments remains an open and difficult challenge.

Interference from Dynamic Objects. Interference is a well-known challenge in wireless sensing systems, and it becomes particularly severe when using communication signals—since these signals propagate throughout the environment, any non-target movement can impact them. Most existing Wi-Fi sensing studies assume that the user is the only dynamic object in the environment. Radar-based systems can help mitigate this issue; for example, RadSee and RadEye demonstrate the ability to avoid interference from objects located one meter away. However, they still suffer from interference when non-target objects are in close proximity. Designing a wireless sensing system that can fully eliminate interference from dynamic objects remains an open research problem.

Multi-Target Recognition Using Communication Signals. Recognizing multiple targets with a wireless sensing system is challenging, as it is difficult to distinguish targets solely based on variations in RF signals. Accurately determining the locations of multiple targets is already difficult, and even when their positions can be estimated, simultaneous motions can cause their signals to interfere with each other. Nonetheless, multi-target recognition is essential for developing practical and robust wireless sensing systems suitable for real-world applications.

Efficient Data Collection. Learning-based approaches heavily rely on data, and acquiring training data efficiently and intelligently remains a significant challenge. Sensing tasks involving human targets are particularly labor-intensive, often requiring repetitive motions and extensive manual labeling. A promising future direction is to develop methods for automatic data collection and labeling, reducing the burden on human effort. Additionally, exploring data augmentation techniques to generate large and diverse datasets from a small amount of collected data will be critical for improving model performance and generalizability.

Lightweight AI Models. Although AI models are highly powerful, they often require substantial computing resources. While our current implementations are deployed on local computers, wireless communication and sensing systems are increasingly expected to run on portable, compact devices. Therefore, designing lightweight AI models that can be efficiently deployed on resource-constrained IoT devices is a critical direction for future research.

BIBLIOGRAPHY

[1] IEEE 802.11ad-2012. https://standards.ieee.org/ieee/802.11ad/4527/, Accessed: 08-July-2022.

[2] Mobile data traffic outlook. https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast, Accessed: 08-July-2022.

[3] 5G massive MIMO. https://res-www.zte.com.cn/mediares/zte/Files/PDF/white_book/202009101153.pdf, Accessed: 09-July-2022.

[4] IEEE 802.11ac-2013. https://standards.ieee.org/ieee/802.11ac/4473/, Accessed: 09-July-2022.

[5] IEEE 802.11ay-2021. https://standards.ieee.org/ieee/802.11ay/6142/, Accessed: 09-July-2022.

[6] IEEE 802.11ad. https://www.ieee802.org/11/Reports/tgad_update.htm, Accessed: 30-July-2022.

[7] Release 15. https://www.3gpp.org/release-15, Accessed: 30-July-2022.

[8] Amazon Alex. https://developer.amazon.com/en-US/alexa, Accessed:11-June-2021.

[9] Decora smart - smart switches. https://www.leviton.com/, Accessed:11-June-2021.

[10] Google home assistant. https://assistant.google.com, Accessed:11-June-2021.

[11] Gosund WiFi smart switch. https://www.gosund.com, Accessed:11-June-2021.

[12] Simplisafe. https://simplisafe.com, Accessed:11-June-2021.

[13] 3GPP. NR and NG-RAN overall description, 2018.

[14] 3GPP Tdoc R1-1901252. Evaluation on TSN requirements, Jan. 2019.

[15] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *Proceedings of 12th USENIX symposium on operating systems design and implementation (OSDI)*, pages 265–283, 2016.

[16] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):1–13, 2015.

[17] Fadel Adib and Dina Katabi. See through walls with wifi! In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pages 75–86, 2013.

[18] AFPRELAXNEWS. Handwriting still has a place in our connected world, now it's a trend on social media. https://tinyurl.com/bddxy57n, April 2024. [Online; accessed 01-April-2024].

[19] Adeel Ahmad, June Chul Roh, Dan Wang, and Aish Dubey. Vital signs monitoring of multiple people using a fmcw millimeter-wave sensor. In *2018 IEEE Radar Conference (RadarConf18)*, pages 1450–1455. IEEE, 2018.

[20] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

[21] Anum Ali, Nuria González-Prelcic, and Robert W Heath. Millimeter wave beam-selection using out-of-band spatial information. *IEEE Transactions on Wireless Communications*, 17(2):1038–1052, 2017.

[22] Kamran Ali, Alex X Liu, Wei Wang, and Muhammad Shahzad. Keystroke recognition using wifi signals. In *Proceedings of the 21st annual international conference on mobile computing and networking*, pages 90–102, 2015.

[23] Ahmed Alkhateeb, Sam Alex, Paul Varkey, Ying Li, Qi Qu, and Djordje Tujkovic. Deep learning coordinated beamforming for highly-mobile millimeter wave systems. *IEEE Access*, 6:37328–37348, 2018.

