Wi-Drone: Wi-Fi-based 6-DoF Tracking for Indoor Drone Flight Control

Guoxuan Chi[1], Zheng Yang[1✉], Jingao Xu[1], Chenshu Wu[2], Jialin Zhang[1], Jianzhe Liang[1], Yunhao Liu[3]

[1]School of Software and BNRist, Tsinghua University
[2]Department of Computer Science, The University of Hong Kong
[3]Department of Automation and Global Innovation Exchange, Tsinghua University
{chiguoxuan,hmilyyz,xujingao13,wucs32,zhjialin16,thssliangjz,yunhaoliu}@gmail.com

ABSTRACT

After years of boom, drones and their applications are now entering indoors. Six-degree-of-freedom (6-DoF) pose tracking is the core of drone flight control, but existing solutions cannot be directly applied to indoor scenarios due to insufficient accuracy, low robustness to adverse texture and light conditions, and signal obstruction in indoor scenarios. To overcome the above limitations, we propose Wi-Drone, a Wi-Fi standalone 6-DoF tracking system for indoor drone flight control. Wi-Drone takes full advantage of both *exteroceptive* and *proprioceptive* measurements of Wi-Fi to estimate the drone's absolute pose and relative motion, and fuse them in a tight-coupling manner to achieve their complementary benefits. We implement Wi-Drone and integrate it into a flight control system. The evaluation results show that Wi-Drone achieves a real-time performance with the average location accuracy of 26.1 cm and the rotation accuracy of 3.8°, which demonstrates its competency of flight control, compared to visual-inertial-based flight control. Such results also outperform existing Wi-Fi-based tracking solutions in terms of both dimensionality and accuracy.

CCS CONCEPTS

• **Networks** → **Location based services**.

KEYWORDS

Wireless Sensing, 6-DoF Pose Tracking, Flight Control, COTS Wi-Fi

1 INTRODUCTION

Recent years have witnessed the advent of a wide range of drone-based indoor applications, such as home security [30], warehouse

---

✉ Zheng Yang is the corresponding author.

management [6, 42], and entertainment [23]. As a key enabler of drone flight control, six-degree-of-freedom (6-DoF) tracking systems provide a drone with its real-time location and orientation in 3D space, realizing the drone's self-balancing, speed adjustment, and path planning.

Existing 6-DoF tracking solutions are various and tangled. To start with, inertial measurement unit (IMU) is widely adopted for drone flight control due to its low cost and ubiquity [36]. Many on-drone 6-DoF tracking systems [24] cooperate the GPS module with IMU and achieve delightful performance in outdoor scenarios. Another prevalent method for drone 6-DoF tracking is to leverage the fusion of vision and IMU [26, 29]. By jointly optimizing the image feature matching and IMU motion estimation results, a high-accuracy 6-DoF pose can be obtained.

Unfortunately, none of the aforementioned 6-DoF pose tracking solutions can be directly applied for indoor drone flight control: 1) Due to the narrow space and complex indoor layout, the tracking error should be controlled within decimeter-level. However, IMU-based solutions suffer from significant drifts due to the hardware noise and the drone motor vibration [20], which are non-trivial to be calibrated [43]. 2) The performance of visual-inertial systems is highly dependent on the light (e.g., RGB cameras cannot work in dark and RGB-D cameras fail with glare and strong illumination) and texture conditions, restricting the working scenarios of indoor drone applications. 3) Due to the obstruction by building walls or other obstacles, the GPS signals and other ground station signals suffer from severe attenuation in indoor environments, making all the GPS-based and ground-station-based tracking unavailable.

The above-mention limitations inspire us to build a Wi-Fi standalone 6-DoF tracking system for indoor drone flight control. Compared with existing methods, a pure Wi-Fi-based solution benefits from the following aspects: 1) Free of cumulative drift. Wi-Fi-based tracking system takes the access point (AP) as the anchor and thus is free of the cumulative drift. 2) Robustness to visual interference. Wi-Fi is agnostic to either illumination change or plain texture, making it suitable for varying indoor environments. 3) Pervasive indoor deployment. Nowadays, Wi-Fi infrastructures have been pervasively deployed in indoor scenarios, making Wi-Fi a ubiquitous solution for indoor device tracking. 4) Last but not least, a pure Wi-Fi-based solution is considerably light-weight in terms of computation overhead, hardware cost, and payload weight.

