August 2020

# Multi-function RF for Situational Awareness

Yang Liu

*Syracuse University*

## Recommended Citation

Liu, Yang, "Multi-function RF for Situational Awareness" (2020). *Dissertations - ALL*. 1241.
https://surface.syr.edu/etd/1241

# ABSTRACT

Radio frequency (RF) communications are an integral part of many situational awareness applications. Sensing data need to be processed in a timely manner, making it imperative to have a robust and reliable RF link for information dissemination. Moreover, there is an increasing need for exploiting RF communication signals directly for sensing, leading to the notion of multi-function RF.

In the first part of this dissertation, we investigate the development of a robust Multiple-Input Multiple-Output (MIMO) communication system suitable for airborne platforms. Three majors challenges in realizing MIMO capacity gain in airborne environment are addressed: 1) antenna blockage due largely to the orientation of the antenna array; 2) the presence of unknown interference inherent to the intended application; 3) the lack of channel state information (CSI) at the transmitter. Built on the Diagonal Bell-Labs Layered Space-Time (D-BLAST) MIMO architecture, the system integrates three key design approaches: spatial spreading to counter antenna blockage; temporal spreading to mitigate signal to interference and noise ratio degradation due to intended or unintended interference; and a simple low rate feedback scheme to enable real time adaptation in the absence of full transmitter CSI. Extensive experiment studies using a fully functioning $4 \times 4$ MIMO system validate the developed system.

In the second part, ambient RF signals are exploited to extract situational awareness information directly. Using WiFi signals as an example, we demonstrate that the CSI obtained at the receiver contains rich information about the propagation environment. Two distinct learning systems are developed for occupancy detection using passive WiFi sensing. The first one is based on deep learning where a parallel convolutional neural network (CNN) architecture is designed to extract useful information from both magnitude and phase of the CSI. Pre-processing steps are carefully designed to preserve human motion induced channel variation while insulating against other impairments and post-processing is applied after CNN to infer presence information for instantaneous motion outputs. To alleviate the need of tedious training efforts involved in deep learning based system, a novel learning problem with contaminated sampling is formulated. This leads to a second learning system: a two-stage solution for motion detection using support vector machines (SVM). A one-class SVM model is first evaluated whose training data are from human free environment only. Decontamination of human presence data using the one-class SVM is done prior to motion detection through a two-class support vector classifier. Extensive experiments using commercial off-the-shelf WiFi devices are conducted for both systems. The results demonstrate that the learning based RF sensing provides a viable and promising alternative for occupancy detection as they are much more sensitive to human motion than passive infrared sensors which are widely deployed in commercial and residential buildings.

# MULTI-FUNCTION RF FOR SITUATIONAL AWARENESS

By

Yang Liu
B.E., University of Electronic Science and Technology of China, 2013

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical & Computer Engineering

Syracuse University
December 2020

# ACKNOWLEDGMENTS

First of all, I want to express my deepest gratitude to my advisor, Dr. Biao Chen. The completion of my dissertation would not have been possible without your guidance. Thank you for leading me to the research world, seeking opportunities for us and seeing potentials in me. It is because of your profound knowledge, patience and encouragement that I can overcome difficulties and reach this milestone. Your integrity, strong sense of responsibility and caring for others also teach me to become a better person in my life.

I am grateful to Dr. Bing Dong, Dr. Senem Velipasalar Gursoy, Dr. Sucheta Soundarajan, Dr. Asif Salekin and Dr. Janek J. Mroczek for serving on my committee. Thank you for your time in reviewing my work and providing insightful suggestions.

I would also like to acknowledge the generous support of the National Science Foundation under award CNS-1731237, the gift fund from Syracuse Research Corporation, Syracuse, New York and the Air Force Research Laboratory under award FA8750-15-1-0045 and FA8750-18-1-0065.

My thanks should go to my fellow labmates, Fangfang, Kapil, Fangrong, Yu, Pengfei, Shengyu and Tiexing. I will always keep in my memory the time we spent together in group meetings and the discussions we had in our labs. Your help and kindness make me feel warm and motivated. The two-year DARPA Spectrum Collaboration Challenges gave me an unique opportunity to work with many great folks including Himanshu, Yuexin, Ziqi, Xin, Jiawei, Komal and Yilin. Thank you for your patience with me and I will never forget the days and nights we worked

together.

My final but everlasting gratitude goes to my parents who gives me unconditional love. Your guidance in my life makes me become who I am and the long distance between us can never stop you from showing your greatest support to me. Special thanks also goes to my beloved half Tiexing Wang who always stand by my side. Thank you for understanding me and inspiring me even in tough times during my Ph.D. adventure.

# TABLE OF CONTENTS

**3 Presence Detection using Deep Learning 34**

**4 Presence Detection with Reduced Training efforts 69**

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Background

Wireless communications play a key role in many situational awareness (SA) applications, both as a means for information dissemination and as a medium for directly harvesting sensing information.

For example, in situational awareness systems involving airborne platforms, a high throughput and low latency wireless link is imperative to ensure timely delivery of SA data from these airborne platforms to information consumers. In the first part of this dissertation, we address challenges in realizing such desirable connections when communication systems involve airborne platforms. Multiple Input Multiple Output (MIMO) technology is a natural candidate due to its superior spectrum efficiency [1,2] and low transmit power requirement. By utilizing parallel spatial channels constructed through pre-processing at the transmitter and equalization at the receiver [3,4], the throughput is shown to scale linearly in the number of transceiver antennas in the high signal to noise ratio (SNR) regime. With airborne platforms, however, MIMO communications face the unique challenge due to the lack of a scattering environment. This, however, can be compensated by spacing out the antennas to take advantage of the large platform size in airborne networks [5,6]. In this part of the work, we design a variable rate MIMO system to tackles challenges in realizing MIMO throughput gain for airborne communications.

Additionally, radio frequency (RF) signals themselves contain rich information about the propagation environment. The ability to harvest SA information from ambient RF signals often provide an appealing alternative to existing solutions that often require extensive infrastructure. For example, WiFi signals have been exploited for various situational awareness tasks, such as detecting, localizing, tracking, and identifying human motion/activities [7–34].

In the second part of the work, we explore the use of RF signals for human motion detection. This provides an infrastructure free solution to human occupancy detection, which can serve as a promising alternative to existing techniques and help to realize the promise of green buildings and smart homes through providing real-time occupancy information.

Two different solutions are proposed for passive RF sensing. In the first one, a convolutional neural network (CNN) based system is designed to achieve reliable presence detection. Using off-the-shelf WiFi devices, the proposed deep learning based RF sensing achieves near perfect presence detection during multiple extended periods of test and exhibits superior performance compared with leading edge passive infrared sensors.

The challenge for the deep learning based approach is in the training process, especially the data collection process. Training data corresponding to human motions need to be collected when continuous motion is expected for the human presence data. This process of collecting training data is laborious and time-consuming and adds significant barriers for wide spread adoption of the proposed solution. To address this challenge, we propose a two-stage solution for which human presence data are collected when subjects carry out normal activities without deliberately introducing continuous motion. First, a one-class support vector machine (SVM) is used to model human absent data environment. This model is then used to remove static RF data collected with human present but remaining still. This is followed by a two-class support vector classifier for motion detection using de-contaminated samples. As such, data parsing and training can be done in an autonomous manner with minimum human intervention.

## 1.2 Robust MIMO Communication for Airborne Platforms

For MIMO, linear scaling of throughput in the number of transceiver antennas is typically attained for a rich scattering environment. The spatial diversity in such environment ensures that the channels between different transmit-receiver element pairs are sufficiently different, leading to a well behaving channel matrix. In an aerial environment, MIMO communications face the unique challenge due to the lack of a scattering environment. For example, with only line of sight transmission, the keyhole effect renders the throughput gain logarithmic instead of linear [35]. The keyhole effect, however, assumes a 'small aperture' system where transceiver array aperture is orders of magnitude smaller than the transmission distance. For large aircraft, large transceiver aperture can be realized and they are typically much larger than hand-held devices. Indeed, using a realistic setting in both aperture size and transmission distance, it was shown that significant capacity gain can still be attained even if only line of sight is assumed [5]. The large antenna spacing mitigates the 'keyhole' effect because of the distinct distances between transmit elements and receive elements. As such, channel gains between different transceiver elements are meaningfully different, leading to the desired spatial diversity even when scatters are scarce [36, 37].

While the potential for throughput gains gives incentive to consider MIMO technology for communications, there are unique challenges that need to be resolved. We address the following three challenges that are most often encountered in aerial communications.

1. Channel impairment due to antenna blockage.

   Antenna blockage is more prevalent when the antennas are conformed to the shape of the aircraft. Clearly, the relative orientation of multiple antennas between communicating nodes often leads to blockage of some antenna elements. This is exacerbated by the lack of scattering as the blockage can not be compensated by indirect paths, leading effectively to antenna outage, i.e., transmissions from blocked transmit antenna elements never reach any receive antenna element or that blocked receive antenna elements never receive any signals from the transmitter.

2. Interference.

   Clearly, unintended or intentional interference will need to be addressed for the system to have robust connectivity in a realistic operating environment. Another complicating factor is the dynamic nature of the RF environment; interference may itself be highly time varying hence requiring a simple yet adaptive mitigation scheme.

3. The lack of transmitter channel state information (CSI).

   In high mobility communications, the mobility of the transmitter and/or receiver results in a highly dynamic channel environment. It is unrealistic to have receiver send the entire CSI to the transmitter, thus the CSI is typically not available at the transmitter side. Additionally, for some applications, the throughput requirement is highly asymmetrical, e.g., video streaming forward link requires much more bandwidth than control signal in the reverse link. As such, channel reciprocity, even if it holds, may not be helpful in such applications.

To address these challenges, we develop a fully adaptive MIMO system built on the Diagonal Bell Labs Layered Space-Time (D-BLAST) architecture. The D-BLAST architecture is chosen as the baseline system primarily due to its ability to operate in the absence of transmitter CSI and its optimality in terms of outage capacity. Through cycling independent data streams across different transmit antennas, a D-BLAST system transmits each data stream over multiple channel states corresponding to multiple transmit antennas. This allows the transceiver to exploit spatial diversity in the absence of CSI at the transmitter.

However, D-BLAST in itself is not sufficient in addressing the loss of degrees of freedom in the presence of antenna blockage or severe signal to interference and noise power ratio (SINR) degradation due to interference.

The developed system integrates three critical signal processing and design components into the baseline D-BLAST system to address the above challenges: 1) spatial spreading to combat rank deficient channel matrix due to antenna blocking; 2) temporal spreading to boost SINR when operating under strong interference, and 3) a low rate feedback scheme for transceiver adaptation.

Notice that these three components are not designed in isolation. For example, while spatial spreading is primarily used to mitigate antenna outage, it can also improve SINR at the receiver. Nevertheless, adaptation of spreading factors in spatial and temporal domains is driven by different channel impairments, hence needs to be handled differently. As shall be seen, antenna blockage leads to ill-conditioned channel matrices, whose remedy requires the reduction of data streams to accommodate for the loss of degrees of freedom. Spatial spreading is thus called for in such a situation. Temporal spreading, on the other hand, is used to boost the SINR at the receiver, hence the actual receiver performance (e.g., measured SINR or bit error rate (BER)) is used to drive temporal spreading adaptation. Finally, choosing the simple spreading based approach as opposed to, say, more sophisticated schemes (e.g., space time code [38–41] to counter antenna blockage) enables the use of a low rate feedback for adaptation: with spreading, the transmitter only needs to be informed by the receiver the desired spreading factors. Together, temporal and spatial spreading collectively provide a range of transmission states that help balance spectrum efficiency with the need for robust connectivity in the presence of various channel impairments.

Another major contribution of the presented work is the development of a fully functioning MIMO system using software defined radios (SDR) that integrates the three design components. We describe in detail the developed MIMO transceiver and provide experimental results on the performance of such a system in the presence of antenna outage or in-band interference. The results validate the robust performance under those channel impairments compared to the conventional D-BLAST system.

## 1.3 Harvest Ambient RF for Presence Detection

Presence detection plays a key role in improving operation efficiency and reducing carbon footprint for office and residential buildings. The use of occupancy information in controlling HVAC and lighting systems has become increasingly prevalent. Additionally, in the age of Covid-19 pandemic, real-time occupancy information can help more accurately identify the need for sanitizing/disinfecting offices and classrooms after their usage. Existing methods for human presence detection include passive infrared (PIR), microwave, $CO_2$, wearable sensors, and cameras [42], among others. Microwave sensors are overly sensitive as they tend to have frequent false alarms, e.g., detecting movements outside of intended coverage areas. $CO_2$ sensors have a slow response time and a high cost barrier. Cameras raise privacy concerns and are sensitive to lighting conditions. Wearable sensors/devices can be intrusive or cumbersome for users. PIR sensors are the most widely deployed method for presence detection. PIR sensors pick up infrared emission using its on-board pyroelectric sensor and detect movement of humans (or objects) through heat variation within the field of view. Its drawback is its low sensitivity and limited coverage and it is mostly used for isolated lighting control.

This part of the work explores the use of RF signals for presence detection. In particular, we use WiFi signals in the current work given its ubiquity in almost all indoor environment. Current and future WiFi systems (i.e., the upcoming WiFi-6) employ multiple-input and multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) at the physical layer. As such, the channel state information (CSI) contains rich information about the ambient environment in spatial, temporal, and frequency domains.

Early work for indoor RF sensing mainly relies on received signal strength indicator

(RSSI) [7–9]. RSSI measures instantaneous attenuation of RF signals at the receiver and its temporal variation can be associated with motion/activities of humans/objects. Recently, more fine-grained features such as CSI have been used for RF sensing. For example, different human activities (e.g., running, walking and eating) or locations can be recognized by analyzing their unique effect on the CSI [15–21]. Other interesting applications, e.g., sign language classification [25] and human identification [43–45], have demonstrated that the CSI from WiFi signals contains subtle and important features in the propagation environment.

There is an important distinction between presence detection and activity detection (e.g., sign language [25] or fall detection [10,46]). For detection of particular activities, one can use either a pattern-based or model-based approach [47–49] as a target activity often imposes identifiable signatures on RF propagation. Thus, hand-crafted features extracted from received signals can be exploited [11, 15] or one can try to establish a CSI model for the target activity [20, 50, 51]. Alternatively, a data-driven approach can be used where collected training data are fed to machine learning algorithms (e.g., a neural network) for discriminating different states (labels) through training. For presence detection, however, there is no defined activities when humans are present thus a pattern-based or model-based approach is typically not adequate. While a data-driven approach appears to be a natural choice, it is unclear *a priori* what would be the best way to collect training data for presence detection. Perhaps the only reasonable assumption that one can make is that humans are not expected to be completely still for an extended period of time. While there exist studies on human presence detection using RF signals through either carefully calibrating the human free environment [22, 23] or breathing detection [24, 52], their performance is highly sensitive to environment change (e.g., room change, furniture move) and human locations (as in the case of breathing detection).

There is prior work on presence detection using CSI. In pattern-based methods, hand-crafted features have been extracted to capture distinctive signal patterns introduced by human presence. Among them, the correlation matrix of CSI time series is most widely explored [24, 27–31]. Features such as eigenvalues/eigenvectors are used as an input to a classifier such as a support vector machine (SVM) to find decision boundaries. Other suggested features include periodicity after performing continuous wavelet transform [53], temporal similarity of CSIs across frequencies [26], histograms of CSI amplitude [54], and statistics from average Doppler spectrum [55]. These pattern-based approaches are susceptible to inconsistent human activities, i.e., when mismatch occurs between human motion to those stored in the profile [47] and are thus prone to performance degradation and limited coverage [27, 54]. Model-based methods [50, 56, 57] often exploit estimated path change to represent Doppler effect induced by moving humans. The performance is highly sensitive to conditions such as moving directions of humans, number of humans in the monitored region, and other ambient factors [50, 57].

Alternatively, data-driven approaches employing learning systems (e.g., neural networks) have also been proposed. Examples include feeding CSI amplitude and phase to multiple layer perceptions (MLP) to classify three common human modes - absence, working and sleeping [58], and Person-in-WiFi [59] that tries to find mapping from CSIs to 2D human image. However, an acute challenge in using deep neural networks is overfitting due to the large amount of trainable parameters compared with often limited training data. This problem becomes even more severe in RF sensing as the wireless environment can evolve over time or change dramatically for different rooms, leading to significant performance degradation. As reported in [58], deploying the pre-trained model to a new environment has resulted in a drop of accuracy from $94.5\%$ to around $50\%$. Similar performance losses have been reported in other systems, see, e.g., [59].

In this dissertation, we present our proposed solutions for RF based presence detection:

1. a WiFi CSI based presence detection system consisting of pre-processing for data representation, a convolutional neural network (CNN) for motion detection, and post-processing for the eventual presence detection.

The proposed parallel CNN architecture separates CSI magnitude and phase in the learning system. This is much more robust than stacking up real and imaginary components as it allows for different pre-processing of CSI magnitude and phase. This is crucial in exploiting motion induced CSI variation in the presence of various channel and hardware impairments. We note that while many existing approaches completely discard CSI phase due to inherent impediments such as carrier frequency offset (CFO) and sampling time offset (STO) [28–31, 54, 60], a more promising way to exploit information from CSI phase is to exploit spatial diversity at the receiver and compute phase difference across receive antennas as CFO and STO are common to all receive antennas [22, 27, 46, 56].

Pre-processing of CSI estimate is carefully designed where spatial, temporal, and frequency domain information is exploited in a holistic manner. The pre-processing takes into consideration how human movement affects CSI while insulating against unintended distortion in RF circuitry (e.g., CFO/STO). Fourier transform is used to localize important motion-induced features in the constructed image, making it more amenable for presence detection as CNN builds its ability for discriminating data through local features (i.e., small kernel size).

2. a one-class SVM based learning approach whose training data consist of CSIs collected in human free environment only.

   In this part, we aim at addressing one major challenge for existing RF based presence detection approaches. Their training process involves significant human intervention. Tedious labelling process is often required (e.g., with the assistance of cameras or manual segmentation) for training data collection [59, 61]. Alternatively, continuous human motion has to be introduced throughout the measurement period for human presence data [26, 27, 31, 58, 62]. The problem becomes even more severe for the neural network based model due to its need of a large amount of samples from both labels.

To simplify the data collection process, support vector data description (SVDD), a special case of one-class SVM, is adopted to construct a compact description of human free data. During test period, any measurements that fall outside of the sphere in the feature space will be considered an outlier, hence classified as motion samples.

3. a two-stage approach whose human presence data are collected when humans are performing their daily activities without introducing any intentional motion behaviours.

   Without any information about human motions, high sensitivity to human presence can only be achieved by the SVDD based model at the sacrifice of reliability in human absent environment. It is feasible that training with some samples from motions as described in [63] can potentially improve the performance. However, collection as well as selection of such motion samples involve human intervention. A much simpler, and more desirable, way is to collect human presence data with human subjects performing daily activities in the environment without deliberately introducing continuous motion.

   However, this simple data collection scheme for RF based sensing can lead to significant performance degradation for learning systems designed under the typical supervised learning framework. This is due to the sample contamination in the human presence data: human presence data include measurement periods when the occupants may be completely or nearly still and the corresponding RF signals are nearly indistinguishable from human absent data. When training is done using the contaminated samples, the trained motion classifier will have significantly elevated false positive.

   We thus propose a two-stage approach to address sample contamination in the training data. First, human free data are used to help remove static measurement in human presence data using SVDD. The subsequent training data are used to train a two-class support vector classifier (SVC) for motion detection. SVC is used in lieu of CNN in [62] since two-class SVC is much less sensitive to residual sample contamination compared with CNN and requires far few training samples.

An important contribution of our work in RF based presence detection is extensive test implemented using commercial off-the-shelf (COTS) WiFi devices. Assuming that humans are not completely still for an extended period of time, our CNN based presence detection system compares much more favorably against that of commercial PIR sensors. Comparison with existing RF based human presence detection also demonstrates its robustness in performance, especially when deployed in a completely new environment. Via extended periods of test, we also demonstrate that the later proposed approaches which are based on SVM provides a viable and promising alternative for presence detection due to the simplified training process.

## 1.4 Summary of Contributions

Our main contributions for this dissertation are listed below.

- We design an adaptive MIMO system for airborne communications which incorporates temporal and spatial spreading into the classical D-BLAST architecture to combat antenna blockage and the presence of interference. The adaptation at the transmitter is driven by a low rate feedback scheme and helps the system to maintain robust connectivity under various channel impairments.
- A fully functioning MIMO system integrating the designed components is developed using USRP and GNU Radio. Experiment results validate the effectiveness of the proposed system in handling antenna blockage and interference.
- We propose a passive WiFi sensing system for indoor presence detection. A CNN with parallel architecture is utilized to exploit presence information from both CSI

magnitude and CSI phase along temporal, frequency and spatial dimensions.

- Pre-processing of CSI estimate is designed to maintain human induced variations effectively while mitigating other unintended fluctuations. Post-processing using the CNN output is designed to give reliable presence detection from instantaneous motion detection outputs according to desired time resolution.
- Extensive experiments using off-the-shelf WiFi devices are conducted to show that the proposed CNN based system achieves reliable presence detection result and is robust against variations in the wireless environment. Comparison studies further demonstrate its performance advantage over PIR sensors and other RF based solutions.
- We develop a two-staged approach whose human presence data for training are collected when humans are performing their daily activities. The proposed model employs SVDD in the first stage followed by a two-class SVC in the second stage to address sample contamination introduced by stationary human measurements.

## 1.5 Organization

The rest of the dissertation is organized as follows. In Chapter 2, we present our designed components for addressing challenges when incorporating MIMO technology into airborne communication environment. Details of the experimental MIMO system and the transmission performance under different test scenarios are also given in this chapter. Chapter 3 and Chapter 4 describe our work in harvesting RF signals for presence detection. In Chapter 3, the design of the sensing system including pre-processing, the proposed CNN architecture, and post-processing is described in details. Extensive experiment studies are presented to analyze the performance of the proposed system. Chapter 4 presents our efforts in simplifying the training process for presence detection: two SVM based models are proposed which do not require introducing continuous human motions throughout the measurement period. Chapter 5 concludes the dissertation by summarizing the efforts and results with some pointers to future work.