[24] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, 2014.

[25] Ahmed Alkhateeb, Geert Leus, and Robert W Heath. Compressed sensing based multiuser millimeter wave systems: How many measurements are needed? In *Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2909–2913. IEEE, 2015.

[26] ALS news today. ALS Facts and Statistics. https://tinyurl.com/bs8rmh3w, September 2024. [Online; accessed 11-September-2024].

[27] Maria Antonieta Alvarez and Umberto Spagnolini. Distributed time and carrier frequency synchronization for dense wireless networks. *IEEE Transactions on Signal and Information Processing over Networks*, 4(4):683–696, 2018.

[28] Amy Yee. Why Do Eye Muscles Function in ALS as Other Muscles Waste Away? https://tinyurl.com/37tk2rm4, September 2024. [Online; accessed 11-September-2024].

[29] Anonymous. Demo video of real-time uplink MU-MIMO mmWave communication. https://youtu.be/Q2Bk7i6O5mg, Accessed:30-July-2022.

[30] Irmak Aykin, Berk Akgun, Mingjie Feng, and Marwan Krunz. MAMBA: A multi-armed bandit framework for beam tracking in millimeter-wave systems. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, pages 1469–1478. IEEE, 2020.

[31] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 65–74, 2017.

[32] Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*, pages 747–754, 2017.

[33] X. Cao, B. Chen, and Y. Zhao. Wi-Wri: Fine-grained writing recognition using Wi-Fi signals. In *2016 IEEE Trustcom/BigDataSE/ISPA*, pages 1366–1373, 2016.

[34] Emanuele Cardillo, Gaia Sapienza, Changzhi Li, and Alina Caddemi. Head motion and eyes blinking detection: A mm-wave radar for assisting people with neurodegenerative disorders. In *2020 50th European Microwave Conference (EuMC)*, pages 925–928, Piscataway, NJ, USA, 2021. IEEE.

[35] Irched Chafaa, Romain Negrel, E Veronica Belmega, and Mérouane Debbah. Self-supervised deep learning for mmwave beam steering exploiting sub-6 ghz channels. *IEEE Transactions on Wireless Communications*, 21(10):8803–8816, 2022.

[36] Zhaoxin Chang, Fusang Zhang, Jie Xiong, Junqi Ma, Beihong Jin, and Daqing Zhang. Sensor-free soil moisture sensing using lora signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–27, 2022.

[37] Yimin Chen, Tao Li, Rui Zhang, Yanchao Zhang, and Terri Hedgpeth. Eyetell: Video-assisted touchscreen keystroke inference from eye movements. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 144–160, Piscataway, NJ, USA, 2018. IEEE.

[38] Haiming Cheng, Wei Lou, Yanni Yang, Yi-pu Chen, and Xinyu Zhang. Twinkletwinkle: Interacting with your smart devices by eye blink. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(2):1–30, 2023.

[39] Umesh Choudhary, Sampada Bhosale, Sonali Bhise, and Purushottam Chilveri. A survey: Cursive handwriting recognition techniques. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 1712–1716. IEEE, 2017.

[40] Tarun S Cousik, Vijay K Shah, Tugba Erpek, Yalin E Sagduyu, and Jeffrey H Reed. Deep learning for fast and reliable initial access in ai-driven 6g mm wave networks. *IEEE Transactions on Network Science and Engineering*, 2022.

[41] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[42] Dassault Systèmes Simulia]. CST Studio Suite. https://www.3ds.com/products-services/simulia/products/cst-studio-suite/, April 2024. [Online; accessed 17-April-2024].

[43] Shivanker Dev Dhingra, Geeta Nijhawan, and Poonam Pandit. Isolated speech recognition using mfcc and dtw. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2(8):4085–4092, 2013.

[44] Murtaza Dhuliawala, Juyoung Lee, Junichi Shimizu, Andreas Bulling, Kai Kunze, Thad Starner, and Woontack Woo. Smooth eye movement interaction using eog glasses. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 307–311, New York, NY, USA, 2016. Association for Computing Machinery.

[45] Muhammad Ehatisham-Ul-Haq, Ali Javed, Muhammad Awais Azam, Hafiz MA Malik, Aun Irtaza, Ik Hyun Lee, and Muhammad Tariq Mahmood. Robust human activity recognition using multimodal feature-level fusion. *IEEE Access*, 7:60736–60751, 2019.

[46] Lijie Fan, Tianhong Li, Yuan Yuan, and Dina Katabi. In-home daily-life captioning using radio signals. In *European Conference on Computer Vision*, pages 105–123. Springer, 2020.

[47] Yuda Feng, Yaxiong Xie, Deepak Ganesan, and Jie Xiong. Lte-based pervasive sensing across indoor and outdoor. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 138–151, New York, NY, USA, 2021. Association for Computing Machinery.