Although numerous Wi-Fi localization and tracking systems [19, 28, 43, 45] have been proposed, translating our intuition into a 6-DoF drone tracking system still faces significant challenges:

• **Dimensionality insufficiency.** Existing Wi-Fi-based localization and tracking systems, even those that leverage large antenna

array or multiple APs, only estimate the motion with no more than 3 degrees of freedom (e.g., 3D location, 2D location with azimuth). In short, there still lacks an algorithm to fully exploit the high-dimensional spatial information hidden in Wi-Fi signals.

- **Accuracy insufficiency.** State-of-the-art Wi-Fi-based systems achieve decimeter-level localization accuracy and millimeter-level tracking accuracy over a short distance. However, their performance degrades when tracking a flying drone due to the following facts: 1) Wi-Fi-based localization systems experience large errors when the target (i.e., drone) moving at a high speed; 2) Wi-Fi-based tracking systems are generally designed for tracking relatively short trajectories, and thus not competent for long-distance tracking. To conclude, none of the existing Wi-Fi-based solutions achieve enough accuracy for real-time and long-distance drone tracking tasks.

To tackle the above challenges, we design and implement Wi-Drone, the first pure Wi-Fi-based 6-DoF device tracking system for drone flight control. The insight behind Wi-Drone is that we find Wi-Fi can provide both *exteroceptive* and *proprioceptive* measurements of the drone's absolute pose and relative motion, respectively. These two distinct measurements can be integrated to achieve complementary benefits.

Specifically, in Wi-Drone, to improve the Wi-Fi-based tracking dimensions, we propose the method of rigid-body coordinate transformation and further derive two 6-DoF tracking algorithms: single-link pose reconstruction algorithm for the *exteroceptive pose estimation* module, and spatial dimension expansion algorithm for the *proprioceptive motion tracking* module. To push the limit of Wi-Fi tracking accuracy, the spatio-temporal phase sanitization algorithm is proposed to reconstruct the accurate phase information by digging deeper into different noise characteristics. Furthermore, we design a joint optimization and fusion framework to fuse the two modules in a tight-coupling manner to ensure tracking accuracy over both short and long trajectories.

We fully design and implement Wi-Drone platform on an industrial drone with commercial Wi-Fi network interface cards (NIC) and further integrate it with ArduPilot [1], one of the most widely used drone flight control systems. We also evaluate Wi-Drone's 6-DoF pose tracking capability on various 3D trajectories with diverse rotation angles, pose transfer speeds, and path lengths. We also compare Wi-Drone with an existing on-drone-deployed 6-DoF tracking solution VINS-Mono [29], as well as state-of-the-art Wi-Fi-based localization and tracking systems including mD-Track [45], SpotFi [18], RIM [43], and MonoLoco [35]. The evaluation results show that Wi-Drone achieves an average location accuracy of 26.1 cm and a rotation accuracy of 3.8°, which is comparable to the visual-inertial-based solution VINS-Mono, showing its capability to serve as a part of the drone flight control system. The location tracking accuracy of Wi-Drone exceeds mD-Track and SpotFi by > 40%, and its azimuth tracking accuracy exceeds RIM and MonoLoco by > 45%, demonstrating Wi-Drone's superiority compared with existing Wi-Fi-based systems.

In summary, our contributions are as follows:

- We propose Wi-Drone, as far as we are aware, the first Wi-Fi-standalone 6-DoF device tracking system. Wi-Drone's superior performance allows it to be used for drone flight control and

other 6-DoF-pose-driven tasks, with any number of APs. Compared with existing approaches, Wi-Drone's Wi-Fi-based tracking solution has its unique advantages mentioned above and will empower a variety of drone-based applications.

- Wi-Drone fully exploits Wi-Fi signals to build up both exteroceptive and proprioceptive measurements of moving devices and fuse them to achieve their complementary benefits. As a result, Wi-Drone achieves a standing out performance in typical operating scenarios for indoor drones with 3D, long-distance, and real-time requirements.

- We fully implement the Wi-Drone on an industrial drone by integrating our design with the ArduPilot flight control system. As a Wi-Fi standalone tracking solution, Wi-Drone's performance is comparable to that of the visual-inertial 6-DoF tracking ones. Our solution, in all or in part, can be directly migrated to a variety of Wi-Fi-ready devices, such as smartphones, tablets, and robots. In addition, some implementation techniques are essential not only for our system but also for the community to build up real-world Wi-Fi sensing applications.