## 1.6 Notation

Throughout the dissertation, we adopt the following notations. Scalars are denoted by either lower or upper case letters, e.g., $a$ and $A$. Column vectors and matrices are denoted by lower and upper case bold letters, e.g., $\mathbf{a}$ and $\mathbf{A}$. The $i$-th entry of a vector $\mathbf{a}$, the $(i, j)$-th entry of $\mathbf{A}$, and the $(i, j, k)$th entry of a 3-D array $\mathbf{A}$ are denoted by $\mathbf{a}_i$, $\mathbf{A}_{i,j}$, and $\mathbf{A}_{i,j,k}$, respectively. $\mathbf{A}_{:,k}$ and $\mathbf{A}_{k,:}$ are used to denote the $k$-th column and the $k$-th row of $\mathbf{A}$. Similarly, $\mathbf{A}_{:,:,k}$ represents the 2-D matrix at depth $k$ of the 3-D array $\mathbf{A}$. $\mathcal{F}$ denotes the discrete Fourier transform (DFT) hence $\mathcal{F}^{-1}$ the inverse DFT. $|A|$ and $\angle A$ denote magnitude and phase of a complex number $A$.

# CHAPTER 2

# VARIATE RATE MIMO FOR ROBUST AERIAL COMMUNICATION

This chapter describes a MIMO transceiver architecture to enable robust and high throughput connectivity for RF communications involving airborne platforms. We are particularly interested in addressing three challenges in airborne MIMO communications: antenna blockage, the presence of interference, and the lack of CSI at the transmitter. Built on D-BLAST MIMO architecture, the proposed system incorporates three major components: spatial spreading to handle antenna blockage, temporal spreading to boost signal to interference and noise ratio (SINR) at the presence of interference and a low rate feedback to drive adaptation at the transmitter. Experimental studies using a developed $4 \times 4$ MIMO system are presented to understand the performance of the proposed transceiver architecture under various channel impairments.

## 2.1 Airborne MIMO System Model

Consider a MIMO system with $n_t$ transmit and $n_r$ receive antennas. We model a narrowband channel by an $n_r \times n_t$ channel matrix $\mathbf{H}$. Thus the received baseband samples $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{2.1}$$

where $\mathbf{x} \in \mathbb{C}^{n_t}$ is the $n_t \times 1$ transmit signal vector, $\mathbf{y} \in \mathbb{C}^{n_r}$ is the $n_r \times 1$ receive signal vector, and $\mathbf{w} \sim \mathbb{C}\mathcal{N}(0, N_0\mathbf{I}_{n_r})$ is the additive white Gaussian noise vector, respectively.

In the absence of CSI (i.e., knowledge of $\mathbf{H}$ at the transmitter), the D-BLAST architecture rotates independent data streams across the transmit antennas, thus effectively averaging out disparate channel states across transmit antenna elements without the need for complete CSI. The D-BLAST architecture, together with successive interference cancellation is known to be optimal in terms of outage capacity [2].

We now describe the two major channel impairments often encountered in aerial communications.

### 2.1.1 Antenna Blockage

Antenna blockage is a unique challenge in airborne MIMO communications; the existence of line-of-sight depends on the antenna orientation which are constantly moving. When a transmit or receive antenna element is completely out of sight, the lack of scattering in an aerial environment often renders the element in effective outage as there may not exist any secondary paths.

Antenna blockage manifests itself in a rank deficient channel matrix. Specifically, if transmit antenna $i$ is in outage, the $i$th column of the channel matrix $\mathbf{H}$ is effectively $0$. Likewise, if receiver antenna $j$ is in outage, the $j$th row of the channel matrix $\mathbf{H}$ is an all zero vector. This is illustrated in Eq. 2.2.

For the original D-BLAST, with transmit element outage, data streams emitted from the element in outage are completely lost. When a receive element is in outage, the receive antenna array may not have sufficient degrees of freedom to resolve independent data streams of the D-BLAST system. In either case, significant performance loss is expected due to the rank deficiency of the channel matrix.

$$\mathbf{H}_{n_r \times n_t} = \begin{bmatrix} h_{1,1} & \cdots & 0 & \cdots & h_{1,n_t} \\ h_{2,1} & \cdots & 0 & \cdots & h_{2,n_t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{n_r,1} & \cdots & 0 & \cdots & h_{n_r,n_t} \end{bmatrix} \tag{2.2}$$

$i^{\text{th}}$ transmit antenna blocking

$j^{\text{th}}$ receive antenna blocking

## 2.1.2 Interference

With a single interference source, the received signal can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{h}^J x^J + \mathbf{w} \tag{2.3}$$

where $x^J \in \mathbb{C}$ is interference signal and $\mathbf{h}^J \in \mathbb{C}^{n_r}$ represents the vector channel from the interference source to the receiver. For this interference model, sophisticated signal processing techniques can be developed to mitigate the interference. For example, if one has knowledge of the channel vector $\mathbf{h}^J$, zero-forcing or other nulling schemes can be applied. This, however, requires knowledge about the interference source which may not be available in a highly dynamic environment. Additionally, multiple interference sources may exist and the receiver array may not have the necessary degrees of freedom to null out all the interference. The presence of interference (or its residue even if pre-processing can be done to partially alleviate their effect) may lead to significant SINR degradation.

## 2.2 Variable Rate MIMO System Design

We describe the variable rate MIMO system intended to combat the channel impairments introduced by antenna blockage and interference. The system consists of three design components integrated into the baseline D-BLAST MIMO system: spatial spreading to counter antenna blockage; temporal spreading to mitigate SINR degradation due to interference; and a low rate feedback to facilitate transceiver adaptation, driven by operating conditions. We describe the three components in details below.

1. Spatial spreading.

   In the absence of transmit CSI, it is infeasible to use any antenna selection scheme in real time for reliable communication. Spatial spreading introduces redundancy across antenna elements, thereby maintaining connectivity even if some transmit elements are in outage. On the other hand, if some receive antenna elements are in outage, spatial spreading at the transmitter reduces the number of effective spatial data streams, ensuring that receiver has sufficient degrees of freedom to resolve independent data streams.

2. Temporal spreading.

   While a myriad of signal processing techniques can be used for mitigating interference, their implementation invariably requires some knowledge of the interference source, e.g., directivity information for zero forcing. With unknown interference, proactive measures need to be taken at the transmitter to boost the effective signal to interference noise ratio (SINR) at the receiver. A simple scheme is the use of temporal spreading (i.e., introducing redundancy in time) to improve SINR at the expense of reduced spectrum efficiency.

3. Low rate feedback for adaptation.

   While spatial and temporal spreading can mitigate the effect of antenna blockage and interference, their integration in the system requires a highly adaptive scheme where the choice of spreading factors matches the operating condition. Since spreading is done at the transmitter, limited feedback is required for transceiver adaptation. A simple low rate feedback is designed whose overhead is much smaller than sending back a quantized version of the CSI.

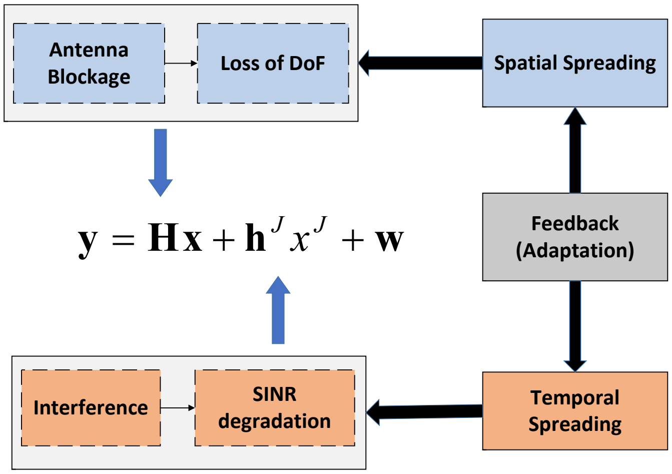
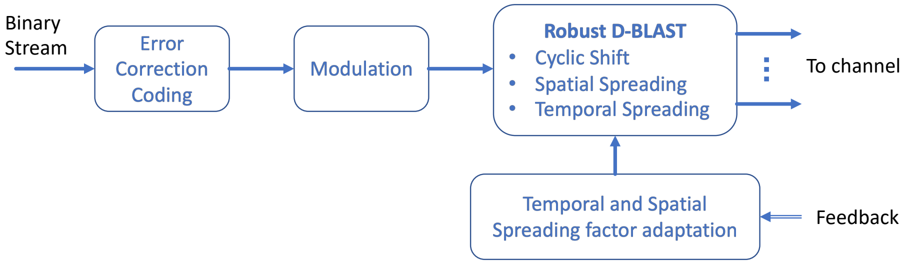


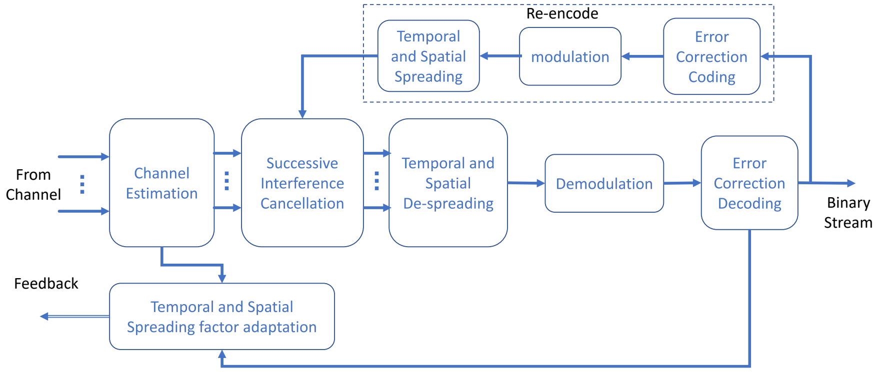
Fig. 1: Overview of the variable rate MIMO system

Fig. 1 summarizes the three design components. Whenever performance loss is observed at the receiver, a low rate feedback scheme is used to drive spreading factor adaptation in order to maintain connectivity. If performance degradation is due to loss of degrees of freedom (i.e., antenna blockage), spatial spreading factor will be increased. On the other hand, if the degradation is due to low SINR (e.g., the presence of interference), temporal spreading factor will be adjusted accordingly.

A high level description of the transceiver architecture of the proposed robust MIMO system is depicted in Fig. 2. At the transmitter side (Fig. 2(a)), spatial and temporal spreading is integrated in D-BLAST processing to introduce redundancy. Spreading factor adaptation is done through a low rate feedback from the receiver. At the receiver side (Fig. 2(b)), current CSI (or the condition number of the channel matrix, which serves as a proxy for



(a) Transmitter



(b) Receiver

Fig. 2: MIMO transceiver block diagram

channel condition) obtained from the channel estimation block is used to determine the desired spatial spreading factor. On the other hand, detection performance (e.g., frame pass rate (FPR), computed with the help of CRC check) or estimated SINR using preambles is used to determine the temporal spreading factor. These spreading factors will be sent back to the transmitter for mode adaptation. We now describe the three components in detail. We will use a $4 \times 4$ MIMO system as an illustration throughout this section, which is consistent with the actual SDR system to be described in Section 2.3.

Table 1: Spatial spreading under $4 \times 4$ MIMO

| SSF | Antenna | Symbols |
|---|---|---|
| 1 | 1 | $x_1$ |
| | 2 | $x_2$ |
| | 3 | $x_3$ |
| | 4 | $x_4$ |
| 2 | 1 | $x_1$ |
| | 2 | $x_1$ |
| | 3 | $x_3$ |
| | 4 | $x_3$ |
| 4 | 1 | $x_1$ |
| | 2 | $x_1$ |
| | 3 | $x_1$ |
| | 4 | $x_1$ |

### 2.2.1 Spatial Spreading

Spatial spreading is a simple repetition scheme where identical symbols are transmitted using multiple transmit antennas. Specifically, antennas are grouped into subgroups; identical symbols are transmitted by antennas within the same subgroup. Grouping can be done strategically by antenna location to maximize diversity among groups. Spatial spreading factor (SSF) is used to determine the number of antennas in a subgroup. With a $4 \times 4$ MIMO system, there are three distinct spreading factors: $1$, $2$ and $4$, and the transmission states corresponding to these SSFs are illustrated in Table 1. Clearly, $\mathrm{SSF} = 1$ amounts to no spreading (i.e., the original D-BLAST) while $\mathrm{SSF} = 4$ amounts to maximum spatial redundancy[1].

### 2.2.2 Temporal Spreading

Temporal spreading is to encode independent data streams using orthogonal codes (e.g., Walsh codes) in order to boost SINR at the receiver. The scheme is simple and requires

---

[1] We note that this simple repetition is different from beamforming as the latter requires CSI that is unavailable at the transmitter.

Table 2: Temporal spreading under $4 \times 4$ MIMO

| TSF | Antenna | Spreading Codes | Symbols |
|---|---|---|---|
| 1 | 1 | $[1]$ | $x_1$ |
|  | 2 | $[1]$ | $x_2$ |
|  | 3 | $[1]$ | $x_3$ |
|  | 4 | $[1]$ | $x_4$ |
| 2 | 1 | $[1, 1]$ | $x_1, x_1$ |
|  | 2 | $[1, -1]$ | $x_2, -x_2$ |
|  | 3 | $[1, 1]$ | $x_3, x_3$ |
|  | 4 | $[1, -1]$ | $x_4, -x_4$ |
| 4 | 1 | $[1, 1, 1, 1]$ | $x_1, x_1, x_1, x_1$ |
|  | 2 | $[1, -1, 1, -1]$ | $x_2, -x_2, x_2, -x_2$ |
|  | 3 | $[1, 1, -1, -1]$ | $x_3, x_3, -x_3, -x_3$ |
|  | 4 | $[1, -1, -1, 1]$ | $x_4, -x_4, -x_4, x_4$ |

no prior information on interference and is driven by the effective receiver SINR through low rate feedback. Temporal spreading factor (TSF) is used to determine the length of the spreading code (hence the level of temporal redundancy) assigned to each antenna. Table 2 describes how temporal spreading is done in the $4 \times 4$ MIMO system under different spreading factors ($1$, $2$ and $4$). Note that temporal spreading factor is not limited by the number of transmit antennas.

The introduced redundancy in the temporal domain improves the effective SINR at the expense of spectrum efficiency. By controlling the spreading factor, the system can balance efficiency and robustness without any knowledge of the interference sources.

### 2.2.3 Low Rate Feedback Enabled Adaptation

The use of spatial and temporal spreading for balancing the need for efficiency and reliability makes it possible to use an extremely low rate feedback to drive the transmission rate adaptation (i.e., selection of spreading factors). With an $n \times n$ MIMO, at most $\log_2 \lceil n \rceil + 1$ SSFs are needed, where $\lceil \cdot \rceil$ is the ceiling function, i.e., the smallest integer greater than or equal to the argument. For example, with $4 \times 4$ system, three SSFs are available: $1$, $2$, and $4$. Thus the total number of bits required to represent those spreading factors is no greater than $\lceil \log_2(1 + \lceil \log_2 n \rceil) \rceil$.

In fact, taking into account the continuity of the operating environment, only three feedback states will be needed for spatial mode adaptation regardless of how large the antenna element number is. Specifically, depending on the channel estimate, the receiver can request that the spreading factor either stays the same, increases, or decreases; thus only two bit feedback is sufficient for spatial mode adaptation regardless of how large $n$ is. The same is true for temporal adaptation, with the difference being that the effective receive SINR (or its proxy such as BER or FPR) is used to determine the temporal spreading factor. Clearly, the required feedback is drastically reduced compared with any system that attempts to feedback quantized CSI to the transmitter.

## 2.3 Experiment

In this section we describe a $4 \times 4$ MIMO communication system that integrates the three design components into the D-BLAST architecture. The software radio implementation utilizes GNU Radio platform [64] (version 3.7) while over the air transmission is enabled by Universal Software Radio Peripheral (USRP) N210 [65]. USRP Hardware Driver (UHD) [66] (version 3.10) is used to facilitate data transfer between host computers and USRPs.

### 2.3.1 Experiment Setup

The $4 \times 4$ MIMO system consists of $8$ USRP N210s, each equipped with an SBX-40 daughterboard (see Fig. 3 for the transmitter and receiver configuration). With Ettus MIMO cables, only two N210s can be connected to form a $2 \times 2$ system. To build a $4 \times 4$ MIMO using N210s, external oscillators and PPS signals are used at both the transmitter and the receiver.

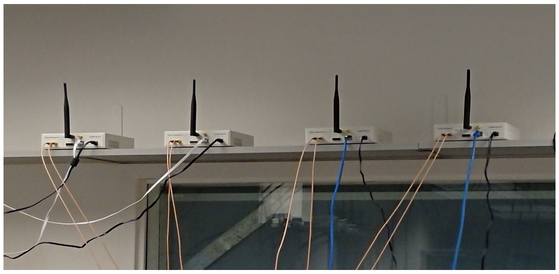
Specifically, an external $10$MHz oscillators from Crystek is used to provide the common frequency reference for the four devices; a micro-controller board Arduino Leonardo generates the PPS signal to synchronize sample time across devices. The transmitter and receiver arrays are placed on shelves at the opposite ends of a rectangular room with a dimension of $39$ feet $\times$ $13$ feet. The shelves are elevated high enough such that there is a strong line of sight component between any transmit and receive antenna element pairs. Throughout this section, the wireless channel for the MIMO experiment uses carrier frequency at $2.45$GHz and bandwidth of $675$ kHz.

A high level abstraction of the transmitter and receiver is given in Fig. 2(a) and Fig. 2(b) respectively. Various processing blocks at the transmitter and receiver are built under the GNU Radio framework. The developed $4 \times 4$ MIMO system operates autonomously when various channel impairments are introduced; a fully functioning feedback mechanism is used to drive the needed adaptation. In addition to spatial and temporal spreading factor adaptation, we have also incorporated modulation adaptation as used in almost all existing wireless systems. Four levels of modulation are incorporated in the system: BPSK, QPSK, 8PSK, and 16QAM. Along with spatial and temporal spreading, these modulations provide even more granularity in balancing the need for spectrum efficiency and robustness[2].

In the following, we describe how the system adapts to various channel conditions and the corresponding communication performance at the receiver. We first examine the bit error rate performance corresponding to various spatial spreading factors under the impairment of antenna blockage with a static radio. Specifically, we repeat the same antenna blockage scenarios for different but fixed spatial spreading factors and examine the corresponding bit error rate performance. The BER difference provides a meaningful guidance on how spatial spreading affects receiver performance when operating condition varies.

---

[2] While coding rate adaptation can also be built into the system, we elect to use a fixed error correction code - BCH$(439, 511)$ code - for simplicity of implementation.



(a) Transmitter

(b) Receiver

Fig. 3: A $4 \times 4$ MIMO setup.

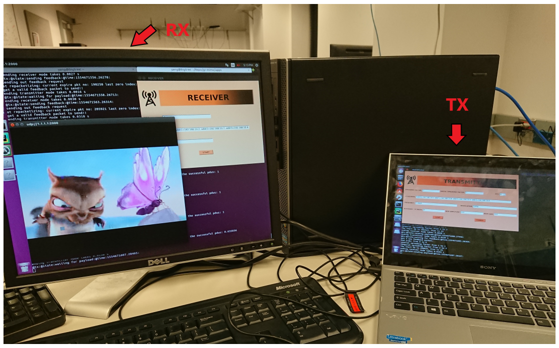


Fig. 4: Streaming setup

### 2.3.2 Bit Error Rates of Different Spreading Factors Under Antenna Blockage

We now examine bit error performance of a MIMO communication system with different but fixed spreading factors under varying antenna blockage scenarios. We emulate antenna blockage by controlling the gain parameters of the software radios that are used to implement the $4 \times 4$ MIMO system. The test environment as described in the previous section is otherwise a static one with stationary scatters. With blocked antenna elements, $\mathbf{H}$ becomes close to rank deficient and the largest singular value becomes increasingly dominant.

We define the following quantity as a measure of how ill-conditioned the channel matrix is:

$$P(\mathbf{H}) = \frac{\sigma_1(\mathbf{H})}{\sum_{i=1}^{4} \sigma_i(\mathbf{H})} \tag{2.4}$$

where $\sigma_i(\mathbf{H})$ for $1 \leq i \leq 4$ are the singular values of $\mathbf{H}$ in decreasing order. Thus $P(\mathbf{H})$ measures the concentration of the channel energy in the dominant eigen channel. The larger it is, the more ill-conditioned the channel matrix becomes. Notice that in this experiment we do not use condition number since it becomes effectively infinity when one or more antennas are taken off-line. Thus the condition number can not distinguish between the cases when either one or two or three antennas are taken off-line.

The experiment starts with no blockage (all antenna elements have equal gains and signal amplitudes). In subsequent phases, antenna gains and signal amplitudes of certain transmit/receive antennas are deliberately controlled such that $P(\mathbf{H})$ gradually increases to its maximum value $1$ (i.e., the channel matrix becomes a rank one matrix). Fig. 5(a) and Fig. 5(b) summarize the performance of three SSFs when antenna blockage happens at the transmitter and receiver respectively. Clearly, spatial spreading helps maintain system connectivity in the presence of antenna outage due to blockage. If we group the antennas strategically according to their relative locations, $\mathrm{SSF} = 2$ can provide reasonable performance when up to two antennas are off-line. At the extreme case, with $\mathrm{SSF} = 4$, while throughput is compromised because of high spatial domain redundancy, connectivity is maintained throughout the communication process despite the lack of knowledge of the exact antenna outage at the transmitter and receiver.

### 2.3.3 A Fully Adaptive System for Video Streaming

We now examine the performance of the overall system through a comprehensive experiment with video streaming and under various channel impairments. The channel impairments are designed to emulate conditions encountered in an aerial communications link. The received signals are then parsed off-line to obtain various performance metrics to fully understand the behavior of the developed MIMO system. For video streaming, the laptop that controls the transmitter (at the right in Fig. 4) uses FFmpeg tool [67] to break down the video file into UDP packets, which are transmitted using the $4 \times 4$ MIMO system. At the receiver (left in Fig. 4), the delivered UDP packets are decoded and displayed using the media player (FFplay tool [67]). Thus the transmission state is represented by the triple (SSF, TSF, Mod) which denote respectively spatial spreading factor SSF$\in \{1, 2, 4\}$, temporal spreading factor TSF$\in \{1, 2, 4\}$, and modulation index Mod$\in \{0, 1, 2, 3\}$ corresponding to BPSK, QPSK, 8PSK, and 16QAM respectively (c.f. Table 3).

### *Impact of Ill-conditioned Channels*

In this section, we describe the video streaming experiment in the presence of antenna blockage. Again, antenna blockage is done through deliberately altering the transmit power of various antenna elements, including taking them completely offline, thus emulating the situation when some antennas may experience partial or complete blockage.