[48] Yuda Feng, Yaxiong Xie, Deepak Ganesan, and Jie Xiong. Lte-based low-cost and low-power soil moisture sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 421–434, 2022.

[49] Andrew Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least square fitting of ellipses. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):476–480, 1999.

[50] Yongjian Fu, Shuning Wang, Linghui Zhong, Lili Chen, Ju Ren, and Yaoxue Zhang. Svoice: Enabling voice communication in silence via acoustic sensing on commodity devices. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 622–636, 2022.

[51] Z. Fu, J. Xu, Z. Zhu, A. X. Liu, and X. Sun. Writing in the air with WiFi signals for virtual reality devices. *IEEE Transactions on Mobile Computing*, 18(2):473–484, 2019.

[52] Zhangjie Fu, Jiashuang Xu, Zhuangdi Zhu, Alex X Liu, and Xingming Sun. Writing in the air with wifi signals for virtual reality devices. *IEEE Transactions on Mobile Computing*, 18(2):473–484, 2018.

[53] GazeRecorder. Online Eye Tracking Software. https://gazerecorder.com/, September 2024. [Online; accessed 6-September-2024].

[54] Yasaman Ghasempour, Claudio RCM Da Silva, Carlos Cordeiro, and Edward W Knightly. IEEE 802.11 ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi. *IEEE Communications Magazine*, 55(12):186–192, 2017.

[55] Nirnimesh Ghose, Loukas Lazos, and Ming Li. SFIRE: Secret-free-in-band trust establishment for COTS wireless devices. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1529–1537, 2018.

[56] Marco Giordani, Michele Polese, Arnab Roy, Douglas Castor, and Michele Zorzi. A tutorial on beam management for 3GPP NR at mmwave frequencies. *IEEE Communications Surveys & Tutorials*, 21(1):173–196, 2018.

[57] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):855–868, 2008.

[58] Yiqun Guo, Zihuan Wang, Ming Li, and Qian Liu. Machine learning based mmwave channel tracking in vehicular scenario. In *Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6. IEEE, 2019.

[59] Z. Guo, F. Xiao, B. Sheng, H. Fei, and S. Yu. WiReader: Adaptive air handwriting recognition based on commercial WiFi signal. *IEEE Internet of Things Journal*, 7(10):10483–10494, 2020.

[60] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social GAN: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, 2018.

[61] Unsoo Ha, Salah Assana, and Fadel Adib. Contactless seismocardiography via deep learning radars. In *Proceedings of the 26th annual international conference on mobile computing and networking*, pages 1–14, 2020.

[62] Muhammad Kumail Haider, Yasaman Ghasempour, Dimitrios Koutsonikolas, and Edward W Knightly. Listeer: mmwave beam acquisition and steering by tracking indicator LEDs on wireless APs. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 273–288, 2018.

[63] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *SIGCOMM Comput. Commun. Rev.*, 41(1):53, January 2011.

[64] Morteza Hashemi, C Emre Koksal, and Ness B Shroff. Out-of-band millimeter wave beamforming and communications to achieve low latency and high energy efficiency in 5G systems. *IEEE Transactions on Communications*, 66(2):875–888, 2017.

[65] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. Fast millimeter wave beam alignment. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 432–445, 2018.

[66] Wenfeng He, Kaishun Wu, Yongpan Zou, and Zhong Ming. Wig: Wifi-based gesture recognition system. In *2015 24th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7. IEEE, 2015.

[67] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):436–453, 2016.

[68] Yuqiang Heng and Jeffrey G Andrews. Machine learning-assisted beam alignment for mmwave systems. *IEEE Transactions on Cognitive Communications and Networking*, 7(4):1142–1155, 2021.

[69] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[70] Yufang Hou. Incremental fine-grained information status classification using attention-based lstms. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1880–1890, 2016.

[71] Jingyang Hu, Hongbo Jiang, Daibo Liu, Zhu Xiao, Schahram Dustdar, Jiangchuan Liu, and Geyong Min. Blinkradar: non-intrusive driver eye-blink detection with uwb radar. In *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, pages 1040–1050, Piscataway, NJ, USA, 2022. IEEE.

[72] Pengfei Hu, Yifan Ma, Panneer Selvam Santhalingam, Parth H Pathak, and Xiuzhen Cheng. Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 11–20. IEEE, 2022.

[73] Hao Huang, Guan Gui, Haris Gacanin, Chau Yuen, Hikmet Sari, and Fumiyuki Adachi. Deep regularized waveform learning for beam prediction with limited samples in non-cooperative mmwave systems. *IEEE Transactions on Vehicular Technology*, 2023.

[74] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *Proceedings of the 5th augmented human international conference*, pages 1–4, New York, NY, USA, 2014. Association for Computing Machinery.