2 OVERVIEW

2.1 Problem Statement

We first briefly introduce the problem of tracking a device's 6-DoF pose, including both the 3D location (l_x, l_y, l_z) and the 3D orientation $(\theta_{\text{Pitch}}, \theta_{\text{Yaw}}, \theta_{\text{Roll}})$. As shown in the left part of Fig. 1 (with a detailed annotation in Fig. 2), there are two reference (a.k.a., coordinate) systems in Wi-Drone: the Drone reference D and the Wi-Fi Access Point (AP) reference A. Since Wi-Fi APs are typically deployed in specific locations with fixed orientations in buildings, the absolute 6-DoF pose x^i of the drone at each timestamp i can be treated as the transformation from reference system D to A (i.e., a combination of the rotation matrix R^i_{AD} and translation vector t^i_{AD}). Note that $\{R_{\text{AD}}, t_{\text{AD}}\}$ and $\{l_x, l_y, l_z, \theta_{\text{Pitch}}, \theta_{\text{Yaw}}, \theta_{\text{Roll}}\}$ are equivalent in representing the drone's 6-DoF pose, and they could be inter-converted by the the Rodrigues' formula [31]. The former representation is adopted in the rest of our paper as it is widely used in drone flight control systems. Specifically, Wi-Drone takes time-series wireless signal $\mathcal{H} = \left\{ H^{i,k} \mid i \in \mathcal{T}, \ k \in \mathcal{S} \right\}$ from the antenna array $\mathcal{S}$ during a certain period of time $\mathcal{T}$ as input, and then derives the drone's real-time pose $x^i = \left\{ R^i_{\text{AD}}, t^i_{\text{AD}} \right\}$.

2.2 Wi-Drone Overview

From the top perspective, we design and implement a platform, namely Wi-Drone, which merely resorts to the ubiquitous Wi-Fi signal for indoor light-weight yet accurate pose tracking. The architecture of Wi-Drone is illustrated in Fig. 1. As seen, Wi-Drone consists of three key components: an *Exteroceptive Pose Estimation* (EPE) module (§3), a *Proprioceptive Motion Tracking* module (§4), and a *Joint Optimization and Fusion* (JOF) module (§5).

In Wi-Drone, both the EPE and PMT extract the channel state information (CSI) from every received Wi-Fi packet as the module input. However, the underlying philosophy of these two modules and their ways to use CSI are significantly distinct. Specifically, the EPE module treats CSI data as **exteroceptive** measurements, which means it takes the AP as the anchor, and thus is free of
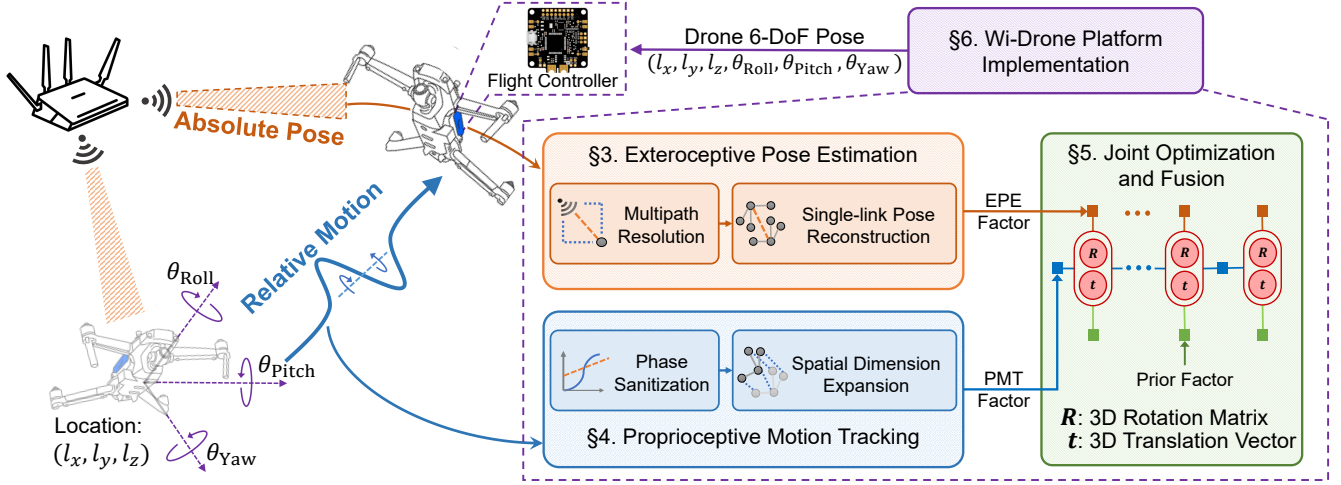
Figure 1: System architecture of Wi-Drone.

cumulative drifts. The EPE module estimates the drone's **absolute pose** by observing the geometric channel parameters (e.g., angle of arrival, time of flight) between the drone and the AP. In EPE, the *single-link pose reconstruction* algorithm is proposed to establish the relationship between the drone's pose and wireless channel parameters.