For this section, we use the condition number $\kappa(H)$ of the estimated channel matrix as a proxy for channel conditions: a small (close to 1) condition number suggests a full-rank channel matrix environment while a large condition number indicates an ill-conditioned channel that is close to rank deficient. The reason that condition number can be used here is because the experiment is controlled in a way that at most one antenna may be taken completely off-line. The experiment is conducted in a single run that lasts about three minutes and consists of five different stages. Each stage corresponds to a different channel condition by altering the transmit power to emulate different level of antenna blockage.

For the receiver performance, we use frame pass rate (FPR) to track the reception performance. While frame duration is fixed, each frame may contain different numbers of UDP packets since modulation order and spreading factors are varying (adapting) depending on the operation condition. FPR directly correlates to the video streaming performance as dropped frames will cause video to freeze or pixelate.

Table 3: Adaptation of (SSF, TSF, Mod) over five stages

| stage | adaptation | $\kappa(H)$ |
|---|---|---|
| 1 | $(1,1,1) \rightarrow (1,2,1) \rightarrow (1,2,2) \rightarrow (1,2,1)$ | $[10,15]$ |
| 2 | $(1,2,1)$ | $[20,35]$ |
| 3 | $(1,2,1) \rightarrow (1,4,1) \rightarrow (1,4,2) \rightarrow (1,4,3)$ | $[30,45]]$ |
| 4 | $(1,4,3) \rightarrow (2,2,3)$ | $[1000,\infty)$ |
| 5 | $(2,2,3) \rightarrow (1,2,3) \rightarrow (1,4,3)$ | $[33,50]$ |

Table 3 gives a summary of the adaptation corresponding to the five stages which we describe in details below.

- Stage 1 This was the initial stage and the four USRPs had equal transmit power (amplitude=$0.2$, gain=$20$dB) in this stage. The transmitter started with both temporal spreading factor and spatial spreading factor equal to $1$ and QPSK is the default modulation scheme. Since $\kappa(H)$ is much larger than $1$ ($[10,15]$), none of the frames can be decoded initially and streaming does not commence with the original setting. Temporal spreading was requested through feedback, and once TSF is increased to $2$, streaming started as FPR improves to nearly $100\%$. As a side note, the system also attempted to move to higher order modulation, leading to a decrease of FPR to around $83\%$. This drove the modulation back to QPSK in subsequent frames through feedback adaptation.

- Stage 2 At the onset of Stage 2, we manually decreased the gain at transmit antenna $1$ by $10$dB; this emulates the situation when partial antenna blockage occurs. However, with TSF equal to $2$, the reception performance was not negatively affected thus the transmission mode stayed the same. Video streaming continued uninterrupted.

- Stage 3 We deteriorated the channel condition further by reducing the tx gain at antenna $4$ by $12$dB. At the onset of this change, the quality of the streaming video was very poor which is consistent with the steep drop of the FPR at the beginning of Stage 3. The system quickly adapted to increase TSF to $4$, which was reduced to effectively transmitting a single data stream. However, once its FPR moved back to $100\%$, modulation order was adapted from QPSK to 8PSK and then to 16QAM which compensated for the throughput loss due to the increased TSF.

- Stage 4 This was a stage that we forced complete antenna outage. Specifically, antenna $1$'s amplitude was set at $0$, which effectively silenced its transmission. As seen from the table, the corresponding condition number for Stage 4 is exceedingly large, indicating a completely rank deficient channel matrix. Spatial spreading was enabled through feedback to compensate for the rank deficiency of the channel matrix. Whenever the system increases spatial spreading, it would first try to lower temporal spreading to balance the need for spectrum efficiency. In this case, it turned out reducing TSF to $2$ can sustain a frame success rate close to $100\%$. Again, video streaming quality directly correlates with the FPR: it deteriorated initially at the beginning of Stage 4 and then recovered once the transmission adaptation was complete.

- Stage 5 This was a stage when the outage element (antenna $1$) was brought back online (i.e., end of blockage). As expected, the adaptation eventually led to the same transmission state as in stage 3 and video streaming continued uninterrupted when the transmission mode stabilized.

Fig. 6 summarizes the performance and spreading adaptation for the entire five stages through parsing of the received signal. The frame pass rate (hence video streaming quality) often suffers at the onset of a change in channel condition or the beginning of a new transmission mode. Nevertheless, the built-in adaptation always leads to the desired transmission mode for successful video streaming, demonstrating the robust connectivity in the presence of dynamic channel impairments.

### *Impact of Interference*

To emulate an interference environment, an additional USRP N210 was placed on the shelf of a side wall as shown in Fig. 7. Its distances to the transmitter and the receiver are respectively 13 feet and 26 feet.

In the experiment, the interference source is transmitting randomly generated QPSK signals. We fix the transmit power of the MIMO transmitter while gradually adjusting the transmission power of the interference source to observe how the system reacts to different interference powers (or equivalently, SINR at the receiver). Denote by $p_s$ the average receive power at a receive USRP when only the MIMO transmitter is transmitting and $p_i$ the average receive power when only the interference source is transmitting. The interference state is represented by the difference $p_s - p_i$.

#### Experiment Part I

We examine in this part how well the MIMO system with different temporal spreading factors perform in combating interference, thus feedback is disabled and the modulation scheme is fixed as QPSK throughout this experiment.

Table 4 summarizes the performance of different configurations in three interference states. We omit SSF since in the lab environment, while interference degrades SINR, no antenna blockage occurs hence no spatial spreading is needed. The first row corresponds to the state that there is no interference. The second and third rows correspond to different interference powers; both temporal spreading factors (2 and $4$) give reasonable performance though $4$ is clearly better than 2.

#### Experiment Part II

Upon observing the system performance with the three states separately, a single experiment run is conducted with three interference states sequentially occurring. Adaptation via feedback is enabled for this part. The results are given in Fig. 8. The SSF stayed at

1 throughout the experiment since no antenna blockage was detected. Also, no adaptation in the modulation scheme was observed (i.e., QPSK is always used). For the initial state (no interference), the system quickly locked into the desired TSF at 2. As the interference state transitioned into a low interference power (i.e., State 2), FPR exhibited some initial fluctuation but quickly stabilized without the need for adjusting the spreading factor. Once the interference power increased further (State 3), TSF was increased to compensate for the decreased SINR. The adaptation is consistent with the results in Table 4.

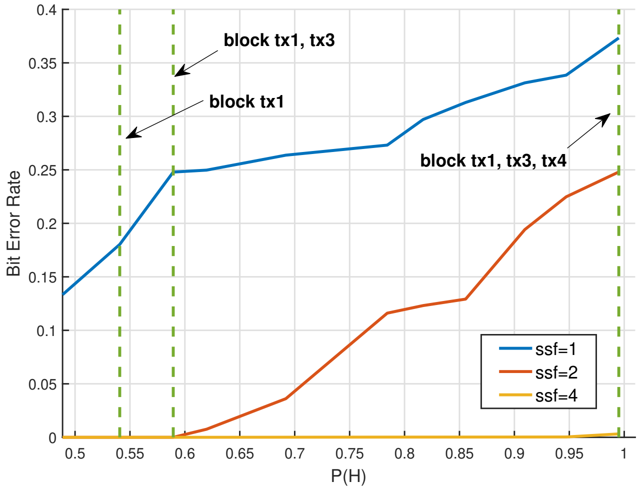
To understand the advantage of the MIMO system against interference, the same experiment was repeated for a single input single output (SISO) system. For the SISO experiment, the transmit antenna is fixed, and the four receive antennas are tested sequentially. The reception performance was quite disparate even with temporal spreading factor at the maximum value 4. Three out of the four receivers were in complete outage (i.e., could not even synchronize with the header) when the interference is present while the fourth one had reasonable performance. Clearly, with SISO, the reception is completely at the mercy of SINR as the signal and interference overlap completely. With MIMO receiver, spatial diversity provides desired immunity to interference.

Table 4: Error performance with interference

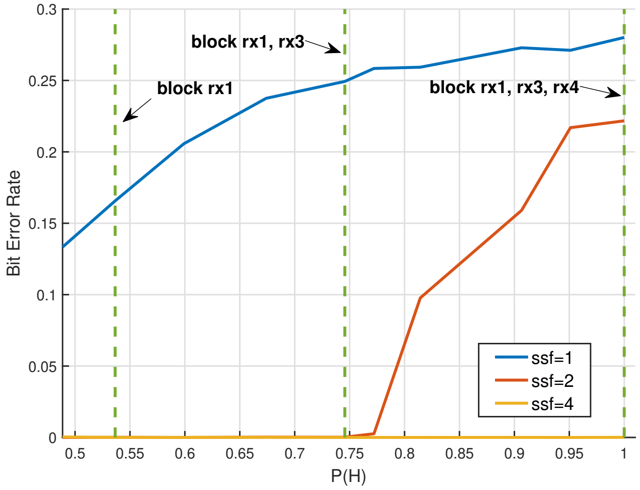
| state index | power diff | TSF | FPR | BER |
|---|---|---|---|---|
| 1 | $N/A$ | 2 | 0.9911 | 0.0003 |
| | | 4 | 1.0 | 0.0 |
| 2 | 9dB-10dB | 2 | 0.9188 | 0.0044 |
| | | 4 | 0.9995 | 0.00002 |
| 3 | 4dB-5dB | 2 | 0.7047 | 0.0187 |
| | | 4 | 0.9528 | 0.0051 |

## 2.4 Summary

This chapter presented a fully functioning MIMO system that integrates three design components for realizing MIMO throughput gain in airborne communication environment. Specifically, antenna blockage, the presence of interference and the lack of CSI at the transmitter are addressed through the design of an adaptive MIMO system that integrates temporal and spatial spreading into the classical D-BLAST architecture, ensuring autonomous operation under dynamically varying channel conditions. Experimental results using the developed MIMO system are presented in the presence of emulated antenna blockage and interference. The performance confirms its distinct advantage over the classical D-BLAST in maintaining persistent connectivity under various channel impairments.



(a) transmit antenna blockage



(b) receive antenna blockage

Fig. 5: Performance of different SSFs under $4 \times 4$ MIMO

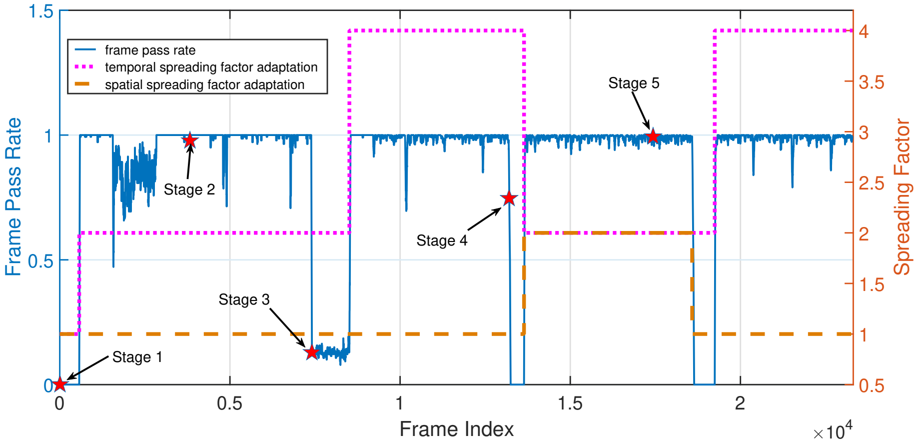


Fig. 6: Reception performance over five different stages of antenna blockage.

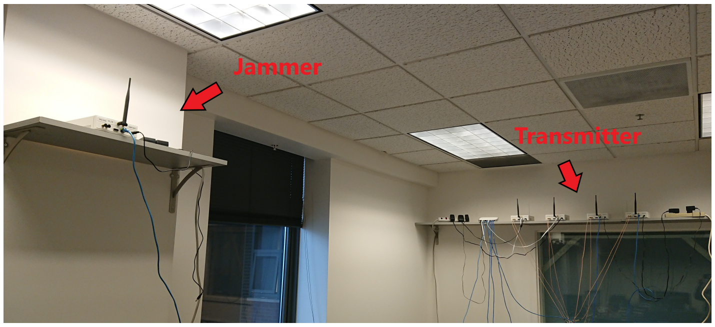


Fig. 7: Interference setup

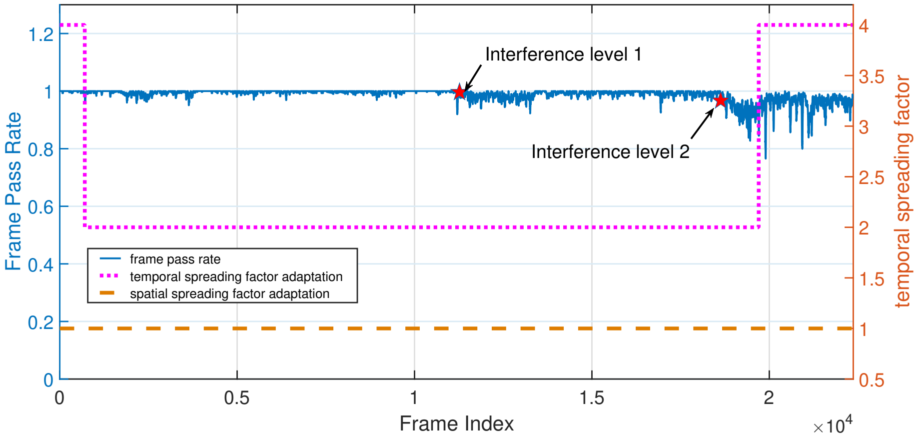


Fig. 8: Reception performance over three stages of interference level.

# CHAPTER 3

# PRESENCE DETECTION USING DEEP LEARNING

This chapter presents our work in leveraging deep learning technique for achieving presence detection using RF signals. The design of the learning system aims at addressing several important challenges in passive RF sensing: how to collect data representing human presence; how to handle high-dimensional and complex CSI time series; and how to mitigate impacts on CSI from factors irrelevant to human presence. In the proposed system, pre-processing is designed to preserve variations induced by human motions. A CNN model with parallel architecture is then utilized to infer human presence from both processed CSI magnitude images and CSI phase images. Based on the assumption that humans are not still for an extended period of time, post-processing accumulates motion detection outputs from CNN and outputs final presence detection results. Extensive experiments are conducted using COTS WiFi devices and demonstrate much more favorable performance compared with that of PIR sensors.

## 3.1 MIMO-OFDM System Model

### 3.1.1 MIMO-OFDM

Consider a MIMO-OFDM system with $N_t$ transmit antennas, $N_r$ receive antennas, and $N_{sc}$ subcarriers. Each physical layer frame consists of $M$ OFDM symbol blocks. Denote by $\mathbf{d}^p[m,i]$ the $m$-th frequency domain OFDM symbol vector in the $i$-th frame sent by the $p$-th transmit antenna. The discrete-time complex baseband signal corresponding to $\mathbf{d}^p[m,i]$ is given by $\mathbf{s}^p[m,i] = \mathcal{F}^{-1}\left(\mathbf{d}^p[m,i]\right)$. At the receiver, after cyclic prefix removal and applying DFT, the complex baseband sample at the $q$-th receive antenna in the frequency domain can be expressed as, for $k = 0, 1, \ldots, N_{sc}-1$,

$$\mathbf{y}_k^q[m,i] = \sum_{p=0}^{N_t-1} \mathbf{h}_k^{q,p}[i]\mathbf{d}_k^p[m,i] + \mathbf{v}_k^q[m,i], \tag{3.1}$$

where $\mathbf{v}_k^q[m,i]$ is the channel noise and $\mathbf{h}_k^{q,p}[i]$ the channel coefficient from the $p$-th transmit antenna to the $q$-th receive antenna on the $k$-th subcarrier, assumed to be a constant within one frame (i.e., for all $M$ OFDM blocks within the $i$-th frame). Expressed in vector form, we have

$$\mathbf{y}_k[m,i] = \mathbf{H}_{k,:,:}[i]\mathbf{d}_k[m,i] + \mathbf{v}_k[m,i],$$

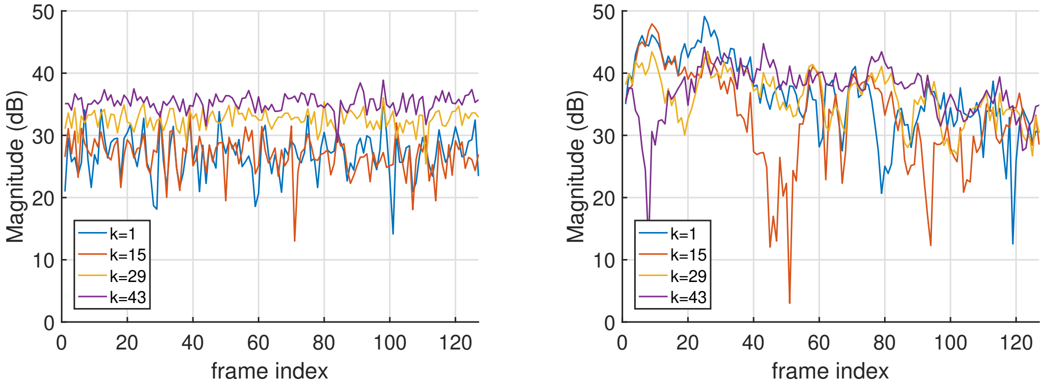
where $\mathbf{H}[i]$ is the 3-D channel array of shape $N_{sc} \times N_r \times N_t$ corresponding to the $i$-th OFDM frame, $\mathbf{H}_{k,q,p}[i] = \mathbf{h}_k^{q,p}$, $\mathbf{d}_k[m,i] = \left[\mathbf{d}_k^0[m,i], \ldots, \mathbf{d}_k^{N_t-1}[m,i]\right]^T$ and $\mathbf{y}_k[m,i] = \left[\mathbf{y}_k^0[m,i], \ldots, \mathbf{y}_k^{N_r-1}[m,i]\right]^T$.

### 3.1.2 Effect of Human Motion on MIMO-OFDM Channel

Human motion leads to CSI variation in both the frequency (across subcarriers) and temporal (across frames) domains. The presence and movement of humans introduce new paths whose delays are affected by human locations, leading to change in path delay profile, hence the change in channel frequency response. Human movement also induces temporal CSI variation as signals may add in-phase or out-of-phase at the receiver depending on the locations of humans. An alternative interpretation is the increase of Doppler spread due to human movement in an otherwise static environment, leading to time-selective channel fading [68]. An example using real WiFi measurement of the variation of $|\mathbf{H}_{k,q,p}[i]|$ over frame index $i$ with and without human movement for fixed $q$ and $p$ is shown in Fig. 9 for four evenly spaced subcarrriers. Clearly, with movement, channel variation both in frequency (across subcarriers) and in time (along the horizontal axis) increases.

The effect of human movement on the CSI in the spatial dimension is more subtle. The human motion induced CSI variation in the temporal and frequency domains applies to every transmit-receive antenna pair. The fact that multiple transceiver pairs (*a.k.a*, spatial diversity) exist in the MIMO-OFDM system should be exploited for enhanced sensing performance. CNN is a natural choice for exploiting such spatial diversity by mapping temporal-frequency CSI corresponding to each transceiver pair to a layer ('channel') in a CNN architecture, much like the way colored images are processed in a CNN where RGB pixels serve as separate channels.

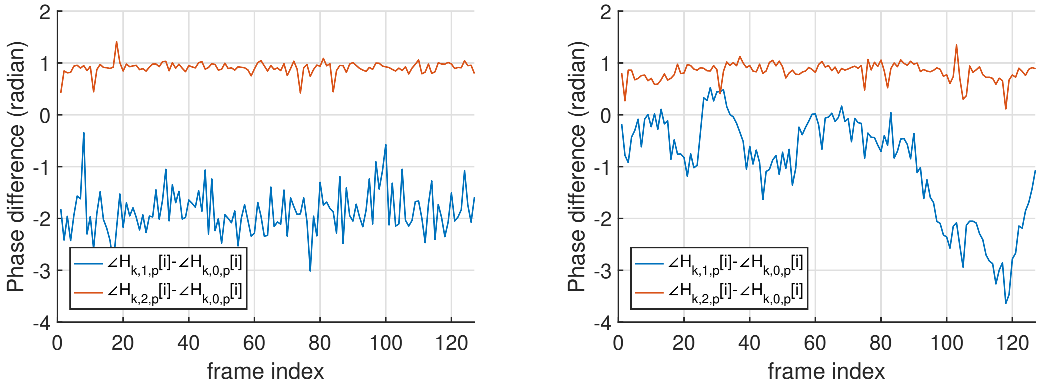
Multiple antennas at WiFi transceivers are also exploited in this paper to make the phase information of CSI estimate much more useful for presence detection. As WiFi devices use a single oscillator for RF circuitry corresponding to different antennas, the CFO, if present, is common to all inputs at different receive antennas. Similarly, sampling is driven by a single clock, hence STO is also identical for all inputs at different receive antennas. Thus instead of using the raw CSI phase measurement, one can use phase differences between receive antennas to remove phase variation due to CFO and STO. While such processing has no effect on digital communication performance (e.g., it does not correct residual CFO/STO for each receive chain for symbol detection), it cleans up the phase information when phase variation due to human movement is of interest. An example of phase differences is given in Fig. 10 with $N_r = 3$ where the CSI phase from the first antenna serves as a reference. Clearly, phase differences stay relatively stable in a human-free environment whereas human motion introduces significant fluctuation to the relative phases across three receive antennas. Fig. 10(b).



(a) Human-free environment

(b) Human movement

Fig. 9: CSI magnitude variation over time for four evenly spaced subcarriers.



(a) Human-free environment

(b) Human movement

Fig. 10: CSI phase difference between antennas variation over time

Figs. 9 and 10 indicate that both magnitude and phase of estimated CSI contain rich information about human motion. While in theory, deep learning trained using labeled data appears to be a straightforward exercise, the challenge is that for presence detection, there is no clearly defined human motion that one tries to detect. As such, collecting labeled training data with human presence needs to be carefully addressed along with the design of the learning system for presence detection.

## 3.2 CNN Based Detection System Design

A high level description of the proposed system is depicted in Fig. 11. Consecutive CSIs are first arranged into CSI magnitude and phase images. They are processed separately and fed into the CNN learning block comprised of two parallel CNNs - one for magnitude images and the other for phase images - followed by fully connected layers (c.f. Fig. 12 and Section 3.2.2). The post-processing block accumulates the instantaneous detection results provided by the CNN and output the final presence detection depending on required time resolution.

