

Tracking from One Side – Multi-Person Passive Tracking with WiFi Magnitude Measurements

Chitra R. Karanam

University of California Santa Barbara
Santa Barbara, California
ckaranam@ece.ucsb.edu

Belal Korany

University of California Santa Barbara
Santa Barbara, California
belalkorany@ece.ucsb.edu

Yasamin Mostofi

University of California Santa Barbara
Santa Barbara, California
ymostofi@ece.ucsb.edu

ABSTRACT

In this paper, we are interested in passively tracking multiple people walking in an area, using only the magnitude of WiFi signals from one WiFi transmitter and a small number of receivers (configured as an array) located on one side of the area. Past works on RF-based tracking either track only a single moving person, use a large number of transceivers surrounding the area to track multiple people, or use additional resources like ultra-wideband signals. Furthermore, magnitude-based tracking provides an attractive feature that additional receiver antennas can easily be added to the antenna array as needed, without the need for phase synchronization, since the magnitude can be measured independently on the different antennas. In this paper, we then propose a new framework that uses only the magnitude of WiFi signals and expresses it in terms of the angles of arrival of signal paths at the receivers as well as the motion parameters of the virtual arrays emulated by the moving people. We then use a two-dimensional Multiple Signal Classification (MUSIC) algorithm to estimate the aforementioned parameters, and further utilize a Particle Filter with a Joint Probabilistic Data Association Filter to track multiple people walking in the area. We extensively validate our proposed framework in both indoor and outdoor areas, through 40 experiments of tracking 1 to 3 people, using only one transmit antenna and three laptops as receivers (a total of four off-the-shelf Intel 5300 WiFi Network Interface Cards (NICs)). Our results show highly accurate tracking (mean error of 38 cm in outdoor areas/closed parking lots, and 55 cm in indoor areas) using minimal WiFi resources on only one side of the area.

CCS CONCEPTS

• **Hardware** → **Wireless devices**; • **Computer systems organization** → **Sensor networks**;

KEYWORDS

Tracking, WiFi, Magnitude-based sensing

ACM Reference Format:

Chitra R. Karanam, Belal Korany, and Yasamin Mostofi. 2019. Tracking from One Side – Multi-Person Passive Tracking with WiFi Magnitude

Measurements. In *The 18th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2019) (IPSN '19)*, April 16–18, 2019, Montreal, QC, Canada. ACM, Montreal, Canada, 12 pages. <https://doi.org/10.1145/3302506.3310399>

1 INTRODUCTION

In recent years, the idea of smart spaces, homes, and buildings is increasingly becoming popular, leading to an increasing demand for easy and effortless interaction of humans with their surroundings and devices. This has also been accompanied by a huge growth in the number of wirelessly-connected devices, leading to ubiquitous RF signals. These RF signals interact with the people and objects in the environment, and implicitly carry information about various attributes of our surroundings. Consequently, there has been a considerable interest in the research community in using RF signals to learn about our surroundings. For instance, several attempts have been made for using RF signals for localization and imaging of static objects [13], localization of devices [14], crowd occupancy estimation [7], tracking peoples' movements [3, 18], and gestures [26]. In particular, passively tracking multiple people that are walking in an area, without relying on them to carry any device, is a challenging problem of considerable interest, due to its importance in many applications such as elderly monitoring, intrusion detection, and retail analytics.

In this paper, we consider the problem of passively tracking multiple people walking in an area, using minimal WiFi resources on only one side. In particular, we are interested in multi-person tracking using only the magnitude of WiFi Channel State Information (CSI) measurements, measured using one WiFi transmitter and a small number of WiFi receiver Network Interface Cards (NICs) located on only one side of the area.

Device-free localization and tracking (DFLT) is a challenging problem that has gained a considerable attention in the recent years. The efforts exerted in DFLT can be broadly categorized into two categories: (a) machine learning and fingerprinting-based approaches, and (b) model-based approaches. Machine learning and fingerprinting approaches (e.g. [23, 28, 30]) require extensive prior calibration and training, which is both time consuming and environment-specific. Model-based approaches, on the other hand, build a model/relationship between the location (or track) of the target and the wireless measurements at the receiver. They then estimate the location and track of the target based on that model. Earlier work in this category assumes the availability of extensive amount of resources. For instance, [3] tracks multiple targets using a link-crossing model based on RSSI, but with a large number of sensors (total of 32) distributed on all sides of the tracking area. Other attempts track multiple targets using specialized hardware

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IPSN '19, April 16–18, 2019, Montreal, QC, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6284-9/19/04...\$15.00

<https://doi.org/10.1145/3302506.3310399>

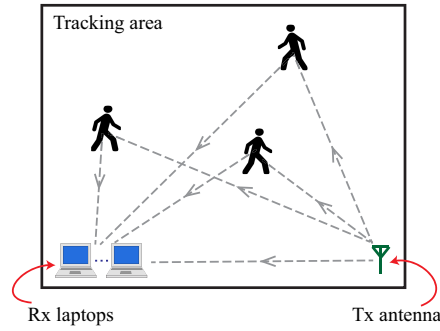


Figure 1: An illustration of our passive multi-person tracking setup. N people are walking in an area. A WiFi link consisting of one Tx antenna and one small Rx array (for instance, from a couple of laptops) is located on one side of this area. We are then interested in estimating the tracks of the N people walking in the area using only the magnitude of the received WiFi signal measurements.

and signals with very large bandwidth [1, 4]. More recently, WiFi CSI has been made available on Commercial Off-The-Shelf (COTS) WiFi devices, such as Intel 5300 and Atheros AR9580 WiFi cards. The availability of CSI on COTS WiFi cards has opened the door for DFLT techniques that require fewer resources than before. Several works then utilized WiFi CSI to track a *single* moving target by estimating different parameters of the wireless signal, e.g. Angle-of-Arrival (AoA), Time of Flight (ToF), and Doppler spread [15, 16, 18]. Two common threads exist among the aforementioned CSI-based device-free tracking approaches. First, all these approaches are designed for single target tracking, and fail to track multiple targets that are moving simultaneously. Second, these approaches rely on the CSI phase information, which can be hard to measure accurately or may be unavailable on other COTS devices. Furthermore, relying on the phase information limits the flexibility of adding more antennas to the receiver, since different WiFi cards need to be synchronized for obtaining meaningful phase information. This will add a considerable synchronization overhead to the system. These two problems motivate the need for a device-free tracking framework, that is able to track *multiple* targets, using only the WiFi *magnitude* measurements.

Contribution Statements: In this paper, we propose a framework for passively tracking *multiple* people walking in an area, without requiring a prohibitive amount of resources (e.g. bandwidth or number of transceivers), which were used previously for such purposes. Our framework uses only the magnitude of WiFi CSI measurements, measured from one side of the area, on a small receiver array. More specifically,

- We propose a new magnitude-based framework to track multiple people walking in an area, using one transmitter and a very small receiver array (for instance, from a couple of laptops), without the need to make any prior measurements in the area of interest. By modeling the tracking problem in terms of only the signal magnitude, our proposed framework can be implemented on any off-the-shelf platform where phase measurements are not reliable or are not easily available. Furthermore, additional Rx NICs can be added to the receiver setup if required, without any need for phase synchronization or calibration. On the other hand, such a receiver antenna extension in the case of relying on phase

measurements would have required an antenna port on each NIC to be used up for the purpose of synchronization, after which the phase is still only accurate to within a median error of 20° [9].

- We propose a two-dimensional signal model for the estimation of various AoA parameters that are functions of the targets' locations and motion directions. By posing our problem as a joint parameter estimation problem in this manner, we show how the ambiguity in individual dimensions can be overcome. We then extend the multi-dimensional MUSIC algorithm to our magnitude-based modeling framework in order to estimate the 2D AoA parameters. Finally, we track multiple targets in the area by using a Particle Filter (PF) with a Joint Probabilistic Data Association Filter (JPDAF).
- We extensively validate our proposed multi-person tracking framework through a total of 40 experiments in 6 different environments, with 1, 2, and 3 people walking on different paths, on different days. We use only one transmit and 3 WiFi receiver NICs on one side of the area to measure WiFi CSI magnitude. Our results show highly accurate tracking with a mean error of 38 cm in outdoor areas/parking lots, and 55 cm in indoor areas.

The rest of this paper is organized as follows. In Sec. 2, we provide a detailed discussion on the state-of-the-art for both single and multiple target tracking. In Sec. 3, we describe our proposed magnitude-based two-dimensional framework for the estimation of AoA parameters. In Sec. 4, we show how to track multiple moving targets using a particle filter with a JPDAF. We experimentally validate our proposed framework for tracking multiple targets in Sec. 5. Finally, we discuss the limitations and future extensions of our proposed framework in Sec. 6, and conclude in Sec. 7.

2 RELATED WORK

In this section, we provide a review of the state-of-the-art on passive target tracking using RF signals. We start by discussing the work that only focused on single target tracking. This is then followed by summarizing the work that enabled passive multiple target tracking. We then place our proposed framework in the context of the state-of-the-art work in tracking and discuss our contributions. A detailed comparison of the different proposed methods (including ours) for single and multiple target tracking is shown in Table 1.