In contrast, the PMT module leverages the CSI phase as **proprioceptive** measurements, which means it infers the drone's **relative motion** change in a finer granularity by perceiving the temporal difference of the received CSI phase. The PMT module first sanitizes the CSI phase and analyzes the phase differences among diverse signal propagation paths and each packet. Further, our proposed *spatial dimension expansion* algorithm is performed to bridge the gap between 1D CSI phase difference and 6D drone's pose change by leveraging these phase differences among multiple antennas.

Finally, the graph-based JOF framework takes the above drone's *exteroceptive absolute pose* and *proprioceptive relative motion* as input and outputs an optimized drone's 6-DoF pose by maximum a posterior (MAP) estimation, which guarantees the fused 6-DoF trajectory is accurate enough while avoiding cumulative drift. It's worth mentioning that the *factor nodes* in JOF are elaborately designed to make them associate with the EPE and PMT.

3 EXTEROCEPTIVE POSE ESTIMATION

In this section, we introduce the proposed *Exteroceptive Pose Estimation* (EPE) module. The EPE module first estimates multipath parameters and then selects the parameters from the direct path to reconstruct the relative pose between the AP reference A (can also be treated as the world coordinate system in our work) and the drone reference D based on a novel *Single-link Pose Reconstruction* algorithm. EPE treats Wi-Fi as an *exteroceptive* sensor and hence provides absolute location and orientation of the drone.

3.1 Multipath Parameter Estimation

The basic idea of resolving multipath is parameter estimation. We first consider a general CSI measurement of the i-th packet, the

j-th subcarrier and the k-th antenna:

$$H(i, j, k) = \sum_{l=1}^{L} \alpha_l e^{-j\phi_l(i,j,k)} + N(i, j, k), \tag{1}$$

where L is the total number of multipaths, α_l and ϕ_l represent the attenuation factor and phase of the l-th propagation path, and N is the additive white Gaussian noise (AWGN). The phase of the l-th path in Eqn. 1 can be written as:

$$\phi_l(i, j, k) = 2\pi(f_c + \Delta f_j)(\tau_l + (\Delta s_k \cdot n_l)/c), \tag{2}$$

where f_c is the carrier frequency, Δf_j is the subcarrier frequency offset, Δs_k is the spatial difference between antenna k and the reference antenna, and c is the speed of light. τ_l and n_l represent the time-of-flight (ToF) and the unit direction vector of the angle of arrival (AoA) of the l-th path respectively. To resolve the path parameters $\theta_l = (\alpha_l, \tau_l, n_l)$ of each path, a maximum likelihood estimator is introduced. Denote $m = (i, j, k)$, the log-likelihood function of $\Theta = (\theta)_{l=1}^{L}$ is:

$$\mathcal{L}(\Theta \mid H(m)) = -\sum_m \left\| H(m) - \sum_{l=1}^{L} P_l(m, \theta_l) \right\|_2^2, \tag{3}$$

where P_l denote the CSI of the l-th path with path parameters θ_l. Therefore, the parameter estimation task is transformed into the following form:

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta \mid H(m)), \tag{4}$$

where all the path parameters are estimated based on the iterative parameter refinement algorithm [45].