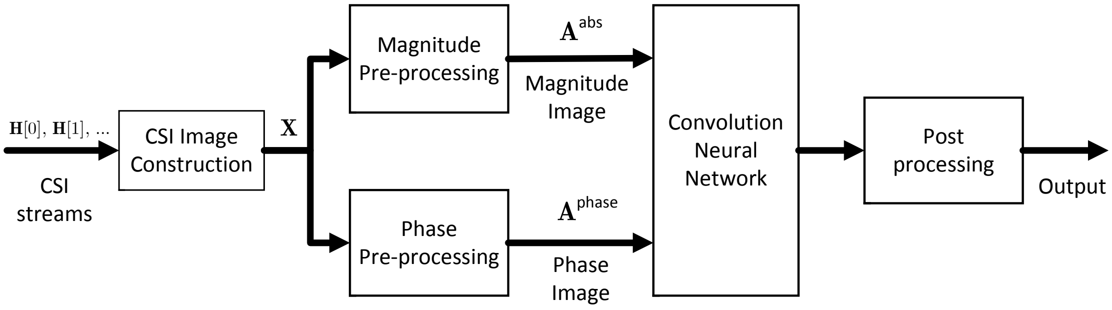


Fig. 11: Flowgraph of the CNN based detection system

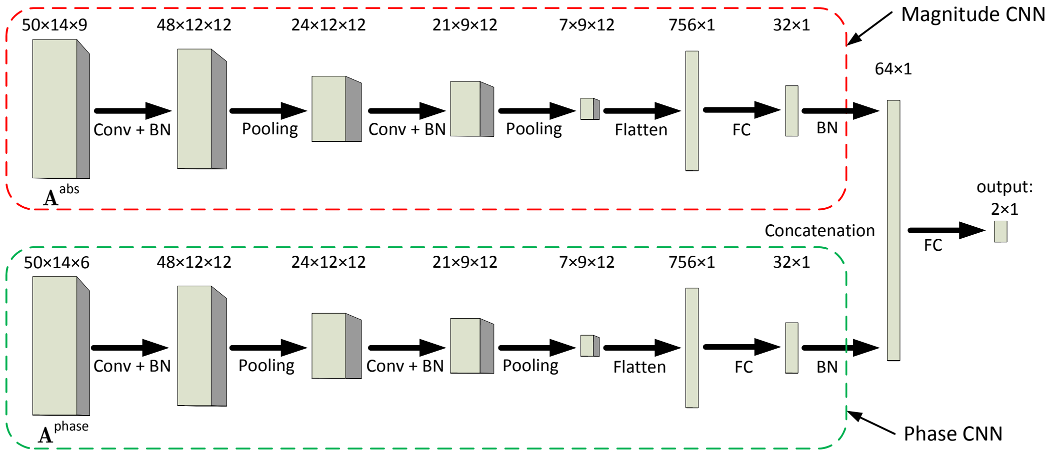


Fig. 12: Architecture of the proposed CNN

### 3.2.1 Input Pre-processing

Recall that $\mathbf{H}[i]$ is an $N_{sc} \times N_r \times N_t$ array consisting of MIMO channel matrices across all subcarriers for the $i$-th frame. Instantaneous motion detection is based on $\mathbf{H}[i]$ collected over $I$ consecutive frames, denoted as $\mathbf{H}[0], \cdots, \mathbf{H}[I-1]$. Here we assume without loss of generality (WLOG) the first CSI array has frame index $0$. For each $\mathbf{H}[i]$, we select $N_f$ evenly spaced subcarriers out of $N_{sc}$ subcarriers, resulting in $\tilde{\mathbf{H}}[i]$ with size $N_f \times N_r \times N_t$. Down selection of subcarriers significantly reduces the data dimension yet does not have any negative effect in sensing performance. This is because the carrier spacing in WiFi signals ($312.5$kHz) is much smaller than the coherent bandwidth in a typical indoor (i.e. low mobility) environment, thus the behaviors of subcarriers that are immediate neighbors closely track each other with or without human motion. The resulting $\tilde{\mathbf{H}}[i]$ are subsequently stacked up along the temporal domain to form a 4-D array $\mathbf{X}$ of size $I \times N_f \times N_r \times N_t$. The magnitude and phase information are then extracted from $\mathbf{X}$ prior to independent pre-processing.

#### *CSI magnitude*

We reshape the 4-D array $|\mathbf{X}|$ into a 3-D array by combining the last two spatial dimensions, i.e., channel matrix for each subcarrier is flattened into a 1-D array. The obtained array, denoted by $\mathbf{X}^{\text{abs}}$ is of size $I \times N_f \times (N_r N_t)$.

Pre-processing $\mathbf{X}^{\text{abs}}$ involves normalization and transformation. Normalization is done to remove dependence of the absolute CSI magnitude on various environment parameters that are irrelevant to presence detection. For example, the dynamic range of $\mathbf{X}^{\text{abs}}$ is highly dependent on the distance between the transmitter and receiver and the existence of line of sight transmission. While various normalization methods can be used, we find through extensive experiments the following offers the most robust performance: for $i = 0, \cdots, I-1$,

$$\tilde{\mathbf{X}}^{\text{abs}}_{i,:,:} = \mathbf{X}^{\text{abs}}_{i,:,:}./\mathbf{X}^{\text{abs}}_{0,:,:}, \tag{3.2}$$

where $./$ denotes element-wise division. Note that $i$ indexes OFDM frame, thus the normalization is done with respect to the first OFDM frame within the $I$ frames contained in $\mathbf{X}^{\text{abs}}$.

Subsequently, a 2-D DFT is applied to $\tilde{\mathbf{X}}^{\text{abs}}_{:,:,j}$ along the temporal (frame) and frequency (subcarrier) dimensions, resulting in the output array for each transceiver antenna pair:

$$\tilde{\mathbf{X}}^{\text{abs-fft}}_{:,;,j} = \mathcal{F}\left(\tilde{\mathbf{X}}^{\text{abs}}_{:,:,j}\right).$$

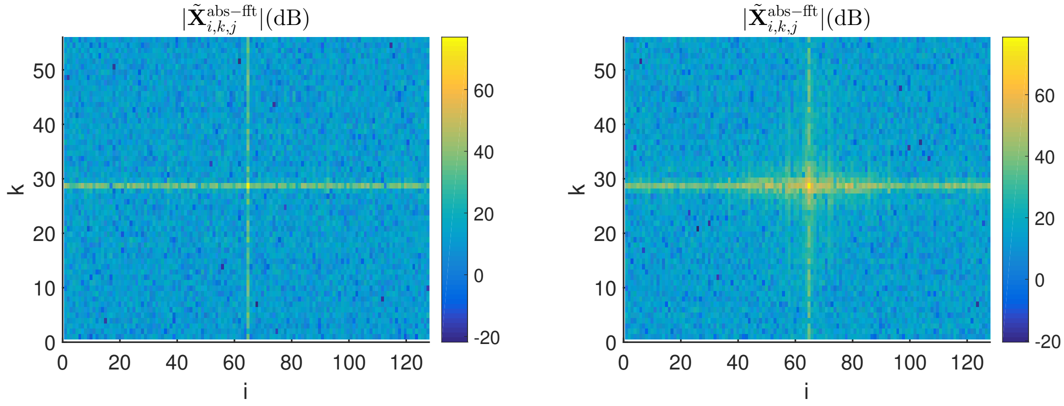
Here the DFT output is properly shifted so that zero frequency is at the center of the array. The use of 2-D DFT serves two purposes. First, human motion induced temporal variation of CSI is continuous in nature. As such, it results in dispersion in the lower frequency region along the temporal dimension. This is in contrast to hardware impairment and channel estimation error when sudden change of CSI may be observed irrespective of human presence. Therefore, high frequency change can be removed by simple cropping of the DFT output along the temporal dimension around zero frequency:

$$\tilde{\mathbf{X}}^{\text{abs-fft-crop}}_{i,:,:} = \left|\tilde{\mathbf{X}}^{\text{abs-fft}}_{\frac{I-T}{2}+i,:,:}\right|, \tag{3.3}$$

where $i = 0, \ldots, T-1$, and $T$ is the cropping window size. Here we assume WLOG that both $I$ and $T$ are even numbers. Cropping also significantly reduces the image size, leading to faster learning and reduced storage requirement. This makes the learning suitable to be implemented on edge devices instead of having to resort to cloud services. Note that further reduction of image size can be achieved by utilizing the conjugate symmetry of the 2-D FFT due to the fact that input to the FFT is real-valued (i.e., magnitude of CSI arrays).

Another reason of using 2-D DFT is its ability to localize motion related CSI variation.

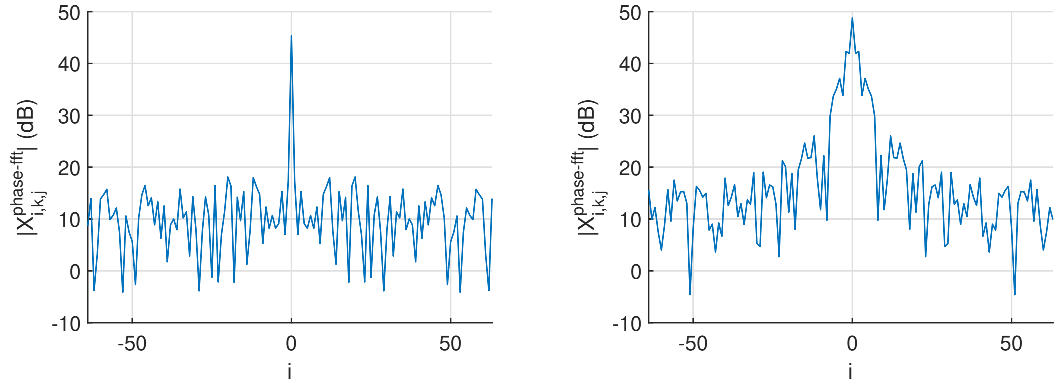
While temporal variation in $\tilde{\mathbf{X}}^{\text{abs}}_{:,:,j}$ is exhibited for the entire $I$ frames, 2-D DFT concentrates such variation into the low frequency region. This is particularly suitable for CNN given its ability to build discriminating ability on local features. Figs. 13(a) and 13(b) provide a sample of $\left|\tilde{\mathbf{X}}^{\text{abs-fft}}\right|$ collected in the same room without and with human motions. One can see that, in the temporal dimension, the $2$-D DFT using data collected in an empty room is dominated by the DC component. With human movement, there is clearly dispersion at low frequency region in the temporal (horizontal) dimension.



(a) Human-free environment

(b) Human movement

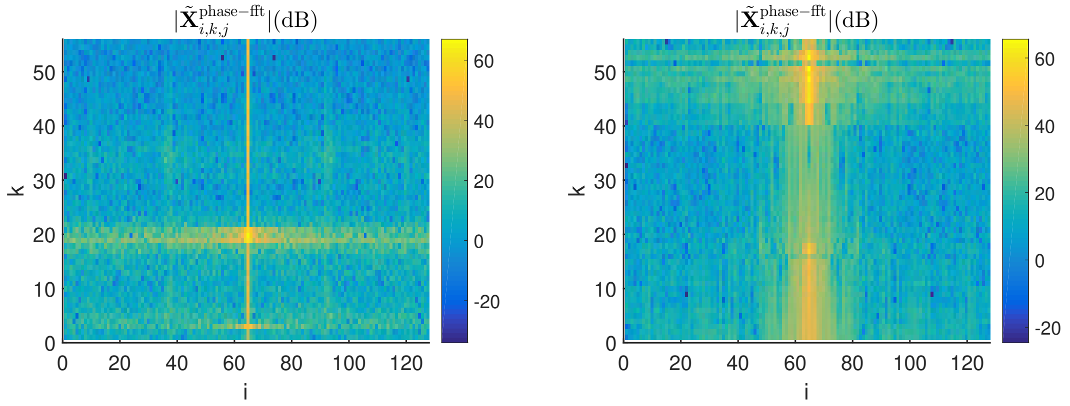
Fig. 13: 2D DFT of CSI magnitude along frame and subcarrier



(a) Human-free environment

(b) Human movement

Fig. 14: DFT of CSI phase difference at a fixed subcarrier



(a) Human-free environment

(b) Human movement

Fig. 15: DFT of CSI phase difference along time at all subcarriers

### *CSI phase*

Even with a completely static environment, the estimated CSI phase will undergo variation (e.g., from residual CFO and STO) which may lead to abrupt changes within $(-\pi, \pi]$. This can be partially resolved by phase unwrapping which removes such abrupt changes. However, phase unwrapping does not remove phase variation introduced by any residual CFO and STO offset, but merely correct discontinuous phase jumps. Thus CSI phases are often discarded for WiFi sensing [28, 29, 31] because of this “noisy" nature.

However, a simple pre-processing that computes phase difference with respect to a reference receive antenna can largely mitigate this problem due to the fact that CFO and STO are common to all receive antennas (see Section 3.1.2). Denote by $\mathbf{X}^{\text{phase}}$ the phase difference between $\tilde{\mathbf{H}}_{k,q,p}[i]$ for different $q$

$$\mathbf{X}^{\text{phase}}_{i,:,q-1,:} = \angle(\tilde{\mathbf{H}}_{:,q,:}[i]./\tilde{\mathbf{H}}_{:,0,:}[i]), \tag{3.4}$$

where $q = 1, \ldots, N_r - 1$. The last two spatial dimensions of $\mathbf{X}^{\text{phase}}$ are then flattened into one dimension and phase is unwrapped along the time axis to remove discontinuity at boundary points $-\pi$ and $\pi$. The obtained result is denoted by $\tilde{\mathbf{X}}^{\text{phase}} \in \mathbb{R}^{I \times N_f \times (N_r - 1)N_t}$. Different from CSI magnitude, only 1-D DFT along the temporal dimension is performed on $\tilde{\mathbf{X}}^{\text{phase}}$ to get $\tilde{\mathbf{X}}^{\text{phase-fft}}_{:,k,j} = \mathcal{F}\left[\tilde{\mathbf{X}}^{\text{phase}}_{:,k,j}\right]$ since phase unwrapping weakens relation of CSI phase across different subcarriers. An example of $\left|\tilde{\mathbf{X}}^{\text{phase-fft}}\right|$ is given in Figs. 14 and 15 where significantly increased dispersion of the DFT output along the temporal dimension can be observed with human movement.

The following steps are similar to how we obtain $\tilde{\mathbf{X}}^{\text{abs-fft-crop}}$, where we shift the zero frequency component to the center and crop out the high frequency components in the temporal domain, leading to the following CSI phase information

$$\tilde{\mathbf{X}}^{\text{phase-fft-crop}}_{i,:,:} = \left|\tilde{\mathbf{X}}^{\text{phase-fft}}_{\frac{I-T}{2}+i,:,:}\right|, \tag{3.5}$$

where $T$ is chosen to be the same as that in (3.3).

### *Image Normalization*

DFT typically results in increased dynamic range of $\tilde{\mathbf{X}}^{\text{abs-fft-crop}}$ and $\tilde{\mathbf{X}}^{\text{phase-fft-crop}}$. Elements with low intensity are easily overwhelmed by those with large values. The logarithmic operator $y = \log_{10}(x+1)$ is applied to each element in both images [69] to reduce such disparity.

$$y = \log_{10}(x+1),$$

where $x \geq 0$. The final input to the two parallel CNNs are

$$\begin{aligned} \mathbf{A}^{\text{abs}[0]} &= \log_{10}(\tilde{\mathbf{X}}^{\text{abs-fft-crop}} + 1), \\ \mathbf{A}^{\text{phase}[0]} &= \log_{10}(\tilde{\mathbf{X}}^{\text{phase-fft-crop}} + 1). \end{aligned} \tag{3.6}$$

Here the superscript "[0]" is introduced to denote the layer index (i.e., input layer, c.f. Section 3.2.2).

### 3.2.2 Architecture of CNN

The architecture of the proposed CNN is shown in Fig. 12. Magnitude and phase images in (3.6) are fed into two parallel CNNs which share the same structure. The output of the two CNN's are then concatenated and fed to the fully connected layers as specified in Fig. 12. Details about each layer are described below. Denote by $\mathbf{A}^{[l]}$ or $\mathbf{a}^{[l]}$ the output of the $l$-th layer depending on whether the output is a matrix or vector for the $l$-th layer. Note that the output of the $l$-th layer is the input of the $(l+1)$-th layer.

#### *Convolution (Conv) layer*

Consider the $l$-th layer of the CNN which is a Conv layer. Assume that the input $\mathbf{A}^{[l-1]}$ has size $(n_h^{[l-1]}, n_w^{[l-1]}, n_c^{[l-1]})$. Denote by $\mathbf{K}^{[l](u)} \in \mathbb{R}^{d_h^{[l]} \times d_w^{[l]} \times d_c^{[l]}}$ for $u = 0, \ldots, U-1$ the kernel functions of the $u$-th output channel, where $n_c^{[l-1]} = d_c^{[l]}$. In the following discussion, $d_c^{[l]}$ will be omitted if no ambiguity arises. The output of the convolution, denoted by $\mathbf{Z}^{[l]}$, is given as

$$\mathbf{Z}_{i,j,u}^{[l]} = \sum_{q,t,d} \mathbf{A}_{is_h^{[l]}+q, js_w^{[l]}+t, d}^{[l-1]} \mathbf{K}_{q,t,d}^{[l](u)} + \mathbf{b}_u^{[l]},$$

where $\mathbf{b}_u^{[l]}$ is the learnable bias term corresponding to $\mathbf{K}^{[l](u)}$. Fig. 16 shows an example of how $\mathbf{Z}^{[l]}$ is calculated, where only one kernel with size $(2, 2)$ and stride $(1, 1)$ is applied to the input and the bias term is assumed to be zero. By applying the same kernel functions to different locations in the input, Conv layer effectively captures similar features across the input image and at the same time significantly reduce the number of parameters to be learned during the training phase. In the proposed system, kernel sizes for the Conv layers 1 and 2 are $(3, 3)$ and $(4, 4)$ respectively and 12 kernels are used for each Conv layer. Due to small image size, stride always equals to $(1, 1)$ for the proposed CNN.

Define $g^{[l]}(\cdot)$ the activation function of the $l$-th Conv layer, to be introduced later. The output of the $l$-th Conv layer is given by

$$\mathbf{A}_{i,j,u}^{[l]} = g^{[l]}\left(\mathbf{Z}_{i,j,u}^{[l]}\right).$$

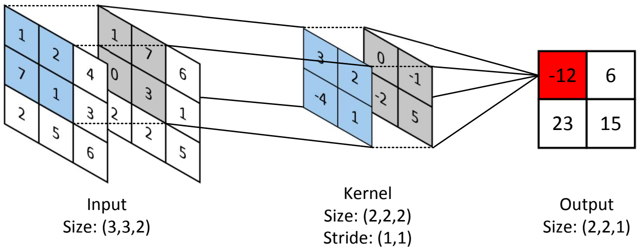


Fig. 16: An example of Conv layers

### *Pooling layer*

In CNN, each Conv layer is usually followed by a pooling layer. In each pooling layer, a pooling window scans through the input image with a pre-defined stride. At each location, the generated output is a single value for each channel. Two common pooling functions are max-pool and average-pool. In the max-pool, maximum value within the rectangular region is kept, while in the average-pool, average value is calculated. Pooling layers not only help reduce the input dimension, but also make the system more robust against variation within small regions in the image since the output only keeps the most dominant or average features. Fig. 17 contains one simple pooling layer where only average value in the window size of $(2, 2)$ is sent to the output.

In the proposed CNN, Conv layers 1 and 2 are followed by average pooling layers with pooling size $(2, 1)$ and $(3, 1)$ respectively.

### *Batch normalization (Batch Norm)*

Due to the large size of training set, input data is divided into disjoint mini-batches with size $N_{mb}$. When normalization is applied, features are normalized by their mean and variance

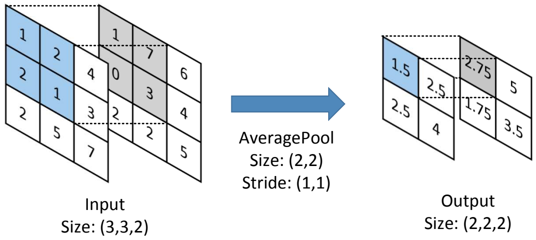


Fig. 17: An example of average pooling layers

in the current batch. Suppose that the $l$-th layer is a Batch Norm layer. Let $\mathbf{a}^{[l-1]}[i]$ denote the $i$-th input sample in the current mini-batch which has $K$ features. Then the normalized samples are given by

$$\mathbf{y}_k^{[l]}[i] = \frac{\mathbf{a}_k^{[l-1]}[i] - \mu_k^{[l-1]}}{\sqrt{\left(\sigma_k^{[l-1]}\right)^2 + \epsilon}}, \tag{3.7}$$

where $\epsilon$ is a small positive number, $k = 1, \ldots, K$, and

$$\mu_k^{[l-1]} = \frac{1}{N_{mb}} \sum_{i=1}^{N_{mb}} \mathbf{a}_k^{[l-1]}[i],$$
$$\left(\sigma_k^{[l-1]}\right)^2 = \frac{1}{N_{mb}} \sum_{i=1}^{N_{mb}} (\mathbf{a}_k^{[l-1]}[i] - \mu_k^{[l-1]})^2.$$

The output of the batch normalization layer is

$$\mathbf{a}^{[l]}[i] = \gamma^{[l]}\mathbf{y}^{[l]}[i] + \beta^{[l]} \tag{3.8}$$

where $i = 1, \cdots, N_{mb}$, and $\gamma^{[l]}$, $\beta^{[l]}$ are learnable parameters.

Batch Norm is added after each layer that has trainable parameter in the proposed system. By centering data, batch norm can speed up training and make the model more robust against variations in outputs of previous layers.

### *Dropout layer*

Dropout is a technique used during the training phase to help prevent overfitting. When dropout is added after layer $l$, some output units of layer $l$ are muted according to predefined dropout probability. This kind of random selection forces weights assigned by layer $l+1$ to spread out across all input neurons instead of focusing on just a small set of them. In the CNN for magnitude/phase images, a dropout layer with dropout probability $0.5$ is inserted before the fully connected layer. At the same time, right after concatenation, a dropout layer is also added to make sure the neural network can learn the contribution from phase and magnitude equally.

### *Fully-Connected (FC) layer*

The high-dimensional outputs of the two CNNs are first flatten into vectors, which serve as inputs to the fully-connected layers where all the input units are directly connected to the hidden neurons. Suppose that the $l$-th layer is a fully-connected layer. Denote by $\mathbf{z}^{[l]}$ the output of the $l$-th layer before the activation function.

$$\mathbf{z}^{[l]} = \mathbf{W}^{[l]}\mathbf{a}^{[l-1]} + b^{[l]},$$

where $\mathbf{W}^{[l]}$ and $b^{[l]}$ are the weights and bias of the $l$th layer. Then the output of the $l$-th layer is given by

$$\mathbf{a}_i^{[l]} = g^{[l]}(\mathbf{z}_i^{[l]}).$$

Our system consists of two FC layers: a hidden layers with $32$ neurons and an output layer with $2$ neurons.