Single target tracking: In [18], the authors proposed Widar2.0 for single target tracking, using one WiFi transmitter and a receiver array to measure multiple wireless signal parameters. These parameters include ToF, AoA, Doppler spread, and attenuation. The authors also rely on successive measurements in time in order to select the single reflected path that fits the time series of these parameters. This reflected path is then used to localize the reflecting target. Widar2.0 achieves an average tracking error of 75 cm. The authors of [16] use a similar setup of one WiFi transmitter and two receiver arrays located on two sides of the area. The receiver arrays independently estimate the Doppler spread and AoA of the single reflected path, and the target is localized accordingly, achieving a median tracking error of 35 cm. In [15], the authors utilize two WiFi links, each with one transmitter and one receiver array, to estimate the AoA and ToF of the reflected path, and then choose the target's location as the one that best fits the measurements on both links, simultaneously. They then achieve a median tracking error in the

Paper	Number of targets	Bandwidth	Magnitude/power information only	Number of devices used	Tracking from one side	Tracking error
Widar2.0 [18]	Single	Narrowband	✗	2 WiFi NICs	✓	75 cm
IndoTrack [16]	Single	Narrowband	✗	3 WiFi NICs	✗	35 cm
DynMusic [15]	Single	Narrowband	✗	4 WiFi NICs	✗	36–62 cm
[12]	Single	Narrowband	✓	4 WiFi NICs	✗	31 cm
WiTrack2.0 [1]	Multiple (4)	FMCW radar	✗	1 FMCW radar	✓	10.6–17.5 cm
[4]	Multiple (2)	UWB radar	✗	1 UWB radar	✓	0.6 cm*
[3]	Multiple (4)	Narrowband	✓	32 ZigBee nodes	✗	26–45 cm
[17]	Multiple (3)	Narrowband	✓	24 ZigBee nodes	✗	31–91 cm
SCPL [29]	Multiple (4)	Narrowband	✓	22 CC1100 nodes [†]	✗	108 cm
This paper	Multiple (3)	Narrowband	✓	4 WiFi NICs	✓	47 cm

* when compared to a colocated LiDAR system

[†] radio transceivers operating in the 909.1 MHz unlicensed band

Table 1: Comparison with the state-of-the-art in target tracking using RF signals.

range of 26–62 cm. The authors of [12] use one WiFi transmitter and 3 WiFi receivers to track a single target using a magnitude-based framework for virtual array AoA estimation and a Particle Filter. They then achieve a mean tracking error of 31 cm in an outdoor environment. All these previously-proposed methods only focus on tracking one target in the area, and thereby estimate the parameters of a single reflected time-varying path. Hence, these methods cannot directly be extended to accommodate multiple targets moving in the area. Furthermore, the algorithms presented in [15, 16, 18] rely on phase information of the received wireless signal, which is not available on many COTS devices, and require calibration and synchronization overhead when using multiple receivers.

Multiple target tracking: Several works have been proposed to track multiple targets simultaneously. In [1], the authors propose WiTrack2.0, a system that uses an FMCW radar, spanning 1.79 GHz of bandwidth, in order to localize passive targets. An FMCW radar produces a very high-resolution ToF profile of the environment. Hence, it is possible to generate a heatmap with a high-resolution in the range of the targets in the area. The ambiguity in cross-resolution is then resolved by using multiple TX-RX pairs for the FMCW radar, achieving a high localization accuracy of 10–17 cm. In [4], the authors propose a Kalman-Filter-based algorithm to track the ranges and velocities of two moving targets using a UWB monostatic radar. Their method achieves a localization accuracy of 0.6 cm when compared to a colocated LiDAR system. Other works utilized off-the-shelf devices to solve the passive multi-target tracking problem. For instance, the authors in [3] utilized many ZigBee nodes (30–33) distributed all around the tracking area, and measuring RSSI information. A target in the area blocks the Line of Sight (LOS) of a subset of the links created by the nodes. The location of the target is then calculated based on the links with the highest changes in RSSI. The tracks of multiple targets are then estimated by means of clustering and Kalman Filter algorithms. The algorithm in [3] then achieves an average error in the range 26–45 cm (depending on the environment, and the number of targets). Similar algorithms that create a mesh network with a very dense node deployment, and a resulting large number of links all around the tracking area, were proposed in [17, 29].

As can be seen from the previous discussion, enabling multiple target tracking has traditionally required a prohibitive increase in

the required resources, either a huge bandwidth and specialized hardware as shown in [1, 4], or a large number of transceivers distributed around the tracking area from all sides as shown in [3, 17, 29]. Our proposed framework in this paper is then the first to enable multiple target tracking with comparable accuracy to the state-of-the-art, without requiring extra resources. We show that multiple target tracking is possible using only magnitude measurements on one receiver array located on the same side of the tracking area as the transmitter. We provide tracking results of up to 3 people, with an average tracking accuracy of 47 cm across 40 experiments in six different tracking areas.

3 PROPOSED 2D FRAMEWORK FOR MULTI-PERSON TRACKING

Consider N people walking in an area, as shown in Fig. 1. A WiFi transmitter (Tx) and a small WiFi receiver (Rx) array (for instance, from a couple of laptops) are located on one side of the area. The WiFi signals in the area are scattered and reflected off of the people and the objects present in the area. When these signals reach the receiver, they implicitly contain information about the people and objects that they interact with, on their path from the transmitter to the receiver. More specifically, as we shall see, these signals can be used to infer the location and the track of the people walking in the area. In this section, we show how to model this interaction of WiFi signals with the people and the environment, in order to obtain valuable information on their whereabouts. We next describe the information that can be extracted from the magnitude of WiFi signals in such a scenario. As mentioned earlier, the advantage of using such a magnitude-based approach is that any number of antennas from different receivers can be added to the array to extend its length if the scenario warrants it, without any need for phase synchronization or phase correction.

3.1 Review of 1D Signal Analysis [12]

In this section, we provide a brief primer on the various target parameters that can be extracted from a one-dimensional signal measured using the framework shown in Fig. 1. More specifically, we first discuss the AoA information that can be extracted using only the magnitude of the signal measured at a receiver antenna array. Then, we discuss the virtual array parameters created by

a person's motion, which can be extracted from the magnitude measurements at a single receiver antenna over time. We henceforth refer to these two dimensions of measurements as the *Rx array angle-of-arrival parameters* and *Motion-induced array parameters*.

3.1.1 Rx Array Angle-of-Arrival Parameters. Consider the receiver array shown in Fig. 2. The baseband received signal (or equivalently, the baseband channel gain) at the Rx array at one time instant due to the WiFi signal transmissions in the area can be written as a function of the distance d along the array as follows [11]:

$$c(d) = \alpha_0 e^{-j \frac{2\pi}{\lambda} d \cos \phi_0} + \sum_{n=1}^N \alpha_n e^{-j \frac{2\pi}{\lambda} d \cos \phi_n} + \eta(d), \quad (1)$$

where α_n is the complex amplitude (or equivalently the gain) of the received signal path from the n^{th} target at the first antenna, ϕ_n is the angle-of-arrival corresponding to the n^{th} path, α_0 and ϕ_0 are the complex amplitude and angle-of-arrival corresponding to the direct signal path from the transmitter to the receiver array, N is the number of targets in the area, λ is the signal wavelength, and $\eta(d)$ is the receiver noise. The Fourier transform of $|c(d)|^2$ for the case of passive targets ($|\alpha_0| \gg |\alpha_n|$) can be derived as,

$$C(f_d) = A\delta(f_d) + \sum_{n=1}^N \alpha_0 \alpha_n^* \delta\left(f_d - \frac{\psi_n^A}{\lambda}\right) + \sum_{n=1}^N \alpha_0^* \alpha_n \delta\left(f_d + \frac{\psi_n^A}{\lambda}\right) + \zeta_d(f_d), \quad (2)$$

where f_d is the spatial frequency, $\delta(\cdot)$ is the Dirac delta function, $A = \sum_{n=0}^N |\alpha_n|^2$, $\psi_n^A = \cos \phi_0 - \cos \phi_n$, and $\zeta_d(f_d)$ is the frequency domain modeling error term. As can be seen, there are peaks in the spectrum $C(f_d)$ at frequencies (normalized with respect to $1/\lambda$) corresponding to $\pm \psi_n^A$. Therefore, given a path with angle-of-arrival ϕ_n , we see two peaks in the spectrum corresponding to the two frequencies $\pm(\cos \phi_0 - \cos \phi_n)$. We then have an ambiguity in the AoA of that path, due to the ambiguity in the sign of ψ_n^A .

In the context of tracking multiple targets, the estimation of the AoAs of the targets from Eq. 2 localizes the targets to a small extent. However, the previously-mentioned ambiguity in the sign of ψ_n^A hinders our ability to accurately estimate these angles for each target. Furthermore, the resolution and the number of angles that can be estimated is limited by the length of the receiver antenna array, which we assume to be small. This is a crucial aspect that we address in this paper, since there could be a relatively large number of signal paths arriving at the receiver due to reflections off of multiple targets as well as static objects in the area.

