

“© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Single-Target Real-Time Passive WiFi Tracking

Zhongqin Wang, J. Andrew Zhang, *Senior Member, IEEE*, Min Xu, *Member, IEEE*,
and Y. Jay Guo, *Fellow, IEEE*

Abstract—Device-free human tracking is an essential ingredient for ubiquitous wireless sensing. Recent passive WiFi tracking systems face the challenges of inaccurate separation of dynamic human components and time-consuming estimation of multi-dimensional signal parameters. In this work, we present a scheme named **WiFi Doppler Frequency Shift (WiDFS)**, which can achieve single-target real-time passive tracking using channel state information (CSI) collected from commercial-off-the-shelf (COTS) WiFi devices. We consider the typical system setup including a transmitter with a single antenna and a receiver with three antennas; while our scheme can be readily extended to another setup. To remove the impact of transceiver asynchronization, we first apply CSI cross-correlation between each RX antenna pair. We then combine them to estimate a Doppler frequency shift (DFS) in a short-time window. After that, we leverage the DFS estimate to separate dynamic human components from CSI self-correlation terms of each antenna, thereby separately calculating angle-of-arrival (AoA) and human reflection distance for tracking. In addition, a hardware calibration algorithm is presented to refine the spacing between RX antennas and eliminate the hardware-related phase differences between them. A prototype demonstrates that WiDFS can achieve real-time tracking with a median position error of 72.32 cm in multipath-rich environments.

Index Terms—WiFi, Tracking, CSI, Doppler Frequency Shift, Hardware Calibration.

1 INTRODUCTION

PASSIVE WiFi tracking is a promising technique that uses WiFi signals to locate people without needing any other sensors or wearable devices. Compared to other wireless signal based solutions like radio frequency identification (RFID) [1], [2], [3], [4], [5] and millimetre wave (mmWave) radar [6], [7], [8], [9], WiFi infrastructures are almost ubiquitous at public work places and homes, thereby avoiding the need of deploying dedicated wireless tracking infrastructure and devices. Wireless sensing techniques are free from light conditions and even perform well in non-line-of-sight (NLOS) scenarios where a target is blocked [10], [11].

Passive WiFi tracking has gained much attention from academic and industrial communities over the past years. Recent works [12], [13], [14], [15] exploit a stable and feature-rich signal parameter for positioning, i.e., channel state information (CSI), which can be extracted from a commercial-off-the-shelf (COTS) WiFi network interface controller (NIC) (e.g., Intel 5300 [16] and Atheros QCA9558 [17]). Such CSI-based localization solutions could achieve finer decimeter-level localization accuracies. To the best of our knowledge, however, most existing approaches with fine tracking accuracy are hard to directly adopt in practical applications due to the following reasons. First, huge computation in each position estimation may jeopardize the system's real-time performance [18]. Second, estimation errors in many Doppler-based continuous tracking solutions [19] may accumulate over time, resulting in trajectory drift. Third, deep learning-based solutions [20], [21] require a huge amount of labeled CSI data in specified scenarios for training. Since dif-

ferent environments have different multipath interference, the trained network may not be universal.

There are also three major challenges in passive WiFi tracking, which have not been well addressed in the literature. These challenges are detailed below.

1) WiFi transceiver clock asynchronization [13], [17], [22], [23] results in time-varying phase shifts in CSI. Many existing works [24], [25], [26] exploit cross-correlation between CSIs of pairwise antennas to address the asynchronous signal processing problem. However, this operation introduces a side product, i.e., the conjugate terms of the dynamic cross-correlation terms of interest. In this case, *DFS ambiguity* is created, meaning that the DFS to be estimated can be the true value or its negative value. A common solution to suppressing the ambiguity is adding a constant to the reference signal and subtracting another one to the rest signals before the cross-correlation [22], [24]. However, this does not always work, especially in multipath-rich scenarios.

2) Separating the dynamic human components is challenging. An existing common method [19] aims to directly separate the dynamic human component from cross-correlation terms by adding a factor to the CSI amplitudes of the reference antenna and subtracting another factor from the CSI amplitudes of other antennas. However, this power adjustment solution cannot completely eliminate the impact of the side product in dynamic component separation. Furthermore, these dynamic components may not always be reliable for position indication. For example, human body is not a perfect reflector and may not reflect signals to a receive antenna array at some sampling time.

3) The difference in WiFi hardware, including WiFi NIC, RX antennas, and RX antenna cables, introduces different phase shifts on each RX antenna, which significantly impact human tracking accuracy. An existing method [27] conducts WiFi hardware calibration using a SMA splitter. However, an unknown π -radians phase ambiguity will be

- Zhongqin Wang, J. Andrew Zhang (Corresponding Author), and Min Xu are with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney 2007, Australia.
E-mail: {zhongqin.wang, andrew.zhang, min.xu}@uts.edu.au
- Y. Jay Guo is with the Global Big Data Technologies Centre, University of Technology Sydney, Sydney 2007, Australia.
E-mail: jay.guo@uts.edu.au

induced. Also, since the RX antenna array is customized, the measured antenna spacing inevitably deviates from the actual value. Removing the π -radians phase ambiguity and estimating the true RX antenna spacing are challenging.

In this paper, we propose a WiFi single-target passive tracking scheme, called **WiFi Doppler Frequency Shift (WiDFS)**, which enables to overcome the above three challenges and can be run in real time at a medium class mini PC. Currently, a standard COTS NIC supports up to three antennas. WiDFS tracks a moving person using a COTS WiFi transceiver with one TX antenna and three RX antennas. Using three-RX-antenna is essential as we need to use their CSIs to estimate the angle-of-arrival (AoA) of incoming signals, and to remove transceiver clock asynchronization and accurately estimate the DFS caused by human movement. Once the DFS estimate is obtained, we then leverage it to separate dynamic human components. WiDFS can then track the person by estimating the AoA in the direction of human reflection and the length of the reflection path from the transmitter to the person and then the receiver. Our main contributions are as follows.

1) We propose a DFS estimation algorithm based on cross-correlation between each RX antenna pair. Compared to existing solutions, WiDFS can eliminate the impact of the side product that is mixed with the dynamic cross-correlation term of interest. WiDFS leverages the static cross-correlation term which is usually abandoned in previous works. This term can be easily obtained by averaging over a CSI sampling window. WiDFS then conducts a straightforward transformation to resolve the DFS ambiguity caused by the side product. Since three RX antennas are separated by less than half a wavelength, it is reasonable to assume that the DFS of each antenna is almost the same. Thus, WiDFS builds a CSI observation matrix to estimate an accurate DFS using a subspace-based MUSIC algorithm in each short-time sampling window.

2) We design a lightweight algorithm that uses the estimated DFS to separate the dynamic human components, which is more robust and accurate than the power adjustment and reference antenna solutions, especially in actual multipath-rich scenarios. In this work, we focus on each antenna's instantaneous power of channel frequency response (CFR) based on self-correlation. The CFR power is also free from the impact of transceiver clock asynchronization. WiDFS first relies on the estimated DFS to refine the CFR power in a CSI sampling window. Then it uses a simple yet effective solution to reconstruct dynamic human components by formulating a linear least-squares problem. We further propose a windowed algorithm that can deal with the problem that RX antennas cannot capture human reflections or can only capture minor reflections. WiDFS combines CSI data from multiple sampling windows for localization parameter estimation. It relies on the estimated DFS to achieve unsupervised motion sensing, which can detect the absence and presence of a moving person.

3) We present a WiFi hardware calibration solution to estimate the hardware-dependent phase shifts and antenna spacing between any two antennas, without any specialized devices. WiDFS collects CSIs by deploying a TX antenna on each side of an RX antenna array and then uses a standard phase-distance model for calibration.

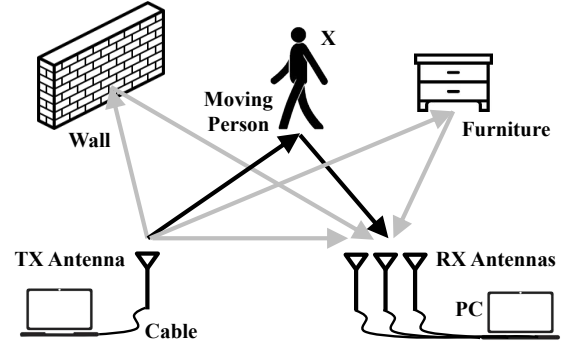


Fig. 1: Signal propagation paths in dynamic environments

A prototype of WiDFS is implemented using a transmitter with a single TX antenna and a receiver with three RX antennas forming a linear uniform antenna array. WiDFS is programmed using Python. The experiments demonstrate that WiDFS achieves median and 90th percentile position errors of 72.31 cm and 170.8 cm, respectively. Such a localization accuracy exceeds a state-of-the-art technique Widar2.0 [25] by about 36 cm and 70 cm. More importantly, WiDFS costs a mean running time of 0.076 s on a mini PC (online) and 0.024 s on a MacBook Pro (offline) to output each position estimate when collecting about 0.1 second CSI data, leading to real-time tracking in practical applications.

2 CSI MODELING IN DYNAMIC ENVIRONMENTS

This section describes our CSI model in a dynamic environment where a person is moving in a typical room. The CSI provides information on the environment as changes of signal propagation environment cause variations of CSI over time. The CSI can be accessed on certain COTS WiFi NICs using 802.11n or Atheros CSI tools [16], [17]. Such a NIC typically supports up to three Tx and Rx antennas. Fig. 1 shows a typical WiFi-based sensing scenario. In this setup with Intel NIC 5300, we use a Tx with a single antenna and a receiver with three RX antennas forming a uniform linear antenna array (ULA). The 3-antenna ULA offers Angle-of-Arrival (AoA) estimation capability for a limited number of multipath signals.

2.1 General CSI model in COTS WiFi systems

The CSI characterizes the channel frequency response (CFR) of a wireless signal propagating from a transmitter to a receiver. In the IEEE 802.11n standard, a channel has a 20 MHz bandwidth with 30 subcarriers. Then the CSI matrix in our scheme contains the number of $1 \times 3 \times 30$ complex channel coefficients at each sampling time. In COTS WiFi systems, the actual CSI measurements always suffer from the additional noise caused by NIC processing imperfection and WiFi hardware diversity.

1) **WiFi NIC Processing Imperfection.** The original CSI may contain *time-varying* terms associated with imperfect signal processing in NIC, such as residual time and frequency offsets due to transmitter-receiver clock asynchronization, a power control uncertainty error and an I-Q imbalance error [13]. These errors produce the same impact on all RX antennas. Let $H_{j,k}^c$ denote the CFR associated with such imperfections at the j -th subcarrier ($j = 1, 2, \dots, 30$) and the

k -th CSI sample (or time-slot, which typically corresponds to the k -th CSI packet, $k = 1, 2, \dots$),

$$H_{j,k}^e = \rho_k^{agc} e^{-J(\varphi_{j,k}^{FO} + \varphi_{j,k}^{TO})}, \quad (1)$$

where ρ_k^{agc} represents the gain set at the automatic gain controller (AGC) at the receiver and it dynamically varies with different channel characteristics; $\varphi_{j,k}^{FO}$ is the time-varying phase shift caused by the frequency offset (FO) at each subcarrier frequency; $\varphi_{j,k}^{TO}$ is the time-varying phase shift caused by the time offset (TO); both $\varphi_{j,k}^{FO}$ and $\varphi_{j,k}^{TO}$ may vary across CSI samples because of the un-locked clocks between Tx and Rx. The CFO may be accurately estimated at the receiver and hence the term $\varphi_{j,k}^{FO}$ can be compensated. However, TO associated with the transmitter and receiver clocking difference is generally hard to be estimated.

2) WiFi Hardware Diversity. The WiFi hardware includes external WiFi antennas, antenna cables, SMA connectors as well as NIC itself. Their impacts on receiving signals at each subcarrier keep unchanged over time. However, manufacturing imperfection may cause different phase shifts on each RX antenna. According to [27], their variations over subcarriers can be ignored, so the WiFi hardware-related CFR H_i^h at the i -th RX antenna ($i = 1, 2, 3$) can be represented as

$$H_i^h = \rho_i^h e^{-J\varphi_i^h}, \quad (2)$$

where ρ_i^h and φ_i are the WiFi hardware-related attenuation and phase shift at the i -th RX antenna.

Accounting for the two multiplicative interference terms in (1) and (2), a general CSI model [28], [29] at the i -th RX antenna, the j -th subcarrier and the k -th CSI sample is represented as

$$CSI_{i,j,k} = H_{j,k}^e H_i^h \sum_{l=1}^L H_{i,j,k}^l[l], \quad (3)$$

where

$$H_{i,j,k}^l[l] = \rho_{i,j,k}^l[l] e^{-J2\pi \frac{f_j}{c} \left(d_k[l] + c \frac{f_{i,k}^D[l]}{f_c} \Delta t + (i-1) \Delta d \sin \theta_k[l] \right)}. \quad (4)$$

Some variables in Eq. 4 are described as follows: f_j is the j -th subcarrier frequency, f_c is the center frequency, c is the speed of light, Δt is the sampling time interval, Δd is the RX antenna spacing no more than half a wavelength. For the l -th multipath, $\rho_{i,j,k}^l[l]$ and $d_k[l]$ are the signal propagation attenuation and length, $f_{i,k}^D[l]$ is the DFS which is introduced to the carrier frequency at the i -th RX antenna due to object movement ($v_{i,k}[l] = c \frac{f_{i,k}^D[l]}{f_c}$ is called the radial speed), $\theta_k[l]$ is the AoA that is the direction from which a reflection is received by the ULA.

According to [28], we define a short-time sampling window which is typically a few milliseconds when tracked objects move at velocities up to several meters per second. In WiDFS, the length of the time window is set to 0.1 seconds during which $N_p = 100$ CSI packets are collected due to the sampling frequency f_s of 1 KHz. In this short-time window, since the indoor environment changes little, the channel characteristic-dependent variables ρ_k^{agc} and $\rho_{i,j,k}^X$ may not have a significant change. And a person can be regarded as moving with a constant speed $v_{i,k}[l]$ within the 0.1-seconds

time window. In this case, we assume that the gain ρ_k^{agc} , attenuation $\rho_{i,j,k}^X$, and DFS $f_{i,k}^D[l]$ in Eq. 4 all remain almost unchanged, and we denote $\rho_k^{agc} = \rho^{agc}$, $\rho_{i,j,k} = \rho_{i,j}$, and $f_{i,k}^D[l] = f_i^D[l]$, respectively. In the following, we will rely on these assumptions to revise the above general CSI model.

2.2 Our CSI model for single-target passive tracking

Here we extend the general CSI model and introduce ours for single-target passive tracking in a dynamic environment in a short-time window.