### *Activation Function*

Activation functions are nonlinear functions added to the output of each neuron. Two activation functions are used in this paper - rectified linear unit (ReLU) and softmax. The former is used for the hidden layer whereas the latter the output layer. The ReLU activation function, with input $x \in \mathbb{R}$, is defined to be

$$g^{\text{ReLu}}(x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Softmax is used to output (estimate) class probabilities for the classification problem. Suppose the number of class to be classified is $C$ and $\mathbf{x} \in \mathbb{R}^C$ are the input to the softmax function. Then the output is given by

$$g_c^{\text{soft}}(\mathbf{x}) = \frac{e^{\mathbf{x}_c}}{\sum_{j=0}^{C-1} e^{\mathbf{x}_j}},$$

where $c = 0, \ldots, C-1$ is the class index. We note that presence detection is a binary classification problem, i.e., $C = 2$. A sigmoid function can be used instead of the softmax. However, our experiment indicates slightly more robust classification using softmax - this can perhaps be attributed to the difference in weight and bias terms between the two: softmax employs two independent sets of weight vector and biases for the two neurons whereas the sigmoid function has a single input to the neuron at the output layer. While mathematically one can show equivalence between the two with $C = 2$ by finding the corresponding parameters, learning such parameters through training may yield some performance difference.

### *Loss function*

The cross-entropy is used as the loss function for the proposed CNN. Suppose there are $N_d$ CSI images. Denote by $p_{i,c}$ the probability that the $i$-th CSI image belongs to the $c$-th class predicted by the proposed CNN. Then the cross-entropy is given by

$$L_o = -\frac{1}{N_d} \sum_{i=1}^{N_d} \sum_{c=0}^{1} \mathbf{y}_c[i] \log(p_{i,c}), \tag{3.9}$$

where $\mathbf{y}[i]$ is a one-hot vector corresponding to the ground truth. That is, with binary classification, $\mathbf{y}[i]$ is length-2 vector whose non-zero entry corresponds to the true label of the $i$-th CSI. For example, $\mathbf{y}_0[i] = 1$ and $\mathbf{y}_1[i] = 0$ imply the true label of the $i$-th sample is $0$ whereas $\mathbf{y}_0[i] = 0$ and $\mathbf{y}_1[i] = 1$ imply the true label of the $i$-th sample is $1$.

**regularization** In addition to cross-entropy, $l_2$ regularization is used in each fully-connected layer to prevent overfitting. Thus, the overall loss function is given as

$$L = L_o + \frac{1}{2N_d} \sum_{l \in B} \lambda_l ||\mathbf{W}^{[l]}||_F^2, \tag{3.10}$$

where $|| \cdot ||_F$ denotes the Frobenius norm of the matrix, $\lambda_l$ is a hyperparameter and set $B$ contains indices of all FC layers.

#### *Output of CNN*

The output of the CNN corresponding to the $i$-th CSI image, $\hat{y}[i]$ is given by

$$\hat{y}[i] = \underset{c \in \{0,\ldots,C-1\}}{\arg\max} \; p_{i,c}.$$

### 3.2.3 Post processing

The design of the post processing block is closely tied with how data collection is conducted. As alluded in the introduction, presence detection differs with detection of certain activities in that one is not looking for a certain activity pattern but rather, whether a room is being occupied or not, assuming that occupants are not completely still for extended periods of time. As such, there are two different ways of collecting training data for the occupant state: one is to collect CSI for the entire duration when occupants are present; an alternative way is to collect CSI only when occupants are moving. While in theory the former seems to be a natural choice - what we try to detect is the presence or absence of humans in a room - doing so leads to significantly high false alarm rate regardless of how many training data are collected. The reason is quite simple: collecting training CSI data when humans are present will include many instances when humans are completely still. Such CSI samples, *albeit* scattered throughout the measurement data (i.e., not for extended period of time), are indistinguishable with that of an empty room. In essence, the training data corresponding to human present are polluted with a large number of data samples that are similar to that training data without human presence.

We elect to use training data corresponding to the CSI instances when there are detectable human movements in the room. While this leads to 'missed detection' corresponding to instances when the occupants are still, simple post processing can be done after the CNN block with tunable parameters such as the resolution with which the presence detection is desired. In short, the training is done so that the CNN attempts to reliably detect human motion of any kinds; complete still human presence thus is likely to be classified as the negative state. Post-processing then applies some averaging operation within a time window, whose duration corresponds to some desired time resolution, for presence detection. This is sufficient in practice since with a truly empty room, the CNN output should contain negative outputs whereas with human present, the output should have significant portion of positive outputs.

## 3.3 Experiment

This section describes the experiment setup where COTS WiFi cards are used to collect WiFi CSI in an indoor environment. Data collection is explained in detail and the presence detection result is compared to that using PIR sensors.

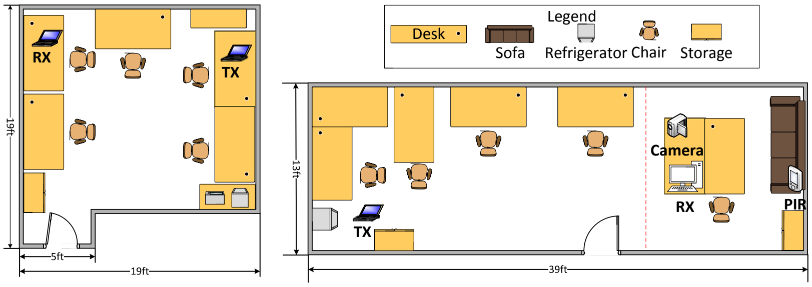
### 3.3.1 Experiment setup

Our WiFi system consists of a laptop (Thinkpad T410) as WiFi access point (AP) and a desktop (Dell OptiPlex 7010) as WiFi client. Atheros 802.11n WiFi chipset, AR9580, and Ubuntu 14.04 LTS with built-in Atheros-CSI-Tool [70] are installed on both computers. The AP sends packets at the rate of $100$ pkts/s. The client records CSIs using Atheros-CSI-Tool, and the CSI sampling interval is roughly $10$ms. With $N_r = 3$ receive antennas, $N_t = 3$ transmit antennas, and $N_{sc} = 56$ subcarriers in a $20$MHz channel operating at channel $6$ in the $2.45$GHz band [71], each CSI instance $\mathbf{H}[i]$ is a $56 \times 3 \times 3$ complex valued array. Down-selecting to $N_f = 14$ evenly spaced subcarriers, the resulting $\tilde{\mathbf{H}}[i]$ is of dimension $14 \times 3 \times 3$.

The indoor environment in which both training data collection and testing are done is sketched in Fig. 18. Three different environments are used: two labs of different size and layout and a typical two-bedroom apartment in a four-story apartment building. In the lab environment, there are multiple monitors/laptops on desks and more chairs on the floor; these are not drawn in the figure and their positions may change in different days. The transmit antennas are placed behind a laptop and the receive antenna array is surrounded by a lot of other computers as shown in Figs. 19(a) and 19(b). Therefore, there is no strong line of sight component between the transmitter and the receiver. For the two-bedroom apartment, only large furniture such as beds and tables are sketched in the plot.

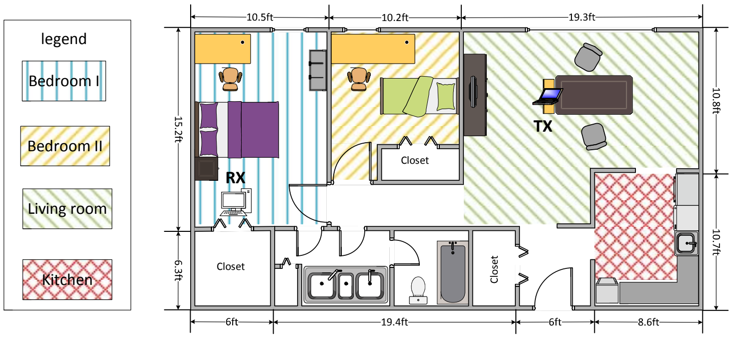
### 3.3.2 Data collection

For the image input to the CNN, we choose $I = 128$ consecutive CSI instances, which lasts for around $1.27$s. This is chosen since one second is sufficiently long for any detectable human motion to induce temporal CSI variation. A CSI image is only used (i.e., considered a valid sample) if it satisfies the following two conditions: 1) Every entry of $|\mathbf{X}|$ is non-zero. This is imposed to remove erroneous CSI estimate - occasionally zero entries will show up in recorded CSI series, potentially due to hardware/firmware problems. 2) The time differ-



(a) Lab I

(b) Lab II



(c) Apartment

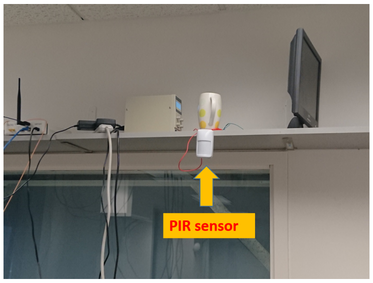
Fig. 18: Indoor space layout

ence between the last and the first frame lies within $1.27 \pm 0.064$s. WiFi scheduling may lead to different frame lengths hence excessively long interval between two CSI estimates. In the experiment, the cropping window size is chosen to be $T = 50$, hence $\mathbf{A}^{\text{abs}}$ and $\mathbf{A}^{\text{phase}}$ in (3.6) are of size $50 \times 14 \times 9$ and $50 \times 14 \times 6$ respectively.

Data collected in the human-free state is labeled $0$. The training data with label $1$ are collected when at least one person is walking randomly in the room. This way, training samples collected when occupants are completely still will not be used. Both human-free and motion data are collected on multiple days since the wireless channels are inherently nonstationary. This prevents CNN from being tuned to features that are irrelevant to presence detection, e.g., different CFO and STO on different days due to frequency drift. Data collection on any given day is also divided into disjoint runs which alternate between human-free and human motion. Finally, training and test data come from completely disjoint days.

The CSIs were collected during $24$ days at three different locations over a period of $8$

(a) Transmitter (b) Receiver



(c) PIR sensor

Fig. 19: Device setup

months, as summarized in Table 5 [1]. Data collected during the first three days were from Lab I, with data for the following $16$ days and last $5$ days from Lab II and the apartment, respectively. Notice that data from Lab II is divided into three parts due to disjoint experiment periods. Spreading out the measurement data over an extended period of time helps us to understand the sensitivity of the learned system to environmental change. There is no deliberate effort to maintain the same furniture arrangement during the measurement duration.

Table 5: Data collection

| Days | Location | Dates |
|---|---|---|
| $1-3$ | Lab I | 09/14/2019 - 09/22/2019 |
| $4-8$ | Lab II | 10/20/2019 - 10/30/2019 |
| $9-16$ | Lab II | 11/26/2019 - 12/06/2019 |
| $17-19$ | Lab II | 12/30/2019 - 01/01/2020 |
| $20-24$ | Apartment | 04/29/2020 - 05/17/2020 |

---

[1] Source code and WiFi data repository can be found on github at `https://github.com/bigtreeyanger/presence_detection_cnn`

The proposed CNN is built under Keras with Tensorflow as backend [72]. Training and off-line testing described in Section 3.3.3 and Section 3.3.4 are performed on a Linux server (Dell PowerEdge R730) with one E5-$2650$ v4 CPU and $128$GB of RAM. On-line detection described in Section 3.3.5 is run on the WiFi receiver (Dell desktop) with one i7-3770 CPU and $8$GB of RAM.

### 3.3.3 Motion Detection

We first evaluate motion detection using the proposed CNN without post-processing, i.e., CNN is trained to classify input CSI images according their labels: $0$ for human free data whereas $1$ for data with someone randomly walking around.

The CNN with $55078$ parameters is trained using data from days $9-14$ in Lab II which span two weeks. The number of training data in each class is summarized in Table 6. Both labels have roughly $40000$ images which correspond to $30-$min data collected each day. The learned model, denoted by model I, is trained for 10 epochs, lasting for a total of $156.30$ seconds. Test data are from days $1-5, 15-16$ and $20-14$; motion data collected on those days have similar motion types as the training data, i.e., they were collected when human introduced large motions such as walking, sitting down, and standing up. Test data were collected at both Lab II (i.e., same as training) and at Lab I and apartment (i.e., different from training).

#### *Testing in the same environment*

Model I is first tested using the remaining data collected in Lab II except days $6-8, 17-19$ which will be evaluated in Section 3.3.5. There are roughly $5000$ CSI images for each label per day corresponding to about half an hour measurements. The results are shown in Fig. 20. The accuracy (vertical axis) is the percentage of test samples that are correctly labelled, and this is similarly defined for all subsequent tests. The performance for both labels is quite consistent (all close to $100\%$) - we note that days $4$ and $5$ were collected

Table 6: Training set composition

| Model name | human free | | human motion | |
|---|---|---|---|---|
| | days | size | days | size |
| model I | $9-14$ | $39866$ | $9-14$ | $41276$ |
| model II | $3, 9-14$ | $24642$ | $9-14$ | $20383$ |
| model III | $9-14$ | $19617$ | $20, 22, 9-14$ | $23831$ |
| model IV | $8, 9-14$ | 50753 | $9-14$ | $41276$ |

one month earlier than the training data, thus lab settings, e.g., the transceiver placement and the location/number of surrounding objects were quite different. The proposed CNN is therefore quite robust to the environment changes over time.

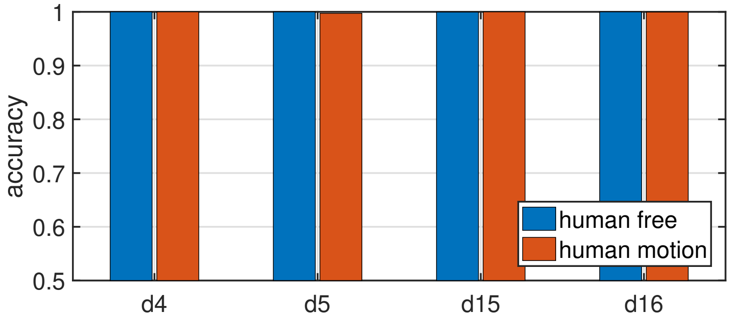


Fig. 20: Performance of Model I in Lab II on different test days

### *Testing in a different environment*

The performance of model I in Lab I is given in Fig. 21. Even though sensitivity to human motions remains high, there is noticeable increase in the false alarm rate. For example, the false alarm rates in day $1$ and day $3$ are $6.05\%$ and $3.21\%$ respectively. A simple remedy is to include human free data collected from Lab I in the training set. This leads to Model II in Table 6 where human free data on day $3$ which comes from 30-min runs are combined with $40000$ randomly chosen samples from the previous training set. From Fig. 21, the new model exhibits noticeable improvement in the first two days' false alarm rate without introducing degradation in the human detection rate.

Model I is further evaluated using data from days $20 - 24$ in a completely different and more complex apartment environment. Different from the above single-room tests, results are now categorized according to where the motions took place. Specifically, human motions can happen in four different rooms: living room, kitchen, bedroom I and bedroom II. The accuracy given in Fig. 22 is averaged over all the test days. Without any data from the unseen environment, the model still has high detection rates (close to $100\%$) in three rooms except the kitchen ($95.6\%$) while at the same time has a low false alarm rate ($0.04\%$). Notice that bedroom II has two walls between the transmitter and receiver, indicating that the model has good through the wall detection capability. To further improve the sensitivity in the kitchen, only data for motions in the kitchen from days $20$ and $22$ are added to the training (day $21$ does not have motion data collected) This portion of data comes from runs lasting for $10$ minunites each day. The Model III given in Table 6 is trained using the new data together with $40000$ randomly chosen samples from the previous training set collected in the lab environment. The average accuracy for motions in the kitchen excluding days $20 - 22$ improved from $95.55\%$ to $99.36\%$ with no noticeable performance changes for other rooms in the apartment.

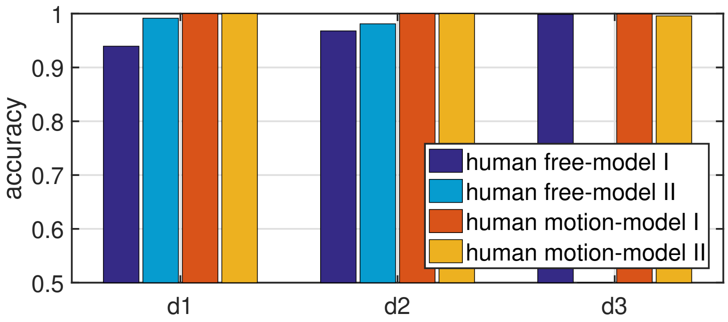


Fig. 21: Performance in Lab I on different test days

### *Mixed state test run*

As we have mentioned in the data collection part, on each day, the experiment was divided into multiple runs. Up till now, for both training and test data, they come from runs that

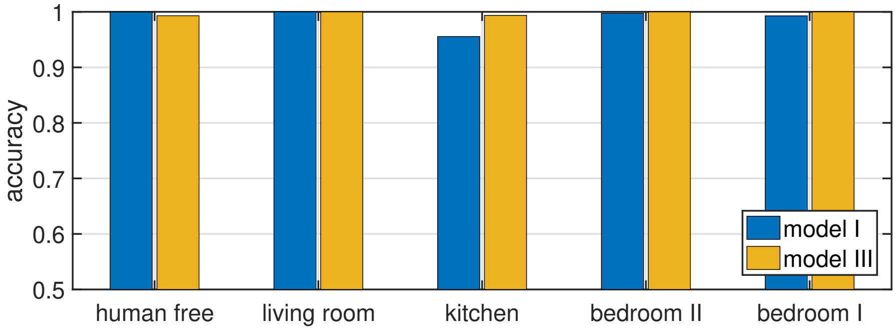
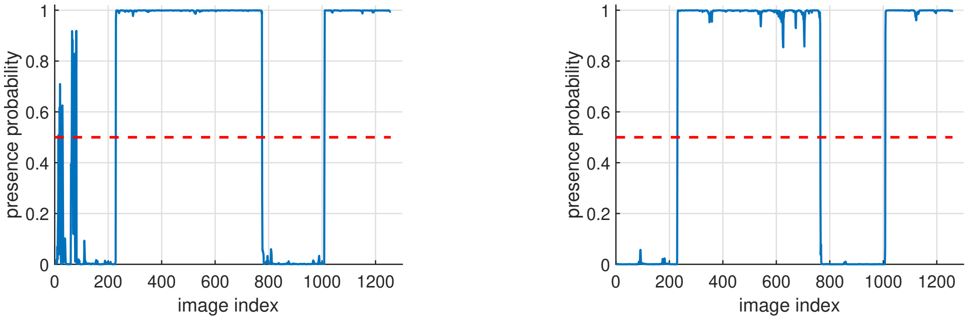


Fig. 22: Performance in the apartment at different locations

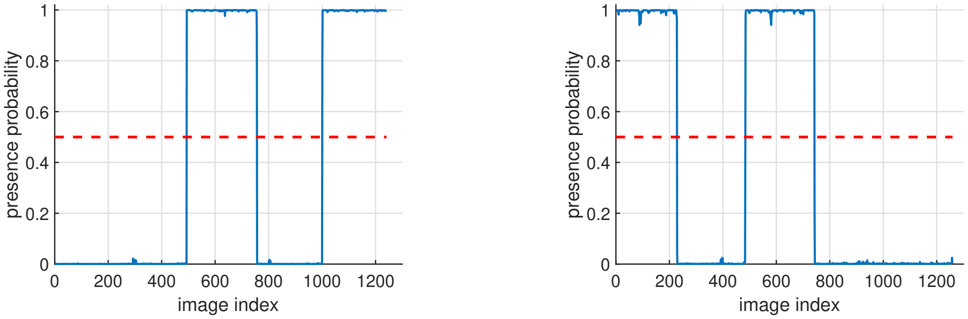
entirely correspond to either an empty room or human movements. That is, data from one run is labeled all as $0$ or $1$. To further rule out the possibility that the proposed system classifies data by the similarity of the hardware status, we conduct some more runs, called mixture runs, on some test days for validation purpose only. Each test run lasts $5$ minutes and is divided into five one-minute intervals. Measurement is done carefully such that each interval has the same state. Detection results of mixture runs on day $1, 3, 4$, and $16$ are given in Fig. 23. We set the CSI step size to be $23$, thus each one-minute interval contains around $260$ images. Instead of using predicted labels, motion probability, $p_{i,1}$ in (3.9), of each image is used to give a more refined detection performance. From Fig. 23, the model can successfully track state change in a single run.

### 3.3.4 Performance Comparison

In this section, we compare the performance of the proposed CNN (Model I) with two recently proposed WiFi based motion detection systems with similar experimental setup: the PADS [27] and R-TTWD [31] systems. PADS leverages the temporal correlation matrix of both CSI amplitudes and phase differences across antenna elements. The first few largest eigenvalues of correlation matrices are adopted as features. R-TTWD exploits CSI amplitude correlation in the frequency domain (i.e., across different subcarriers). To increase detection of moving humans at different side of a wall, R-TTWD proposes to evaluate means of first-order difference of the first few eigenvectors. Both PADS and R-TTWD use



(a) day 1 (ground truth: 01101) (b) day 3 (ground truth: 01101)



(c) day 4 (ground truth: 00101) (d) day 16 (ground truth: 10100)

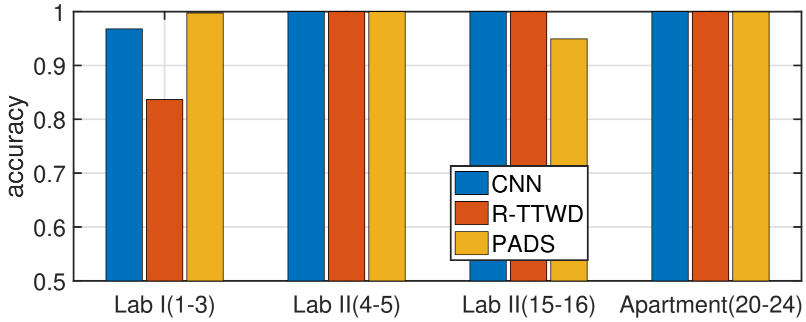
Fig. 23: Mixture run detection result

Support Vector Machine (SVM) [73] after feature extraction to find the decision boundary in the feature space for presence detection.