Specifically, with the number of paths L to be determined, the following steps are performed: 1) we first treat all the signal except the strongest $P_1(m, \theta_1)$ as noise, and then only estimate the first path parameter θ_1; 2) from the residual $r_1 = H(m) - P_1(m, \theta_1)$, the θ_2 path parameters can be estimated; 3) the iteration continues until the power of residual r_L is lower than the level of noise N, and the residual is treated as the updated noise $N' = r_L$. Once the number of paths L is determined, the refinement process begins: 1) the pseudo CSI corresponding to each path l is reconstructed as $P_l + N'$; 2) the
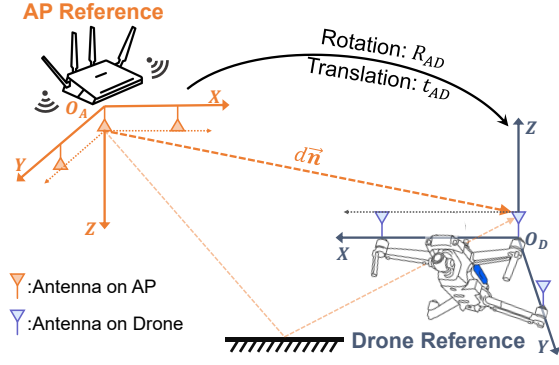
Figure 2: Illustration of the *Single-link Pose Reconstruction* algorithm. EPE leverages the AoA vector n and the distance d of the direct path measured by both AP and drone.

system re-estimates the each path parameter θ'_l, and update the residual and noise as $N'' = r'_L = H(m) - P'_1(m, \theta'_1) - \cdots - P'_L(m, \theta'_L)$; 3) the above iterations continues until convergence, i.e., each path parameter θ_l doesn't change anymore. To accelerate the estimation process, the generalized expectation maximization algorithm [14] are adopted.

The above parameter estimation process can be considered as an over-determined optimization problem. In Eqn. 4, the unknown path parameters Θ contains $5L$ unknown dimensions, since each path parameter θ_l consists of a 2D complex attenuation α_l, a 1D ToF, and a 2D AoA vector. In contrast, the dimension of each CSI measurement $H(m)$ is generally much higher.

3.2 Single-link Pose Reconstruction

The key insight of *Single-link Pose Reconstruction* is to leverage the direct path parameter (i.e., the θ_l with the largest $|\alpha_l|$) on both sides of the AP and the drone to derive the coordinate transformation between drone reference D and AP reference A.

As shown in Fig. 2, denote the ToF measured by the AP and the drone as τ_A and τ_D, and the unit direction vector of AoA as n_A and n_D. According to the principle of 3D coordinate transformation, the following relations are established:

$$\begin{cases} R_{AD} n_D = -n_A, \\ t_{AD} = c\tau_D n_D = -c\tau_A R_{AD}^\mathsf{T} n_A, \end{cases} \tag{5}$$

where R_{AD} indicates the orientation difference between AP and drone, and t_{AD} refers to the location difference. The intuition behind Eqn. 5 is two-fold: 1) by performing a rotation transformation, the orientation of two reference systems should be aligned, which means the AoA direction vector observed by the AP and the drone should be exactly the opposite; 2) the relative location difference between the two coordinates can be jointly determined by the ToF and the AoA.

To solve the rotation R_{AD}, we have the following equations:

$$\begin{aligned} \hat{R}_{AD} &= \arg\min_{R_{AD}} \|R_{AD} n_D + n_A\|_2^2 \\ &= \arg\min_{R_{AD}} \left(n_D^\mathsf{T} R_{AD}^\mathsf{T} R_{AD} n_D + 2 n_A^\mathsf{T} R_{AD} n_D + n_A^\mathsf{T} n_A \right) \\ &= \arg\min_{R_{AD}} n_A^\mathsf{T} R_{AD} n_D, \end{aligned} \tag{6}$$

which can be solved by singular value decomposition (SVD) [2]:

$$\hat{R}_{AD} = V\Sigma U^\mathsf{T}, \quad U\Sigma V^\mathsf{T} = n_D n_A^\mathsf{T}. \tag{7}$$

As for the translation t_{AD}, we have:

$$\hat{t}_{AD} = c(\tau_D + \tau_A) n_D / 2. \tag{8}$$

By performing the aforementioned steps, the drone's pose real-time pose $\hat{x}_{AD} = \{\hat{R}_{AD}, \hat{t}_{AD}\}$ can be estimated from the direct path parameter of each CSI measurement.

4 PROPRIOCEPTIVE MOTION TRACKING

Distinguished from EPE which estimates the drone's absolute yet rough pose, PMT exploits CSI phase samples as *proprioceptive* measurements to track the drone's relative motion. Compared with ToF and AoA that are used in EPE, the CSI phase is more competent to track the tiny movements of objects with centimeter-level or even millimeter-level accuracy. In this section, we present the design of PMT, especially *how to derive the drone's instant 6-DoF motion in 3D space from the merely 1D CSI phase*.