When a transmitter emits WiFi signals to space, the receiver receives two categories of signals. One is the *direct* signal that travels along the direct path from the transmitter to the receiver. Another is the *reflected* signals that bounce off different objects like the floor, wall, furniture, and the tracked person. Furthermore, we divide these signals into *static* and *dynamic* signals: (1) the former includes a direct signal that propagates from the transmitter to the receiver, and the signals that reflect off surrounding static objects to the receiver; (2) the latter contains the reflected signals that directly reflect off the human body to the receiver, and those that firstly bounce off the human body to other surrounding objects and then travel back to the receiver. Let $H_{i,j}^S$ be the CFR corresponding to the *static* signals between the TX antenna and the i -th RX antenna at the j -th subcarrier, called *static component*. Specially, the signal strength of the direct path between the TX and RX antennas is much higher than other static multipath signals. Let $H_{i,j,k}^X$ be the CFR corresponding to the reflected signals off the moving person X , called *dynamic human component*. Thus, we rewrite the CSI model in Eq. 3 to formulate the problem of single-target passive tracking as

$$CSI_{i,j,k} = H_{j,k}^e H_i^h \left(H_{i,j}^S + H_{i,j,k}^X \right), \quad (5)$$

where

$$\begin{cases} H_{i,j}^S = \rho_{i,j}^S e^{-J2\pi \frac{f_j}{c} d_i^S} + \mathbb{N}_{i,j}^S \\ H_{i,j,k}^X = \rho_{i,j}^X e^{-J2\pi \frac{f_j}{c} d_{i,k}^X} + \mathbb{N}_{i,j,k}^X \end{cases}. \quad (6)$$

In the above, $\rho_{i,j}^S$ and $\rho_{i,j}^X$ are the propagation attenuations of the direct and reflected path; d_i^S is the distance of the direct path between the transmitter and the i -th RX antenna, which can be manually measured in advance; $d_{i,k}^X$ is the distance of the reflected path where the transmit signal is reflected from the tracked person X to the i -th RX antenna at the k -th sampling point (for simplicity, we call it *human reflection distance*); $\mathbb{N}_{i,j}^S$ and $\mathbb{N}_{i,j,k}^X$ are the noise terms caused by other minor multipath reflections.

In a short-time window as defined above, the human reflection distance $d_{i,k}^X$ can be represented as

$$d_{i,k}^X \approx d_{i,1}^X + c \frac{f_j^D}{f_c} (k-1) \Delta t, \quad (7)$$

where $d_{i,1}^X$ is the human reflection distance at the initial position in the window. Further, the distance $d_{i,k}^X$, $i \in \{2, 3\}$, can be rewritten using an initial AoA θ_1^X ,

$$\begin{aligned} d_{i,k}^X &\approx d_{1,k}^X + (i-1) \Delta d \sin \theta_k^X \\ &\approx d_{1,1}^X + c \frac{f_1^D}{f_c} (k-1) \Delta t + (i-1) \Delta d \sin \left(\theta_1^X + \Delta \theta_k^X \right). \end{aligned} \quad (8)$$

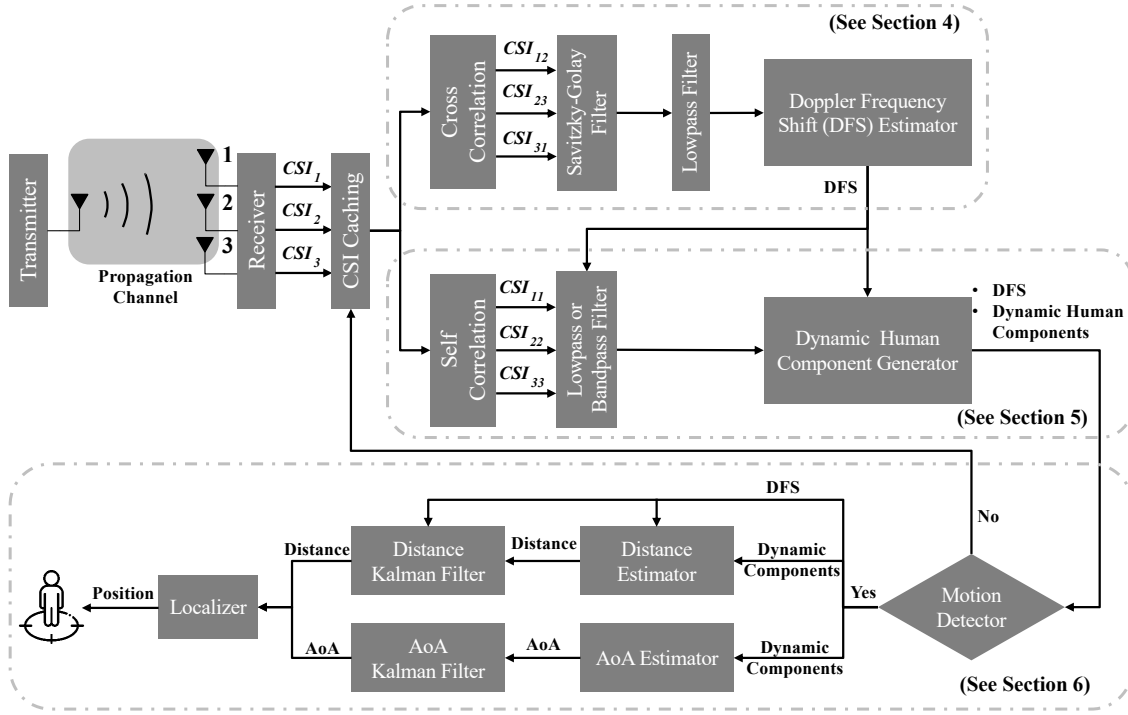


Fig. 2: WiDFS workflow

where $\Delta\theta_k^X$ is the AoA spacing between the first and the k -th CSI samples.

In indoor environments, since the person does not moves too fast, $\Delta\theta_k^X$ would be very small in the time window of 0.1 seconds. Then the derivative of $\sin \theta_k^X$ is

$$\lim_{\Delta\theta_k^X \rightarrow 0} \frac{\sin(\theta_1^X + \Delta\theta_k^X) - \sin \theta_1^X}{\Delta\theta_k^X} = \cos \theta_1^X. \quad (9)$$

Then Eq. 8 can be approximately rewritten as

$$\begin{aligned} d_{i,k}^X &\approx d_{1,1}^X + c \frac{f_1^D}{f_c} (k-1) \Delta t + \\ &(i-1) \Delta d \left(\sin \theta_1^X + \Delta\theta_k^X \cos \theta_1^X \right) \\ &= d_{1,1}^X + (i-1) \Delta d \sin \theta_1^X + \frac{c}{f_c} \left(f_1^D + f_i^{AoA} \right) (k-1) \Delta t, \end{aligned} \quad (10)$$

where $f_i^{AoA} = (i-1) \Delta d \frac{\Delta\theta_k^X \cos \theta_1^X}{k-1} \frac{f_c}{c} f_s$ is called the *AoA frequency shift (AFS)* in this work. Here we note that the DFS at the i -th RX antenna ($i = 2, 3$) is $f_i^D = f_1^D + f_i^{AoA}$. We can see that the DFS has a slight difference among different RX antennas. In practice, when a person moves at a relatively low speed, the value of f_i^{AoA} is very small and thereby we can ignore its impact on DFS in our model. In Section 9.4, we will conduct an experiment to verify the impact of the AFS on the DFS estimation accuracy. In the rest of this work, we denote $f^D = f_1^D \approx f_2^D \approx f_3^D$. And we use θ^X and d^X to denote the AoA and human reflection distance for simplicity.

Therefore, the dynamic human component $H_{i,j,k}^X$ can be revised based on the three key signal parameters (DFS f^D , AoA θ^X and human reflection distance d^X), and we have

$$H_{i,j,k}^X = \rho_{i,j}^X e^{-j2\pi \frac{f_j}{c} \left[d^X + c \frac{f_j^D}{f_c} (k-1) \Delta t + (i-1) \Delta d \sin \theta^X \right]} + \mathbb{N}_{i,j,k}^X. \quad (11)$$

2.3 Challenges

Our CSI model reveals that, to be able to track a single moving object, we need to estimate $\{f^D, \theta^X, d^X\}$ in the presence of multiple static multipath signals, the time-varying phase shifts $\varphi_{j,k}^{TO}$ and $\psi_{j,k}^{FO}$ in $H_{j,k}^e$ and the WiFi hardware-dependent phase shift φ_i . Specifically, the time-varying phase shifts are major hurdles for jointly exploiting CSIs across time in tracking, and they will also cause ambiguity in parameter estimation if not being removed. The large number of multipath will cause inefficiency in parameter estimation, particularly in AoA estimation given the limited number of Rx antennas. These issues cause that conventional localization and tracking algorithms cannot be directly applied. To achieve fine-grained passive tracking, we firstly need to minimize the impact of transceiver asynchrony, which is a critical and challenging problem in WiFi tracking; we can then separate the dynamic human component $H_{i,j,k}^X$ from the CSI measurements and then estimate, ideally, only the parameters associated with the single dynamic path reflected from the object. In the following, we aim to achieve this goal by proposing a novel scheme.

3 SYSTEM OVERVIEW

This work presents a scheme WiDFS that enables single-target real-time passive tracking using COTS WiFi devices. In WiDFS, a COTS Intel 5300 WiFi NIC that supports up to three antennas is used in a transmitter and receiver for CSI collection. The transmitter has one TX antenna while the receiver connects to a linear array consisting of three RX antennas. Such a three-antenna deployment is necessary due to the design of our algorithm in addressing the impact of transceiver asynchronization, as well as for estimating AoA. WiDFS firstly applies CSI cross-correlation between each pair of RX antennas to remove the time-varying phase shifts

in CSIs. Then WiDFS obtains an unambiguous DFS estimate from the calculated cross-correlation terms in a short-time window. After that, WiDFS adopts CSI self-correlation of each RX antenna to acquire each antenna's CFR power and then separates the dynamic human component using a simple yet effective DFS-based separation algorithm. When WiDFS detects the presence of a moving person, it enables to estimate the AoA of the tracked person relative to the RX array and the human reflection distance from the person to the TX and RX antennas. Finally, WiDFS combines the estimated signal parameters to achieve real-time tracking. The workflow of WiDFS is illustrated in Fig. 2. The processing in the main modules is summarized below and will be detailed later.

1) Doppler Frequency Shift Estimation. WiDFS firstly adopts cross-correlation between CSIs of any two antennas to remove time-varying phase shifts. The corresponding cross-correlation terms are CSI_{12} , CSI_{23} , and CSI_{31} , respectively. In a short-time window containing 100 CSI samples (about 0.1 s due to the sampling frequency of 1 kHz), WiDFS cleans high-frequency noise via Savitzky-Golay and lowpass filters. The DFS estimator outputs an unambiguous DFS estimate based on these filtered CSI cross-correlation terms. This part is described in detail in Section 4.

2) Dynamic Human Component Separation. In this part, WiDFS adopts self-correlation to calculate the CFR power of each antenna and obtain CSI_{11} , CSI_{22} and CSI_{33} , respectively. After cleaning the CFR power terms via a low-pass/bandpass filter, WiDFS combines the refined dynamic CFR powers with the estimated DFS to separate the dynamic human component. The details will be provided in Section 5.

3) Moving Person Detection and Tracking. Since a human body does not act as a perfect reflector, the RX array may just capture a few signal reflections. In this part, WiDFS combines the estimated dynamic human components over multiple CSI sampling windows. A DFS-based motion detector is designed to determine the absence and presence of a moving person. When the human movement is present, WiDFS separately performs AoA and distance estimation and then uses a Kalman filter to refine each parameter. Finally, the localizer uses optimized parameters to locate the person being tracked. Section 5 will present the details.

4) WiFi Hardware Diversity Calibration. In addition, we propose a one-time WiFi hardware calibration algorithm in Section 6 to calibrate and compensate for the difference in hardware-related phase shifts between RX antennas. At the same time, the algorithm can also calibrate the actual spacing between two adjacent RX antennas, which may be a little different from our manually measured value in our

4 ESTIMATING DOPPLER FREQUENCY SHIFT

To achieve passive human tracking, three key signal parameters, i.e., DFS f^D , AoA θ^X , and human reflection distance d^X , will be estimated based on our CSI model. This section mainly introduces how to apply CSI cross-correlation to estimate the DFS f^D in each sampling window.

4.1 Random phase shift removal via cross-correlation

Recall that $CSI_{i,j,k}$ is the reported CSI by a COTS WiFi system at i -th antenna, j -th subcarrier and k -th CSI sampling time. The time-varying phase shifts $\varphi_{j,k}^{FO}$ and $\varphi_{j,k}^{TO}$ in

CSI across packets are unknown. To remove them, WiDFS adopts CSI cross-correlation between RX antennas by multiplying a CSI for a RX antenna (e.g., $CSI_{1,j,k}$) by the conjugate of a CSI for another antenna (e.g., $CSI_{2,j,k}$) at the same subcarrier. Different to previous works [22], [24], we do not apply the approach of adding and subtracting constants. Although this approach aims to suppress the imaging components in the cross-correlation output, it is not always effective and can introduce more interfering terms.

Here let us take Antenna 1 and Antenna 2 for example. The CSI cross-correlation between them is

$$\begin{aligned} & CSI_{1,j,k} \overline{CSI}_{2,j,k} \\ &= \left(H_{j,k}^e H_1^h \right) \left(H_{1,j}^S + H_{1,j,k}^X \right) \left(\overline{H}_{j,k}^e \overline{H}_{i=2}^h \right) \left(\overline{H}_{2,j}^S + \overline{H}_{2,j,k}^X \right) \\ &= \underbrace{\|H_{j,k}^e\|^2 H_1^h \overline{H}_2^h H_{1,j}^S \overline{H}_{2,j}^S}_{\text{Static Cross-correlation Term}} + \\ & \quad \underbrace{\|H_{j,k}^e\|^2 H_1^h \overline{H}_2^h \left(H_{1,j}^S \overline{H}_{2,j,k}^X + H_{1,j,k}^X \overline{H}_{2,j}^S + H_{1,j,k}^X \overline{H}_{2,j,k}^X \right)}_{\text{Dynamic Cross-correlation Term}}, \end{aligned} \quad (12)$$

where the function $\|\cdot\|$ denotes the operator of computing the amplitude and $\|H_{j,k}^e\| = \rho^{agc}$. Likewise, we also obtain the cross-correlation terms, i.e., $CSI_{2,j,k} \overline{CSI}_{3,j,k}$ and $CSI_{3,j,k} \overline{CSI}_{1,j,k}$, from other RX antenna pairs.