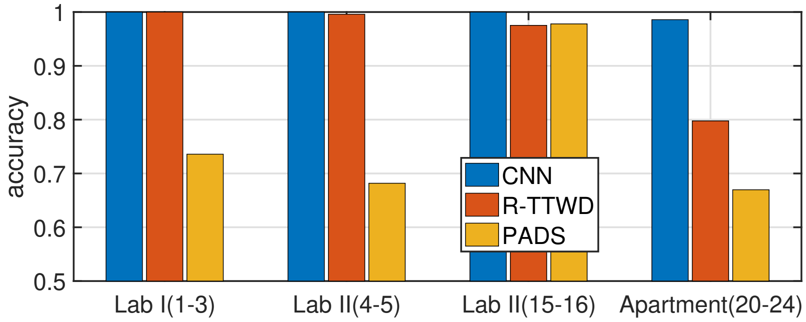
For a fair comparison, the same input CSIs are given to three systems with the only difference in the number of subcarriers $N_f$. PADS and R-TTWD use $28$ evenly spaced subcarriers instead of $14$ in the proposed CNN to maintain the desired frequency resolution (both systems use all $30$ subcarriers extracted from off-the-shelf NIC, Intel 5300, in the original papers). Besides, all the pre-processing steps are conducted as described in [27, 31].

The average accuracy of the three systems under four different test scenarios are shown in Fig. 24 with test data from days listed in the parentheses of the horizontal axis. In Lab II, i.e., the same environment as the training data, our model and R-TTWD give comparable result. However, when tested in Lab I, R-TTWD has considerably higher false alarm rate compared with CNN. Performance deteriorates even further in the apartment for R-TTWD as its miss detection rate increases dramatically. For the PADS system, even though its false alarm rate remains consistently low (Fig. 24(a)), its motion detection suffers significantly with environment change (Fig. 24(b)).

Clearly, compared to both PADS and R-TTWD systems, the proposed learning system performs consistently better and exhibits much more robust performance when tested in a completely new environment.



(a) Human free



(b) Human motion

Fig. 24: Comparison with PADS and R-TTWD

### 3.3.5 Presence Detection

Before presenting our presence detection results conducted in Lab II, let us first examine how the CNN trained with walking data performs when more subtle human motion is used for testing. These data are collected on days $6-8$ in Lab II with various small scale motion (turning in chairs, arm waving, etc.). The results are summarized in Table 7. Clearly, with training data coming from exclusively random walking for label $1$, the CNN can still reliably detect other motion types.

Table 7: Test accuracy for small scale motion

| Days | human free | | human motion | |
|---|---|---|---|---|
| | size | model I | size | model I |
| 6 | 5396 | 100% | 4992 | 99.94% |
| 7 | 5447 | 100% | 5044 | 98.47% |
| 8 | 10887 | 99.98% | 4837 | 96.90% |

The actual model used for presence detection (model IV in Table 6) is obtained by further augmenting training data with label $0$ data (i.e., human free) collected on day 8. This is done since the output motion probabilities of human-free data collected on day $8$ are closer to $0.5$ than other days. Thus adding these data for CNN training provides more sample diversity. Model IV is then deployed at the WiFi receiver, along with post-processing, for real-time presence detection.

As a comparison study, presence detection is conducted concurrently using a PIR sensor. We chose Honeywell DT8035 [74], a leading edge PIR sensor with a coverage range of $40$ft $\times$ $56$ft (our lab dimension is $13$ft $\times$ $39$ft). A camera is used in the lab to provide ground truth. The PIR sensor is mounted on the shelf at one side of the room at a height of $6.8$ft (see Fig. 19(c)). Note that DT8035 also has a microwave sensor which was disabled for this experiment. Throughout the experiment, human activities are restricted to the left side of the room (left of the red dash line in Fig. 18b) to avoid blind spot of the PIR sensor as its coverage is in a conical shape.

The post-processing is an averaging process on the motion detection outputs of the CNN. Each new CSI instance is used to construct CSI images with the previous 127 CSI estimates, i.e., a sliding window with step size one is applied to the CSI series. This results in a CNN output at a rate of about one per $10$ms. Presence detection output occurs every second. Each second is divided into five subintervals, each of duration $200$ms. For each subinterval, a positive motion detection occurs when at least $10$ CSI images have output label $1$. A positive detection is declared for the one second period if at least three out of the five subintervals have positive motion detection. Finally, given that the PIR sensor outputs its detection result between $2$ to $5$ times each second, we choose the detection resolution to be $1$ second for both WiFi and PIR: a presence is detected for each second if there is at least one positive detection within the one second period for PIR.

### *False positive test*

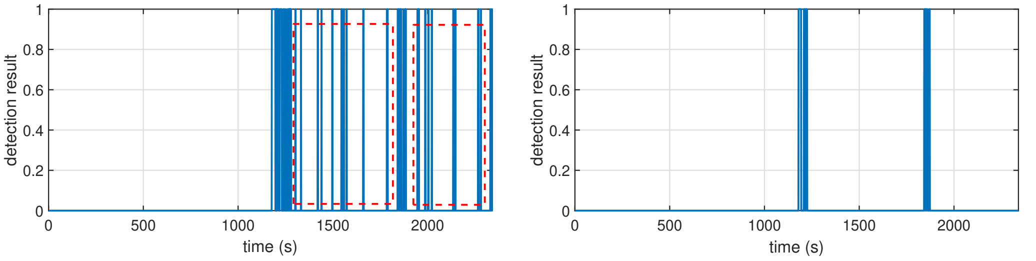
The test is done over a $3$ day period (days $17 - 19$ from 12/30/2019 to 01/01/2020), when Lab II is empty. Results shown in Table 8 are the numbers of one-second intervals in which presence is detected by CNN and PIR sensor. To avoid interruption to normal lab activities, a single test run on certain days can not last for very long. For example, on day $17$, the entire test is broken into three periods, with the shortest one lasting for $20$ minute during lunch break. The entire test lasts for about $46.5$ hours, the proposed system only report false positive four times, yielding a false alarm rate $2.4 \times 10^{-5}$. The PIR sensor has zero false alarm rate and the results are comparable given that isolated one second positive can be easily ruled out for occupancy detection.

Table 8: False alarm counts (in seconds) in an empty room

| day | index | duration | CNN | PIR |
|---|---|---|---|---|
| 17 | 1 | 8hrs | 3s | 0s |
| | 2 | 20mins | 0s | 0s |
| | 3 | 9hrs | 1s | 0s |
| 18 | 1 | 8hrs | 0s | 0s |
| | 2 | 9hrs | 0s | 0s |
| 19 | 1 | 12hrs | 0s | 0s |

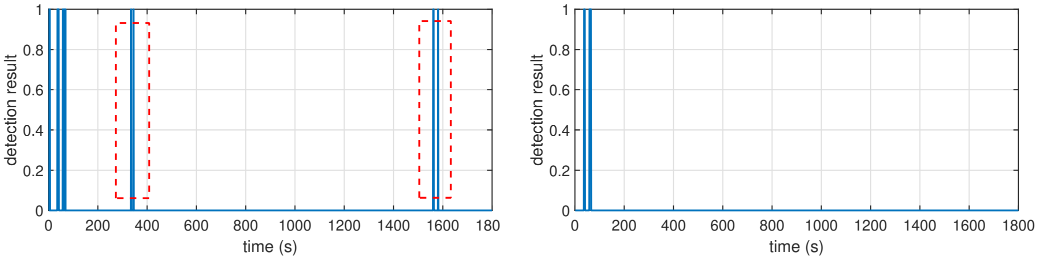
### *Sensitivity test*

This part evaluates the sensitivity of the system to human presence. The experiments are done when people go about with their daily activities in the lab without introducing inten-



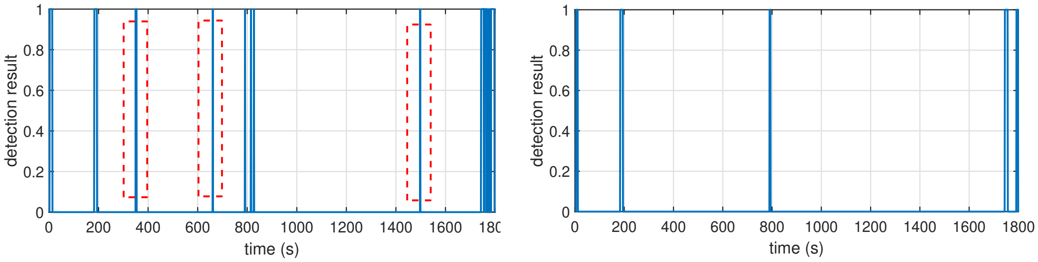
(a) Test 1: CNN

(b) Test 1: PIR



(c) Test 2: CNN

(d) Test 2: PIR



(e) Test 3: CNN

(f) Test 3: PIR

Fig. 25: Comparison with PIR sensor

tional motions. In most of the time, occupants would just sit in front of their computers and occasionally engaged in normal conversations. Five tests are done in days 17 and 18. Duration of each test and presence counts reported by CNN and PIR are summarized in Table 9. Fig. 25 shows detection results of 3 tests done in day 18. All tests were done with at least one person present in the lab from the beginning to the end except test 1 on day 18 as in Figs. 25(a) and 25(b) when the lab is empty for the first 20 minutes. Human activities detected by CNN but not by PIR are marked using red rectangular boxes in Fig. 25. Notice that we only highlight parts when there is not a single positive detection output from PIR sensor for the entire duration of the box. For example, at around 1200s in Figs. 25(a) and 25(b), CNN can detect much longer human presence than the PIR sensor but is not marked

Table 9: Presence counts (in seconds)

| day | test index | duration | CNN | PIR |
|---|---|---|---|---|
| 17 | 1 | 1800s | 212s | 119s |
|  | 2 | 1800s | 76s | 26s |
| 18 | 1 | 2340s | 117s | 48s |
|  | 2 | 1800s | 19s | 11s |
|  | 3 | 1800s | 68s | 41s |

in the figure for clarity of presentation. To compare the sensitivity of two systems more accurately, we summarize presence counts in Table 9. Each count corresponds to a positive detection for a one second period during the entire test run. WiFi sensing consistently outperforms PIR in all runs. By cross reference with video recordings, we find that the presence detected in the highlighted ranges (i.e., those detected by WiFi but not by PIR) in Fig. 25 is associated with subtle human movement, such as stretching while sitting, adjusting sitting postures and conversing with each other without excessive movement. These subtle movements are often missed by PIR but can be easily picked up by WiFi sensing. We emphasize again that model IV is trained with only random walking for label 1 data, i.e., no small scale motion is included.

It is worth noting that there are still movements missed by both WiFi sensing and PIR. The most important example is when occupants are typing on keyboards but otherwise remain completely still. Such movement appears to be too subtle to be detected by even WiFi sensing. A possible remedy is to deliberately add those keyboard typing measurement data to the motion training set yet it is likely to increase the false alarm rate given the subtlety of such movements.

### 3.3.6 Discussions

We discuss in this section the impact of various design parameters on the WiFi sensing performance.

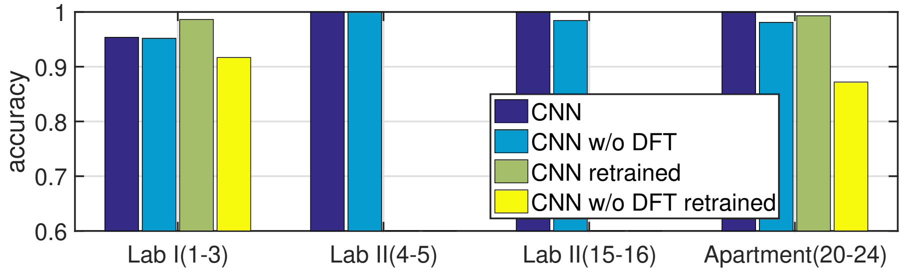
#### *CSI sampling interval*

The CSI sampling interval is set at $10$ms in all the experiments reported above. Retraining model I under two more sampling intervals, $20$ms and $40$ms, slight degradation occurs for motion detection but with negligible effect on the presence detection with properly designed post-processing, *provided that training and testing are done in the same lab space*. With training and testing done at different labs, slower sampling rate leads to noticeable performance loss.

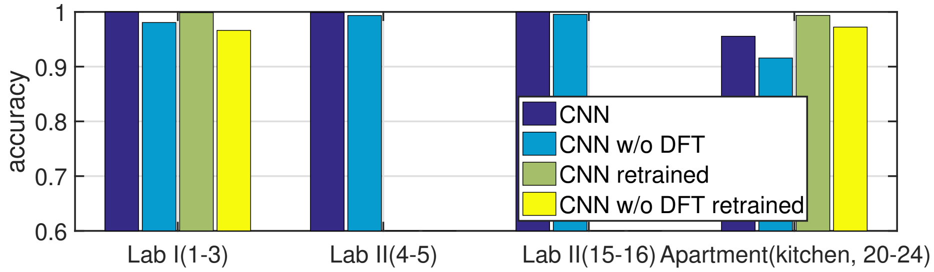
#### *Pre-processing*

An important pre-processing step in the proposed system is applying 2-D DFT and 1-D DFT to CSI magnitude images and CSI phase images respectively. We now compare the performance with one that does not use DFT for pre-processing. The CNN architecture for inputs without DFT is modified to achieve best training and test performance to guarantee a fair comparison. In particular, without DFT, the kernel size of the first convolutional layer is changed from $(3, 3)$ to $(5, 3)$ and the pooling size of the following pooling layers is changed from $(2, 1)$ to $(4, 1)$ and $(3, 1)$ to $(4, 1)$ respectively. This small adjustment leads to noticeable improvement in false alarm rate while maintaining similar sensitivity in test set compared with the one used for input with DFT. The motivation of increased kernel size to due to the fact that DFT pre-processing localizes CSI dispersion into low frequency region, hence allows the use of significantly smaller kernel size. In the absence of DFT, large kernel size helps capture such channel dispersion due to human motion.

As in Section 3.3.4, models with and without DFT are evaluated under $4$ test scenarios whose results are given in Fig. 26. Apartment test only reports motions in the kitchen area as performance in other three locations are similar to the lab environment. In each test scenario, the first two bars correspond to models trained with the same training set as model I, while the last two bars for Lab I and apartment show results after retraining as described in Section 3.3.3. In Lab II, both models give comparable results for motion detection, while CNN with DFT has lowest false alarm rate ($1.5\%$ lower false alarm rate in Lab II (day 15-16)). When tested in Lab I, both models have noticeable degradation in false alarm performance. With retraining using human free data from day $3$, the CNN with DFT has significantly lower false alarm rate while the CNN without DFT does not have any improvement. For the kitchen area in the apartment, both models give higher false negatives than the lab environment. The simple retraining method does not help CNN without DFT since it increases the motion sensitivity in the kitchen at the expense of further elevating false positive in a human free environment (false alarm rate rises by $10.9\%$). With DFT, however, learning becomes more immune to CSI features *not* attributed to human motions, leading to noticeable improvement in performance when retrained with additional data.



(a) Human free



(b) Human motion

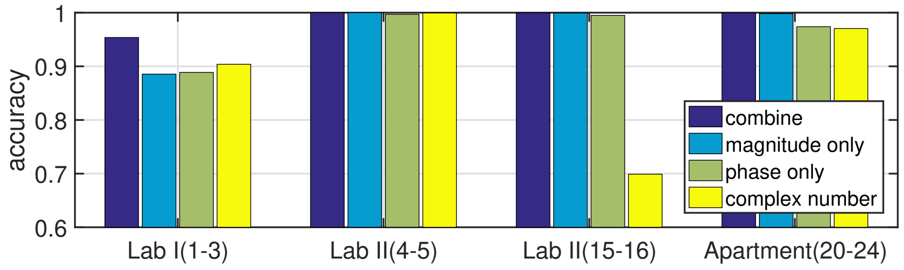
Fig. 26: Pre-processing comparison

### *Input*

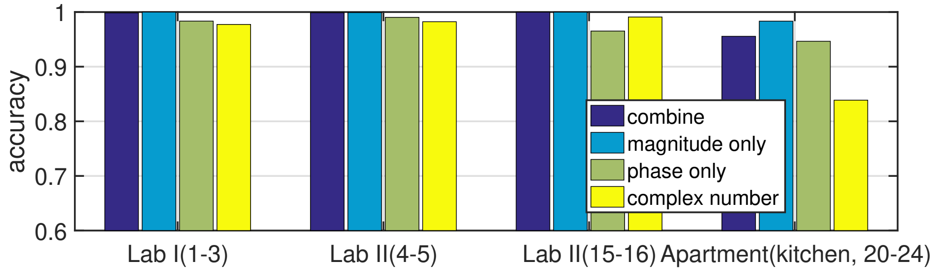
The proposed system harvests presence information from CSI magnitude and phase in an explicit manner. In this part, we investigate the impact of different inputs on detection performance: a) CSI magnitude only; b) CSI phase only; c) complex CSI input without explicit decomposing into magnitude and phase. For c), we combine the last two spatial dimensions of the 4-D CSI array $\mathbf{X} \in \mathbb{C}^{128\times14\times3\times3}$ into one and apply normalization as in (3.2) but to complex entries. The resulting real and imaginary components are then stacked up along the last dimension, resulting in inputs to a CNN model of shape $128 \times 14 \times 18$. The sizes of convolutional/pooling layers are adjusted as in Section 3.3.6 for a fair comparison due to enlarged input size.

The results of the three models plus Model I are given in Fig. 27. With information from Lab II only, as seen in Fig. 27(a), Model I, which explicitly uses both CSI magnitude and phase results in the most robust model in a human free environment. In contrast, the performance using complex CSI input without decomposing into magnitude and phase varies wildly for different test days. While it has low false positive rate on days 4 and 5, it has highly elevated false positive rate on days 15 and 16. When tested in Lab I, which is different from where the training data were collected, the three models all lead to higher false false detection compared with Model I.

Motion detection results are given in Fig. 27(b) where, as before, for the apartment scenario, only the test result in the kitchen is plotted. Complex CSI input is consistently worse than Model I and CSI magnitude only in all four test scenarios. Overall, CSI phase only is not as sensitive to motions as CSI magnitude only. Indeed, the performance of Model I in the apartment is negatively affected by the CSI phase, which can be compensated by retraining Model I with data collected for kitchen from day $20, 22$ (c.f. model III in Table 6). These additional training leads to an improvement in detection performance from $95.55\%$ to $99.36\%$ with combined magnitude and phase input.



(a) Human free



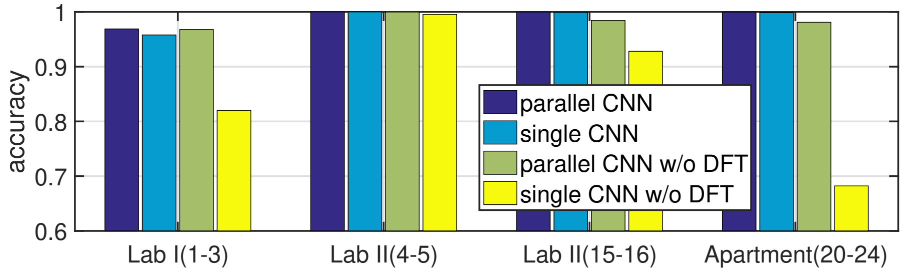
(b) Human motion

Fig. 27: Input comparison

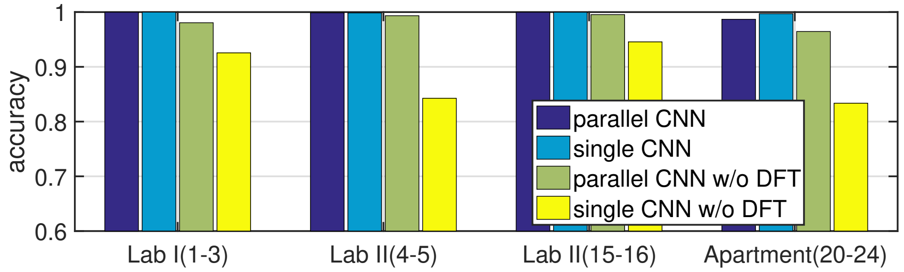
### *Architecture design*

The proposed CNN uses a parallel structure to extract information from CSI magnitude and phase images separately. An alternative architecture is to use a single conventional CNN by concatenating the magnitude and phase images along the last axis. Thus, the inputs to the CNN are of size $50 \times 14 \times 15$. The performance of the parallel CNN and the single CNN are given in Fig. 28. Notice that in this part, detection accuracy in the apartment is averaged over all four locations since performance of the models differ in most areas. With training data from Lab II only, these two models have comparable performance in Lab II. When tested in Lab I, the parallel CNN has slightly less false positives in the human absent environment; in the apartment, however, the single CNN results is more sensitive to human motions. Overall, the parallel CNN and the single CNN have comparable performance.

Finally, we evaluate the impact of pre-processing for the two different CNN architectures. Specifically, we evaluate the parallel and single CNN architectures with input without DFT on the CSI images. The sizes of convolutional/pooling layers are adjusted accordingly as in Section 3.3.6 to ensure a fair comparison. As seen in Fig. 28, there exists a noticeable performance gap between the parallel and single CNN without DFT for all test scenarios. Without DFT, the single CNN architecture tends to overfit training data, thus yielding lower accuracy in test data.



(a) Human free



(b) Human motion

Fig. 28: Architecture comparison

## 3.4 Summary

This chapter introduces a parallel convolutional neural network architecture which is designed to harvest occupancy information in CSI estimates along temporal, frequency, and spatial dimensions. With judicious pre-processing to remove hardware/system impairments and post-processing to infer presence information from motion detection output, the proposed learning system provides a viable and promising alternative for real time occupancy detection. Extensive experiments were conducted using commercial off-the-shelf WiFi devices. It was demonstrated that system is much more sensitive to human presence than PIR sensors and exhibit desired robustness against environment variation compared with existing RF based presence detection systems.

# CHAPTER 4

# PRESENCE DETECTION WITH REDUCED TRAINING EFFORTS

This chapter describes a new learning approach for occupancy detection without the need to introduce deliberate and continuous motion during the training process. A new learning problem is formulated, namely classification with contaminated samples. A two-stage solution using support vector machines is proposed to achieve reliable motion detection.

## 4.1 Motivation and Formulation

One major challenge in RF based occupancy detection system is collecting training data corresponding to human presence. Existing approaches often put strenuous requirement on the data collection process. For example, while the CNN based system presented in the Chapter 3 achieves superior presence detection performance, it requires continuous human motions throughout the measurement period for human motion data collection. Other existing works in this area often involve tedious labelling (e.g., with the assistance of cameras or manual segmentation) [59, 61] if continuous motions are not needed. Such training data collection process is cumbersome and prevents any proposed solution from wide-spread deployment.