3.1.2 Motion-Induced Array Parameters. Next consider the scenario of measuring the time series of the received signal at a single antenna of the array shown in Fig. 2. As the targets move in the area, they create equivalent virtual antenna arrays when the signal receptions at the antenna are considered over time. The temporal received signal in such a case can be written as,

$$c(t) = \alpha_0 + \sum_{n=1}^N \alpha_n e^{-j \frac{2\pi}{\lambda} \psi_n^M t} + \eta(t), \quad (3)$$

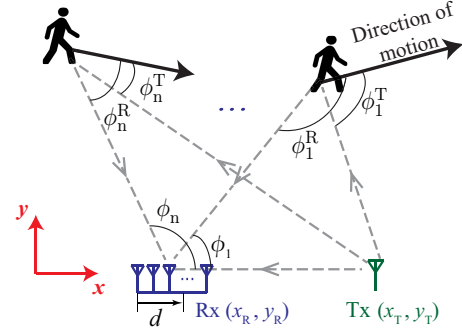


Figure 2: Signal model for the multi-target tracking problem. One temporal snapshot of the measurements at the small receiver array can estimate the array-based angles-of-arrival of the targets, and measurements over time at one antenna of the array can estimate the motion-induced array parameters of a moving target, using only the magnitude of the received signals.

where α_n is the complex amplitude of the path arriving from the n^{th} moving target at time $t = 0$, $\psi_n^M = v_n(\cos \phi_n^R + \cos \phi_n^T)$ is the motion-induced array parameter that arises from the virtual antenna array created by the motion of the n^{th} target, ϕ_n^R and ϕ_n^T are angles with respect to the direction of motion as shown in Fig. 2, v_n is the speed of the n^{th} target, and $\eta(t)$ is the receiver noise. Consequently, the magnitude of the signal can be used to estimate the motion-induced array parameter ψ_n^M , which contains information about the location of the corresponding target. The spectrum of $|c(t)|^2$ generated with respect to the variable t can then be written as follows:

$$C(f_t) = A\delta(f_t) + \sum_{n=1}^N \alpha_0 \alpha_n^* \delta\left(f_t - \frac{\psi_n^M}{\lambda}\right) + \sum_{n=1}^N \alpha_0^* \alpha_n \delta\left(f_t + \frac{\psi_n^M}{\lambda}\right) + \zeta_t(f_t), \quad (4)$$

where f_t is the frequency variable, and $\zeta_t(f_t)$ is the modeling error term in the spectrum. In the spectrum in Eq. 4, we see peaks at locations $\pm \psi_n^M$, thereby exhibiting ambiguity regarding the sign of ψ_n^M for the n^{th} target. Furthermore, the value ψ_n^M itself does not localize the n^{th} target, since different locations, headings and speeds of the target can result in the same value of ψ_n^M . However, it is a function of the targets' motion parameters, which is still informative and can be utilized to track the moving targets over time [12].

In summary, both the Rx array angle-of-arrival parameters (ψ^A) and motion-induced array parameters (ψ^M) measure different quantities related to the targets' locations and headings, but are ambiguous in the sign of the respective measurements as well as the locations they correspond to, in the area of interest. We next propose a framework to jointly estimate both quantities, and show that this joint estimation additionally eliminates the ambiguity in the signs of the individual measurements. Sec. 4 then shows how to resolve the residual location ambiguity and fully track the targets.

3.2 Multi-Dimensional Signal Analysis for Target Tracking

So far, we have seen that the Rx array angle-of-arrival (ψ^A) and motion-induced array parameters (ψ^M) contain different kinds of

information about the targets in the area. In this section, we propose to estimate these parameters jointly, by using the magnitude-based framework to generate a joint spectrum. Consider the multi-dimensional received signal $c(t, d)$, which is a function of time t and distance d along the array, written as follows:

$$c(t, d) = \alpha_0 e^{-j\frac{2\pi}{\lambda} \cos \phi_0} + \sum_{n=1}^N \alpha_n e^{-j\frac{2\pi}{\lambda} \psi_n^M t} e^{-j\frac{2\pi}{\lambda} d \cos \phi_n} + \eta(t, d), \quad (5)$$

where $\eta(t, d)$ is the receiver noise. The two parameters ψ^A and ψ^M then appear jointly in the two-dimensional spectrum generated from $|c(t, d)|^2$. More specifically, the 2D spectrum of $|c(t, d)|^2$ can be written as

$$\begin{aligned} C(f_t, f_d) = & A\delta(f_t) + \sum_{n=1}^N \alpha_0 \alpha_n^* \delta(f_t - \frac{\psi_n^M}{\lambda}, f_d - \frac{\psi_n^A}{\lambda}) \\ & + \sum_{n=1}^N \alpha_0^* \alpha_n \delta(f_t + \frac{\psi_n^M}{\lambda}, f_d + \frac{\psi_n^A}{\lambda}) + \zeta_{t,d}(f_t, f_d), \end{aligned} \quad (6)$$

where $\delta(\cdot, \cdot)$ is the 2D Dirac delta function, and $\zeta_{t,d}(f_t, f_d)$ represents the modeling error term in the 2D spectrum. The locations of the peaks in this 2D spectrum then give the corresponding pairs of ψ^A and ψ^M values for each of the moving targets. By using a joint estimation framework, the chance of two targets resulting in the same peak considerably decreases. For instance, two targets could have the same ψ^A values, but they could be different in their ψ^M values, or vice versa. Such scenarios are now well separated in the 2D spectrum.

Note that in the joint spectrum in Eq. 6, we still obtain two peaks corresponding to each target in the area. For instance, the n^{th} target generates peaks in the spectrum at (ψ_n^M, ψ_n^A) and $(-\psi_n^M, -\psi_n^A)$. However, by choosing the location of the transmitter appropriately, we can eliminate this ambiguity. To this end, we propose to place the transmitter at one extreme of the angle-of-arrival space of the receiver array ($\phi_0 = 0^\circ$ or $\phi_0 = 180^\circ$). Without loss of generality, suppose that we place the transmitter such that $\phi_0 = 0^\circ$, as shown in Fig. 2. Then, $\psi_n^A = 1 - \cos \phi_n$, which is a quantity that lies in the interval $[0, 2]$. This implies that $-\psi_n^A$ lies in $[-2, 0]$. Since these two intervals are disjoint, we can restrict the search space of ψ^A in the spectrum to the $[0, 2]$ interval. Then, the n^{th} target generates only one peak in the limited spectrum at (ψ_n^M, ψ_n^A) , thereby eliminating the ambiguity in the sign of both the ψ parameters. We henceforth use this configuration in all the discussions in this paper. A similar analysis can be derived for the case when the transmitter is located such that $\phi_0 = 180^\circ$. Thus, our proposed joint framework eliminates the ambiguity in the peaks and provides a larger search space for multiple targets in the spectrum. Fig. 3 shows an example of a 2D spectrum with 3 peaks corresponding to 3 targets in the region $-2 \leq \psi^M \leq 2$, and $0 \leq \psi^A \leq 2$. The locations of the peaks in the (ψ^M, ψ^A) space are $(1.2, 1.4)$, $(1.2, 0.6)$, and $(-1.2, 0.6)$. In the 1D analysis for ψ^M , all the peaks would not be resolvable since they have the same absolute value of 1.2. On the other hand, in the 1D analysis for the ψ^A dimension, two of the peaks would not be resolvable due to having the same value of 0.6. However, as can be seen in Fig. 3, all three peaks are resolvable in the joint 2D spectrum.

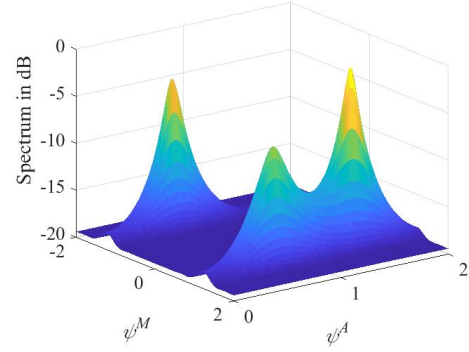


Figure 3: A sample 2D spectrum with 3 peaks corresponding to 3 targets in an area. The locations of the peaks in the (ψ^M, ψ^A) space are $(1.2, 1.4)$, $(1.2, 0.6)$, and $(-1.2, 0.6)$. The peaks are resolvable only in the joint 2D spectrum, but not in the individual dimensions.

So far, we have shown how the peaks in the joint spectrum in Eq. 6 contain information about the targets in the area. While this joint spectrum can easily be obtained by using a 2D Fourier Transform on $|c(t, d)|^2$, in practice, we would need a long antenna array to get a reasonable resolution in the f_d dimension in the 2D spectrum. Thus, we next discuss how we can efficiently estimate the required 2D spectrum, using MUSIC, even with a small Rx antenna array, and subsequently use that information to track moving targets in the area.

REMARK 1. Note that in equations 5 and 6, the static multipath does not affect the locations of the peaks of the moving targets in the spectrum. All the signal paths corresponding to the static multipath get lumped at $\psi^M = 0$ in the spectrum. Thus, by removing the temporal mean of $|c(t, d)|^2$, we can eliminate the effect of the static multipath.

3.3 Multi-Dimensional Parameter Estimation - 2D MUSIC

In this section, we describe our framework to estimate the 2D spectrum from the raw spatio-temporal magnitude-squared measurements $|c(t, d)|^2$. We can then estimate the positions of the spectrum peaks, which constitute a set of (ψ^M, ψ^A) pairs that carry information about the locations and tracks of the N moving targets. Then, in Sec. 4, we show how this set of pairs can subsequently be used to track the N moving targets in the area.

Spectral content estimation of time or space signals is a well-explored problem in the literature, and several methods have been proposed to this end. Examples of these methods include, but are not limited to, Fourier Transform [24], Multiple Signal Classification (MUSIC) [20], and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [19]. In this paper, we propose to use 2D MUSIC spectral estimation for the problem of Sec. 3.2, due to its simplicity and high-resolution capability. Another advantage of using MUSIC for the joint estimation of parameters is that the resolvability of paths in each dimension depends on the length of the arrays in both dimensions [25]. For instance, while a longer time window better resolves paths in the dimension of time, it can also help resolve paths in the dimension of space, i.e. paths that have the same ψ^M but different ψ^A . This is particularly crucial in

the context of multi-target tracking, since we need to clearly distinguish the peaks in the spectrum, whereas for the case of single target tracking, one would only be concerned with the location of the single-largest peak. In our framework of multi-person tracking, we are then interested in the joint estimation of parameters (ψ^M, ψ^A) from the multi-dimensional signal model shown in equations 5 and 6. We next show how we can utilize 2D MUSIC for a magnitude-based signal model in order to estimate the spectrum and the corresponding peaks.