[75] Grant Jenks. wordsegment. https://grantjenks.com/docs/wordsegment/, April 2024. [Online; accessed 17-April-2024].

[76] Neeta Jha, Amrita Mishra, Jyotsna Bapat, and Debabrata Das. Fast beam search with two-level phased array in millimeter-wave massive MIMO: A hierarchical approach. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1371–1376. IEEE, 2022.

[77] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. Towards environment independent device free human activity recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 289–304, 2018.

[78] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. Towards 3d human pose construction using wifi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.

[79] Kyle Jamieson Jie Xiong, Karthikeyan Sundaresan. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 537–549, New York, NY, USA, 2015. Association for Computing Machinery.

[80] Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. A review on video-based human activity recognition. *Computers*, 2(2):88–131, 2013.

[81] Yongning Ke, Hui Gao, Wenjun Xu, Lixin Li, Li Guo, and Zhiyong Feng. Position prediction based fast beam tracking scheme for multi-user uav-mmwave communications. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2019.

[82] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 303–316, 2014.

[83] Eamonn J Keogh and Michael J Pazzani. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 285–289, 2000.

[84] Sara Khosravi, Hossein Shokri-Ghadikolaei, and Marina Petrova. Learning-based handover in mobile millimeter-wave networks. *IEEE Transactions on Cognitive Communications and Networking*, 7(2):663–674, 2020.

[85] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, pages 269–282, 2015.

[86] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, pages 269–282, 2015.

[87] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, Piscataway, NJ, USA, 2016. IEEE.

[88] Shajahan Kutty and Debarati Sen. Beamforming for millimeter wave communications: An inclusive survey. *IEEE Communications Surveys & Tutorials*, 18(2):949–973, 2015.

[89] Antoinelame Antoine Lamé. Gaze Tracking. https://github.com/antoinelame/GazeTracking, September 2024. [Online; accessed 11-September-2024].

[90] Brian J Lee, Ronald D Watkins, Chen-Ming Chang, and Craig S Levin. Low eddy current rf shielding enclosure designs for 3t mr applications. *Magnetic resonance in medicine*, 79(3):1745–1752, 2018.

[91] Kyoung-Min Lee, Annie P Lai, James Brodale, and Arthur Jampolsky. Sideslip of the medial rectus muscle during vertical eye rotation. *Investigative ophthalmology & visual science*, 48(10):4527–4533, 2007.

[92] Bin Li, Zheng Zhou, Weixia Zou, Xuebin Sun, and Guanglong Du. On the efficient beamforming training for 60ghz wireless personal area networks. *IEEE Transactions on Wireless Communications*, 12(2):504–515, 2012.

[93] Chenning Li, Manni Liu, and Zhichao Cao. WiHF: Enable user identified gesture recognition with wifi. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pages 586–595. IEEE, 2020.

[94] Chenning Li, Zheng Liu, Yuguang Yao, Zhichao Cao, Mi Zhang, and Yunhao Liu. Wi-fi see it all: generative adversarial network-augmented versatile wi-fi imaging. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 436–448, 2020.

[95] Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, and Dina Katabi. Making the invisible visible: Action recognition through walls and occlusions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 872–881, 2019.

[96] Xiaopeng Li, Fengyao Yan, Fei Zuo, Qiang Zeng, and Lannan Luo. Touch well before use: Intuitive and secure authentication for iot devices. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[97] Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. Hashtag recommendation with topical attention-based lstm. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3019–3029, 2016.

[98] Zhengxiong Li, Fenglong Ma, Aditya Singh Rathore, Zhuolin Yang, Baicheng Chen, Lu Su, and Wenyao Xu. Wavespy: Remote and through-wall screen attack via mmwave sensing. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 217–232. IEEE, 2020.

[99] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)*, 35(4):1–19, 2016.

[100] Sun Hong Lim, Sunwoo Kim, Byonghyo Shim, and Jun Won Choi. Deep learning-based beam tracking for millimeter-wave communications under mobility. *IEEE Transactions on Communications*, 69(11):7458–7469, 2021.

[101] Weiyao Lin, Ming-Ting Sun, Radha Poovandran, and Zhengyou Zhang. Human activity recognition for video surveillance. In *2008 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2737–2740. IEEE, 2008.

[102] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.

[103] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. Blinklistener: " listen" to your eye blink using your smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–27, 2021.

[104] Mengxi Liu, Sizhen Bian, and Paul Lukowicz. Non-contact, real-time eye blink detection with capacitive sensing. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers*, pages 49–53, New York, NY, USA, 2022. Association for Computing Machinery.

[105] Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[106] Steven Loria. TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/en/dev/, April 2024. [Online; accessed 17-April-2024].