In this chapter, we relax such assumption and develop a learning model for occupancy detection without the need of introducing deliberate and continuous motion in data collection. Instead, human presence data can now be collected while the occupants carry out normal activities. Such a data collection process complicates the learning problem since the CSI sequences corresponding to human present while remaining completely still are indistinguishable from human free data. A new learning problem, termed as classification with sample contamination, is formulated that captures the effect of mixed states with human presence: those corresponding to various motions and those corresponding to human(s) remaining completely still.

A one-class support vector machine, the so-called SVDD [63], is trained using the human free data. The motion detection problem can thus be formulated as an outlier detection problem. A more refined approach is a two-stage learning model: the trained SVDD model is then used to de-contaminate the training samples for human presence data. This is followed by a two-class SVC [75] for final motion detection using the static (human free) data and the outlier (human motion) data as a result of applying SVDD. The choice of support vector based learning also reduces the training data requirement. The CNN based motion detection requires a large amount of training data, a consequence of the model complexity itself. Support vector classifiers, on the other hand, requires a much less training data to achieve reasonable detection performance.

## 4.2 Pre-processing

The CNN based human presence detection described in Section 3.1 exploits CSI changes in temporal, frequency, and spatial dimensions in the presence of human motions. Judicious pre-processing is needed to help extract features useful for human motion detection. Initial pre-processing for this chapter is similar to that described in Section 3.2.1. To capture CSI variation in time, $I$ consecutive $\tilde{\mathbf{H}}[i]$'s are stacked up along the temporal domain, resulting in a 4-D array $\mathbf{X}$ of shape $I \times N_f \times N_r \times N_t$. While both magnitude and phase of $\mathbf{X}$ are shown to contain rich information about human presence [27, 46, 62], we focus in this chapter on the use of CSI magnitude for presence detection. Our goal is to show that by formulating and solving new learning problem, the training data requirement can be much more relaxed compared with the CNN based approach.

The extracted CSI magnitude is denoted by $\mathbf{X}^{\text{abs}}$ whose shape is $I \times N_f \times (N_r N_t)$ after combining last two spatial dimensions of $|\mathbf{X}|$. The following normalization method is applied to remove the dependency of each $\mathbf{X}^{\text{abs}}$ on the absolute power level: for $i = 0, \cdots, I-1$,

$$\tilde{\mathbf{X}}^{\text{abs}}_{i,:,:} = \mathbf{X}^{\text{abs}}_{i,:,:} ./ \mathbf{X}^{\text{abs}}_{0,:,:},$$

where $./$ denotes element-wise division.

Human motion introduces temporal CSI variations due to increased Doppler spread and frequency CSI variations due to altered propagation paths. For each transmit-receive antenna pair (i.e., fixing the last index of $\tilde{\mathbf{X}}^{\text{abs}}$), a 2-D discrete Fourier transform (DFT) is applied to the first two dimensions of $\tilde{\mathbf{X}}^{\text{abs}}$:

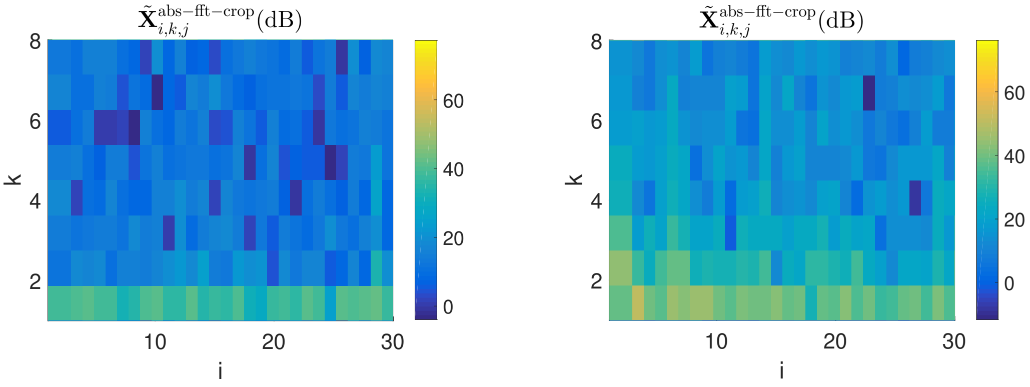
$$\tilde{\mathbf{X}}^{\text{abs-fft}}_{:,;,j} = \mathcal{F}\left(\tilde{\mathbf{X}}^{\text{abs}}_{:,:,j}\right),$$

where $\mathcal{F}(\cdot)$ denotes the 2-D DFT operation. This is followed by a 2-D cropping with window size $(T, D)$ centered around zero frequency:

$$\tilde{\mathbf{X}}^{\text{abs-fft-crop}}_{i,k,:} = \left|\tilde{\mathbf{X}}^{\text{abs-fft}}_{i,k,:}\right|,$$

where $i = 0, \ldots, T-1$ and $k = 0, \ldots, D-1$. Fig. 29 provides a sample of $\tilde{\mathbf{X}}^{\text{abs-fft-crop}}$ collected in the same room with and without human movement at a fixed antenna pair. This cropping helps preserve variations induced by human motions in the temporal-frequency dimensions as they are concentrated in the low frequency region in the 2-D DFT domain while removing high frequency variations of CSIs that are likely induced by other factors such as hardware impairment and channel estimation errors. The added benefit is the reduced dimensionality of the input to learning modules, which can significantly decreases the demand on the sample size for training.

The 2-D DFT of the CSI can lead to increased dynamic range. The logarithmic operator $y = \log_{10}(x+1)$ is thus applied to each element of $\tilde{\mathbf{X}}^{\text{abs-fft-crop}}$. To further make use of the spatial diversity provided by the multiple antennas, we take the mean value calculated across different transceiver pairs. The final 2-D sample is then flattened into a vector $\mathbf{v}[i] \in \mathbb{R}^{TD}$ to be the input sample to subsequent learning.



(a) Human-free environment

(b) Human movement

Fig. 29: 2D-DFT of CSI magnitude along frame and subcarrier

## 4.3 Training with human-free data only

### 4.3.1 Outlier Detection Using SVDD

SVDD is a one-class SVM that constructs a compact description of data presumably coming from a single distribution. With human-free training samples $\mathbf{v}[i]$, SVDD finds a sphere centered at $\mathbf{m}$ with radius $R$ in the feature space by solving the following optimization problem:

$$\begin{aligned} \min_{R,\mathbf{m},\boldsymbol{\xi}} \quad & R^2 + C\sum_{i=0}^{N-1} \xi_i \\ \text{subject to} \quad & ||\phi(\mathbf{v}[i]) - \mathbf{m}||^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0,\ i = 0, 1, \ldots, N-1. \end{aligned} \tag{4.1}$$

Here $N$ is the sample size; $\phi(\cdot)$ is the feature map which is often implicitly defined by the choice of the kernel function $k(\mathbf{v}[i], \mathbf{v}[j]) = \langle \phi(\mathbf{v}[i]), \phi(\mathbf{v}[j]) \rangle$; $\xi_i$ is the slack variable for the $i$th sample to allow for a soft boundary; $C$ is a regularization parameter with $C \in (0, 1]$ with $C = 1$ corresponding to hard-margin SVDD (i.e., all $\xi_i = 0$). Notice that while the above formulation uses $\phi(\cdot)$ explicitly, the formulation can be conveniently transformed using the kernel function as the quantities involved in the constraints depend on $\phi(\cdot)$'s only through corresponding inner product.

Finding the right value for $C$ is key to ensure good detection performance. Typical cross validation for SVDD often leads to a large $C$ (close to 1), and the resulting SVDD is close to a hard margin one when nearly every sample is inside the sphere. The consequence is that the obtained model will suffer from reduced sensitivity to human motions. If, however, $C$ is chosen to be too small, the model can have elevated false alarms.

### 4.3.2 Experiment Results

#### *Experiment Setup*

Experiments are conducted in a typical lab environment - Lab II whose layout is given in Fig. 18. The setup of the transmitter and the receiver is identical to Figs. 19(a) and 19(b). After downing selecting subcarriers from 56 to 14, the resulting CSI array at a given time $i$, $\tilde{\mathbf{H}}[i]$, is of dimension $14 \times 3 \times 3$.

Throughout the experiment, the CSI sampling interval is roughly 10ms, and each detection window consists of $I = 128$ consecutive CSI samples, lasting for around 1.27s. The input $\mathbf{X}^{\text{abs}}$ to the pre-processing block is of shape $128 \times 14 \times 9$. The cropping window is chosen to have size $(8, 3)$, resulting in the input vector $\mathbf{v}$ to the learning modules as a length-$24$ vector.

### *Data collection*

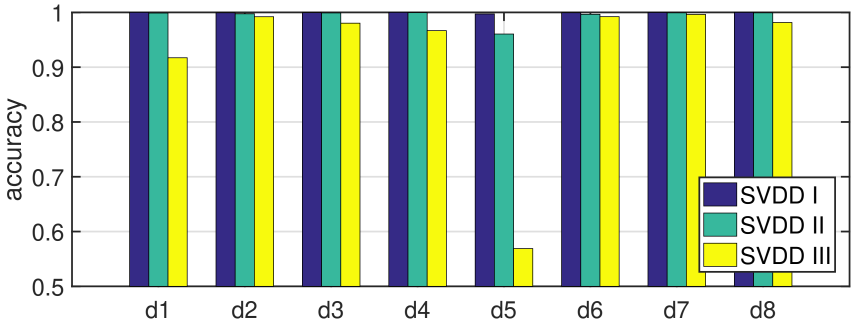
The CSIs were collected during $10$ days, indexed from $1$ to $10$, over a period of $1.5$ months. Training data comes from days $9$ and $10$. On these two consecutive days, collection for human free date lasted for around $40$ mins on each day. The number of instances used to train the SVDD modules is $N = 5000$.

All the data from days $1 - 8$ are used as test data to evaluate the detection performance. On these $8$ days, human presence data were collected when human motions were continuously introduced. Specifically, on days $4 - 8$, large scale motions were included such as walking, sitting down and standing up. On days $1 - 3$, small scale motions (e.g., turning in chairs, arm waving, etc) were also introduced. There are roughly $4500$ vectors generated for each label each day corresponding to around $35$-min measurements.

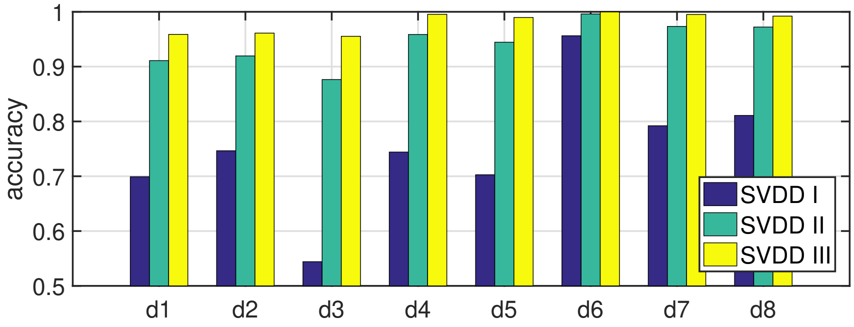
### *Training and test results*

All the models are trained using a radial basis function (RBF) kernel $k(\mathbf{v}[1], \mathbf{v}[2]) = \exp\left(-\gamma \|\mathbf{v}[1] - \mathbf{v}[2]\|^2\right)$ with $\gamma = 0.008$. The regularization parameter $C$ in (4.1) are chosen by a $5$-fold cross validation scheme [76]. The upper bound value imposed on $C$ can determine the model's robustness as well as sensitivity. Three models, SVDD I, SVDD II and SVDD III, are obtained with upper bound of $C$ equal to $1, 0.1, 0.01$ respectively. Their performance on each test day is given in Fig. 30. SVDD I has consistently low false positive rates in human-free environment as shown in Fig. 30(a) but suffers from reduced sensitivity to human presence in Fig. 30(b). The highest detection rate $0.9563$ is achieved on day 6; for the other test days, motion detection rates are all at or below $0.8$. The reason is that independent cross-validation for SVDD leads to a large sphere around human free data. As such, samples with small-scale motions are likely to fall inside the sphere, leading to degraded motion detection performance. As we decrease the upper bound of $C$, the model is allowed to make more errors on the boundary and thus finds a smaller sphere around training data. From Fig. 30(b), more motions are detected by SVDD II and SVDD III. However, such performance boost in detection is achieved at the sacrifice of false alarm rates. For example, SVDD III which has the highest sensitivity reports half of the samples as positive in human-absent environment on day 5.

Different from CNN based model in Chapter 3, to achieve reliable presence detection by training a SVDD model using human free data only, one needs to take into account the increased false positive rates in human-absent environment. It is feasible that training with some negative samples (including some motion samples) as described in [63] can potentially address such trade-off.



(a) Human free



(b) Human motion

Fig. 30: Performance of SVDD models on different test days

### *Training with motion samples*

With training data from one class only, SVDD learns to find a closed boundary around the training samples. As described in Section 4.3.2, its performance in detecting outliers depends on the tightness of the boundary (sphere) in the feature space. According to [63], incorporating negative samples (objects which should be rejected) in the training can help improve the description.

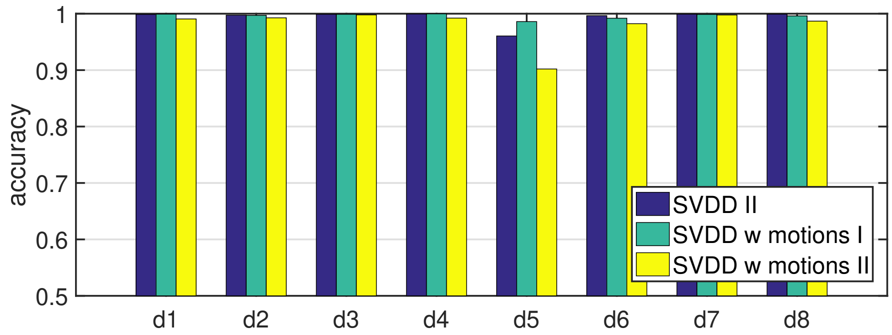
- SVDD with negative samples

  Suppose there are $N^{(0)}$ samples from human-free data, and $N^{(1)}$ samples from human motions. With human-free data and motion data enumerated by indices $i, j$ respectively, the optimization problem in (4.1) is modified as

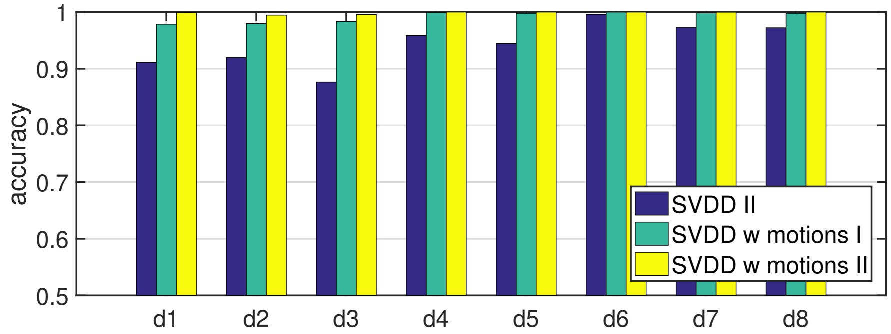
$$\begin{aligned} \min_{R,\mathbf{m},\boldsymbol{\xi}} \quad & R^2 + C_0 \sum_{i=0}^{N^{(0)}-1} \xi_i + C_1 \sum_{j=0}^{N^{(1)}-1} \xi_j \\ \text{subject to} \quad & ||\phi(\mathbf{v}[i]) - \mathbf{m}||^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0,\ i = 0, 1, \ldots, N^{(0)} - 1, \\ & ||\phi(\mathbf{v}[j]) - \mathbf{m}||^2 \geq R^2 + \xi_j, \\ & \xi_j \geq 0,\ j = 0, 1, \ldots, N^{(1)} - 1. \end{aligned} \tag{4.2}$$

- Performance analysis

  In the experiment, we let $N^{(0)}$ stay as $5000$ and include $100$ motion samples into the training set. Similarly, the regularization parameters $C_0$ and $C_1$ in (4.2) are determined by $5$-fold cross validation scheme. An upper bound smaller than $1$ imposed on $C_0$ is identical to that for SVDD II for a fair comparison. The performance of SVDD II and the newly obtained model, SVDD w motions I, are given in Fig. 31. After adding motions samples, sensitivity of the model has noticeable improvement while maintaining similar level of robustness in the human free room. Actually, SVDD w motions I reports less false positives on day 5 compared with SVDD II.



(a) Human free



(b) Human motion

Fig. 31: Performance of SVDD models with motions samples on different test days

The choice of motion samples incorporated in the training set can have impact on the performance. We replace $100$ motions samples used to get SVDD w motions I with another set which has more smaller scale motions, and the new model is named as SVDD w motions II. From Fig. 31(b), SVDD w motions II has the highest motion detection rates among these three models. However, its false alarm rates suffer from certain degradation especially on day $5$.

- Conclusion

  Adding a small set of motion samples into the training data can help improve the description of human free data and address the trade-off issue inherent in SVDD. However, such performance improvement replies on the quality of the included motion samples. Collecting and selecting optimal motion samples require human intervention which can be troublesome and time-consuming for users. A simpler, and more desirable, way to collect human presence data is needed.

## 4.4 Training with contaminated presence data

In this section, we present our designed two-stage approach for presence detection. The proposed method has nearly perfect occupancy detection result with a much less demanding data collection process. Specifically, we include both human presence data and human free data in the training set. The collection of human presence data happens when human subjects are performing normal activities in the monitored region without introducing motions deliberately. Subsequent data parsing and training are done in an autonomous manner with minimum human intervention.

A high level description of the proposed system is depicted in Fig. 32. The training samples $\mathbf{v}[i]$ (constructed as given in Section 4.2) corresponding to human free data are first used to remove samples from human presence data that correspond to static (i.e., motionless) measurements. This is done using SVDD: human free data are used to build a support vector model (referred to as Model I) represented by a sphere boundary in the feature space. This model is then used to remove samples from presence data that fall within the boundary (i.e., considered similar to human free data). Subsequently, the de-contaminated presence data and the human free data are used to train a two-class SVC to obtain Model II that is used for the actual motion detection. Testing on new input data is done using Model II. The final occupancy detection can be achieved similarly to that in Section 3.2.3 through post-processing the motion detection output.

### 4.4.1 Sample De-Contamination Using SVDD

Since SVDD is used to describe data from a single distribution, we train it using human-absent data and then apply the obtained model for removing static measurements in presence data. Regularization term $C$ in (4.1) determines the tightness of the sphere around the

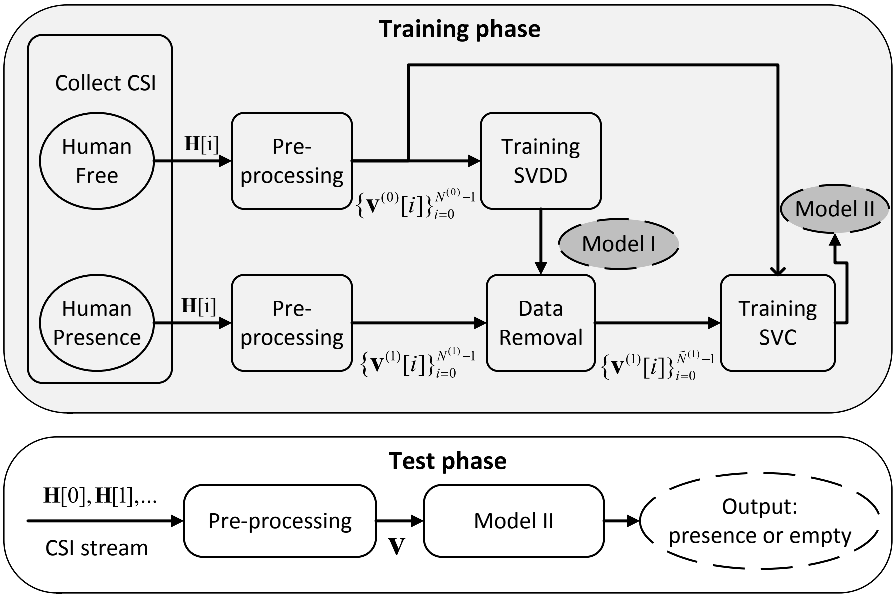


Fig. 32: Flowgraph of the proposed two-stage approach

training data. Different from Section 4.3 where $C$ balances the false alarm rate and sensitivity directly, here the choice of $C$ decides how many samples are moved from the presence data, thus impacting the performance of the following classification model. A large $C$ results in discarding a big portion of presence data, leading to significantly decreased sensitivity in motion detection. $C$ with very small values, however, increases model II's false positive rates due to too many residual contamination samples in the presence data.

### 4.4.2 SVC-Based Motion Detector

The de-contaminated human presence data, in which samples corresponding to motionless presence were removed, are used along with human free data to train a two-class SVC. The SVC finds a hyperplane as the classification boundary in the feature space by solving the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\omega},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C\sum_{i=0}^{N-1}\xi_i \\ \text{subject to} \quad & y_i(\boldsymbol{\omega}^T\phi(\mathbf{v}[i]) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0,\ i = 0, 1, \ldots, N-1. \end{aligned} \tag{4.3}$$

Here $y_i \in \{0, 1\}$ is the class label of training vector $\mathbf{v}[i]$. $C > 0$ is the regularization parameter that provides a trade-off between in-sample error and margin width of the SVC.

The training phase involves two stages: Model I is first trained with human absence data to decontaminate samples from human presence data; Model II is subsequently trained using the decontaminated data. Testing only requires the use of Model II: CSI measurement is pre-processed according the procedure described in Section 4.2 and then fed to Model II for classification into motion or static data. Again, the final occupancy detection can be done by post-processing the motion detection output using a simple counting process.

### 4.4.3 Experimental Results

The experiment setup and data collection are provided in Section 4.3.2. Presence data for training also come from days $9$ and $10$. Collection for human presence data happened when there are at least one person in the lab, going about normal activities without introducing intentional motions. The duration of collecting human presence data is around $1$ hour on each day. The number of instances used to train the SVM modules is $N^{(0)} = 5000$ and $N^{(1)} = 10000$.