Consider the scenario where a receiver array contains M_A antennas with inter-antenna spacing of d_{ant} . The antennas of the array sample the received signal at a rate of $1/T_s$ samples/sec for a duration T_{win} . The number of samples in space and time are thus M_A and $M_T = \lfloor T_{\text{win}}/T_s \rfloor$, respectively. Denote by \mathbf{C} the $M_A \times M_T$ matrix of magnitude-squared measurements in the spatio-temporal window:

$$\mathbf{C} = \begin{bmatrix} |c_{1,1}|^2 & |c_{1,2}|^2 & \dots & |c_{1,M_T}|^2 \\ |c_{2,1}|^2 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ |c_{M_A,1}|^2 & \dots & \dots & |c_{M_A,M_T}|^2 \end{bmatrix}, \quad (8)$$

where $c_{i,j} = c((i-1)d_{\text{ant}}, (j-1)T_s)$ is the measured 2D received signal described in Eq. 5.

In order to estimate the 2D spectral content of the measurements in \mathbf{C} , we define the steering vector $s(\psi^M, \psi^A)$ as shown in Eq. 7. Then, it is straightforward to show that the vectorized form of \mathbf{C} can be written in terms of the steering vectors of the paths arriving at the Rx array as follows:

$$\tilde{\mathbf{C}} = \mathbf{S}\mathbf{A} + \tilde{\eta}, \quad (9)$$

where $\tilde{(\cdot)}$ denotes the vectorized form of a matrix, \mathbf{S} is an $M_A M_T \times N$ matrix whose n^{th} column is $s(\psi_n^M, \psi_n^A)$, and $\mathbf{A} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$. The MUSIC algorithm calculates the eigen-decomposition of the correlation matrix R_c of the measurement vector $\tilde{\mathbf{C}}$ [20],

$$R_c = E\{\tilde{\mathbf{C}}\tilde{\mathbf{C}}^H\} = \mathbf{S}\mathbf{R}_A\mathbf{S}^H + R_\eta, \quad (10)$$

where $R_A = E\{\mathbf{A}\mathbf{A}^H\}$, $R_\eta = E\{\eta\eta^H\}$, and $E\{\cdot\}$ is the expectation operator. It can be shown that the eigenvectors of R_c are divided into bases of a *signal subspace*, whose dimension is equal to the rank of R_A , and bases of a *noise subspace*, which is orthogonal to all the steering vectors corresponding to the N signal paths arriving at the receiver array. Therefore, we can define a pseudospectrum $P(\psi^M, \psi^A)$ as

$$P(\psi^M, \psi^A) = \frac{1}{s^H(\psi^M, \psi^A)E_N E_N^H s(\psi^M, \psi^A)}, \quad (11)$$

where E_N is a matrix whose columns constitute the bases for the noise subspace. $P(\psi^M, \psi^A)$ peaks at the locations of (ψ_n^M, ψ_n^A) , $n =$

$1, \dots, N$, since the steering vectors corresponding to these locations are orthogonal to the noise subspace E_N . Hence, extracting the locations of the peaks of $P(\psi^M, \psi^A)$ provides the required (ψ_n^M, ψ_n^A) pairs needed for tracking the N targets.

A critical assumption in the MUSIC algorithm is that the matrix R_A is full rank, i.e., all the different N signals are uncorrelated. Such an assumption is not valid in many practical scenarios where scattering and multipath propagation are involved. Then, in order to uncorrelate the signals, spatial smoothing is a technique commonly used in the literature [5]. In spatial smoothing, the correlation matrix R_c is calculated by averaging the correlation matrices of different subsets of the antenna array, given that each of the subsets is a set of contiguous antennas. Then, to address the correlation of signals in our 2D framework, we extend spatial smoothing to *spatio-temporal smoothing MUSIC* for our scenario. We divide the matrix \mathbf{C} into overlapping sub-matrices \mathbf{C}^{sub} of size $M_A^{\text{sub}} \times M_T^{\text{sub}}$ each. The correlation matrix R_c is then calculated as the average of the correlation matrices R_c^{sub} of the sub-matrices \mathbf{C}^{sub} . Similar spatio-temporal smoothing techniques have been proposed for the JADE MUSIC problem in the literature [2].

After computing the pseudospectrum $P(\psi^M, \psi^A)$, we next find the locations of the peaks of the pseudospectrum as

$$\Psi = \{\psi_j = (\psi_j^M, \psi_j^A), j = 1, \dots, J\},$$

where J is the number of detected peaks in the pseudospectrum. As we shall see next, this information is then used to estimate the tracks of the N targets.

4 MULTIPLE TARGET TRACKING

In this section, we show how we can use the extracted information from the 2D spectrum to track multiple targets. In order to extract the information about the targets' locations and headings at time t , we apply the aforementioned 2D spatio-temporal smoothing MUSIC algorithm on the data $|c(t, d)|^2$ in a time window of duration T_{win} starting at time t , to extract the set of peaks Ψ_t at time t . We first list the problems that arise when relying directly on Ψ_t (with cardinality J_t) for tracking the N targets. Then, we present our solutions to overcome these problems and reconstruct the targets' tracks using Ψ_t .

Two main problems arise when using Ψ_t for tracking:

- **Ambiguity:** As previously mentioned, while the 2D joint parameter estimation resolves a few ambiguities that exist when estimating each parameter individually, the pair of (ψ^M, ψ^A) does not give sufficient information about the location of the target that resulted in a particular measurement. For instance, Fig. 4 shows an example of two different valid solutions to a target's location and heading for a measurement of $\psi^M = 0.187$

$$s(\psi^M, \psi^A) = \left[\underbrace{1, e^{-\frac{j2\pi}{\lambda}\psi^A d_{\text{ant}}}, \dots, e^{-\frac{j2\pi}{\lambda}\psi^A (M_A-1)d_{\text{ant}}}}_{\text{array measurements at } t=0}, \underbrace{e^{-\frac{j2\pi}{\lambda}\psi^M T_s}, e^{-\frac{j2\pi}{\lambda}(\psi^M T_s + \psi^A d_{\text{ant}})}, \dots, e^{-\frac{j2\pi}{\lambda}(\psi^M T_s + \psi^A (M_A-1)d_{\text{ant}})}}_{\text{array measurements at } t=T_s}, \dots, \underbrace{e^{-\frac{j2\pi}{\lambda}\psi^M (M_T-1)T_s}, e^{-\frac{j2\pi}{\lambda}(\psi^M (M_T-1)T_s + \psi^A d_{\text{ant}})}, \dots, e^{-\frac{j2\pi}{\lambda}(\psi^M (M_T-1)T_s + \psi^A (M_A-1)d_{\text{ant}})}}_{\text{array measurements at } t=(M_T-1)T_s} \right]^T \quad (7)$$

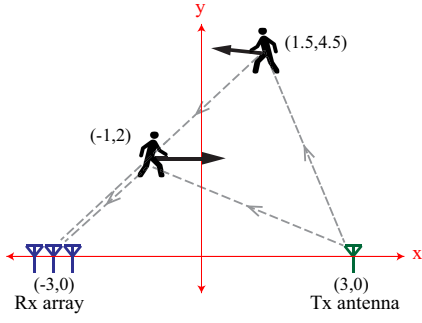


Figure 4: Example of ambiguity resulting from the measurement of $\psi = (0.187, 0.707)$. Two targets result in the same measurement: one at location $(-1, 2)$ with heading of 0° , the other at location $(1.5, 4.5)$ with heading of 173° . Both targets have a speed of 1 m/s.

and $\psi^A = 0.707$, thus showing the ambiguity prevalent in each (ψ^M, ψ^A) measurement.

- **Association:** At each time instant, we extract a set of J_t measurements from the 2D spectrum. However, we lack the knowledge of the subset of these J_t measurements that are actual detections from the moving targets, and the complimentary subset of false alarms. Furthermore, for the subset of actual detections, we would need an association profile of which detections correspond to which targets. Such an issue does not arise and is thus not addressed in a single target tracking framework. Thus, the methods proposed for single-target tracking cannot be directly utilized for multi-target tracking in this paper.

In order to overcome these problems, we exploit the fact that the targets are moving and model the measurements associated with the targets' motion as a nonlinear dynamical system [12]. We further utilize a Particle Filter (PF) with a Joint Probabilistic Data Association Filter (JPDAF) to solve this dynamical system and obtain an estimate for the track of each target, as we shall see next.