After the cross-correlation operation, we can see that the phase shifts $\varphi_{j,k}^{FO}$ and $\varphi_{j,k}^{TO}$ in $H_{j,k}^e$ are eliminated. Recall that a short-time sampling window of about 0.1s is defined. WiDFS collects 3 RX antennas \times 30 subcarriers \times N_p CSI samples in this window, where $N_p = 100$. Then all CSI cross-correlation terms between each pair of RX antennas free from the impact of time-varying phase shifts are

$$\begin{cases} CSI_{12} = \{ CSI_{1,j,1} \overline{CSI}_{2,j,1}, \dots, CSI_{1,j,N_p} \overline{CSI}_{2,j,N_p} \} \\ CSI_{23} = \{ CSI_{2,j,1} \overline{CSI}_{3,j,1}, \dots, CSI_{2,j,N_p} \overline{CSI}_{3,j,N_p} \} \\ CSI_{31} = \{ CSI_{3,j,1} \overline{CSI}_{1,j,1}, \dots, CSI_{3,j,N_p} \overline{CSI}_{1,j,N_p} \} \end{cases} \quad (13)$$

4.2 High-frequency noise removal

We then input the CSI cross-correlation terms CSI_{12} , CSI_{23} and CSI_{31} into a Savitzky-Golay (SG) smoothing filter followed by a lowpass filter, which aims to remove high-frequency noise from these terms.

According to [24], the high-frequency noise comes from WiFi NICs and varies much faster than the dynamic human component of interest. In the SG filter, the polynomial order and frame length are set to 3 and 5, respectively. And we assume the maximum motion speed is 3.5 m/s in indoor scenarios, so the passband of the lowpass filter is set to $f_{pass} = \frac{3.5f_c}{c} \approx 60$ Hz, where the center carrier frequency is $f_c = 5.32$ GHz in our WiFi system. Here let CSI'_{12} , CSI'_{23} and CSI'_{31} be the filtered cross-correlation terms.

4.3 Doppler frequency shift estimation

In the following, we use the filtered cross-correlation terms to estimate a DFS f^D in a short-time sampling window.

Static and Dynamic Cross-correlation Term Separation. The static component $H_{i,j}^S$ is a constant in the window and its power is much higher than that of the dynamic

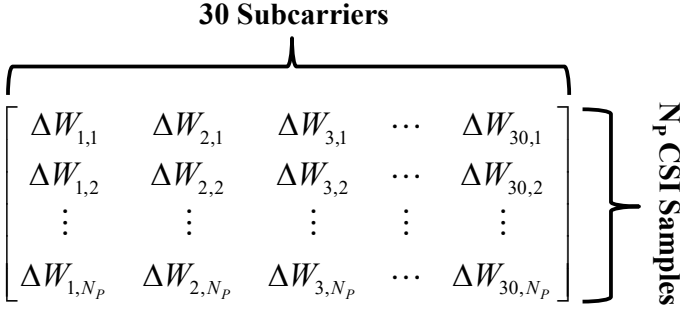


Fig. 3: Observation matrix for unambiguous DFS estimation

human component $H_{i,j,k}^X$, so the static cross-correlation term (denoted as $U_{12,j}$) can be separated by calculating the mean value of the cross-correlation terms. The remaining dynamic cross-correlation term (denoted as $V_{12,j,k}$) at each CSI sample is obtained by subtracting the mean value,

$$\begin{cases} \|H_{j,k}^e\|^2 H_1^h \bar{H}_2^h H_{1,j}^S \bar{H}_{2,j}^S = U_{12,j} \\ \|H_{j,k}^e\|^2 H_1^h \bar{H}_2^h (H_{1,j}^S \bar{H}_{2,j,k}^X + H_{1,j,k}^X \bar{H}_{2,j}^S + H_{1,j,k}^X \bar{H}_{2,j,k}^X) = V_{12,j,k} \end{cases}, \quad (14)$$

where

$$\begin{cases} U_{12,j} = \frac{1}{N_p} \sum_{k=1}^{N_p} CSI'_{1,j,k} \overline{CSI'}_{2,j,k} \\ V_{12,j,k} = CSI'_{1,j,k} \overline{CSI'}_{2,j,k} - U_{12,j} \end{cases}. \quad (15)$$

Since the amplitude of $H_{i,j}^S$ may be far larger than that of $H_{i,j,k}^X$, we ignore the product term $H_{1,j,k}^X \bar{H}_{2,j,k}^X$ in Eq. 14.

DFS Estimation using Subspace-based Method. From Eq. 14, we can see that the dynamic cross-correlation term of interest (e.g., $H_{1,j,k}^X \bar{H}_{2,j}^S$) is mixed with a side product (e.g., $\bar{H}_{2,j,k}^X H_{1,j}^S$), so the estimated DFS f^D would be an actual one or its negative. This phenomenon is called the *DFS ambiguity*. To tackle this issue, many existing works [19], [25], [26] adjust the CSI amplitude of each antenna by adding or subtracting a real value. However, this intuitive method seems to be signal-dependent and is not always effective in multipath-rich scenarios. In this work, WiDFS achieves unambiguous DFS estimation as follows:

First, we divide the equations in Eq. 14 to remove the unknown term $\|H_{j,k}^e\|^2 H_{1,j}^h \bar{H}_{2,j}^h$,

$$\frac{H_{1,j,k}^X}{H_{1,j}^S} + \frac{\bar{H}_{2,j,k}^X}{\bar{H}_{2,j}^S} = \frac{V_{12,j,k}}{U_{12,j}}. \quad (16)$$

Likewise, we can obtain

$$\frac{H_{2,j,k}^X}{H_{2,j}^S} + \frac{\bar{H}_{3,j,k}^X}{\bar{H}_{3,j}^S} = \frac{V_{23,j,k}}{U_{23,j}}, \quad (17a)$$

$$\frac{H_{3,j,k}^X}{H_{3,j}^S} + \frac{\bar{H}_{1,j,k}^X}{\bar{H}_{1,j}^S} = \frac{V_{31,j,k}}{U_{31,j}}. \quad (17b)$$

By subtracting Eq. 17a from the conjugate of Eq. 17b and subtracting Eq. 17b from the conjugate of Eq. 16, we have

$$\begin{cases} \frac{H_{1,j,k}^X}{H_{1,j}^S} - \frac{H_{2,j,k}^X}{H_{2,j}^S} = \frac{\bar{V}_{31,j,k}}{\bar{U}_{31,j}} - \frac{V_{23,j,k}}{U_{23,j}} = \Delta W_{12,j,k} \\ \frac{H_{2,j,k}^X}{H_{2,j}^S} - \frac{H_{3,j,k}^X}{H_{3,j}^S} = \frac{\bar{V}_{12,j,k}}{\bar{U}_{12,j}} - \frac{V_{31,j,k}}{U_{31,j}} = \Delta W_{23,j,k} \end{cases}. \quad (18)$$

Then we compute the difference of the above equations,

$$\begin{aligned} \Delta W_{j,k} &= \Delta W_{12,j,k} - \Delta W_{23,j,k} \\ &= \frac{H_{1,j,k}^X}{H_{1,j}^S} - 2 \frac{H_{2,j,k}^X}{H_{2,j}^S} + \frac{H_{3,j,k}^X}{H_{3,j}^S} \\ &= \frac{H_{2,j,k}^X}{H_{2,j}^S} \left(\frac{H_{1,j,k}^X}{H_{2,j,k}^X} \frac{H_{2,j}^S}{H_{1,j}^S} + \frac{H_{3,j,k}^X}{H_{2,j,k}^X} \frac{H_{2,j}^S}{H_{3,j}^S} - 2 \right) \end{aligned}. \quad (19)$$

In free space without multipath interference, we can apply far-field and narrowband assumptions to obtain

$$\begin{cases} \|H_{1,j,k}^X\| \approx \|H_{2,j,k}^X\| \approx \|H_{3,j,k}^X\| \\ \|H_{1,j}^S\| \approx \|H_{2,j}^S\| \approx \|H_{3,j}^S\| \\ \angle \frac{H_{1,j,k}^X}{H_{2,j,k}^X} \approx \angle \frac{H_{2,j,k}^X}{H_{3,j,k}^X} \\ \angle \frac{H_{1,j}^S}{H_{2,j}^S} \approx \angle \frac{H_{2,j}^S}{H_{3,j}^S} \end{cases}, \quad (20)$$

where the function \angle denotes the operator of computing the phase shift of a complex number. Thus, we have

$$\Gamma_{j,k} = \frac{H_{1,j,k}^X}{H_{2,j,k}^X} \frac{H_{2,j}^S}{H_{1,j}^S} + \frac{H_{3,j,k}^X}{H_{2,j,k}^X} \frac{H_{2,j}^S}{H_{3,j}^S} \approx 2 \cos \left(\angle \frac{H_{1,j,k}^X}{H_{2,j,k}^X} \frac{H_{2,j}^S}{H_{1,j}^S} \right) \quad (21)$$

Unfortunately, there exist multiple signal propagation paths in an actual environment. However, for the following two reasons, we can assume that $\Gamma_{j,k}$ keeps unchanged, i.e., $\Gamma_{j,k} = \Gamma_j$. (1) $H_{i,j}^S$ is static environment-related and remain unchanged, then $H_{1,j}^S/H_{2,j}^S$ and $H_{2,j}^S/H_{3,j}^S$ are approximately constant. (2) Recall that we ignore the change in the amplitude of $H_{i,j,k}^X$ in the window. The values of $H_{1,j,k}^X/H_{2,j,k}^X$ and $H_{2,j,k}^X/H_{3,j,k}^X$ also do not vary over time. Thus, $\Delta W_{j,k}$ can be rewritten as

$$\Delta W_{j,k} \approx \frac{\Gamma_j - 2}{H_{2,j}^S} [\rho_{2,j}^X e^{-J[2\pi \frac{f_j}{c} d_2^X + 2\pi f^D (k-1)\Delta t]} + \mathbb{N}_{2,j,k}^X]. \quad (22)$$

The Fast Fourier Transform (FFT) is commonly used to identify signal frequencies. However, given the small sample size (i.e., 100 CSI samples), FFT may not provide enough high estimation accuracy. Instead, WiDFS uses root multiple signal classification (Root-MUSIC) algorithm [30] to estimate f^D . It outputs frequency estimates along with the corresponding signal power estimates. As shown in Fig. 3, we build an observation matrix using $\Delta W_{j,k}$ in Eq. 22 at all subcarriers. Each column represents a separate observation. The number of snapshots is 30. To determine the subspace dimension, we calculate the maximum eigenvalue eig_{max} of the correlation matrix and then find the number of eigenvalues above an empirical threshold $0.6 eig_{max}$. Finally, we select a frequency estimate associate with the maximum signal power as an optimal one. Fig.4 shows the measured

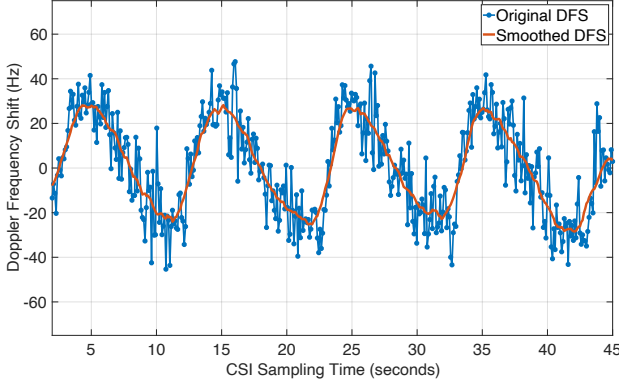


Fig. 4: Measured DFS when a person walks four times along an approximate elliptical trajectory

DFSs when a person walks four times along an elliptical trajectory. The results are consistent with human movement, and the smoothed results clearly reveal the DFS variation.

5 SEPARATING DYNAMIC HUMAN COMPONENT

This section uses the estimated DFS and CSI self-correlation to build the complex-valued dynamic human components.

5.1 Random phase shift removal via self-correlation

First of all, the time-varying phase shifts $\varphi_{j,k}^{FO}$ and $\varphi_{j,k}^{TO}$ in $H_{j,k}^e$ are removed by applying a self-correlation operation in WiDFS, that is, we multiply $CSI_{i,j,k}$ by its complex conjugate $\overline{CSI}_{i,j,k}$ and obtain $\|CSI_{i,j,k}\|^2$. The product is also called the CFR power, which is

$$\begin{aligned} \|CSI_{i,j,k}\|^2 &= \left(H_{j,k}^e H_i^h\right) \left(H_{i,j}^S + H_{i,j,k}^X\right) \left(\overline{H}_{j,k}^e \overline{H}_i^h\right) \left(\overline{H}_{i,j}^S + \overline{H}_{i,j,k}^X\right) \\ &= \underbrace{\|H_{j,k}^e H_i^h H_{i,j}^S\|^2}_{\text{Static Power Term}} + \\ &\quad \underbrace{\|H_{j,k}^e H_i^h\|^2 \left[2 \|H_{i,j}^S H_{i,j,k}^X\| \cos \angle \left(\overline{H}_{i,j}^S H_{i,j,k}^X\right) + \|H_{i,j,k}^X\|^2\right]}_{\text{Dynamic Power Term}}, \end{aligned} \quad (23)$$

In the pre-defined sampling window with 100 CSI samples, the calculated self-correlation terms at each RX antenna ($i = 1, 2, 3$) and subcarrier ($j = 1, 2, \dots, 30$) are denoted as $CSI_{ii} = \{\|CSI_{i,j,1}\|^2, \dots, \|CSI_{i,j,N_p}\|^2\}$.

5.2 Dynamic power separation and refinement

Next, WiDFS uses the calculated CSI_{11} , CSI_{22} and CSI_{33} to separate dynamic power terms in the window and then refines them using the estimated DFS f^D .

Separation of Static and Dynamic Power Terms. We follow the same operation in Eq. 15 to calculate the static and dynamic power terms (denoted as $u_{i,j}$ and $v_{i,j,k}$, respectively)

$$\begin{cases} \|H_{j,k}^e H_i^h H_{i,j}^S\|^2 = u_{i,j} \\ \|H_{j,k}^e H_i^h\|^2 \left[2 \|H_{i,j}^S H_{i,j,k}^X\| \cos \angle \left(\overline{H}_{i,j}^S H_{i,j,k}^X\right) + \|H_{i,j,k}^X\|^2\right] \approx v_{i,j,k} \end{cases}, \quad (24)$$

where

$$\begin{cases} u_{i,j} = \frac{1}{N_p} \sum_{k=1}^{N_p} \|CSI_{i,j,k}\|^2 \\ v_{i,j,k} = \|CSI_{i,j,k}\|^2 - u_{i,j} \end{cases}. \quad (25)$$

The term $\|H_{i,j,k}^X\|^2$ can be ignored since it is much smaller than other terms.

Filtering of Dynamic Power Terms of Interest. Since the dynamic power terms $v_{i,j,k}$ of interest may be around the estimated DFS f^D , we refine $v_{i,j,k}$ using a lowpass filter with the passband frequency of $|f^D| + \Delta f$, where the empirical frequency Δf is set to 10 Hz in our scheme. And if $|f^D| > 15$ Hz, we further apply a highpass filter with the passband frequency of $|f^D| - \Delta f$. The refined dynamic power terms at each RX antenna and subcarrier in the window are denoted as $v'_{ii} = \{v'_{i,j,1}, \dots, v'_{i,j,N_p}\}$. For single-object tracking, it is reasonable to assume that there is only a dominant human reflection in each $v'_{i,j,k}$.