[107] Lina Ma, Yangtao Ye, Changzhan Gu, and Junfa Mao. High-accuracy contactless detection of eyes' activities based on short-range radar sensing. In *2022 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, pages 266–268, Piscataway, NJ, USA, 2022. IEEE.

[108] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. Signfi: Sign language recognition using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–21, 2018.

[109] George R MacCartney, Sijia Deng, and Theodore S Rappaport. Indoor office plan environment and layout-based mmWave path loss models for 28 GHz and 73 GHz. In *Proceedings of the IEEE 83rd vehicular technology conference (VTC Spring)*, pages 1–6. IEEE, 2016.

[110] Aamir Mahmood, Muhammad Ikram Ashraf, Mikael Gidlund, Johan Torsner, and Joachim Sachs. Time synchronization in 5G wireless edge: Requirements and solutions for critical-MTC. *IEEE Communications Magazine*, 57(12):45–51, 2019.

[111] Christos Masouros and Gan Zheng. Exploiting known interference as green signal power for downlink beamforming optimization. *IEEE Transactions on Signal processing*, 63(14):3628–3640, 2015.

[112] I McCowan, D Moore, J Dines, D Gatica-Perez, M Flynn, P Wellner, and H Bourlard. On the use of information retrieval measures for speech recognition. Technical report, tech. rep., IDIAP Research Institute, Martigny, Switzerland, 2005.

[113] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. Echoflex: Hand gesture recognition using ultrasound imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1923–1934, 2017.

[114] Jonas Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37. Springer Science & Business Media, 2012.

[115] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. *arXiv preprint arXiv:1610.09513*, 2016.

[116] Duy HN Nguyen, Long Bao Le, Tho Le-Ngoc, and Robert W Heath. Hybrid MMSE precoding and combining designs for mmWave multiuser systems. *IEEE Access*, 5:19167–19181, 2017.

[117] Thomas Nitsche, Carlos Cordeiro, Adriana B Flores, Edward W Knightly, Eldad Perahia, and Joerg C Widmer. Ieee 802.11 ad: directional 60 ghz communication for multi-gigabit-per-second wi-fi. *IEEE Communications Magazine*, 52(12):132–141, 2014.

[118] Thomas Nitsche, Adriana B Flores, Edward W Knightly, and Joerg Widmer. Steering with eyes closed: mm-wave beam steering without in-band measurement. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, pages 2416–2424. IEEE, 2015.

[119] Kai Niu, Fusang Zhang, Jie Xiong, Xiang Li, Enze Yi, and Daqing Zhang. Boosting fine-grained activity sensing by embracing wireless multipath effects. In *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, pages 139–151, 2018.

[120] Song Noh, Michael D Zoltowski, and David J Love. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. *IEEE Transactions on Wireless Communications*, 16(9):5689–5701, 2017.

[121] Michele Polese, Francesco Restuccia, and Tommaso Melodia. Deepbeam: Deep waveform learning for coordination-free beam management in mmwave networks. In *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 61–70, 2021.

[122] Swadhin Pradhan, Eugene Chai, Karthikeyan Sundaresan, Lili Qiu, Mohammad A Khojastepour, and Sampath Rangarajan. Rio: A pervasive rfid-based touch gesture interface. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 261–274, 2017.

[123] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38, 2013.

[124] Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Christopher KI Williams. Approximation methods for gaussian process regression. In *Large-scale kernel machines*, pages 203–223. MIT Press, 2007.

[125] Colin Raffel and Daniel PW Ellis. Feed-forward networks with attention can solve some long-term memory problems. *arXiv preprint arXiv:1512.08756*, 2015.

[126] Rima-Maria Rahal and Susann Fiedler. Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology*, 85:103842, 2019.

[127] Dinesh Ramasamy, Sriram Venkateswaran, and Upamanyu Madhow. Compressive tracking with 1000-element arrays: A framework for multi-Gbps mmWave cellular downlinks. In *Proceedings of 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 690–697. IEEE, 2012.

[128] Maryam Eslami Rasekh, Zhinus Marzi, Yanzi Zhu, Upamanyu Madhow, and Haitao Zheng. Noncoherent mmwave path tracking. In *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*, pages 13–18, 2017.

[129] Mattia Rebato, Jihong Park, Petar Popovski, Elisabeth De Carvalho, and Michele Zorzi. Stochastic geometric coverage analysis in mmWave cellular networks with realistic channel and antenna radiation models. *IEEE Transactions on Communications*, 67(5):3736–3752, 2019.

[130] Sajad Rezaie, Elisabeth De Carvalho, and Carles Navarro Manchón. A deep learning approach to location-and orientation-aided 3d beam selection for mmwave communications. *IEEE Transactions on Wireless Communications*, 21(12):11110–11124, 2022.

[131] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results. *IEEE Communications Magazine*, 52(2):106–113, 2014.