Consider the scenario where a Tx is located at (x_T, y_T) and a Rx array is centered at (x_R, y_R) such that its array axis is parallel to the x-axis, as shown in Fig. 2. We define the state of the n^{th} target at time t as a 4-dimensional vector \mathbf{x}_t^n that carries information about the target's location, heading, and speed. More specifically, $\mathbf{x}_t^n = [x_n(t), y_n(t), \theta_n(t), v_n(t)]^T$, where $x_n(t), y_n(t)$ define the location of the n^{th} target at time t , $\theta_n(t)$ is its direction of motion, measured with respect to the x-axis, and $v_n(t)$ is its speed. Furthermore, we define a measurement process $\psi_n(t)$ as the pair $(\psi_n^M(t), \psi_n^A(t))$, which can be related to the target's state as follows:

$$\begin{aligned} \psi_n^M(t) &= v_n(t) \left(\frac{(x_R - x_n(t)) \cos(\theta_n(t)) + (y_R - y_n(t)) \sin(\theta_n(t))}{\sqrt{(x_R - x_n(t))^2 + (y_R - y_n(t))^2}} \right) \\ &+ v_n(t) \left(\frac{(x_T - x_n(t)) \cos(\theta_n(t)) + (y_T - y_n(t)) \sin(\theta_n(t))}{\sqrt{(x_T - x_n(t))^2 + (y_T - y_n(t))^2}} \right) \\ &+ \eta_M(t), \end{aligned} \quad (12)$$

and

$$\psi_n^A(t) = 1 - \left(\frac{x_n(t) - x_R}{\sqrt{(x_R - x_n(t))^2 + (y_R - y_n(t))^2}} \right) + \eta_A(t), \quad (13)$$

where η_M and η_A are measurement noise processes with variances $\sigma_{\eta_M}^2$ and $\sigma_{\eta_A}^2$, respectively. On the other hand, we assume a simple motion dynamics model for the targets, in which a target maintains the same direction of motion with probability P_c , and occasionally changes that direction with probability $1 - P_c$. More specifically, we assume the state of the n^{th} target evolves with time according to the model $\mathbf{x}_{t+1}^n = g_n(\mathbf{x}_t^n)$ as follows:

$$\begin{aligned} x_n(t+1) &= x_n(t) + v_n(t) \cos(\theta_n(t)) + \eta_{x_n}(t+1), \\ y_n(t+1) &= y_n(t) + v_n(t) \sin(\theta_n(t)) + \eta_{y_n}(t+1), \\ \theta_n(t+1) &= \eta_{\theta_n}(t+1) + \begin{cases} \theta_n(t) & \text{w.p. } P_c \\ \sim \mathcal{U}(0, 2\pi) & \text{w.p. } 1 - P_c \end{cases}, \\ v_n(t+1) &= v_n(t) + \eta_{v_n}(t+1), \end{aligned} \quad (14)$$

where $\eta_{x_n}, \eta_{y_n}, \eta_{\theta_n}$, and η_{v_n} are all dynamics noise processes with variances $\sigma_{\eta_{x_n}}^2, \sigma_{\eta_{y_n}}^2, \sigma_{\eta_{\theta_n}}^2$, and $\sigma_{\eta_{v_n}}^2$, respectively, and $\mathcal{U}(0, 2\pi)$ is the uniform distribution in the interval $[0, 2\pi)$.

For the estimation of the state of the n^{th} target \mathbf{x}_t^n at time t , we propose to compute the filtering Probability Density Function (PDF) $p(\mathbf{x}_t^n | \Psi_{1:t})$ of the n^{th} target's state at time t given all the measurements up to time t . Then, we use the mean of this PDF as the estimate of the target's state $\hat{\mathbf{x}}_t^n = E\{\mathbf{x}_t^n | \Psi_{1:t}\}$. To this end, we propose to use a Particle Filter (PF) for the computation of the filtering PDF of the n^{th} target [21]. The underlying principle of PFs is that they approximate any probability distribution using samples (or particles) drawn from that distribution. Such a representation is favorable in many scenarios, especially when nonlinear random variable transformations are involved. The steps of the PFs used in our problem are summarized in Algorithm 1. The PF for the n^{th} target starts by drawing a total of I samples/particles $\mathbf{x}_1^{[i,n]}, i = 1, \dots, I$ from an initial distribution $\chi_1^n(\mathbf{x}_1^n)$, which can depend on any prior information we have about the initial state of the n^{th} target. Then, these particles are given *importance weights* $w_1^{[i,n]}$ which represent how well they fit the current set of measurements Ψ_1 (step 4 in Algorithm 1). However, the aforementioned association problem hinders the completion of this step, since the PF lacks the knowledge of which of the measurements in Ψ_1 was generated by the n^{th} target. To overcome this, we propose to utilize a Joint Probabilistic Data Association Filter (JPDAF) to calculate the importance weights [22]. We will discuss the details of the JPDAF later in this section. After the importance weights are calculated, a *resampling* step (step 9) is performed in order to neglect the low-weight particles and retain particles that have a high probability of producing the current measurement set. The resampled particles then evolve according to the motion model in Eq. 14 and the whole process is repeated for consecutive time instants. More details on PFs can be found in [21].

The JPDAF, on the other hand, deals with the problem of associating measurements to targets. Consider the set of measurements $\Psi_t = \{\psi_j, j = 1, \dots, J_t\}$ measured at time t . Some of these measurements can be false alarms that are not associated with any target, arising due to the modeling errors. We denote the probability of such false alarms as P_{FA} . Furthermore, some target measurements can be missing from the set Ψ_t , for instance, due to blockage by other targets. We denote the probability of a target miss as $1 - P_D$, where P_D is the detection probability. The underlying principle

of the JPDAF is then to calculate the probabilities of all possible *association profiles* given the current set of measurements and particles [22]. An association profile ω matches each target to one of the J_t measurements. In other words, an association profile ω is a set of N pairs (k, l) where $l = 1, 2, \dots, N$, $k \in \{0, 1, \dots, J_t\}$, and a pair (k, l) represents assigning the measurement ψ_k to the l^{th} target.¹ Afterwards, the probability of the n^{th} target generating the measurement ψ_j can be computed by summing the probabilities of all the association profiles which assign the measurement ψ_j to the n^{th} target. We denote the set of all such association profiles by Ω_{jn} , i.e., $\Omega_{jn} = \{\omega; (j, n) \in \omega\}$. The details of the JPDAF calculation of the importance weights are shown in Algorithm 2.

REMARK 2. Note that in the case of tracking one person, we still utilize the JPDAF in the calculation of the particle weights in the PF. In such a case, the main function of the JPDAF is to distinguish false alarm measurements from the actual measurement corresponding to the target's motion.

Algorithm 1 Particle Filter for Motion Tracking

Input: Total tracking time T , Number of particles I , Number of moving people N , Measurements $\Psi_{1:T}$

Output: Estimate of the target states $\hat{\mathbf{x}}_{1:T}^n$, $n = 1, 2, \dots, N$

- 1: Initialize $t = 1$
 - 2: **for** $1 \leq n \leq N$ **do**
 - 3: Sample $\mathbf{x}_1^{[i,n]} \sim \chi_1^n(\mathbf{x}_1^n)$ for $i = 1, 2, \dots, I$
 - 4: **end for**
 - 5: Compute the importance weights $\tilde{w}_1^{[i,n]}$ using the JPDAF in Algorithm 2, and normalize $w_1^{[i,n]} = \frac{\tilde{w}_1^{[i,n]}}{\sum_{i=1}^I \tilde{w}_1^{[i,n]}}$
 - 6: Estimate the initial state of the n^{th} target as $\hat{\mathbf{x}}_1^n = E\{\mathbf{x}_1^n | \Psi_1\} = \sum_{i=1}^I w_1^{[i,n]} \mathbf{x}_1^{[i,n]}$
 - 7: **for** $2 \leq t \leq T$ **do**
 - 8: **for** $1 \leq n \leq N$ **do**
 - 9: Sample $\tilde{\mathbf{x}}_{t-1}^{[i,n]}$, for $i = 1, \dots, I$, from the distribution defined by $p(\tilde{\mathbf{x}}_{t-1}^{[i,n]} = \mathbf{x}_{t-1}^{[i,n]} | \Psi_{1:t-1}) = w_{t-1}^{[i,n]}$
 - 10: Sample $\mathbf{x}_t^{[i,n]} \sim g_n(\tilde{\mathbf{x}}_{t-1}^{[i,n]})$
 - 11: **end for**
 - 12: Compute the importance weights $\tilde{w}_t^{[i,n]}$ using the JPDAF in Algorithm 2, and normalize $w_t^{[i,n]} = \frac{\tilde{w}_t^{[i,n]}}{\sum_{i=1}^I \tilde{w}_t^{[i,n]}}$
 - 13: Estimate the state of the n^{th} target as $\hat{\mathbf{x}}_t^n = \sum_{i=1}^I w_t^{[i,n]} \mathbf{x}_t^{[i,n]}$
 - 14: **end for**
-

5 EXPERIMENTAL RESULTS

In this section, we present the experimental results of our proposed magnitude-based framework for multi-person tracking, using WiFi CSI magnitude measurements from one side of the area. We first discuss our experimental setup and the practical considerations that arise in these experiments. We then show the performance

¹Note that the pair $(k = 0, l)$ represents the case of no measurement associated to the l^{th} target, which can happen with probability $(1 - P_D)$, where P_D is the detection probability.

Algorithm 2 Joint Probabilistic Data Association Filter for Particle Weight Calculation

Input: All current particles $\mathbf{x}^{[i,n]}$, current measurement set Ψ , P_D, P_{FA}

Output: The particles' importance weights $\tilde{w}^{[i,n]}$

- 1: Calculate the number of current measurements $J = |\Psi|$
- 2: Calculate $\gamma_j^{[i,n]} = p(\psi_j | \mathbf{x}^{[i,n]})$, which denotes the probability of the measurement ψ_j being generated by the n^{th} target having a state $\mathbf{x}^{[i,n]}$, according to Eq. 12 and Eq. 13
- 3: Generate all possible association profiles ω , where $\omega = \{(k, l); k \in \{0, 1, \dots, J\}, l = \{1, \dots, N\}\}$, and (k, l) is a pair assigning the measurement ψ_k to the l^{th} target
- 4: Calculate the probability of each association profile as

$$p(\omega | \Psi) = P_{FA}^{|\omega|} P_D^{|\omega| - |\omega_o|} (1 - P_D)^{|\omega_o|} \prod_{\substack{(k,l) \in \omega \\ k \neq 0}} \frac{1}{I} \sum_{i=1}^I p(\psi_k | \mathbf{x}^{[i,l]}) \quad (15)$$

where ω_o is a subset of ω with targets not being assigned to any of the measurements, i.e., $\omega_o = \{(k, l); (k, l) \in \omega, k = 0\}$

- 5: Calculate the probability that a measurement ψ_j is caused by the n^{th} target β_{jn} by summing over all association profiles making such an assignment,

$$\beta_{jn} = \sum_{\omega \in \Omega_{jn}} p(\omega | \Psi) \quad (16)$$

- 6: Calculate the importance weights

$$\tilde{w}^{[i,n]} = \frac{1}{\sum_{j=0}^J \beta_{jn}} \left(\beta_{0n} + \sum_{j=1}^J \beta_{jn} p(\psi_j | \mathbf{x}^{[i,n]}) \right) \quad (17)$$

of our tracking framework through extensive experiments (40 in total) carried out in six different environments, with various levels of clutter. Finally, we discuss the impact of several experimental parameters on the results, and compare with the state-of-the-art tracking algorithms.