5.3 Dynamic human component reconstruction

Here WiDFS depends on the filtered v'_{ii} and the estimated DFS f^D to reconstruct the complex-valued dynamic human component at each RX antenna and subcarrier.

Static-Dynamic Associated Component Estimation. At first, we remove the unknown term $\|H_{j,k}^e H_i^h\|$ by

$$\frac{v'_{i,j,k}}{u_{i,j}} = 2 \left\| \frac{H_{i,j,k}^X}{H_{i,j}^S} \right\| \cos \angle \left(\overline{H}_{i,j}^S H_{i,j,k}^X \right). \quad (26)$$

We substitute the proposed dynamic human component $H_{i,j,k}^X$ described in Section 2.2 into the above equation (the minor noise term is removed after filtering in Section 5.2),

$$\begin{aligned} \frac{v'_{i,j,k}}{u_{i,j}} &= 2 \left\| \frac{H_{i,j,k}^X}{H_{i,j}^S} \right\| \cos \left(\angle \overline{H}_{i,j}^S + \angle H_{i,j,k}^X \right) \\ &= 2 \left\| \frac{H_{i,j,k}^X}{H_{i,j}^S} \right\| \times \\ &\quad \cos \left\{ \angle \overline{H}_{i,j}^S - 2\pi \frac{f_j}{c} \left[d^X + c \frac{f^D}{f_c} (k-1) \Delta t + (i-1) \Delta d \sin \theta^X \right] \right\} \\ &= x_{i,j} \cos \left[2\pi f^D (k-1) \Delta t \right] + y_{i,j} \sin \left[2\pi f^D (k-1) \Delta t \right], \end{aligned} \quad (27)$$

where

$$\begin{cases} x_{i,j} = 2 \left\| \frac{H_{i,j,k}^X}{H_{i,j}^S} \right\| \cos \left\{ \angle \overline{H}_{i,j}^S - 2\pi \frac{f_j}{c} \left[d^X + (i-1) \Delta d \sin \theta^X \right] \right\} \\ y_{i,j} = 2 \left\| \frac{H_{i,j,k}^X}{H_{i,j}^S} \right\| \sin \left\{ \angle \overline{H}_{i,j}^S - 2\pi \frac{f_j}{c} \left[d^X + (i-1) \Delta d \sin \theta^X \right] \right\} \end{cases}. \quad (28)$$

Specially, since the subcarrier frequency f_j is very close to the center frequency f_c , we have $f_j/f_c \approx 1$.

Given $N_p = 100$ CSI samples at a RX antenna and subcarrier in the window, we write Eq. 27 in a matrix form,

$$\begin{bmatrix} 1 & 0 \\ \cos(2\pi f^D \Delta t) & \sin(2\pi f^D \Delta t) \\ \vdots & \vdots \\ \cos[2\pi f^D (N_p - 1) \Delta t] & \sin[2\pi f^D (N_p - 1) \Delta t] \end{bmatrix} \begin{bmatrix} x_{i,j} \\ y_{i,j} \end{bmatrix} = \begin{bmatrix} v'_{i,j,1} \\ u_{i,j} \\ v'_{i,j,2} \\ u_{i,j} \\ \vdots \\ v'_{i,j,N_p} \\ u_{i,j} \end{bmatrix}. \quad (29)$$

A least-squares method is applied to easily solve $x_{i,j}$ and $y_{i,j}$, so the static-dynamic associated component $Z_{i,j}^{S,X}$ at the i -th antenna and the j -th subcarrier in the window is

$$\begin{aligned} Z_{i,j}^{S,X} &= w_{i,j} e^{J \text{atan2}(y_{i,j}, x_{i,j})} \\ &= w_{i,j} e^{J \left\{ \angle \overline{H}_{i,j}^S - 2\pi \frac{f_j}{c} \left[d^X + (i-1) \Delta d \sin \theta^X \right] \right\}}, \end{aligned} \quad (30)$$

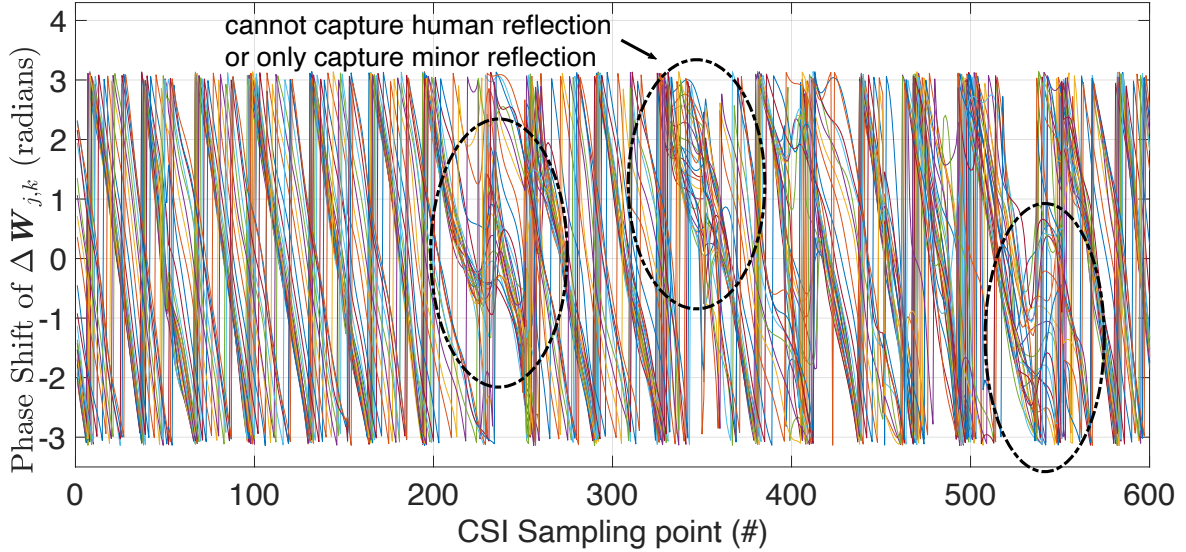


Fig. 5: Signal reflections off human body. Recall from Eq. 19, $\Delta W_{j,k}$ represents the dynamic component caused by human motion. Since human body has a complex surface, the transmit signals may not always be reflected to RX antennas.

where $w_{i,j}$ denotes the weight derived from the sum of squared residuals. When the residual is minimum, we normalize the weight to be maximum (or vice versa).

Dynamic Human Component Separation. Further, let $\Delta\varphi_{12}^h$ and $\Delta\varphi_{31}^h$ be the phase differences caused by WiFi hardware diversity, which can be pre-estimated using our proposed approach to be detailed in Section 7,

$$\begin{cases} \Delta\varphi_{12}^h = \angle(H_1^h \bar{H}_2^h) \\ \Delta\varphi_{23}^h = \angle(H_2^h \bar{H}_3^h) \\ \Delta\varphi_{31}^h = \angle(H_3^h \bar{H}_1^h) \end{cases} \quad (31)$$

We combine them with the calculated static cross-correlation terms $U_{12,j}$ and $U_{31,j}$ in Section 4 to conduct a transformation on $Z_{2,j}^{S,X}$ and $Z_{3,j}^{S,X}$,

$$\begin{cases} w_{2,j} e^{\mathcal{J}[\angle \bar{H}_{1,j}^S - 2\pi \frac{f_j}{c} (d^X + \Delta d \sin \theta^X)]} = Z_{2,j}^{S,X} e^{-\mathcal{J}(\angle U_{12,j} - \Delta\varphi_{12}^h)} \\ w_{3,j} e^{\mathcal{J}[\angle \bar{H}_{1,j}^S - 2\pi \frac{f_j}{c} (d^X + 2\Delta d \sin \theta^X)]} = Z_{3,j}^{S,X} e^{\mathcal{J}(\angle U_{31,j} - \Delta\varphi_{31}^h)} \end{cases}, \quad (32)$$

where

$$\begin{cases} \angle U_{12,j} - \Delta\varphi_{12}^h = -(\angle H_{2,j}^S + \angle \bar{H}_{1,j}^S) \\ \angle U_{31,j} - \Delta\varphi_{31}^h = \angle H_{3,j}^S + \angle \bar{H}_{1,j}^S \end{cases}. \quad (33)$$

Thus, the dynamic human components $Z_{1,j}^X$, $Z_{2,j}^X$, and $Z_{3,j}^X$ at the j -th subcarrier in the window are reconstructed as

$$\begin{cases} Z_{1,j}^X = w_{1,j} e^{-\mathcal{J} 2\pi \frac{f_j}{c} d^X} = e^{\mathcal{J} \angle H_{1,j}^S} Z_{1,j}^{S,X} \\ Z_{2,j}^X = w_{2,j} e^{-\mathcal{J} 2\pi \frac{f_j}{c} (d^X + \Delta d \sin \theta^X)} = e^{\mathcal{J} \angle H_{1,j}^S} Z_{2,j}^{S,X} e^{-\mathcal{J}(\angle U_{12,j} - \Delta\varphi_{12}^h)}, \\ Z_{3,j}^X = w_{3,j} e^{-\mathcal{J} 2\pi \frac{f_j}{c} (d^X + 2\Delta d \sin \theta^X)} = e^{\mathcal{J} \angle H_{1,j}^S} Z_{3,j}^{S,X} e^{\mathcal{J}(\angle U_{31,j} - \Delta\varphi_{31}^h)} \end{cases} \quad (34)$$

where the distance $d_{1,j}^S$ between the transmitter and the receiver Antenna 1 can be measured in advance, so we can get the approximation $\angle H_{1,j}^S \approx -\text{mod}(2\pi \frac{f_j}{c} d_{1,j}^S, 2\pi)$.

6 DETECTING AND TRACKING MOVING PERSON

This section describes how to leverage the estimated DFS f^D to detect the presence of a moving person and introduces

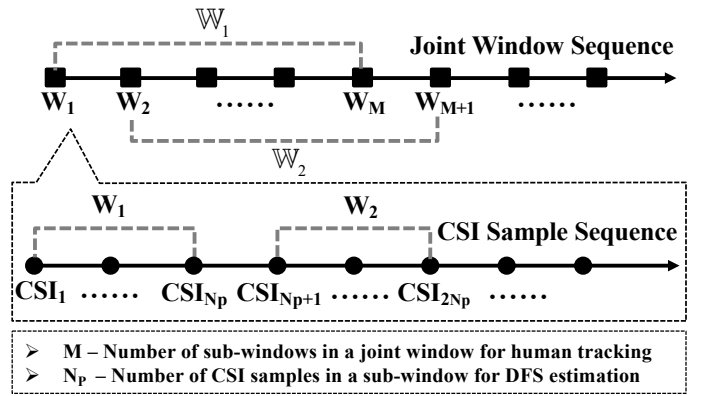


Fig. 6: Combining multiple CSI sampling sub-windows for reliable motion detection and human tracking

how to track a moving person using the calculated dynamic human components $Z_{i,j}^X$ over multiple sampling windows.

6.1 Capture of Human body reflection

When the transmitter sends WiFi signals to space, the RX antenna array may not capture all reflection signals off human since signal propagation is complicated by reflections from the human body surface. To intuitively illustrate this phenomenon, we plot the phase change of $\Delta W_{j,k}$ in Eq. 19 at 30 subcarriers in Fig. 5. We collect 6 seconds of CSI data in a dynamic scenario that a person moves at a speed of about 1 m/s away from the receiver. It shows the phase variation cannot match our expectation in some sampling time due to the fact that some reflections may be invisible to the receiver.

To achieve robust passive human tracking, we combine the human reflections over multiple CSI sampling sub-windows \mathbb{W} (as shown in Fig. 6), where each sub-window has $N_P = 100$ CSI samples as described before. And the window \mathbb{W} consisting of the amount of M sub-windows is called the *joint window*. The two adjacent joint windows overlap with $(M - 1)$ sub-windows. The average computation time should be less than 0.1 s to enable real-time tracking.

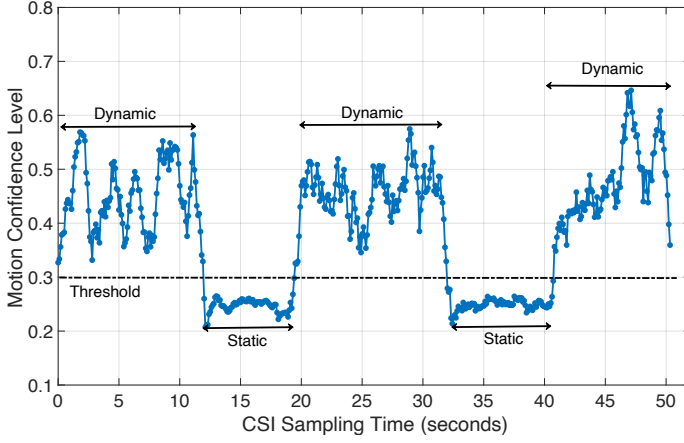


Fig. 7: Detecting the absence or presence of human motion

6.2 Dynamic environment sensing

WiDFS performs passive human tracking only when a moving person is present. It uses the estimated f^D and $\Delta W_{j,k}$ in Eq. 19 in Section 4 to determine the absence or presence of a moving person in the region of interest by

$$\mathbb{P} = \frac{1}{30 \times M N_p} \sum_{l=1}^M \sum_{j=1}^{30} \left\| \sum_{k=1}^{N_p} e^{j \angle \Delta W_{j,k}[l]} e^{j 2\pi f^D[l](k-1)\Delta t} \right\|, \quad (35)$$

where $\Delta W_{j,k}[l]$ and $f^D[l]$ represent the corresponding estimates in the l -th sub-window. The range of \mathbb{P} is within (0, 1). In a static scenario, the extracted $\Delta W_{j,k}$ is a noise term, so \mathbb{P} maintains in a low level. However, when a person moves in the region, \mathbb{P} will be close to 1. Our judging criteria is that if \mathbb{P} is lower (or higher) than a pre-defined threshold, it indicates the absence (or presence) of a moving person.

In Fig. 7, we use an example to show the result of dynamic environment sensing. A person moves randomly and then sits on a chair for a while. The volunteer repeats the two motions. We find that the sensing function could effectively indicate the human motion status via a threshold. In Section 9.4, a comprehensive experiment will be conducted to determine this threshold in different scenarios.

6.3 Localization parameter estimation

WiDFS performs tracking using two fundamental parameters. One is the AoA θ^X , the direction that the human reflections are received from at the RX antenna array. Another is the reflection distance d^X of a transmitted signal being reflected by the tracked person to an RX antenna. To guarantee our system's real-time performance, we separately calculate the two parameters.