[132] Prasun Roy, Subhankar Ghosh, and Umapada Pal. A cnn based framework for unistroke numeral recognition in air-writing. In *2018 16th international conference on frontiers in handwriting recognition (ICFHR)*, pages 404–409. IEEE, 2018.

[133] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[134] Cian Ryan, Brian O'Sullivan, Amr Elrasad, Aisling Cahill, Joe Lemley, Paul Kielty, Christoph Posch, and Etienne Perot. Real-time face & eye tracking and blink detection using event cameras. *Neural Networks*, 141:87–97, 2021.

[135] Batool Salehi, Mauro Belgiovine, Sara Garcia Sanchez, Jennifer Dy, Stratis Ioannidis, and Kaushik Chowdhury. Machine learning on camera images for fast mmwave beamforming. In

IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), pages 338–346. IEEE, 2020.

[136] Science Daily. Eye muscles are resilient to ALS. https://tinyurl.com/cy4kxhv2, September 2024. [Online; accessed 11-September-2024].

[137] Sangho Shin and Henning Schulzrinne. Measurement and analysis of the voip capacity in ieee 802.11 wlan. *IEEE Transactions on Mobile Computing*, 8(9):1265–1279, 2009.

[138] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[139] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. Multipath triangulation: Decimeter-level WiFi localization and orientation with a single unaided receiver. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*, pages 376–388, 2018.

[140] Youngwook Son, Seongwon Kim, Seongho Byeon, and Sunghyun Choi. Symbol timing synchronization for uplink multi-user transmission in IEEE 802.11ax WLAN. *IEEE access*, 6:72962–72977, 2018.

[141] Jiho Song, Junil Choi, and David J Love. Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays. *IEEE Transactions on Communications*, 65(4):1859–1872, 2017.

[142] Kunzhe Song, Qijun Wang, Shichen Zhang, and Huacheng Zeng. Siwis: Fine-grained human detection using single wifi device. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1439–1454, New York, NY, USA, 2024. Association for Computing Machinery.

[143] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[144] Nielen Stander and KJ Craig. On the robustness of a simple domain reduction scheme for simulation-based optimization. *Engineering Computations*, 19(4):431–450, 2002.

[145] Statista. Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2033. https://www.statista.com/statistics/1194682/iot-connected-devices-vertically/, Accessed:08-May-2025.

[146] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Learning-by-synthesis for appearance-based 3d gaze estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1821–1828, Piscataway, NJ, USA, 2014. IEEE.

[147] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 77–89, 2015.

[148] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. WiFi-assisted 60 GHz wireless networks. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 28–41, 2017.

[149] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. Towards scalable and ubiquitous millimeter-wave wireless networks. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 257–271, 2018.

[150] Sanjib Sur, Xinyu Zhang, Parmesh Ramanathan, and Ranveer Chandra. Beamspy: Enabling robust 60 ghz links under blockage. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI '16)*, pages 193–206, 2016.

[151] Koh Tadokoro, Toru Yamashita, Yusuke Fukui, Emi Nomura, Yasuyuki Ohta, Setsuko Ueno, Saya Nishina, Keiichiro Tsunoda, Yosuke Wakutani, Yoshiki Takao, et al. Early detection of cognitive decline in mild cognitive impairment and alzheimer's disease with a novel eye tracking test. *Journal of the neurological sciences*, 427:117529, 2021.

[152] Tzu-Chun Tai, Kate Ching-Ju Lin, and Yu-Chee Tseng. Toward reliable localization by unequal AoA tracking. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '19, page 444–456, New York, NY, USA, 2019. Association for Computing Machinery.

[153] Hanqing Tao, Shiwei Tong, Hongke Zhao, Tong Xu, Binbin Jin, and Qi Liu. A radical-aware attention-based model for chinese text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5125–5132, 2019.

[154] Hans-Leo HM Teulings and Arnold JWM Thomassen. Computer-aided analysis of handwriting movements. *Visible Language*, 13(3):218, 1979.

[155] Texas Instruments. TI AWR6843. https://www.ti.com/tool/AWR6843ISK, September 2024. [Online; accessed 11-September-2024].

[156] Tobill. TOBII PRO SPECTRUM. https://www.tobii.com, September 2024. [Online; accessed 11-September-2024].

[157] Y Ming Tsang, Ada SY Poon, and Sateesh Addepalli. Coding the beams: Improving beamforming training in mmwave communication system. In *Proceedings of IEEE Global Telecommunications Conference-GLOBECOM 2011*, pages 1–6. IEEE, 2011.

[158] Deepak Vasisht, Swarun Kumar, and Dina Katabi. Decimeter-level localization with a single WiFi access point. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, pages 165–178, Santa Clara, CA, March 2016. USENIX Association.