5.1 Experimental Setup

For the data collection process, we use laptops with Intel 5300 WiFi NICs for both transmission and reception. For the Tx, a tripod-mounted antenna is connected to one port of an Intel card that broadcasts WiFi packets on channel 36 in the 5 GHz band. We then use the WiFi cards of three laptops as receivers, with each WiFi card providing two antenna ports. In other words, we use WiFi NICs of three laptops and connect two WiFi ports of each laptop to two antennas mounted on a tripod, as shown in Fig. 5.² The 3 Rx WiFi NICs log the packets transmitted on the WiFi channel. We then process the measured data offline using Csitool [10] to extract the CSI measurements and track the moving subjects. As previously

²Note that while each Intel 5300 NIC has 3 antenna ports available, we observed that the signal on port 3, which is located between port 1 and 2 on the NIC, is sometimes corrupted due to crosstalk (as is reported by other users [8]). Hence, we use only ports 1 and 2 on each WiFi Rx. In the future, if one could obtain clean measurements on all the three ports, then one would only need 2 Rx laptops with Intel 5300 NICs to achieve the results of this paper.

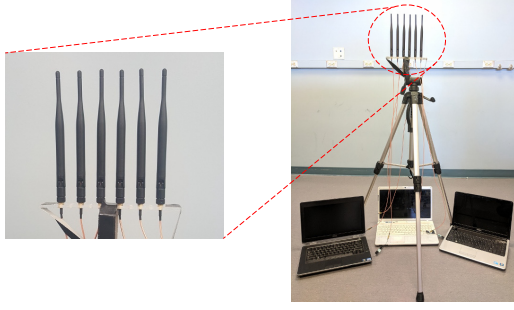


Figure 5: Receiver setup: WiFi cards of 3 laptops are used, resulting in 6 total antennas that we space $\lambda/2$ apart on a tripod as shown.

mentioned, since we rely only on the magnitude of the CSI measurements, the Rx NICs do not need any phase synchronization. Thus, our proposed framework is also flexible to facilitate further addition of antennas to the array as needed, without any additional calibrations. We next discuss some practical considerations that arise in our experiments.

- **Spatio-temporal sampling rates:** As shown in Sec. 3, the reflected signal from the n^{th} person results in a peak in the 2D spectrum at $(\psi_n^M, \psi_n^A) = (v_n(\cos \phi_n^R + \cos \phi_n^T), 1 - \cos \phi_n)$. Hence, the maximum frequency content for f_t and f_d are $\frac{2v_{\max}}{\lambda}$ and $2/\lambda$ respectively, where v_{\max} is the maximum possible human walking speed. Then, according to the Nyquist sampling theorem, the sampling rates for the 2D received signal in time and space should be greater than $\frac{4v_{\max}}{\lambda}$ and $4/\lambda$ respectively. In the temporal dimension, we set the sampling rate to 1000 packets/sec, which is much higher than the required sampling rate of 139 packets/sec (assuming a v_{\max} of 2 m/s). However, for the spatial dimension, fixing the antennas $\lambda/4$ meters (1.45 cm) apart is difficult due to the relatively large physical dimensions of the antennas. Hence, we place the antennas $\lambda/2$ apart, which leads to aliasing in the f_d dimension of the spectrum. In order to overcome such aliasing effect, we propose to place the Rx array in a corner of the tracking area, so that the ϕ_n s for all the targets are less than 90° . Hence, the maximum possible value of f_d in this case is $1/\lambda$, and such a $\lambda/2$ -spaced array configuration does not suffer from aliasing problems.
- **Data clean-up process:** Raw CSI measurements on commodity WiFi cards can suffer from noise due to the internal state transitions in the Tx and Rx WiFi NICs [27]. To reduce the noise in the raw CSI measurements, we utilize two denoising schemes.
 - (1) Principal Component Analysis (PCA): The Intel 5300 NIC reports CSI measurements on 30 different subcarriers. It has been shown in [27] that the changes in CSI due to human movements on different subcarriers are correlated. Hence, the reflected signal can be separated from noise by performing PCA on the data from the 30 subcarriers.
 - (2) Wavelet denoising: Discrete Wavelet Transform (DWT)-based noise suppression techniques have been shown to outperform traditional denoising schemes such as band-pass filters [6]. Hence, we apply wavelet denoising on the PCA-denoised signal in order to suppress residual noise.
- **2D MUSIC parameters:** We choose the array parameters of the 2D MUSIC algorithm described in Sec. 3 as follows: $T_{\text{win}} =$

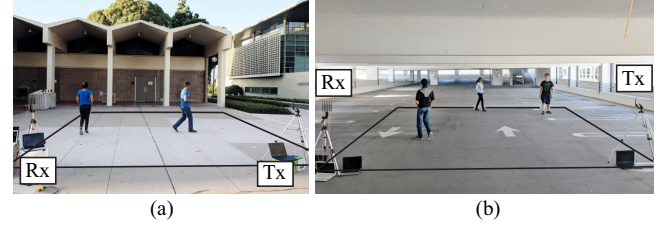


Figure 6: Tracking experimental setup in outdoor areas in (a) an open area and (b) a closed parking lot. The boundaries of the workspace are marked with a solid black line.

0.5s, $T_{\text{win}}^{\text{sub}} = 0.25\text{s}$, $M_A = 6$, and $M_A^{\text{sub}} = 5$. Note that a small T_{win} implies that people can take any track in our framework and are not limited to walk on straight lines. In order to detect peaks in the pseudospectrum, we define a peak as a point in the pseudospectrum whose value is greater than its neighbors, and greater than an empirically predefined threshold $p_{\text{th}} = 0.6 \times P_{\min, \text{dB}}$, where $P_{\min, \text{dB}}$ is the minimum value in the normalized pseudospectrum in dB (with the maximum value being 0 dB in the normalized pseudospectrum).

- **Particle filter parameters:** In order to set the parameters of the PF, we collect a few prior measurements (not in the same area of the experiment) and estimate the values for the noise variances and probabilities of detection and false alarms. These parameters are then used in all the different experiments in different areas. The parameters are then set as follows: $\sigma_{\eta_M} = 0.1$, $\sigma_{\eta_A} = 0.07$, $\sigma_{\eta_{x_n}} = \sigma_{\eta_{y_n}} = 1 \text{ cm}$, $\sigma_{\eta_{\theta_n}} = 1^\circ$, $\sigma_{\eta_{v_n}} = 2.5 \times 10^{-3}$, $P_c = 0.9$, $P_D = 0.85$, $P_{FA} = 0.25$ for outdoor areas, $P_{FA} = 0.35$ for indoor areas, and $I = 5000$. Note that the probability of false alarm is higher in indoor environments due to the stronger multipath.

5.2 Tracking Results

In this section, we show how our proposed framework can track multiple moving people in an area, based on only the WiFi CSI magnitude measurements of 3 laptops that are located on one side of the area. We carry out tracking experiments in six different environments, with up to three people walking simultaneously in the area. We categorize the areas into outdoor and indoor scenarios. Fig. 6 shows the outdoor areas, where Fig. 6 (a) is an open area with minimal clutter, and Fig. 6 (b) is a parking lot which has considerable multipath due to the walls and the low ceiling beams. The top row of Fig. 8 then shows some of the indoor areas, which are more challenging than the outdoor areas due to higher extent of clutter (e.g. furniture, walls) and the resulting multipath. In all experiments, we ask the subjects to walk on predefined tracks defined by floor markers in a $7 \text{ m} \times 7 \text{ m}$ area, and time-stamp their motion at the markers in order to obtain the ground-truth locations of the subjects. Furthermore, since we cannot know the exact point of reflection on the person's body at which the signal bounces off at each time instant, we approximate a person as a cylindrical object of radius 25 cm. We then calculate the tracking error, at any time instant, as the minimum distance between the estimated location and the surface of that cylinder. Such a method of error calculation has previously been adopted in similar contexts in the literature [12, 16].