6.3.1 AoA

Based on the calculated dynamic human component $Z_{i,j}^X$, the AoA θ^X in a joint window is estimated by

$$\operatorname{argmax}_{\theta^X \in [-90^\circ, 90^\circ]} \sum_{l=1}^M \sum_{j=1}^{30} \left\| \sum_{i=1}^3 Z_{i,j}^X[l] e^{j 2\pi \frac{f_j}{c} (i-1) \Delta d \sin \theta^X} \right\|. \quad (36)$$

Due to the 2π phase ambiguity, the searching range of AoA should be within $[-90^\circ, 90^\circ]$. In this way, the tracked person is required to move on one side of the receive antenna array. The angular searching spacing is set as 1° in WiDFS.

6.3.2 Human reflection distance

The human reflection distance d^X in a joint window is calculated by

$$\operatorname{argmax}_{d^X \in (d_{min}^X, d_{max}^X]} \sum_{l=1}^M \sum_{i=1}^3 \left\| \sum_{j=1}^{30} Z_{i,j}^X[l] Z^D[l] e^{j 2\pi \frac{f_j}{c} d^X} \right\|, \quad (37)$$

where

$$\begin{cases} Z^D[l] = e^{-j 2\pi \sum_{\ell=l}^{\epsilon-1} f^D[\ell] N \Delta t}, & l < \epsilon \\ Z^D[l] = 1, & l = \epsilon \\ Z^D[l] = e^{j 2\pi \sum_{\ell=\epsilon}^{l-1} f^D[\ell] N \Delta t}, & l > \epsilon \end{cases}, \quad (38)$$

Here $\epsilon = \frac{M+1}{2}$ and M is an odd number. The minimum distance d_{min}^X is equal to the length d^S of the direct path between the TX and RX antennas since the human reflection distance is larger than d^S . The maximum distance d_{max}^X is 20 m, which is sufficiently large for indoor scenarios. The distance searching spacing is set as 5 cm in WiDFS.

6.4 Localization parameter refinement

Since the estimated θ^X and d^X may be relatively noisy, we refine them using Kalman smoothers.

Let $\theta^X[L]$ and $d^X[L]$ be the current AoA and human reflection distance in the L -th joint window after outlier rejection. The corresponding Kalman model is built as follows,

$$\begin{cases} \sin(\theta^X[L+1]) \approx \sin(\theta^X[L]) \\ d^X[L+1] = d^X[L] + c \frac{f^D[L] N_p}{f_c f^s} \end{cases}, \quad (39)$$

where $f^D[L]$ is the median of $\{f^D[1], f^D[2], \dots, f^D[M]\}$ in the L -th joint window. In the AoA Kalman smoother, the noise covariance of the sine of AoA is set to 0.2. In the distance Kalman smoother, the noise covariance of the reflection distance and DFS are set as 20 cm and 5 Hz, respectively. Fig. 8 and Fig. 9 show the estimated AoA and human reflection distance without and with Kalman smoothing.

6.5 Human localization

So far, the two core localization parameters, i.e., human reflection distance d^X and AoA θ^X , have been calculated. The person position in a 2D coordinate system can then be estimated using these estimated parameters.

As shown in Fig. 10, let $d^{X \rightarrow TX}$ and $d^{X \rightarrow RX}$ be the distance of the tracked person to the transmitter and receiver, respectively, where $d^X = d^{X \rightarrow TX} + d^{X \rightarrow RX}$. And let θ^S be the AoA of the transmitter relative to the RX antenna array, i.e., $\Delta d \sin \theta^S \approx d_2^S - d_1^S \approx d_3^S - d_2^S$. Since the phase differences caused by WiFi hardware diversity are pre-estimated, the AoA θ^S is obtained by

$$\operatorname{argmax}_{\theta^S \in [-90^\circ, 90^\circ]} \sum_{l=1}^M \sum_{j=1}^{30} \left\| (Z_{12}^S[l] + Z_{23}^S[l]) e^{-j \Delta d \sin \theta^S} \right\|, \quad (40)$$

where

$$\begin{cases} Z_{12}^S[l] = e^{j(\angle U_{12,j}[l] - \Delta \varphi_{12}^h)} \\ Z_{23}^S[l] = e^{j(\angle U_{23,j}[l] - \Delta \varphi_{23}^h)} \end{cases}. \quad (41)$$

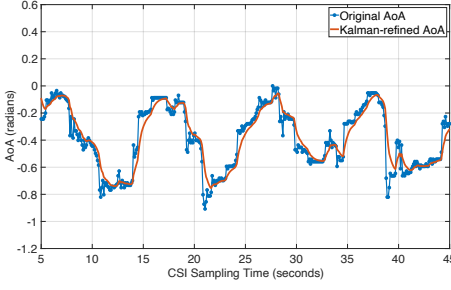


Fig. 8: Human AoA relative to receiver

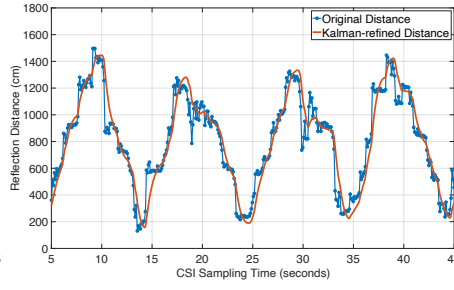


Fig. 9: Human reflection distance

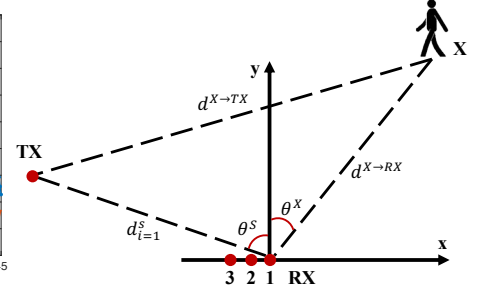
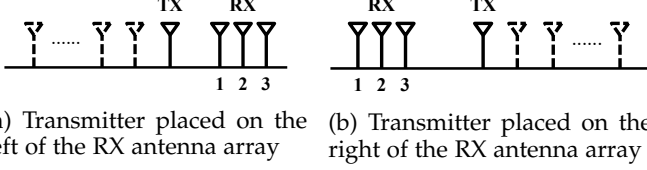


Fig. 10: Human localization



(a) Transmitter placed on the left of the RX antenna array (b) Transmitter placed on the right of the RX antenna array

Fig. 11: Static component measurement on the both sides of receive antennas for hardware diversity calibration

According to the cosine formula, the distance $d^{X \rightarrow RX}$ can be calculated by

$$d^{X \rightarrow RX} = \frac{(d^X)^2 - (d_1^S)^2}{2[d^X - d_1^S \cos(\theta^X - \theta^S)]}. \quad (42)$$

Thus, the person's position in a x-y coordinate system is

$$\begin{cases} x = d^{X \rightarrow RX} \sin \theta^X \\ y = d^{X \rightarrow RX} \cos \theta^X \end{cases}. \quad (43)$$

7 WiFi HARDWARE DIVERSITY CALIBRATION

In [27], a method was proposed to calibrate the phase difference between antennas caused by manufacturing imperfection. However, the calibrated results are always wrapped by an unknown π -radians ambiguity, meaning that the estimated phase difference is the actual value or the actual value plus π radians. Also, this method cannot calibrate the antenna spacing. For example, a spacing error of 0.5 cm will cause a phase deviation of about 0.557 radians, which will significantly impact the AoA estimation accuracy. Here we develop a novel WiFi hardware calibration algorithm. It is a one-time procedure after the system is set up.

Here we use RX Antenna 1 and Antenna 2 as an example. As shown in Fig. 11, we firstly place the TX antenna at the right side of the RX antenna array. They are deployed in a straight line. We measure multiple static components denoted as $\{U_{12,j}^R[1], U_{12,j}^R[2], \dots\}$. We then place the TX antenna at the left side of the array and obtain $\{U_{12,j}^L[1], U_{12,j}^L[2], \dots\}$. The data collection is conducted in a very low-multipath environment. Let Δd_{12} be the spacing between Antenna 1 and Antenna 2, then

$$\begin{cases} \angle U_{12,j}^L = \Delta \varphi_{12}^h + \frac{2\pi f_c}{c} \Delta d_{12} + 2\pi K_{12}^L \\ \angle U_{12,j}^R = \Delta \varphi_{12}^h - \frac{2\pi f_c}{c} \Delta d_{12} + 2\pi K_{12}^R \end{cases}, \quad (44)$$

where K_{12}^L and K_{12}^R are unknown integers.

Then the antenna spacing Δd_{12} is calculated by

$$\Delta d_{12} = \frac{c}{2\pi f_c} \left[\frac{1}{2} \angle \left(\sum_l e^{J(\angle U_{12,j}^L[l] - \angle U_{12,j}^R[l])} \right) + \pi K_{12} \right], \quad (45)$$

where K_{12} is an integer making Δd_{12} closest to our predefined spacing (half a wavelength of about 2.8 cm).

Once Δd_{12} is determined, we substitute it into Eq. 44 to estimate hardware-related phase difference $\Delta \varphi_{12}^h$,

$$\Delta \varphi_{12}^h = \angle \sum_l \left[e^{J(\angle U_{12,j}^L[l] - \frac{2\pi f_c}{c} \Delta d_{12})} + e^{J(\angle U_{12,j}^R[l] + \frac{2\pi f_c}{c} \Delta d_{12})} \right]. \quad (46)$$

Similarly, the distance Δd_{23} and phase difference $\Delta \varphi_{23}^h$ can also be measured based on the above process.

8 IMPLEMENTATION & EVALUATION

Implementation: (1) *Hardware.* We implement WiDFS using two computers separately equipped with an Intel 5300 NIC, shown in Fig. 12. One is served as a transmitter and has one WiFi antenna, while another is a receiver that has three external antennas to form a linear uniform antenna array. These antennas are all omnidirectional and have 2 dBi gain. (2) *Software.* The operating system of each laptop is Ubuntu 14.04 LTS with 3.16.0-30-generic Linux kernel version. They are configured in the monitor mode via *Linux 802.11n CSI Tool* [16], [31], [32]. The CSI sampling rate is 1 kHz. The center frequency is 5.32 GHz. WiDFS is programmed by *Python 3.8* and implemented on a Mini PC with Intel(R) Core CPU i5-7300U 2.6 GHz $\times 4$ and 3.8G memory. We adopt *csiread package*¹ to parse CSI data in real time.

WiFi Hardware Diversity Calibration. The manually measured antenna spacing in the RX antenna array is 2.8 cm. However, our hardware calibration algorithm outputs that the spacing between Antenna 1 and Antenna 2 is 2.618 cm while the spacing between Antenna 2 and Antenna 3 is 2.391 cm. It also shows the hardware-related phase difference between Antenna 1 and Antenna 2 is 5.956 radians while that between Antenna 2 and Antenna 3 is 1.418 radians.

Default Configuration. The transmitter and receiver are placed at the same height, and their separation distance is 235 cm. The AoA of the transmitter relative to the receiver is about -70° . A CSI sampling window contains $100 \times 30 \times 3$ samples for 3 RX antennas and 30 subcarriers, and a joint window contains 9 CSI sampling sub-windows.

Evaluation. We evaluate WiDFS in a multipath-rich office environment with various types of strong reflectors such as tables, chairs, metal lockers, computers, large-size

1. <https://github.com/citysu/csiread>

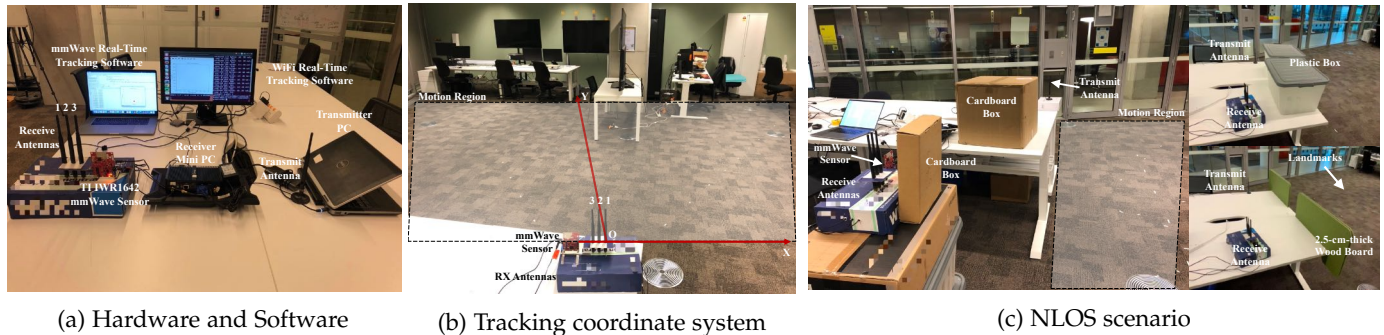


Fig. 12: WiDFS Implementation

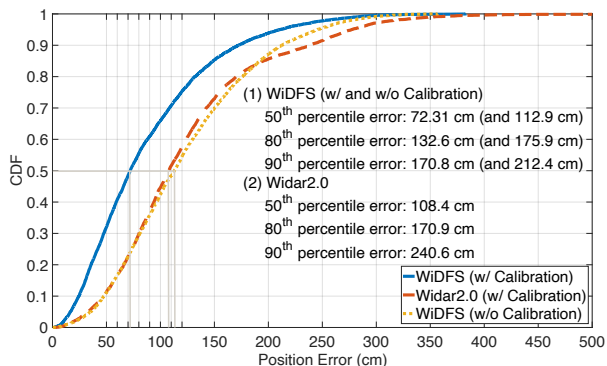


Fig. 13: CDF of tracking accuracy of WiDFS and Widar2.0

displays, concrete ceiling, and tempered glass/hollow walls. In each experiment, we ask a person to walk along with three types of trajectories (shown in Fig. 14a, Fig. 15a and Fig. 16a) in different office regions under different multipath interference. For NLOS experiments, we firstly use cardboard boxes to block RX antennas, shown in Fig. 12c (the scene on the left), so there is no LOS path between the tracked person and the transmitter to the receiver. In this scene, a TI IWR1642 mmWave radar besides the receiver can penetrate the cardboard box to capture the ground-truth positions of the tracked person. In addition, we also place other two materials, i.e., plastic boxes and wooden boards, in front of the transmitter and receiver to block the LOS path between the person and the receiver, shown in Fig. 12c (the scene on the right). However, the mmWave radar cannot penetrate the two materials to obtain the ground truth. In the two NLOS experiments, we deploy several landmarks on the floor and ask the person to move around a landmark in each experiment, where the ground truth is obtained by the mmWave radar in LOS.