4.1 Spatio-temporal Phase Sanitization

Due to the imperfect NIC hardware design, various types of errors are introduced into CSI measurement. For errors varying across each packet (e.g., packet boundary detection (PBD) error, sampling frequency offset (SFO), and carrier frequency offset (CFO)), there is a need for novel phase sanitization algorithms [44]. Consider an erroneous version of CSI phase received by antenna k with a propagation path l, which suffers from both the frequency offset ϵ_f caused by CFO and the timing offset ϵ_t caused by PBD and SFO:

$$\widetilde{\phi}_k^l = 2\pi(f_c + \Delta f_j + \epsilon_f)(\tau_k^l + \epsilon_t). \tag{9}$$

To recover the accurate phase from disturbed CSI, the PMT module performs the following three steps.

Spatial correlation. Our key observation is, though varying across different packets, ϵ_f and ϵ_t are spatial-invariant, which means given one packet, these errors keep stationary across different propagation paths [18]. Therefore, PMT first extracts the CSI H_k of each antenna k, and then calculates the spatial correlation term C_k by conjugate multiplication, which can be written into the following form:

$$\begin{aligned} C_k &= H_k H_k^* = \left(H_k^{\text{LoS}} + \sum_{l \in \mathcal{L}_N} H_k^l \right)\left(H_k^{\text{LoS}*} + \sum_{l \in \mathcal{L}_N} H_k^{l*} \right) \\ &= \underbrace{H_k^{\text{LoS}} H_k^{\text{LoS}*}}_{\text{semi-static term } C_k^S} + \underbrace{\sum_{l \in \mathcal{L}_N} H_k^l \sum_{l \in \mathcal{L}_N} H_k^{l*}}_{\text{higher order minima } C_k^O} \\ &+ \underbrace{H_k^{\text{LoS}} \sum_{l \in \mathcal{L}_N} H_k^{l*} + H_k^{\text{LoS}*} \sum_{l \in \mathcal{L}_N} H_k^l}_{\text{dynamic term } C_k^D}, \end{aligned} \tag{10}$$

where the $\mathcal{L}_N$ denotes the NLoS paths (i.e., non-direct paths) collection. As seen, each C_k consists of three terms: the semi-static term C_k^S, the dynamic term C_k^D, and the higher order minima C_k^O.
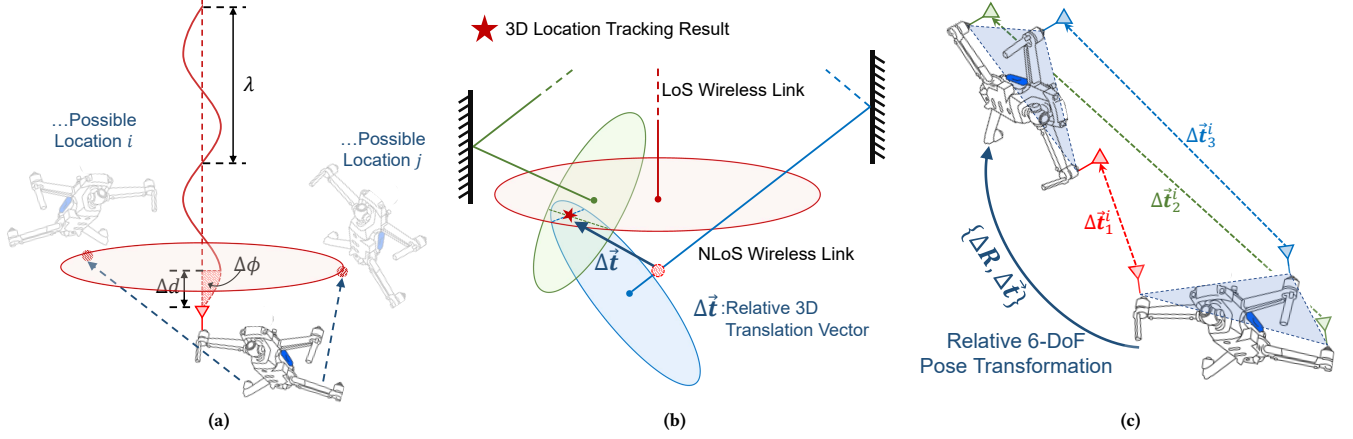
Figure 3: A step-by-step illustration of the *Spatial Dimension Expansion* algorithm. (a) The phase difference encountered by each antenna indicates the radial distance change. However, the drone's 3D location would be ambiguous with only 1D distance information. (b) By embracing multipath, PMT uniquely determines the 3D location of one specific antenna. (c) Finally, putting all the measured 3D translations of antennas together, the 6-DoF motion of the drone can be resolved using the rigid-body constraint among multiple antennas.
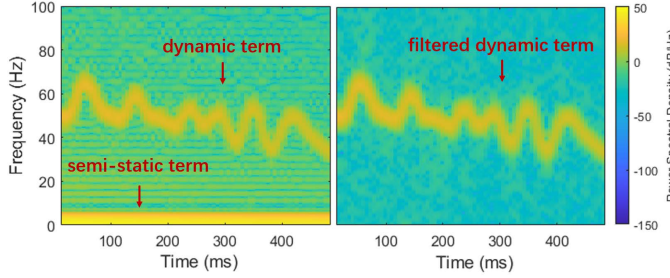