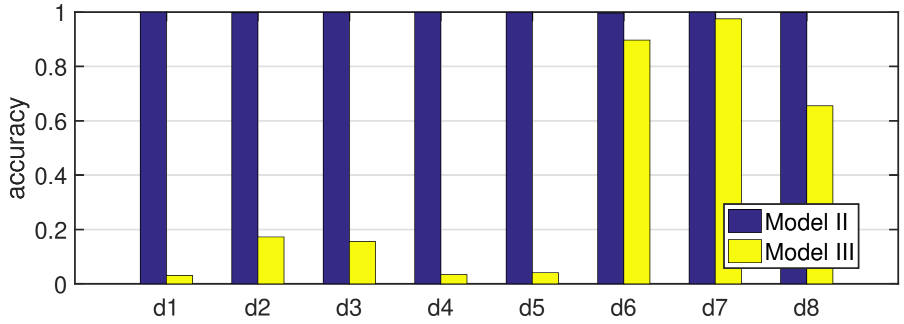
#### *Training and test results*

All the models are trained using a RBF kernel with $\gamma = 0.008$. The regularization parameter $C$ in (4.1) and (4.3) are chosen by a 5-fold cross validation scheme. Notice that to improve sensitivity to human presence, an upper bound smaller than $1$ is imposed on $C$ in (4.1) during cross-validation. How to set the value of the upper bound is guided by the performance of Model II on training set, in particular the false alarm rate on human free data, along with the percentage of samples in human presence data that are considered to be motion data. In our experiment, the upper bound of $C$ is chosen with the assumption that $20\%$ of the samples in human presence data contains motion while the rest correspond to motionless presence. Notice that such percentage value is not fixed for different presence data sets due to varying human mobility levels.

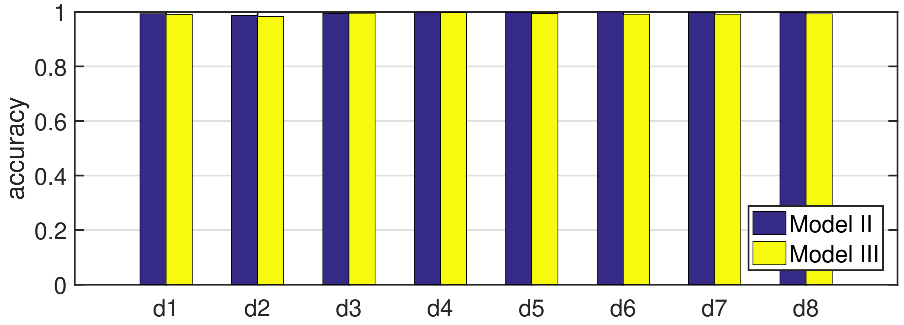
The average accuracy of Model II trained using the proposed two stage approach under different test days are shown in Fig. 33. Clearly, Model II achieves near perfect detection performance on all test days for both presence and absence test data. Given that the motion types are different over the $8$ test days, the near perfect result demonstrates that the trained model is sensitive to different types of human motions. Data collected on days $1-5$ were more than 1 month apart from the training data collected on days $9$ and $10$. The lab settings in terms of furniture locations were quite different. This demonstrates that the trained Model II is robust against environment variations over time.

### *Comparison with SVDD Based Classification*

Since the same test sets are used in the experiment for SVDD models as given in Section 4.3.2, to see the performance difference between the proposed two-stage approach and SVDD models, one can compare Fig. 33 and Fig. 30 directly. By including presence data into the training set, Model II can have high sensitivity to human motions on all test days while raising low false positives when there is nobody around. Therefore, the proposed approach doesn't suffer from the trade-off issue in SVDD where one needs to adjust the regularization term $C$ in (4.1) according to requirements. Training SVDD with motion samples can improve its performance as shown in Section 4.3.2. However, collecting motion samples for optimal performance requires human efforts and is not as easy as obtaining presence data for getting Model II.



(a) Human free



(b) Human motion

Fig. 33: Performance of models on different test days

### *Comparison with C-SVC Based Classification*

As an another baseline for performance comparison, we include Model III which is obtained by training a two-class SVC using all the training data without any de-contamination effort. From Fig. 33, Model III has significantly high false positive rates on all test days except day 7. The elevated false positive rates are due to the fact that the training is done with a large amount of human presence data behaving similar to human present data. As such, new samples corresponding to human absence can be easily mis-classified as human presence.

## 4.5 Summary

In this chapter, we introduced two SVM based presence detection approaches which have reduced training efforts especially in training data collection. The SVDD based method in Section 4.3 regards samples corresponding to human motions as outliers and training is done using human free data only. The drawback of this approach is that it does not have built-in flexibility to achieve desired trade-off between false alarm rates and sensitivity to human motions. For example, a close to $1$ upper bound imposed on the regularization term $C$ in cross-validation can lead to a trained model with low false positives but elevated false negatives. Adding a few samples from human motions to the training set can mitigate this problem through SVDD with negative samples. However, the collection and selection of such motion data can impact the performance of the obtained model.

In the second part of this chapter, a two stage approach was proposed which has a more desirable way of collecting human presence data. That is, during the training data collection, human subjects are allowed inside the monitored regions and to perform their daily activities without introducing intentional motions. To address the challenge of sample contamination in human presence data brought by such simple collection scheme, a SVDD based model trained by human free data is adopted for removing measurements which involve only stationary humans. The remaining training data is then fed into a two-class SVC for motion detection. By using off-the-shelf WiFi devices, it was demonstrated that the proposed method has low false alarm rates in a human absent environment, while maintaining high sensitivity to human motions. Through extended periods of test, the method is further shown to be robust against temporal variations in the wireless environment.

# CHAPTER 5

# CONCLUSION AND FUTURE RESEARCH

## 5.1 Conclusion

Wireless communications are an integral part of any situational awareness applications. In this dissertation, we investigated challenges in adopting RF communications for situational awareness applications from two difference aspects. When used as a medium for communicating digital data, RF systems should be able to provide a reliable and robust link for timely information dissemination. Besides, the ever exploding wireless devices and services that we became increasingly dependent upon suggest that those RF signals are entrenched in our lives in both a figurative and in a literal sense. While technological advances have made it possible to get us connected anytime anywhere through wireless communications, it is also quite reasonable to suggest that more can be done with these RF signals than simply treating them as a carrier for data propagation. Recently, the field of exploiting RF signals directly for sensing has gained more and more attentions.

When situational awareness systems involve airborne platforms, their success relies on the ability to deliver high throughput and low latency data reliably from these airborne platforms. MIMO technique seems to be a nature candidate to help achieve this goal. Chapter 2 is dedicated to tackle a slew of challenges in realizing MIMO throughput gain for a real airborne communication system. In the designed MIMO system, spatial spreading and temporal spreading approaches are incorporated into the baseline D-BLAST architecture to address antenna blockage and the presence of interference respectively. Due to the adoption of such simple spreading techniques, a low rate feedback scheme is feasible to drive adaptation at the transmitter when experiencing channel impairments in the airborne communication environment.

Chapter 3 and Chapter 4 are focused on RF sensing. Chapter 3 presents a passive WiFi sensing system for indoor occupancy detection. Carefully designed pre-processing steps including DFT extract motion induced variation from CSI time series while mitigating the impact from factors irrelevant to human presence. A CNN with parallel architecture is adopted to harvest presence information from both magnitude and phase of the CSI effectively. The final post-processing module helps achieve reliable presence detection from instantaneous motion detection outputs. Through extensive experiments, including long-term tests at three different environment and comparison studies with both PIR sensors and existing RF based solutions, we proved that the proposed system is much more sensitive than PIR sensors, while maintaining desired robustness in human free environment.

Even though data-driven approaches are more prevalent in the domain of inferring presence information from RF signals, one of the major challenges which can potentially prevent their application is that collection of human presence data involves significant human intervention. Chapter 4 addressed such challenge by presenting two SVM based models. The first proposed model is based on one-class SVM and requires training data from human free environment only. However one needs to carefully adjust the regularization term in the model to achieve balance between sensitivity to human presence and false alarm. In the second proposed model, human presence data are incorporated into the training set to boost performance and their collection process requires neither tedious labeling nor continuous human motions. Due to induced sample contamination from stationary human measurements, a SVDD model is first trained to remove contaminated samples and is followed by a two-class SVC to achieve motion detection.

## 5.2 Future Research

We conclude this dissertation by listing a few future research directions.

- The presented work in Chapter 2 assumes a narrowband MIMO system where spatial and temporal diversities are explored to combat various channel impairments. For a wideband system, the frequency domain diversity can also be utilized. For example, with MIMO-OFDM, channel diversity across different sub-carriers can be exploited in addition to (or in place of) spatial or temporal diversities.
- Existing presence detection system can not distinguish moving objects. Therefore, it can not separate a dog from a human as both will result in motion-present features in CSI. Applying various supervised learning approaches for human/pets classification using passive WiFi signals is of interest to future research. To the best of our knowledge, none of the previous works exploit CSI information for detecting animals such as dogs [77]. Due to differences in body shapes and occupied space for activities, humans and pets may introduce distinct variation patterns in the signal propagation paths. This results in distinguishable changes across frequency domain. Also, pets like dogs usually move faster than humans which may lead to high temporal variation due to larger doppler spread.

# REFERENCES

[1] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell labs technical journal*, vol. 1, no. 2, pp. 41–59, 1996.

[2] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[3] E. Telatar, “Capacity of multi-antenna gaussian channels,” *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[4] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[5] M. J. Gans, “Aircraft free-space mimo communications,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2009, pp. 663–666.

[6] Gans, Michael J, “System for airborne communications,” U.S. Patent 9419330B1, Aug. 2016.

[7] H. Abdelnasser, M. Youssef, and K. A. Harras, “Wigest: A ubiquitous wifi-based gesture recognition system,” in *Proc. IEEE Conf. on Comput. Commun. (INFOCOM)*, Hong Kong, China, Apr. 2015, pp. 1472–1480.

[8] Y. Gu, F. Ren, and J. Li, “Paws: Passive human activity recognition based on wifi ambient signals,” *IEEE Internet Things J.*, vol. 3, no. 5, pp. 796–805, 2015.

[9] M. Moussa and M. Youssef, “Smart cevices for smart environments: Device-free passive detection in real environments,” in *Proc. IEEE Int. Conf. Pervasive Computing and Commun.*, Galveston, TX, USA, Mar. 2009, pp. 1–6.

[10] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, “Falldefi: Ubiquitous fall detection using commodity wi-fi devices,” *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2018.

[11] Q. Xu, Y. Han, B. Wang, M. Wu, and K. R. Liu, “Indoor events monitoring using channel state information time series,” *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4977–4990, 2019.

[12] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, “Design and implementation of a csi-based ubiquitous smoking detection system,” *IEEE/ACM Trans. Networking*, vol. 25, no. 6, pp. 3781–3793, 2017.

[13] S. Li, X. Li, K. Niu, H. Wang, Y. Zhang, and D. Zhang, “Ar-alarm: An adaptive and robust intrusion detection system leveraging csi from commodity wi-fi,” in *Proc. Int. Conf. Smart Homes Health Telematics*, Paris, France, Aug. 2017, pp. 211–223.

[14] X. Liu, J. Cao, S. Tang, and J. Wen, “Wi-sleep: Contactless sleep monitoring via wifi signals,” in *Proc. IEEE Real-Time Syst. Symp.*, Rome, Italy, Dec. 2014, pp. 346–355.

[15] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, “E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures,” in *Proc. 20th Annu. Int. Conf. Mobile Computing Networking*, Hawaii, USA, Sep. 2014, pp. 617–628.

[16] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, “Wi-chase: A wifi based human activity recognition system for sensorless environments,” in *Proc. IEEE 18th Int. Symp. A World of Wireless, Mobile Multimedia Networks (WoWMoM)*, Macao, China, Jun. 2017, pp. 1–6.

[17] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, “Device-free human activity recognition using commercial wifi devices,” *IEEE J. Select. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, 2017.

[18] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, “Csi-based device-free wireless localization and activity recognition using radio image features,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10 346–10 356, 2017.

[19] Q. Xu, F. Zhang, B. Wang, and K. R. Liu, “Time reversal indoor tracking with centimeter accuracy,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 6433–6437.

[20] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, “Indotrack: Device-free indoor human tracking with commodity wi-fi,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.

[21] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, “Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi,” in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Networking Computing*, Chennai, India, Jul. 2017, pp. 1–10.

[22] E. Soltanaghaei, A. Kalyanaraman, and K. Whitehouse, “Peripheral wifi vision: Exploiting multipath reflections for more sensitive human sensing,” in *Proc. Int. Workshop Physical Analytics*, Niagara Falls, NY, USA, Jun. 2017, pp. 13–18.

[23] Y. Zhao, N. Patwari, J. M. Phillips, and S. Venkatasubramanian, “Radio tomographic imaging and tracking of stationary and moving people via kernel distance,” in *Proc. ACM/IEEE Int. Conf. Inform. Process. Sensor Networks (IPSN)*, Philadelphia, PA, USA, Apr. 2013, pp. 229–240.

[24] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, “Non-invasive detection of moving and stationary human with wifi,” *IEEE J. Select. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, 2015.

[25] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, “Signfi: Sign language recognition using wifi,” *Proc. ACM Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–21, 2018.

[26] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, “Freedetector: Device-free occupancy detection with commodity wifi,” in *Proc. IEEE Int. Conf. Sensing, Commun. and Networking (SECON Workshops)*, San Diego, CA, USA, Jun 2017, pp. 1–5.

[27] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, “Enabling contactless detection of moving humans with dynamic speeds using csi,” *ACM Trans. Embedded Computing Syst. (TECS)*, vol. 17, no. 2, pp. 1–18, 2018.

[28] L. Gong, W. Yang, Z. Zhou, D. Man, H. Cai, X. Zhou, and Z. Yang, “An adaptive wireless passive human detection via fine-grained physical layer information,” *Ad Hoc Networks*, vol. 38, pp. 38–50, 2016.

[29] S. Palipana, P. Agrawal, and D. Pesch, “Channel state information based human presence detection using non-linear techniques,” in *Proc. 3rd ACM Int. Conf. Syst. Energy-Efficient Built Environments*, Palo Alto, CA, USA, Nov. 2016, pp. 177–186.

[30] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, “Fimd: Fine-grained device-free motion detection,” in *IEEE 18th Int. Conf. Parallel Distributed Syst.* IEEE, 2012, pp. 229–235.

[31] H. Zhu, F. Xiao, L. Sun, R. Wang, and P. Yang, “R-ttwd: Robust device-free through-the-wall detection of moving human with wifi,” *IEEE J. Select. Areas in Commun.*, vol. 35, no. 5, pp. 1090–1103, 2017.

[32] S. Kianoush, S. Savazzi, V. Rampa, and M. Nicoli, “People counting by dense wifi mimo networks: Channel features and machine learning algorithms,” *Sensors*, vol. 19, no. 16, p. 3450, 2019.

[33] S. Savazzi, S. Sigg, F. Vicentini, S. Kianoush, and R. Findling, “On the use of stray wireless signals for sensing: a look beyond 5g for the next generation of industry,” *Comput.*, vol. 52, no. 7, pp. 25–36, 2019.

[34] P. M. Holl and F. Reinhard, “Holography of wi-fi radiation,” *Physical review letters*, vol. 118, no. 18, pp. 183 901(1)–183 901(5), 2017.

[35] P. Almers, F. Tufvesson, and A. F. Molisch, “Keyhole effect in mimo wireless channels: Measurements and theory,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3596–3604, 2006.

[36] M. J. Gans, K. M. Borle, B. Chen, T. Freeland, D. McCarthy, R. Nelson, D. Overrocker, and P. Oleski, “Enhancing connectivity of unmanned vehicles through mimo communications,” in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Las Vegas, USA, Sept. 2013, pp. 1–5.

[37] Y. Liu, K. M. Borle, J. J. Mroczek, B. Chen, M. J. Gans, P. J. Oleski, and R. J. Michalak, “Variable rate mimo communication with airborne platforms: An experimental study,” in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Baltimore, MD, USA, Nov. 2016, pp. 1167–1172.

[38] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[39] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[40] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[41] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, 2000.

[42] T. Teixeira, G. Dublon, and A. Savvides, "A survey of human-sensing: Methods for detecting presence, count, location, track, and identity," *ACM Computing Surveys*, vol. 5, no. 1, pp. 59–69, 2010.

[43] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Computing*, Maui, Hawaii, USA, Sept. 2016, pp. 363–373.

[44] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. J. Spanos, "Wifi-based human identification via convex tensor shapelet learning," in *AAAI Conf. Artificial Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 1711–1718.

[45] S. D. Regani, Q. Xu, B. Wang, M. Wu, and K. R. Liu, "Driver authentication for smart car using wireless sensing," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2235–2246, 2019.

[46] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, 2017.

[47] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free wifi human sensing: From pattern-based to model-based approaches," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 91–97, 2017.

[48] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[49] B. Guo, Y. J. Chen, N. Lane, Y. Liu, and Z. Yu, “Behavior recognition based on wi-fi csi: Part 1,” *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 90–90, 2017.

[50] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, “Freesense: a robust approach for indoor human detection using wi-fi signals,” *Proc. ACM Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–23, 2018.

[51] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, “Understanding and modeling of wifi signal based human activity recognition,” in *Proc. 21st Annu. Int. Conf. Mobile Computing Networking*, Paris, France, Sept. 2015, pp. 65–76.

[52] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, “Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas,” *Proc. ACM Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, 2019.

[53] J. Lv, D. Man, W. Yang, L. Gong, X. Du, and M. Yu, “Robust device-free intrusion detection using physical layer information of wifi signals,” *Appl. Sci.*, vol. 9, no. 1, pp. 175–191, 2019.

[54] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, “Omnidirectional coverage for device-free passive human detection,” *IEEE Trans. Parallel Distributed Systs*, vol. 25, no. 7, pp. 1819–1829, 2013.

[55] S. Di Domenico, M. De Sanctis, E. Cianca, and M. Ruggieri, “Wifi-based through-the-wall presence detection of stationary and moving humans analyzing the doppler spectrum,” *IEEE Aerospace Electron. Syst. Mag.*, vol. 33, no. 5-6, pp. 14–19, 2018.

[56] X. Li, D. Zhang, J. Xiong, Y. Zhang, S. Li, Y. Wang, and H. Mei, “Training-free human vitality monitoring using commodity wi-fi devices,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–25, 2018.

[57] S. Li, Z. Liu, Y. Zhang, X. Niu, L. Wang, and D. Zhang, “A real-time and robust intrusion detection system with commodity wi-fi,” in *Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Computing Proc. ACM Int. Symposium Wearable Comput.*, London, UK, Sept. 2019, pp. 316–319.

[58] S.-H. Fang, C.-C. Li, W.-C. Lu, Z. Xu, and Y.-R. Chien, “Enhanced device-free human detection: Efficient learning from phase and amplitude of channel state information,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3048–3051, 2019.

[59] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, “Person-in-wifi: Fine-grained person perception using wifi,” in *Proc. IEEE Int. Conf. Comput. Vision*, Seoul, Korea, Nov. 2019, pp. 5452–5461.

[60] J. Lv, D. Man, W. Yang, X. Du, and M. Yu, “Robust wlan-based indoor intrusion detection using phy layer information,” *IEEE Access*, vol. 6, pp. 30 117–30 127, 2017.

[61] H. Huang and S. Lin, “Widet: Wi-fi based device-free passive person detection with deep convolutional neural networks,” *Computer Communications*, vol. 150, pp. 357–366, 2020.

[62] Y. Liu, T. Wang, Y. Jiang, and B. Chen, “Harvesting ambient rf for presence detection through deep learning,” *arXiv preprint arXiv:2002.05770*, 2020.

[63] D. M. Tax and R. P. Duin, “Support vector data description,” *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[64] GNU Radio, [Online]: http://www.gnuradio.org.

[65] USRP: Universal Software Radio Perpheral, accessed September 2013. [Online]. Available: http://www.ettus.com.

[66] UHD: USRP Hardware Driver, [Online]. Available: https://kb.ettus.com/UHD.

[67] FFmpeg, [Online]. Available: https://ffmpeg.org/.

[68] T. S. Rappaport, *Wireless Communication - Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[69] C. Gonzalez and E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1991.

[70] Y. Xie, Z. Li, and M. Li, “Precise power delay profiling with commodity wifi,” in *Proc. 21st Annu. Int. Conf. Mobile Computing and Networking*, ser. MobiCom ’15. New York, NY, USA: ACM, 2015, p. 53–64. [Online]. Available: http://doi.acm.org/10.1145/2789168.2790124

[71] IEEE Std. 802.11n-2009: Enhancements for higher throughput, 2009. [Online]. Available: http://www.ieee802.org.

[72] F. Chollet *et al.*, “Keras,” https://keras.io, 2015.

[73] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[74] Honeywell PIR sensor DT8035. [Online]. Available: https://www.security.honeywell.com/product-repository/dt8035.

[75] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[76] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, 2nd ed. New York, NY, USA: Springer series in statistics, 2009.

[77] Y. Ma, G. Zhou, and S. Wang, “Wifi sensing with channel state information: A survey,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

# VITA

NAME OF AUTHOR: Yang Liu

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, Syracuse, NY, USA, 2013-2020

University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 2009-2013

DEGREES AWARDED:

B.E., 2013, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

PUBLICATIONS:

Y. Liu, T. Wang and B. Chen, "Harvesting Ambient RF for Presence Detection Through Deep Learning", *IEEE Trans. Neural Networks, Learning Syst.*, accepted Nov., 2020.

J.J. Mroczek, Y. Liu, and B. Chen, Provisional Patent, “Spatial diversity, to counter loss of radio frequency signal when one or more antenna elements are in outage”, filed Nov., 2020.

Y. Liu, T. Wang, Y. Jiang and B. Chen, "Passive RF Sensing for Motion Detection With Contaminated Samples", *submitted to Proc. IEEE ICASSP*, Toronto, Ontario, Canada, Jun 2021.

T. Wang, Y. Liu and B. Chen, "On Exponentially Consistency of Linkage-based Hierarchical Clustering Algorithm Using Kolmogrov-Smirnove Distance", *Proc. IEEE ICASSP*, Barcelona, Spain, May 2020.

Y. Liu, B. Chen, J.J. Mroczek, J.E. Malowicki and R.J. Michalak, "Robust MIMO communications against antenna blockage and interference", *Proc, IEEE MILCOM*, Norfolk, VA, Nov. 2019.

Y. Liu, K. Borle, J.J. Mroczek, B. Chen, M.J. Gans, P.J. Oleski, and R.J. Michalak, "Variable rate MIMO communication with airborne platforms: an experimental study", *Proc, IEEE MILCOM*, Baltimore, MD, Nov. 2016.