Baseline. We compare WiDFS to a state-of-the-art unsupervised method Widar2.0 [25] with its implementation software downloaded from the website². The baseline has the same system setup as ours. It also works with three parameters, i.e., DFS, AoA, and human reflection distance. However, our scheme WiDFS uses significantly different key techniques compared with those used in Widar2.0. (1) Widar2.0 firstly suppresses the image component from cross-correlation terms by adding a constant factor to the CSI amplitudes of the reference antenna and subtracting another constant factor from the CSI amplitudes of other antennas; (2) Widar2.0 applies a nonlinear optimization

solution to jointly estimate multiple signal parameters from the separated dynamic human components, including DFS, AoA, ToF (Time-of-Flight corresponding to the human reflection distance in our work) and attenuation; and (3) Widar2.0 designs a path matching algorithm to optimize these signal estimates and thereby achieves human tracking using the optimized ToF and AoA. Overall, our scheme can achieve better image suppression and more accurate estimation at lower complexity, as will be seen from the experimental results next. In addition, the Widar2.0 software does not support real-time implementation, and hence its tracking results are obtained offline. For fairness, we substitute hardware calibration parameters into Widar2.0 and use its algorithm in Matlab to track a person.

Ground Truth. We use a TI IWR1642 mmWave radar to measure the person’s trajectory as ground truth. Such a radar could achieve centimeter-level tracking accuracy. It is placed beside the WiFi RX antenna array. We modify the project³ in Python to perform single-target real-time mmWave tracking as follows. We firstly obtain point cloud data and use DBSCAN clustering algorithm to eliminate some noisy points. We then adopt a Gaussian mixture to capture the density center of refined cloud data as the tracked target’s position. The localization results of WiFi and mmWave are synchronized using 1-D data interpolation based on their real-world sampling timestamps.

9 RESULTS

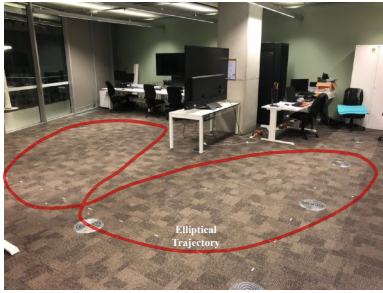
9.1 Comparison to state-of-the-art

We start by comparing WiDFS’s performance to Widar2.0. Fig. 13 shows the CDF of the tracking error of WiDFS and Widar2.0 after hardware diversity calibration. It shows that Widar2.0 achieves a median tracking error of 108.4 cm and a 90th percentile error of 240.6 cm. Comparatively, the median and 90th percentile error of WiDFS are reduced to 72.31 cm and 170.8 cm, respectively, which could achieve about 36 cm and 70 cm improvement over Widar2.0. It is further noted that Widar2.0 performs backward smoothing by using posterior localization estimates to inversely refine the previous, which could result in large processing delay and may not be used in a real-time tracking system.

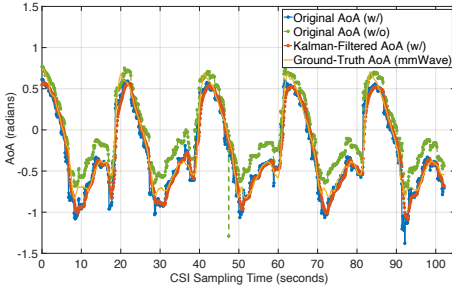
To demonstrate the significance of WiFi hardware diversity calibration, we plot the results of WiDFS without conducting calibration for comparison. In this case, a median accuracy of 112.9 cm and a 90th percentile error of 212.4 cm are

2. <http://tns.thss.tsinghua.edu.cn/wifiradar/Widar2.0Project.zip>

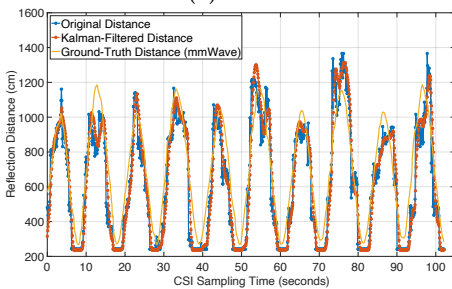
3. <https://github.com/ibaiGorordo/AWR1642-Read-Data-Python-MMWA-VE-SDK-2>



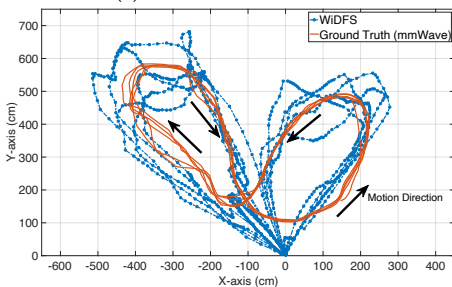
(a) Elliptical trajectory



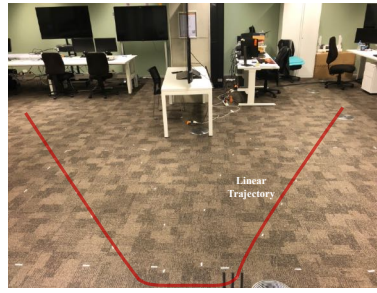
(b) AoA



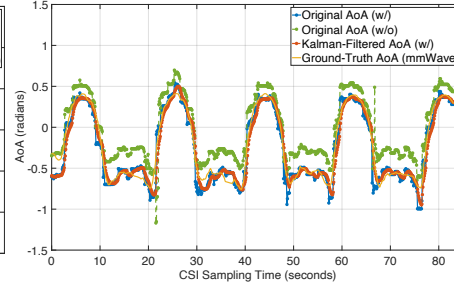
(c) Reflection distance



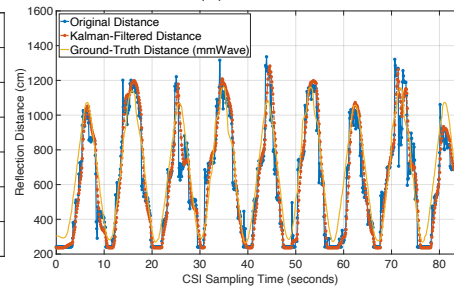
(d) Estimated ellipse trajectory



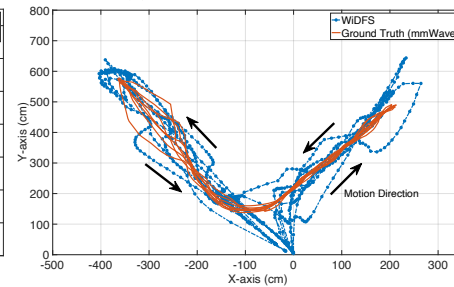
(a) Linear trajectory



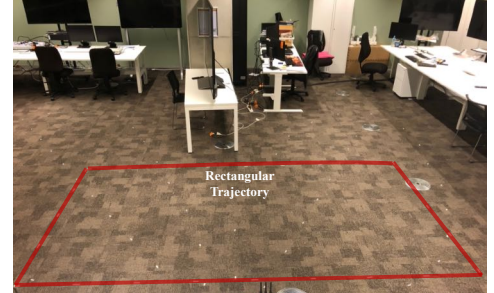
(b) AoA



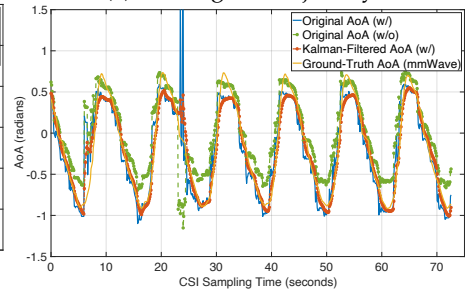
(c) Reflection distance



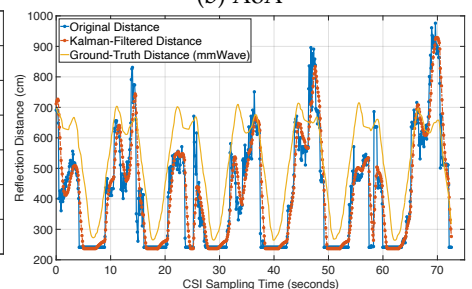
(d) Estimated line trajectory



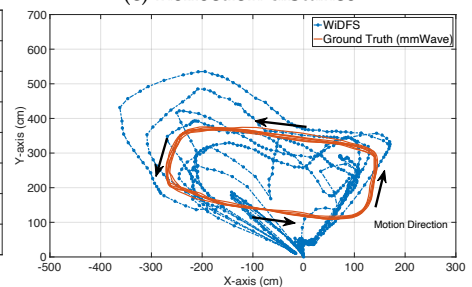
(a) Rectangular trajectory



(b) AoA



(c) Reflection distance



(d) Estimated rectangle trajectory

Fig. 14: A case of ellipse path tracking

Fig. 15: A case of line path tracking

Fig. 16: A case of rectangle path tracking

achieved. The tracking error is significantly higher than that after applying calibration. Fig. 14, Fig. 15 and Fig. 16 show three cases of the trajectories estimated by WiDFS. Most localization results could match the ground-truth trajectory. It can be observed when the tracking target turns around (at peak positions in these figures), some estimated reflection distances exhibits larger errors than other cases. This may mainly be attributed to that part of human reflections are not be reflected to RX antennas. In addition, we know that the human reflection distance should be larger than the LOS distance between the transmitter and the receiver. However, when the tracked person is close to the receiver, we find that some estimates of the human reflection distance may become smaller than the LOS distance due to the impact of measurement noise. In this case, we set the person to be at the origin in the 2D coordinate system in WiDFS.

9.2 Passive tracking accuracy in LOS and NLOS

Next, Fig. 17a and Fig. 17b plots the CDF of the tracking error along the x- and y-axis dimensions in LOS and NLOS cases. At first, Fig. 17a shows the median errors along x- and y-axis in a LOS scenario (on the left in Fig. 12c without cardboard boxes) are less than 32 cm and 56 cm, while Fig. 17b shows the results in another LOS scenario (on the right in Fig. 12c without plastic boxes and wood boards) are less than 37 cm and 49 cm. The 90th percentile errors along x- and y-axis in the two LOS cases are less than (91 cm, 151 cm) and (85 cm, 106 cm), respectively. Obviously, the x-axis has a lower error than the y-axis, which is likely attributed to that three RX antennas form a linear array along the x-axis.

In the NLOS scenario where cardboard boxes are used to block the receiver, the tracking performance is very close to the LOS case, shown in Fig. 17a. In this scenario, WiFi sig-

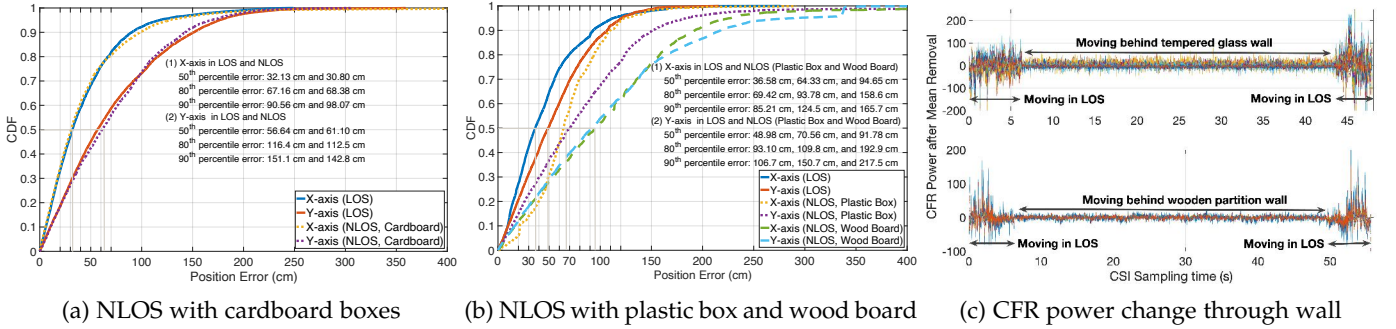


Fig. 17: Tracking performance in LOS and NLOS

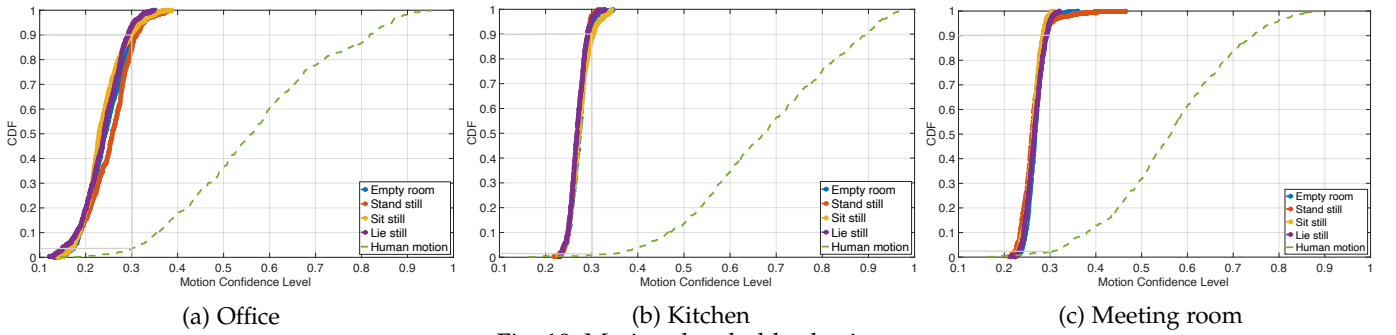


Fig. 18: Motion threshold selection

nals may easily penetrate the cardboard boxes and the obstacles may have little effect on WiFi signals. However, when other materials are used to block the LOS path between the person and the transceiver, Fig. 17b the median errors along x- and y-axis are increased to (64 cm, 71 cm) for plastic-box obstacles and (95 cm, 92 cm) for wood-board obstacles. The thicker or denser obstacles will reduce reflected signal strength even more. Moreover, recall from Section 5 that the phase shift of the static component $H_{1,j}^s$ at each subcarrier is approximately calculated based on the manually-measured distance of the transmitter to the receiver. However, the obstacles may generate relatively strong reflection signals, which cannot be ignored. They may reduce the efficiency of removing the impact of the static component, thereby introducing more position errors. Ideally, as long as the receiver could capture reliable human body reflections in NLOS cases, we believe that WiDFS enables performing accurate tracking since an obstacle may introduce a same phase shift on each subcarrier [33], [34].

We also conduct experiments to test WiDFS in NLOS scenarios where a person moves behind a glass and partition wall. Fig. 17c shows the change in CSI amplitudes at different subcarriers (after lowpass filtering and mean removal). The amplitudes change with human movement in the LOS region, while they almost keep constant in NLOS. This means the receiver may not capture human reflections due to higher signal attenuation through the wall. We also find that the corresponding motion confidence levels calculated based on Eq. 35 are almost below the predefined motion threshold (detailed in Section 9.3) so that WiDFS cannot sense the presence of a moving person.