Figure 4: Time-frequency analysis of C_k.

The dynamic term can be presented as follows:

$$C_k^D = \sum_{l \in \mathcal{L}_N} \left(\alpha_k^{\text{LoS}} \alpha_k^{l*} e^{j\Delta\phi_k^l} + \alpha_k^{\text{LoS*}} \alpha_k^l e^{-j\Delta\phi_k^l} \right), \quad (11)$$

where $\Delta\phi_k^l$ indicates the phase difference between the LoS path (i.e., the direct path) and the NLoS path l and can be written as:

$$\Delta\phi_k^l = 2\pi(f_c + \Delta f_j + \epsilon_f)(\tau_k^l - \tau_k^{\text{LoS}}), \quad (12)$$

where the timing offset ϵ_t has been eliminated.

Temporal analysis. The remaining problem is how to separate the dynamic term C_k^D from C_k in Eqn. 10. Our key insight is to leverage the temporal relationship across the CSI packets. Specifically, as the drone moves: 1) the semi-static term C_k^S reflects the attenuation changes over a short period, which is nearly constant, since centimeter-level distance change doesn't cause any obvious signal strength fluctuation; 2) the dynamic term C_k^D changes dramatically since even a centimeter-level slight motion incurs a periodical change in CSI phase in Eqn. 12; 3) the high-order term C_k^O can be omitted since the reflected signals from the non-direct path are orders weaker than the LoS signal [28]. Therefore, as shown in Fig. 4, after performing the time-frequency analysis on a series of C_k, a high-pass filter can be used to separate the dynamic term series.

Spectral fitting. With both the spatial correlation and the temporal analysis mentioned above, the timing offset ϵ_t can be eliminated,

leaving the frequency offset ϵ_f to be dealt with. To reconstruct accurate CSI phase, the phase in Eqn. 12 is treated as a linear function of Δf_j across all subcarriers, and both the common phase slope $\beta_k^l = \tau_k^l - \tau_k^{\text{LoS}}$ across the subcarriers and the corresponding overall attenuation $\eta_k^l = \alpha_k^{\text{LoS}} \alpha_k^{l*}$ are jointly estimated [35]:

$$\left\{ \hat{\beta}_k^l, \hat{\eta}_k^l \right\} = \arg\min_{\beta_k^l, \eta_k^l} \left(C_k^D - \sum_{l \in \mathcal{L}_N} \left(\eta_k^l e^{j\Delta\phi_k^l} + \eta_k^{l*} e^{-j\Delta\phi_k^l} \right) \right), \quad (13)$$

$$\Delta\phi_k^l = 2\pi(f_c + \Delta f_j + \epsilon_f)\beta_k^l,$$

and the phase could be reconstructed as:

$$\Delta\hat{\phi}_k^l = 2\pi(f_c + \Delta f_j)\hat{\beta}_k^{\,l}, \quad (14)$$

So far, both ϵ_f and ϵ_t are eliminated, leaving the sanitized phase $\Delta\hat{\phi}_k^l$ in Eqn. 14, which indicates the phase difference between the NLoS path l and the LoS path.

4.2 Spatial Dimension Expansion

To infer the drone's 6-DoF relative pose from the 1D sanitized phase, PMT performs *Spatial Dimension Expansion* algorithm.