Outdoor Tracking: In this section, we show our tracking results for the outdoor areas shown in Fig. 6. The first location, shown in

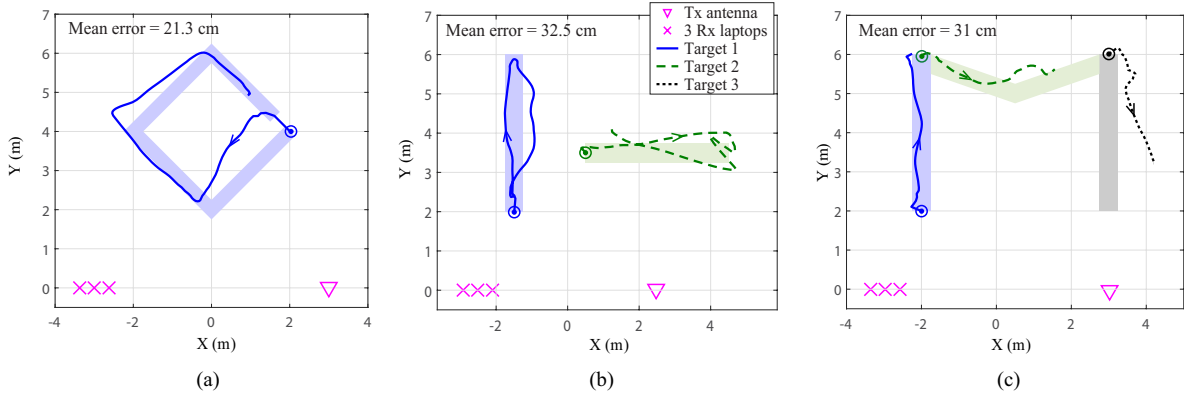


Figure 7: Sample multi-person tracking results in the outdoor areas shown in Fig. 6 – (a) One person walking along a diamond-shaped route in the area of Fig. 6b, (b) two persons walking back and forth on perpendicular straight lines in the area of Fig. 6a, and (c) three persons walking on different parts of an M-shaped route in the area of Fig. 6b. The light background patches represent the actual tracks, while the \odot symbols represent their starting points.

Fig. 6 (a), is a relatively open area with little to no clutter, resulting in minimal multipath. On the other hand, the second location, shown in Fig. 6 (b), is a parking structure where the walls and ceiling beams generate considerable multipath. In both cases, the Tx antenna and the 3 Rx laptops are fixed to the corners on one side of the $7\text{ m} \times 7\text{ m}$ area of interest, as shown in Fig. 6. Overall, we ran 17 experiments in these 2 areas of Fig. 6, with 1, 2, and 3 people on several different days, walking in different paths. In all the experiments, we initialize the PF with particles that are uniformly distributed in a $3\text{ m} \times 3\text{ m}$ square around the locations where the targets start their motion. Fig. 7 then shows a few sample results of our tracking framework for these two areas. It can be seen that our proposed framework estimates the track of the people with a high accuracy in all the cases. Overall, we achieve a mean tracking error of 38 cm (median of 29 cm) when considering all the 17 experiments.

Indoor Tracking: In this section, we show our tracking results for the indoor areas of conference rooms, a classroom, and a lounge area, shown in the top row of Fig. 8. In all the locations, the walls, ceiling, and furniture constitute clutter which makes the effect of multipath more significant. While we can remove the effect of the static multipath by subtracting the temporal mean of the received signal as described in Remark 1, higher order reflections involving both a moving target and a static object, although weaker, still affect the received signal, and consequently the 2D spectrum. This results in a higher number of false alarms as mentioned in Sec. 5.1.

Similar to the outdoor areas, we fix the Tx antenna and the 3 Rx laptops to the corners on one side of the area of interest. We also initialize the PF with particles that are uniformly distributed in a $3\text{ m} \times 3\text{ m}$ square around the locations where the targets start their path. Overall, we ran 23 experiments in 4 different areas (the three areas shown in Fig. 8 and one additional conference room) with 1, 2, and 3 people on several different days, walking in different paths. The bottom row of Fig. 8 then shows a few sample tracking results for these locations. It can be seen that our proposed magnitude-based framework achieves a good accuracy of tracking multiple people in indoor environments as well, with an overall mean tracking error of 55 cm (median of 39 cm) across all the 23 different experiments. It should be noted that we do not utilize any information about the clutter (e.g., the furniture) in the track

estimation framework. If the information about the locations of the furniture was known apriori, it can be used, for example, to prohibit any particles in the PF from appearing on their locations, thereby improving the track estimation accuracy.

5.3 Discussion

In this section, we investigate the impact of some of the experimental parameters on the performance of our proposed tracking framework, and compare the performance of our proposed framework with the state-of-the-art.

Effect of environment: As previously mentioned, indoor environments are more challenging than outdoor ones because of the stronger multipath resulting from the clutter. The noisier spectrum and higher false alarm probability affect the performance of the tracking framework and increase the tracking error. To quantify this effect, Fig. 9 (a) shows the Cumulative Distribution Function (CDF) of the tracking error for both indoor and outdoor environments. As can be seen, the performance in indoor environments is slightly worse compared to outdoor ones, as expected. More specifically, the indoor environments have an overall mean tracking error of 55 cm, in comparison to 38 cm for outdoor areas.

Effect of the number of people: We also test the performance of our tracking framework by varying the number of people being tracked. Fig. 9 (b) shows the CDF curves of the tracking error for different number of people. It can be seen that the performance is comparable in all the cases of tracking 1 to 3 people. While tracking multiple people, there is a higher chance that a measurement corresponding to one of the persons disappears momentarily, if that person is blocked by other people. However, the JPDAF (with an appropriate P_D setting) accounts for that missing detection and keeps track of the blocked target, thereby preserving the accuracy of the framework even in the presence of blocking effects.

Effect of the closeness to Tx or Rx: Fig. 9 (c) shows the box-plot distribution of the point-wise tracking error of our framework over all the tracks in all the 40 experiments, as a function of the logarithm of the ratio between the distance of the target to the Tx and its distance to the Rx. Negative values to the left side of the figure correspond to targets that are closer to the Tx than the Rx, while positive values to the right correspond to targets that are

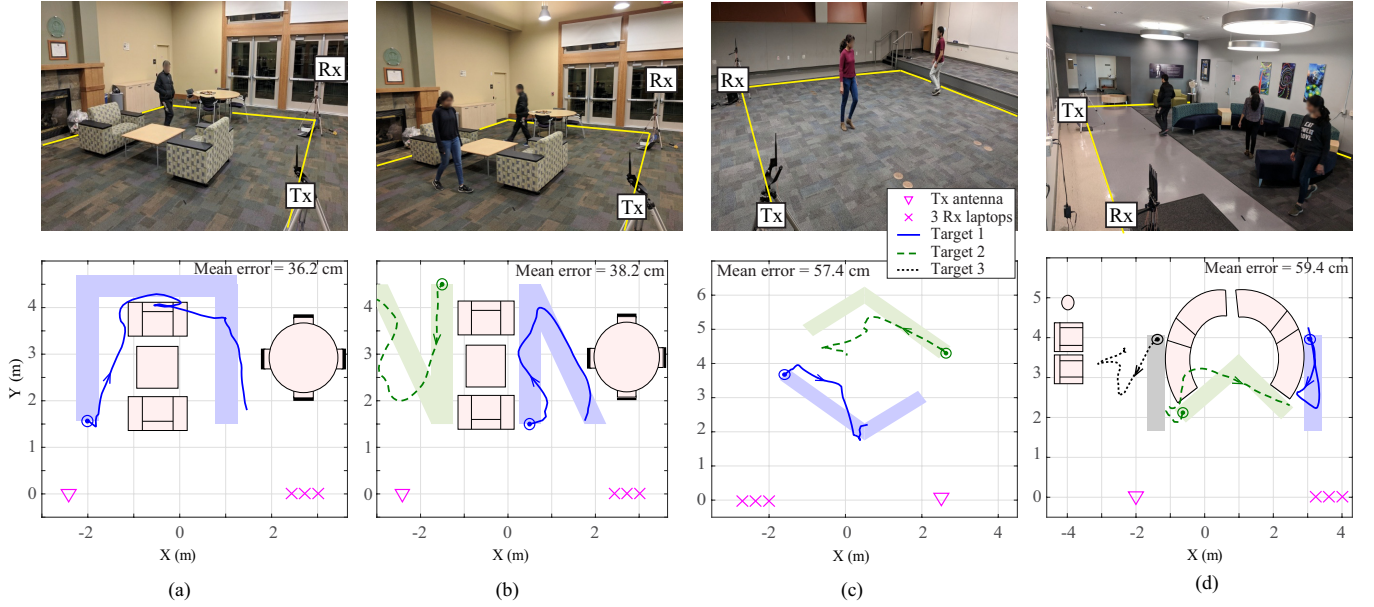


Figure 8: (Bottom) Sample multi-person tracking results in (top) corresponding indoor areas with various degrees of clutter – (a) One person walking along a U-shaped route in an area including tables, chairs, and futons, (b) two persons walking along two V-shaped routes in the same area, (c) two persons walking along two checkmark-shaped routes in a classroom, and (d) three persons walking on different lines in an area containing multiple chairs, sofas, and light fixtures, where the targets 1 and 3 walk in a back-and-forth fashion along the marked route. The yellow lines in the area pictures represent the tracking area boundary. The light background patches on the figures represent the actual tracks, while the \odot symbols represent their starting points.

closer to the Rx. It can be seen that the error tends to be lower when the target is closer to the Rx, since the reflections off of the target's body are more likely to reach the Rx array. On the other hand, targets farther away from the Rx would be scattering in different directions and the reflections are less likely to reach the Rx array. Hence, if the antenna dimensions of the Rx permit placing them $\lambda/4$ apart, it is recommended to place the Rx array in the midpoint of the link side of the tracking area (see Sec. 5.1).

Effect of the distance between targets: Fig. 9 (d) shows the box-plot distribution of the point-wise tracking error of our framework as a function the distance between the targets in all the multi-target tracking experiments. It can be seen that such a distance has little to no effect on the tracking performance of our framework. This is primarily because a small distance between two targets does not imply that they are indistinguishable in the measurement domain (ψ_M, ψ_A), since two close targets with different moving directions have different ψ_M s.