9.3 Motion threshold selection

This experiment aims to verify our motion detection method in different scenarios, including office, kitchen, and meeting room. We first ask each of the volunteers to separately stand

TABLE 1: 90th-percentile motion confidence level

	Office	Kitchen	Meeting Room
Empty room	0.2998	0.2979	0.2915
Stand still	0.3094	0.2923	0.2881
Sit still	0.2992	0.3027	0.2847
Lie still	0.2932	0.2914	0.2922

still, sit still, and lie still. And we also keep the office empty for testing. Table. 1 shows their 90th-percentile motion confidence levels, which are almost less than 0.3. Then each volunteer moves randomly at different velocities and motion directions. Fig. 18 shows that only about 3th-percentile, 1th-percentile, and 2th-percentile motion confidence levels in the three scenarios are less than 0.3. In other words, when the motion confidence level of 0.3 is selected in WiDFS to detect whether there is a moving person, the false negative rate (when WiDFS fails to detect human motion) is less than about 3% and the false positive rate (when WiDFS fails to detect static scenarios) is less than about 10%. The result shows that the fixed motion threshold could be applied to different environments. In the presence of human motion, WiDFS will automatically perform our tracking algorithm.

In addition, we also test WiDFS in a meeting room, shown in Fig. 19. The distance between the transmitter and receiver is about 120 cm. Fig. 20 plots the CDF of the tracking accuracy along x- and y-axis. We can see that the median errors are 44.20 cm and 64.98 cm, while the 90th percentile errors are 111.6 cm and 170.2 cm. The tracking accuracy is close to that in the office scenario. These results demonstrate the robustness of WiDFS to different environments.

9.4 Impact of AFS on DFS estimation accuracy

This experiment is conducted to evaluate the impact of the AFS when estimating the DFS. Firstly, we use the AoA estimates in Fig. 8 to calculate the AoA difference $\Delta\theta_k^X$ at successive sampling times. We then set $\Delta d = c/(2f_c)$ (i.e., half a wavelength) and $\cos\theta_1^X = 1$ (which is maximal).

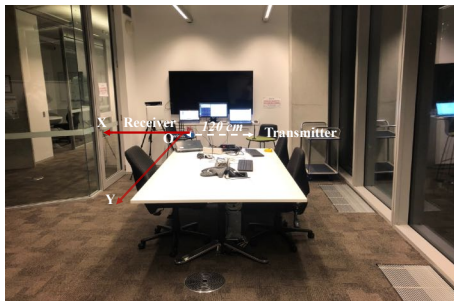


Fig. 19: System setup in meeting room

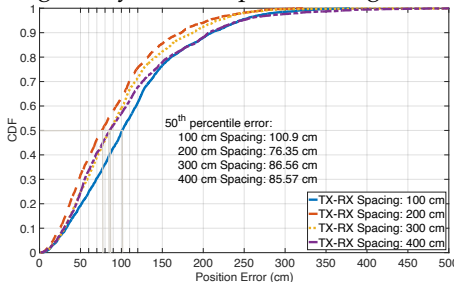
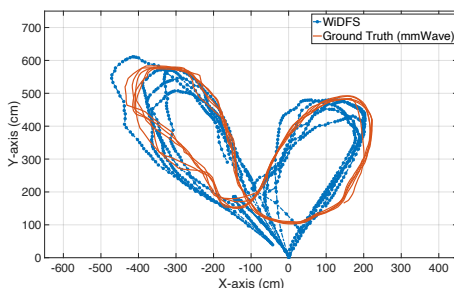


Fig. 22: Impact of TX-RX distance



(a) Elliptical Trajectory

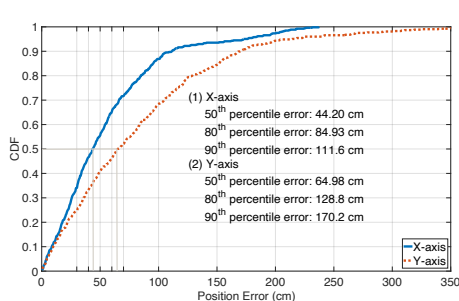


Fig. 20: Tracking error in meeting room

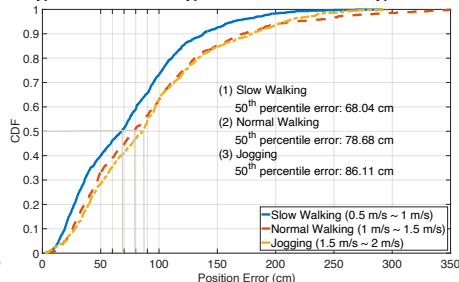
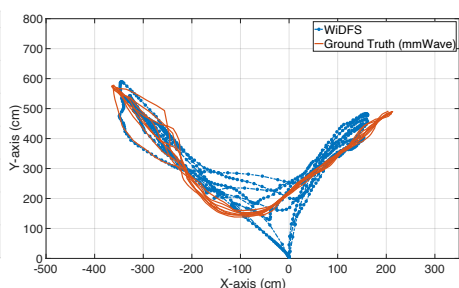


Fig. 23: Impact of motion speed



(b) Linear Trajectory

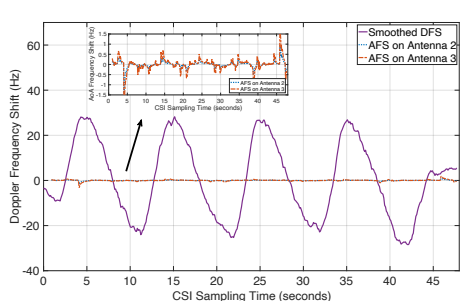


Fig. 21: Impact of AFS on DFS accuracy

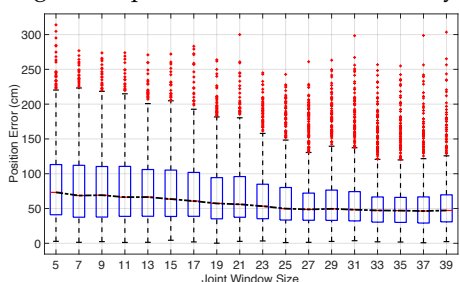
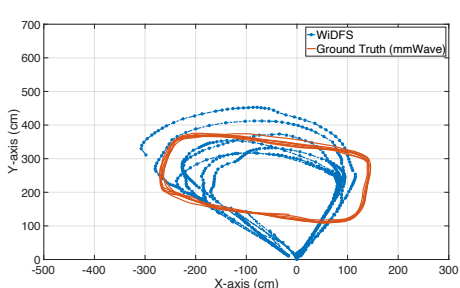


Fig. 24: Impact of joint window size



(c) Rectangular Trajectory

Fig. 25: Three trajectories under a joint window containing 31 CSI sampling sub-windows. Compared to the results in Section 9.1, they could better match the ground truth. However, WiDFS needs more CSI samples for initialization to form a joint window. The delay in each estimation will be increased to about 3s, but it will not accumulate over time.

TABLE 2: Running time of WiDFS and Widar2.0

Algorithm	Language	Platform	Mean	Std.
WiDFS	Python	Mini PC	0.076 s	0.018 s
WiDFS	Python	MacBook Pro 2019	0.024 s	0.002 s
WiDFS	Matlab	MacBook Pro 2019	0.016 s	0.002 s
Widar2.0	Matlab	MacBook Pro 2019	0.136 s	0.021 s

According to the definition of the AFS in Section 2.2, we use these parameters to calculate the AFS. Fig. 21 shows that the AFS values for Antenna 2 and 3 are far smaller than the DFS, so we can ignore the impact of the AFS in our model.

9.5 Realtime performance

WiDFS outputs a position estimate when receiving $100 \times 30 \times 3$ CSI samples from 3 antennas and 30 subcarriers. The sampling interval is about 0.1 s. Any computation delay larger than this upper bound may affect WiDFS's real-time performance. Table. 2 shows that the mean running time of WiDFS (in Python) is 0.076 s on a Mini PC. We also run WiDFS on a MacBook Pro with Intel Core i7 2.6 GHz $\times 6$. The csiread package is used to offline parse CSI data. The running time is reduced to 0.024 s. Also, we program WiDFS using Matlab. We parse raw CSI data to a .MAT file and then exploit it to perform WiDFS and Widar2.0. The computation

time of WiDFS on MacBook Pro is 0.016 s, which is about 10 times faster than Widar2.0. This reduction benefits from the fact that WiDFS does not need to simultaneously estimate multi-dimensional localization parameters.

9.6 Accuracy vs. different parameters

The following will evaluate WiDFS's tracking accuracy as a function of different system parameters: transmitter-to-receiver distance, motion speed, joint window size and person position.

9.6.1 Impact of transmitter-to-receiver distance

We vary the transmitter-to-receiver distance from 100 cm to 400 cm at a spacing of 100 cm. The RX antenna array is fixed. We only change the position of the TX antenna in this experiment. In each case, a person moves along the same trajectory. Fig. 22 shows the minimum tracking error occurs at a distance of 200 cm. This shows that there exist a tradeoff between the separation distance and tracking accuracy. When the transmitter-to-receiver separation distance is enlarged, the geometry of the tracked person relative to the transmitter and receiver is good, which means that the sensitivity to the distance measurement error will accordingly

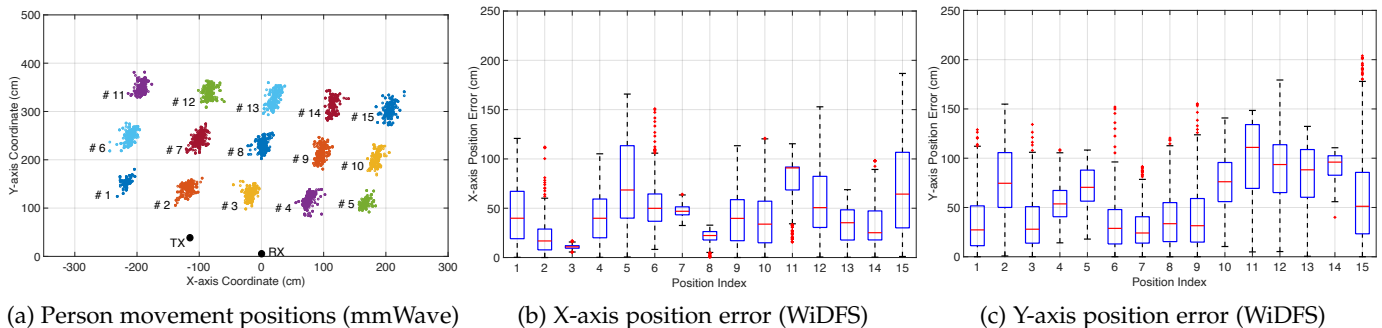


Fig. 26: Impact of person position

decrease. However, if the separation distance is too large, our assumption that there exists one dominant direct signal between the transmitter and receiver may become violated. Surrounding objects produce static multipath signals that we cannot neglect. In contrast, when the transmitter and receiver are placed too closely, the strength of the direct signal between them would overwhelm dynamic multipath.

9.6.2 Impact of motion speed

Fig. 23 shows the median tracking error when a person moves along the same trajectory at different speeds, i.e., Slow walking (about 0.5~1 m/s), Normal walking (1~1.5 m/s), and Jogging (1.5~2 m/s). The median error increases gradually with the higher velocity. Recall from our algorithm that we ignore the difference in DFS among different RX antennas in a CSI sampling window. However, the difference will increase at a higher movement speed. Our future solution is to extend WiDFS to detect the presence of a target moving very quickly and estimate the AFS. In this way, the DFS for each RX antenna would be separately estimated, thereby achieving more accurate tracking.

9.6.3 Impact of joint window size

We next vary the size of a joint window from 5 to 39 CSI sampling sub-windows. Fig. 24 shows that the tracking error gradually decreases as the number of sub-windows increases. Such a result is expected since RX antennas may not capture enough reliable human reflections in some sub-windows, and it is challenging for WiDFS to dilute the impact of the inaccurate dynamic components in a smaller joint window size. This problem can be mitigated by utilizing a larger joint window size. Fig. 25 shows three trajectories used in Section 9.1 for evaluation when a joint window contains 31 sub-windows. It shows that these trajectories are all refined and can better match the ground truth. However, a 3 s delay is introduced for each estimate under the joint window with 31 sub-windows. *Note that since the mean estimation time is less than a sampling sub-window time, the delay will not be accumulated over time.*

9.6.4 Impact of person position

In the experiment, we ask a person to move to different positions relative to the transceiver, and we test the impact of different distances and directions on the localization accuracy. We set 15 landmarks on the floor in the region of interest. The person moves at a landmark each time. We use the mmWave sensor to capture the ground-truth positions. Fig. 26a shows all positions of the person moving at each

landmark. Fig. 26b and Fig. 26c shows the position errors along the x- and y-axis of WiDFS. We can see that the localization results at the middle landmarks are generally better than others. When the person is closer to the transceiver, the receiver can capture more reliable human reflections and thereby WiDFS can achieve better localization accuracy.

10 RELATED WORK

This section reviews various techniques of WiFi-based tracking and sensing, including *signal model-based active tracking*, *signal model-based passive tracking*, and *deep learning-based passive tracking*. The signal model-based solutions are designed by purely analyzing CSI models, while the deep learning-based approaches generally require the use of pre-collected training data. The difference between *active* and *passive* systems is that the active tracking requires a target carrying a WiFi device while the passive tracking is free from this limitation and achieved only based on human body reflections. Compared to previous works, WiDFS is the first unsupervised real-time WiFi tracking system that leverages novel DFS estimation and dynamic component separation techniques to enable passive tracking.

10.1 Signal model-based active WiFi tracking

The topic of WiFi localization has attracted much attention in past years. Typically, ArrayTrack [27] constructs a specialized MIMO-based WiFi receiver platform to estimate AoA from a transmitter's incoming signal, which could pinpoint the transmitter at decimeter-level accuracy. However, its localization error is highly dependent on the number of receivers and the relative geometry of the receivers to the transmitter. SpotFi [22] is also a AoA-based solution using COTS WiFi NICs. It estimates AoA and ToF of a transmitter's signal arriving at a three-antenna receiver via smoothed MUSCI algorithm. However, this ToF is not an actual value distorted by unknown time and frequency shifts between the transmitter and receiver. Chronos [35] is a ToF-based localization solution via COTS devices, which achieves decimeter-level localization only using a receiver equipped with multiple RX antennas. It works by combining multiple frequency bands of 2.4 GHz and 5 GHz to compute the ToF between the transmitter to each antenna of the receiver. In practice, however, it is challenging to span multiple frequency bands for COTS WiFi NICs. WiCapture [36] is designed based on COTS WiFi devices. It estimates AoA and complex attenuation of each signal propagation path and then combines them to calculate the relative displacement

of a transmitter. It just estimates a transmitter's relative trajectory rather than absolute positions. Navid et al. [13] proposes a series of pre-processing methods to eliminate random phase shifts due to transceiver clock asynchronization. Then a MUSIC algorithm is applied to obtain ToF estimates and thereby achieve decimeter-level localization. Our work focuses on passive WiFi tracking and leverages 'bad' multipath interference to track a moving target's trajectory.