[159] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS'17*, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[160] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. Multi-user gesture recognition using WiFi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 401–413, 2018.

[161] Chao Wang, Feng Lin, Zhongjie Ba, Fan Zhang, Wenyao Xu, and Kui Ren. Wavesdropper: Through-wall word detection of human speech via commercial mmwave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–26, 2022.

[162] Chuyu Wang, Lei Xie, Yuancan Lin, Wei Wang, Yingying Chen, Yanling Bu, Kai Zhang, and Sanglu Lu. Thru-the-wall eavesdropping on loudspeakers via rfid by capturing sub-mm level vibration. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–25, 2021.

[163] Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. Person-in-wifi: Fine-grained person perception using wifi. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5452–5461, 2019.

[164] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. *IEEE Transactions on Mobile Computing*, 16(2):511–526, 2016.

[165] Jue Wang, Deepak Vasisht, and Dina Katabi. RF-IDraw: Virtual touch screen in the air using RF signals. *ACM SIGCOMM Computer Communication Review*, 44(4):235–246, 2014.

[166] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 851–860, 2016.

[167] Xuyu Wang, Chao Yang, and Shiwen 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)*, pages 1230–1239. IEEE, 2017.

[168] Yao Wang, Wandong Cai, Tao Gu, and Wei Shao. Your eyes reveal your secrets: An eye movement based password inference on smartphone. *IEEE transactions on mobile computing*, 19(11):2714–2730, 2019.

[169] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615, 2016.

[170] Yong Wang, Yuhong Shu, and Mu Zhou. A novel eye blink detection method using frequency modulated continuous wave radar. In *2021 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM)*, pages 1–3, Piscataway, NJ, USA, 2021. IEEE.

[171] Yuxi Wang, Kaishun Wu, and Lionel M Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, 2016.

[172] Zhongqin Wang, Fu Xiao, Ning Ye, Ruchuan Wang, and Panlong Yang. A see-through-wall system for device-free human motion sensing based on battery-free rfid. *ACM Transactions on Embedded Computing Systems (TECS)*, 17(1):1–21, 2017.

[173] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. Acoustic eavesdropping through wireless vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 130–141, 2015.

[174] Teng Wei and Xinyu Zhang. mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 117–129, 2015.

[175] Zhiqing Wei, Yuan Wang, Liang Ma, Shaoshi Yang, Zhiyong Feng, Chengkang Pan, Qixun Zhang, Yajuan Wang, Huici Wu, and Ping Zhang. 5g prs-based sensing: A sensing reference signal approach for joint sensing and communication system. *IEEE Transactions on Vehicular Technology*, 72(3):3250–3263, 2022.

[176] Wikipedia. Substitution cipher. https://en.wikipedia.org/wiki/Substitution_cipher, Accessed:13-June-2021.

[177] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[178] Genta Indra Winata, Onno Pepijn Kampman, and Pascale Fung. Attention-based lstm for psychological stress detection from spoken language using distant supervision. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6204–6208. IEEE, 2018.

[179] Chenshu Wu, Feng Zhang, Yusen Fan, and KJ Ray Liu. RF-based inertial measurement. In *Proceedings of the ACM Special Interest Group on Data Communication*, pages 117–129. 2019.

[180] Zhenyu Xiao, Lipeng Zhu, Zhen Gao, Dapeng Oliver Wu, and Xiang-Gen Xia. User fairness non-orthogonal multiple access (NOMA) for millimeter-wave communications with analog beamforming. *IEEE Transactions on Wireless Communications*, 18(7):3411–3423, 2019.

[181] Jiahong Xie, Hao Kong, Jiadi Yu, Yingying Chen, Linghe Kong, Yanmin Zhu, and Feilong Tang. mm3dface: Nonintrusive 3d facial reconstruction leveraging mmwave signals. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, pages 462–474, New York, NY, USA, 2023. Association for Computing Machinery.

[182] Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity WiFi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 53–64, New York, NY, USA, 2015. ACM.

[183] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. Md-track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[184] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. md-track: Leveraging multi-dimensionality for passive indoor wi-fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[185] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, pages 71–84, Lombard, IL, April 2013. USENIX Association.

[186] Hao Xue, Du Q Huynh, and Mark Reynolds. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1186–1194. IEEE, 2018.

[187] Hongfei Xue, Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shiyang Wang, Ye Yuan, Shuochao Yao, Aidong Zhang, and Lu Su. Deepmv: Multi-view deep learning for device-free human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–26, 2020.

[188] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmmesh: towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 269–282, 2021.

[189] Zhenyu Yan, Qun Song, Rui Tan, Yang Li, and Adams Wai Kin Kong. Towards touch-to-access device authentication using induced body electric potentials. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[190] Edwin Yang, Qiuye He, and Song Fang. Wink: Wireless inference of numerical keystrokes via zero-training spatiotemporal analysis. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 3033–3047, 2022.