Comparison to the state-of-the-art: Table 1 shows the tracking accuracy of the state-of-the-art as well as our framework. It can be seen that our framework achieves a decimeter-level tracking accuracy that is comparable to the state-of-the-art, for both single and multiple target tracking, but without requiring any extra bandwidth or several transceivers that were previously required for multiple target tracking, or phase measurements that were previously required for single target tracking.

6 LIMITATIONS AND FUTURE EXTENSIONS

In this paper, we have proposed a new framework for multi-person tracking using only WiFi magnitude measurements from a WiFi

link on only one side of the tracking area. We next discuss some of the limitations and possible future extensions of this work:

Assuming knowledge of the number of people: As is common in the multi-person tracking literature, our framework assumes the knowledge of the number of people in the area, in order to initialize the same number of PFs. In practical scenarios, such an assumption can be realized by monitoring the entrance of the area of interest (such as the door(s) in the room), and initializing a new particle filter with particles around the entrance area whenever an entry is detected. However, for scenarios where such an initialization is not possible, we plan to further incorporate an occupancy estimation algorithm to estimate the number of people in the area.

Tracking through walls: Our proposed tracking framework assumes that the Tx and Rx are in the same room/area as the people. Tracking through walls or other static obstacles is challenging, since the signal undergoes various transformations as it passes through objects. Furthermore, we do not collect any prior measurements in the same area in the absence of targets, which considerably increases the complexity of the through-wall tracking problem. As part of future work, we plan to model these through-wall propagation artifacts and account for them in the tracking pipeline.

7 CONCLUSIONS

In this paper, we have considered the problem of passively tracking multiple persons using only WiFi magnitude measurements on a small number of WiFi receivers, located on one side of the tracking area. We have proposed a framework based on the joint estimation of multi-dimensional parameters of the received WiFi signal. More specifically, our framework jointly estimates the angles-of-arrival from the targets to the receiver array, as well as the parameters of

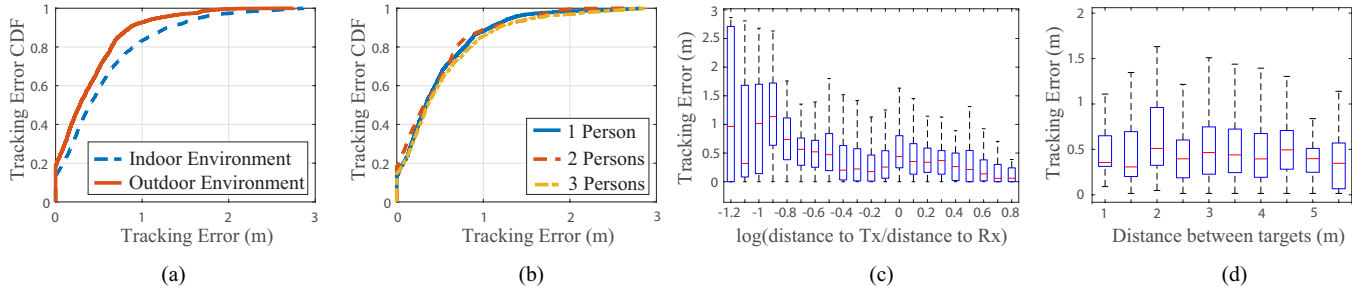


Figure 9: Tracking error analysis over 40 different experiments in 6 different areas (five area pictures shown in this paper and one additional indoor area not shown) and various tracking routes. (a) CDF of tracking errors in outdoor vs indoor environments from tracking 1, 2, and 3 people walking on different tracks, on different days. Performance is better in outdoor environments due to less multipath, as expected. (b) CDF of tracking errors for different number of people. Comparable performance is seen for all cases of 1, 2, or 3 people. (c) Box plot of the distribution of point-wise tracking error over all the experiments as a function of the logarithm of the ratio between the distance of the target to the Tx and its distance to Rx. Targets closer to Rx tend to have lower errors. (d) Box plot of the distribution of point-wise tracking error as a function of the inter-target distance showing little to no effect.

the arrays induced by the targets' motion. Furthermore, we have utilized Particle Filters and Joint Probabilistic Data Association Filter (JPDAF) in order to associate the estimated parameters to the targets in the area, and consequently reconstruct the tracks of these targets. We have experimentally validated our framework through extensive experiments (total of 40) in six different environments (indoor and outdoor). Our experimental results show high tracking accuracy with a mean tracking error of only 38 cm in outdoor areas/closed parking lots, and 55 cm in indoor areas.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers and the shepherd for their valuable comments and helpful suggestions. This work is funded in part by NSF CCSS award # 1611254 and in part by NSF NeTS award # 1816931.

REFERENCES

- [1] F. Adib, Z. Kabelac, and D. Katabi. 2015. Multi-Person Localization via RF Body Reflections. In *USENIX NSDI*. 279–292.
- [2] A. Bazzi, D. Slock, and L. Meilhac. 2016. On spatio-frequency smoothing for joint angles and times of arrival estimation of multipaths. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3311–3315.
- [3] M. Bocca, O. Kallio, N. Patwari, and S. Venkatasubramanian. 2014. Multiple target tracking with RF sensor networks. *IEEE Transactions on Mobile Computing* 13, 8 (2014), 1787–1800.
- [4] S. Chang, R. Sharan, M. Wolf, N. Mitsumoto, and J. W. Burdick. 2009. UWB radar-based human target tracking. In *IEEE Radar Conference*. IEEE, 1–6.
- [5] Q. Chen and R. Liu. 2011. On the explanation of spatial smoothing in MUSIC algorithm for coherent sources. In *International Conference on Information Science and Technology*. 699–702.
- [6] T. Z. Chowdhury. 2018. *Using Wi-Fi channel state information (CSI) for human activity recognition and fall detection*. Ph.D. Dissertation. University of British Columbia.
- [7] S. Depatla and Y. Mostofi. 2018. Crowd Counting Through Walls Using WiFi. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–10.
- [8] Github. 2015. linux-80211n-csitol-supplementary. (2015). <https://github.com/dhalperi/linux-80211n-csitol-supplementary/issues/124#issuecomment-160837750>
- [9] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson. 2014. Phaser: Enabling phased array signal processing on commodity WiFi access points. In *Proceedings of the 20th annual International Conference on Mobile Computing and Networking*. ACM, 153–164.
- [10] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. 2011. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR* 41 (2011).
- [11] W. Jakes and D. Cox. 1994. *Microwave mobile communications*. Wiley-IEEE Press.
- [12] C. R. Karanam, B. Korany, and Y. Mostofi. 2018. Magnitude-based angle-of-arrival estimation, localization, and target tracking. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks*. 254–265.
- [13] C. R. Karanam and Y. Mostofi. 2017. 3D through-wall imaging with unmanned aerial vehicles using WiFi. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. ACM, 131–142.
- [14] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. 2015. Spotfi: Decimeter level localization using wifi. In *ACM SIGCOMM Computer Communication Review*, Vol. 45. ACM, 269–282.
- [15] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei. 2016. Dynamic-music: accurate device-free indoor localization. In *Proceedings of the ACM International Conference on Pervasive and Ubiquitous Computing*. ACM, 196–207.
- [16] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Yue Zhang, and H. Mei. 2017. IndoTrack: Device-free indoor human tracking with commodity Wi-Fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 72.
- [17] S. Nannuru, Y. Li, Y. Zeng, M. Coates, and B. Yang. 2013. Radio-frequency tomography for passive indoor multitarget tracking. *IEEE Transactions on Mobile Computing* 12, 12 (2013), 2322–2333.
- [18] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu. 2018. Widar2.0: Passive human tracking with a single Wi-Fi link. *Proceedings of ACM MobiSys* (2018).
- [19] R. Roy and T. Kailath. 1989. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 7 (1989), 984–995.
- [20] R. Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas & Propagation* 34 (1986), 276–280.
- [21] T. B. Schön. 2015. Nonlinear system identification using particle filters. *Encyclopedia of Systems and Control* (2015), 882–890.
- [22] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers. 2003. People tracking with mobile robots using sample-based joint probabilistic data association filters. *The International Journal of Robotics Research* 22, 2 (2003), 99–116.
- [23] S. Shi, S. Sigg, L. Chen, and Y. Ji. 2018. Accurate Location Tracking from CSI-based Passive Device-free Probabilistic Fingerprinting. *IEEE Trans. on Vehicular Technology* (2018).
- [24] B. D. Van Veen and K. M. Buckley. 1988. Beamforming: A versatile approach to spatial filtering. *IEEE Acoustics, Speech, and Signal Processing (ASSP) magazine* 5, 2 (1988), 4–24.
- [25] M. C. Vanderveen, A. J. Van der Veen, and A. Paulraj. 1998. Estimation of multipath parameters in wireless communications. *IEEE Transactions on Signal Processing* 46, 3 (1998), 682–690.
- [26] R. H. Venkatnarayan, G. Page, and M. Shahzad. 2018. Multi-User Gesture Recognition Using WiFi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 401–413.
- [27] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the International Conference on Mobile Computing and Networking*. ACM, 65–76.
- [28] K. Wu, Y. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni. 2013. CSI-based indoor localization. *IEEE Trans. on Parallel and Distributed Systems* 24, 7 (2013).
- [29] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. 2013. SCPL: indoor device-free multi-subject counting and localization using radio signal strength. In *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 79–90.
- [30] J. Zhang, W. Xiao, S. Zhang, and S. Huang. 2017. Device-free localization via an extreme learning machine with parameterized geometrical feature extraction. *Sensors Journal* 17, 4 (2017), 879.