10.2 Signal model-based passive WiFi tracking

The device-free sensing is a promising technique since tracked targets do not require carrying any sensors. LiFS [37] formulates the relationship between CSIs and a target location based on a signal power fading model. Then it can rely on the change in CSI amplitude to determine the absence or presence of a person in the first Fresnel zone and locate the target. However, this work requires deploying a large number of receivers in advance (11 receiver PCs are used). MaTrack [38] is a AoA-based solution which uses 2D MUSIC algorithm to estimate absolute AoAs and relative ToFs of static and dynamic paths. However, multiple receivers are required to be deployed apart around a surveillance region and the localization accuracy is subject to the precision of random phase shift removal. Rui Zhou et al. [39] construct a CSI fingerprint database by modeling the relationship between CSI fingerprints and target locations. When a set of new CSI samples comes, this method performs localization by finding the most similar sample from the fingerprint database. However, constructing such a fingerprint database is a time-consuming task. Widar [40] and IndoTrack [19] are DFS-based solutions. They both adopt a reference antenna approach to coarsely separate each antenna's dynamic component from cross-correlation terms. However, the tracking error will accumulate over time. FingerDraw [41] is a system for finger trajectory tracking using WiFi signals. In this work, the authors use a CSI-quotient model to remove the impact of random gain and phase shifts in CSI. This method is shown to be able to achieve improved sensitivity due to the increased SNR. However, since the CSI-quotient based solution cannot separate the static and dynamic components, it may only achieve relative localization. In our scheme WiDFS, the cross-correlation and self-correlation based solution can extract the dynamic human component and achieve absolute localization. Widar2.0 [25], mD-Track [18], and WiPolar [15] are joint parameter estimation solutions by simultaneously estimating signal attenuation, AoA, ToF, and DFS from CSIs. Widar2.0 could completely mitigate random phase shifts by CSI cross correlation between RX antennas. However, its separated dynamic component is not accurate enough. The localization accuracy of mD-Track and WiPolar is determined by the accuracy of random phase shift measurement. Moreover, the three solutions require huge computations in each estimation and cannot achieve real-time tracking. Our proposed WiDFS scheme only uses a COTS transmitter (with 1 TX antenna) and receiver (with 3 RX antennas). It completely removes the impact of random phase shifts via cross correlation. A novel DFS estimation algorithm is developed to obtain unambiguous DFSs. On this basis, WiDFS separately estimates the AoA and reflection distance to achieve low-cost and real-time passive tracking.

10.2.1 Deep Learning-based WiFi tracking

Recently, many applications have been designed with the help of deep learning technique to achieve fine-grained localization/tracking. FreeTrack [42] is a deep neural network (DNN)-based tracking system by feeding into CSI amplitude fingerprints. To reduce the impact of environment change, FreeTrack needs to fine-tune the pre-trained DNN with a few CSI samples from new environments. DLoc [43] designs a novel DL-based framework by treating WiFi localization as an image translation problem. It transforms CSI data into an image and labels training data with a customized robotic mapping platform. FiDo [14] could significantly reduce the number of training data in new scenarios. Even at the same position, different people may induce different CSI data. FiDo adopts a domain-adaptive solution and can pinpoint different persons only using few labelled CSI data. Compared to these data-driven systems that are sensitive to the changes in deployment environment and transceiver position, WiDFS is an unsupervised system and can still achieve high-precision passive tracking.

11 CONCLUSION

This paper introduces a device-free WiFi tracking system that can track a moving person in real time at sub-meter position accuracy. The key design is a novel DFS algorithm that enables measuring an unambiguous DFS when CSI cross-correlation is exploited to mitigate the impact of transceiver asynchronization. And novel algorithms of dynamic human component separation and localization parameter estimation are proposed to achieve high-precision and real-time tracking. Our system prototype runs in a practical multipath-rich scenario without requiring any environment-specific training. Some video clips demonstrating real-time tracking results are available from our website⁴.

One of the important future works is to extend our method to achieve multi-person tracking. To fulfil this task, three TX antennas can be utilized to build a multiple-input-multiple-output (MIMO) system. The AoA resolution of the MIMO WiFi system with 3 TX antennas and 3 RX antennas may be achieved to be equivalent to that of a single-input-multiple-output (SIMO) system (i.e., our WiDFS) with 9 RX antennas. Therefore, the MIMO setup would be a promising way to improve the accuracy in multi-person tracking.

ACKNOWLEDGMENTS

This research was supported partially by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP210101411) and by the NSW Defence Innovation Network and the NSW State Government through Pilot Project grant DINPP-19-20 10.01.

REFERENCES

- [1] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed rfid tags," in *ACM MobiCom*, 2017, pp. 248–260.
- [2] Y. Ma, N. Selby, and A. Fadel, "Drone relays for battery-free networks," in *ACM SIGCOMM*, 2017, pp. 335–347.

4. <https://sites.google.com/view/andrewzhang/research/wifisensing?authuser=0>

- [3] Z. Luo, Q. Zhang, Y. Ma, M. Singh, and F. Adib, "3d backscatter localization for fine-grained robotics," in *USENIX NSDI*, 2019, pp. 765–782.
- [4] H. Xu, D. Wang, R. Zhao, and Q. Zhang, "Adarf: Adaptive rfid-based indoor localization using deep learning enhanced holography," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–22, 2019.
- [5] Z. Wang, M. Xu, N. Ye, F. Xiao, W. Ruchuan, and H. Huang, "Computer vision-assisted 3d object localization via cots rfid devices and a monocular camera," *IEEE Transactions on Mobile Computing*, vol. 3, no. 20, pp. 893–908, 2021.
- [6] T. Wei and X. Zhang, "mtrack: High-precision passive tracking using millimeter wave radios," in *ACM MobiCom*, 2015, pp. 117–129.
- [7] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," in *IEEE DCOSS*, 2019, pp. 33–40.
- [8] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM*, 2020, pp. 2400–2409.
- [9] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "Human tracking and identification through a millimeter wave radar," *Ad Hoc Networks*, vol. 116, p. 102475, 2021.
- [10] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *ACM SIGCOMM*, 2018, pp. 267–281.
- [11] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *IEEE CVPR*, 2018, pp. 7356–7365.
- [12] C. R. Karanam, B. Korany, and Y. Mostofi, "Tracking from one side: multi-person passive tracking with wifi magnitude measurements," in *IEEE IPSN*, 2019, pp. 181–192.
- [13] N. Tadayon, M. T. Rahman, S. Han, S. Valaee, and W. Yu, "Decimeter ranging with channel state information," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3453–3468, 2019.
- [14] X. Chen, H. Li, C. Zhou, X. Liu, D. Wu, and G. Dudek, "Fido: Ubiquitous fine-grained wifi-based localization for unlabelled users via domain adaptation," in *Proceedings of The Web Conference*, 2020, pp. 23–33.
- [15] R. H. Venkatnarayan, M. Shahzad, S. Yun, C. Vlachou, and K.-H. Kim, "Leveraging polarization of wifi signals to simultaneously track multiple people," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–24, 2020.
- [16] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM*, vol. 40, no. 4, pp. 159–170, 2010.
- [17] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wi-fi," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1342–1355, 2018.
- [18] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "md-track: Leveraging multi-dimensionality for passive indoor wi-fi tracking," in *ACM MobiCom*, 2019, pp. 1–16.
- [19] 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," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.
- [20] X. Wang, L. Gao, S. Mao, and S. Pandey, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2016.
- [21] M. Abbas, M. Elhamshary, H. Rizk, M. Turki, and M. Youssef, "Wideep: Wifi-based accurate and robust indoor localization system using deep learning," in *IEEE PerCom*, 2019, pp. 1–10.
- [22] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *ACM SIGCOMM*, 2015, pp. 269–282.
- [23] Y. Zhuo, H. Zhu, H. Xue, and S. Chang, "Perceiving accurate csi phases with commodity wifi devices," in *IEEE INFOCOM*, 2017, pp. 1–9.
- [24] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity wi-fi for interactive exergames," in *ACM CHI*, 2017, pp. 1961–1972.
- [25] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single wi-fi link," in *ACM MobiSys*, 2018, pp. 350–361.
- [26] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su, "Towards 3d human pose construction using wifi," in *ACM MobiCom*, 2020, pp. 1–14.
- [27] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *USENIX NSDI*, 2013, pp. 71–84.
- [28] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath Jr, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *arXiv preprint arXiv:2102.12780*, 2021.
- [29] Z. Ni, J. A. Zhang, X. Huang, K. Yang, and J. Yuan, "Uplink sensing in perceptive mobile networks with asynchronous transceivers," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1287–1300, 2021.
- [30] B. D. Rao and K. S. Hari, "Performance analysis of root-music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.
- [31] dhalperi, "linux-80211n-csitool," December 2014. [Online]. Available: <https://github.com/dhalperi/linux-80211n-csitool>
- [32] dhalperi, "linux-80211n-csitool-supplementary," December 2014. [Online]. Available: <https://github.com/dhalperi/linux-80211n-csitool-supplementary>
- [33] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *USENIX NSDI*, 2014, pp. 317–329.
- [34] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via rf body reflections," in *USENIX NSDI*, 2015, pp. 279–292.
- [35] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in *USENIX NSDI*, 2016, pp. 165–178.
- [36] M. Kotaru and S. Katti, "Position tracking for virtual reality using commodity wifi," in *IEEE CVPR*, 2017, pp. 68–78.
- [37] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "Lifs: low human-effort, device-free localization with fine-grained subcarrier information," in *ACM MobiCom*, 2016, pp. 243–256.
- [38] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-music: accurate device-free indoor localization," in *ACM UbiComp*, 2016, pp. 196–207.
- [39] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with svm and csi fingerprinting," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7990–7999, 2017.
- [40] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi," in *ACM Mobihoc*, 2017, pp. 1–10.
- [41] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, "Fingerdraw: Sub-wavelength level finger motion tracking with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–27, 2020.
- [42] R. Zhou, M. Tang, Z. Gong, and M. Hao, "Freetrack: Device-free human tracking with deep neural networks and particle filtering," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2990–3000, 2019.
- [43] R. Ayyalasomayajula, A. Arun, C. Wu, S. Sharma, A. R. Sethi, D. Vasisht, and D. Bharadia, "Deep learning based wireless localization for indoor navigation," in *ACM MobiCom*, 2020, pp. 1–14.



Dr Zhongqin Wang is currently a research engineer in the School of Electrical and Data Engineering at University of Technology Sydney. He received Ph.D. degree from University of Technology Sydney, Australia in 2021. He received M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China in 2014. His research interest includes wireless sensing based on WiFi, RFID, and mmWave.



Dr J. Andrew Zhang (M'04-SM'11) received B.Sc. degree from Xi'an JiaoTong University, China, in 1996, M.Sc. degree from Nanjing University of Posts and Telecommunications, China, in 1999, and Ph.D. degree from the Australian National University, in 2004. Currently, He is an Associate Professor in the School of Electrical and Data Engineering, University of Technology Sydney, Australia. Dr. Zhang's research interests are in the area of signal processing for wireless communications and sensing, and autonomous

vehicular networks. He has published 200+ journal and conference papers, and has won 5 best paper awards.



Min Xu (M'10) is currently an Associate Professor at University of Technology Sydney. She received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2000, the M.S. degree from National University of Singapore, Singapore, in 2004, and the Ph.D. degree from University of Newcastle, Callaghan NSW, Australia, in 2010. Her research interests include multimedia data analytics, computer vision and machine learning. She has published over 100 research papers in high quality international journals and conferences. She has been invited to be a member of the program committee for many international top conferences, including ACM Multimedia Conference and reviewers for various highly-rated international journals, such as IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology and much more. She is an Associate Editor of Journal of Neurocomputing.



Y. Jay Guo (Fellow'2014) received a Bachelor Degree and a Master Degree from Xidian University in 1982 and 1984, respectively, and a PhD Degree from Xian Jiaotong University in 1987, all in China. His research interest includes antennas, mm-wave and THz communications and sensing systems as well as big data technologies. He has published four books and over 600 research papers including over 280 IEEE Transactions papers, and he holds 26 international patents. He is a Fellow of the Australian Academy of Engineering and Technology, a Fellow of IET, and was a member of the College of Experts of Australian Research Council (ARC, 2016-2018). He has won a number of most prestigious Australian Engineering Excellence Awards (2007, 2012) and CSIRO Chairman's Medal (2007, 2012). He was named one of the most influential engineers in Australia in 2014 and 2015, respectively, and one of the top researchers across fields in Australia in 2020 and 2021.

He is a Distinguished Professor and the Director of Global Big Data Technologies Centre (GBDTC) at the University of Technology Sydney (UTS), Australia. He is the founding Technical Director of the New South Wales Connectivity Innovation Network. Prior to joining UTS in 2014, he served as a Director in CSIRO for over nine years. Before joining CSIRO, he held various senior technology leadership positions in Fujitsu, Siemens and NEC in the U.K.

Prof Guo has chaired numerous international conferences and served as guest editors for a number of IEEE publications. He is the Chair of International Steering Committee, International Symposium on Antennas and Propagation (ISAP). He has been the International Advisory Committee Chair of IEEE VTC2017, General Chair of ISAP2022, ISAP2015, iWAT2014 and WPMC'2014, and TPC Chair of 2010 IEEE WCNC, and 2012 and 2007 IEEE ISCIT. He served as Guest Editor of special issues on "Low-Cost Wide-Angle Beam Scanning Antennas", "Antennas for Satellite Communications" and "Antennas and Propagation Aspects of 60-90GHz Wireless Communications," all in IEEE Transactions on Antennas and Propagation, Special Issue on "Communications Challenges and Dynamics for Unmanned Autonomous Vehicles," IEEE Journal on Selected Areas in Communications (JSAC), and Special Issue on "5G for Mission Critical Machine Communications", IEEE Network Magazine.

Prof Guo has chaired numerous international conferences and served as guest editors for a number of IEEE publications. He is the Chair of International Steering Committee, International Symposium on Antennas and Propagation (ISAP). He has been the International Advisory Committee Chair of IEEE VTC2017, General Chair of ISAP2022, ISAP2015, iWAT2014 and WPMC'2014, and TPC Chair of 2010 IEEE WCNC, and 2012 and 2007 IEEE ISCIT. He served as Guest Editor of special issues on "Low-Cost Wide-Angle Beam Scanning Antennas", "Antennas for Satellite Communications" and "Antennas and Propagation Aspects of 60-90GHz Wireless Communications," all in IEEE Transactions on Antennas and Propagation, Special Issue on "Communications Challenges and Dynamics for Unmanned Autonomous Vehicles," IEEE Journal on Selected Areas in Communications (JSAC), and Special Issue on "5G for Mission Critical Machine Communications", IEEE Network Magazine.