[191] Lei Yang, Qiongzheng Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. See through walls with cots rfid system! In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 487–499, 2015.

[192] Xi Yang, Michail Matthaiou, Jie Yang, Chao-Kai Wen, Feifei Gao, and Shi Jin. Hardware-constrained millimeter-wave systems for 5G: challenges, opportunities, and solutions. *IEEE Communications Magazine*, 57(1):44–50, 2019.

[193] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.

[194] Wenfang Yuan, Simon MD Armour, and Angela Doufexi. An efficient and low-complexity beam training technique for mmwave communication. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 303–308. IEEE, 2015.

[195] Shichao Yue, Yuzhe Yang, Hao Wang, Hariharan Rahul, and Dina Katabi. Bodycompass: Monitoring sleep posture with wireless signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–25, 2020.

[196] H Zamani, A Abas, and MKM Amin. Eye tracking application on emotion analysis for marketing strategy. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(11):87–91, 2016.

[197] Jiansong Zhang, Zeyu Wang, Zhice Yang, and Qian Zhang. Proximity based IoT device authentication. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, 2017.

[198] L. Zhang, J. Wang, Q. Gao, X. Li, M. Pan, and Y. Fang. Letfi: Letter recognition in the air using CSI. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2018.

[199] Shichen Zhang, Bo Ji, Kai Zeng, and Huacheng Zeng. Realizing uplink mu-mimo communication in mmwave wlans: Bayesian optimization and asynchronous transmission. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2023.

[200] Shichen Zhang, Pedram Kheirkhah Sangdeh, Hossein Pirayesh, Huacheng Zeng, Qiben Yan, and Kai Zeng. Authiot: A transferable wireless authentication scheme for iot devices without input interface. *IEEE Internet of Things Journal*, 9(22):23072–23085, 2022.

[201] Shichen Zhang, Qijun Wang, Maolin Gan, Zhichao Cao, and Huacheng Zeng. Radsee: See your handwriting through walls using fmcw radar. In *Network and Distributed Systems Security (NDSS) Symposium*, 2025.

[202] Shichen Zhang, Qijun Wang, Kunzhe Song, Qiben Yan, and Huacheng Zeng. Radeye: Tracking eye motion using fmcw radar. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2025.

[203] Shichen Zhang, Qiben Yan, Tianxing Li, Li Xiao, and Huacheng Zeng. Tbp: Temporal beam prediction for mobile millimeter-wave networks. *IEEE Internet of Things Journal*, 2024.

[204] Tianfang Zhang, Zhengkun Ye, Ahmed Tanvir Mahdad, Md Mojibur Rahman Redoy Akanda, Cong Shi, Yan Wang, Nitesh Saxena, and Yingying Chen. Facereader: Unobtrusively mining vital signs and vital sign embedded sensitive info via ar/vr motion sensors. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pages 446–459, New York, NY, USA, 2023. Association for Computing Machinery.

[205] Xi Zhang, Yu Zhang, Zhenguo Shi, and Tao Gu. mmfer: Millimetre-wave radar based facial expression recognition for multimedia iot applications. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, New York, NY, USA, 2023. Association for Computing Machinery.

[206] Xiaotong Zhang, Zhenjiang Li, and Jin Zhang. Synthesized millimeter-waves for human motion sensing. In *Proceedings of the Twentieth ACM Conference on Embedded Networked Sensor Systems (SenSys)*, page 377–390, 2022.

[207] Xinze Zhang, Walid Brahim, Mingyang Fan, Jianhua Ma, Muxin Ma, and Alex Qi. Radar-based eyeblink detection under various conditions. In *Proceedings of the 2023 12th International Conference on Software and Computer Applications*, pages 177–183, New York, NY, USA, 2023. Association for Computing Machinery.

[208] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4511–4520, Piscataway, NJ, USA, 2015. IEEE.

[209] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7356–7365, 2018.

[210] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human mesh recovery using radio signals. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10113–10122, 2019.

[211] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. Rf-based 3d skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 267–281, 2018.

[212] Renjie Zhao, Timothy Woodford, Teng Wei, Kun Qian, and Xinyu Zhang. M-cube: A millimeter-wave massive MIMO software radio. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.

[213] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 313–325, 2019.

[214] Anfu Zhou, Xinyu Zhang, and Huadong Ma. Beam-forecast: Facilitating mobile 60 ghz networks via model-driven beam steering. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, pages 1–9. IEEE, 2017.

[215] Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. Attention-based lstm network for cross-lingual sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 247–256, 2016.

[216] Wangjiang Zhu and Haoping Deng. Monocular free-head 3d gaze tracking with deep learning and geometry constraints. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3143–3152, Piscataway, NJ, USA, 2017. IEEE.