For ease of notion, we first consider the geometric meaning of the CSI phase. As shown in Fig. 3a, the phase difference $\Delta\phi$ between two consecutive packets reveals the radial distance change $\Delta d = \frac{\Delta\phi}{2\pi}\lambda$, where the λ indicates the wavelength of the radio signal. However, the distance change is only of one dimension. Therefore, the spatial dimension expansion algorithm is proposed to bridge the dimensional gap in two steps: 1) bridge the gap between 1D distance change and 3D location change based on multipath, and 2) bridge the gap between 3D location change and 6D pose change based on multiple antennas.

From 1D distance to 3D location. The PMT module leverages the sanitized phase in Eqn. 14 to derive the 3D location change. For the i-th packet, suppose there are L NLoS paths and one LoS path, aggregate the sanitized phase in Eqn. 14 as $\Delta\hat{\boldsymbol{\phi}}_{i,k} = \left(\Delta\hat{\phi}_{i,k}^1, \cdots, \Delta\hat{\phi}_{i,k}^L \right)^\mathsf{T}$,

which refers to the phase difference between all L NLoS paths and the LoS path. Correspondingly, subtract the direction vector of the LoS path $\boldsymbol{n}^0$ from the direction vector of L NLoS paths to get $\boldsymbol{N}_i = \left(\boldsymbol{n}_i^1 - \boldsymbol{n}_i^0, \cdots, \boldsymbol{n}_i^L - \boldsymbol{n}_i^0\right)^{\mathsf{T}}$. Thus, we have the following approximate relationship between the phase change and the 3D location change $\Delta \boldsymbol{d}_{i,k} = \boldsymbol{d}_{i,k} - \boldsymbol{d}_{i-1,k}$:

$$\Delta \hat{\boldsymbol{\phi}}_{i,k} - \Delta \hat{\boldsymbol{\phi}}_{i-1,k} = 2\pi \frac{\boldsymbol{N}_i \Delta \boldsymbol{d}_{i,k}}{\lambda}, \tag{15}$$

which indicates that the phase change on each propagation path is induced by the projection of the drone's 3D location change on each corresponding direction, as shown in Fig. 3b.

From 3D location to 6D pose. To derive the drone's relative rotation and translation, our key insights are as follows:

- As the drone rotates, different antennas may experience distinctive location changes, as shown in Fig. 3c.
- Multiple antennas on the drone are rigidly attached. Therefore, PMT infers macroscopic pose change from the microscopic motion of each antenna based on rigid body kinematics [50].

Denote location changes of antennas as $\Delta \boldsymbol{D}_i = \left(\Delta \boldsymbol{d}_{i,1}, \cdots, \Delta \boldsymbol{d}_{i,K}\right)$. With preknown antenna arrangement $\Omega = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_K)$, the equation below can be derived:

$$\Delta \boldsymbol{D}_i = (\Delta \boldsymbol{R}_i \Omega + \Delta \boldsymbol{t}_i) - \Omega, \tag{16}$$

indicating the location change of each antenna is induced by an overall rotation and translation of the array.

So far, the basic ideas of the spatial dimension expansion algorithm have been illustrated. To infer the relative 6-DoF pose change of the drone, we gather all sanitized phase of L NLoS path and K antennas into an $L \times K$ matrix $\Delta \Phi_i = \left(\Delta \hat{\boldsymbol{\phi}}_{i,1}, \cdots, \Delta \hat{\boldsymbol{\phi}}_{i,K}\right)$, and define $\Delta^2 \Phi_i = \Delta \Phi_i - \Delta \Phi_{i-1}$ as the phase difference between packet i and $i-1$. By combining Eqn. 15 and Eqn. 16, the relative pose $\Delta \boldsymbol{x}_i = \{\Delta \boldsymbol{R}_i, \Delta \boldsymbol{t}_i\}$, which indicates the relative rotation and translation between packet i and $i-1$ can be solved as follows:

$$\Delta \hat{\boldsymbol{x}}_i = \underset{\{\Delta \boldsymbol{R}_i, \Delta \boldsymbol{t}_i\}}{\arg\min} \left\| \frac{2\pi \boldsymbol{N}_i \left((\Delta \boldsymbol{R}_i - \boldsymbol{I}) \Omega + \Delta \boldsymbol{t}_i\right)}{\lambda} - \Delta^2 \Phi_i \right\|_2^2. \tag{17}$$

5 JOINT OPTIMIZATION AND FUSION

So far, two independent 6-DoF tracking modules have been derived. However, EPE and PMT suffer from severe pose tracking bias due to the limited ToF resolutions, and cumulative drift, respectively, as depicted in Fig. 5. To further push the limit of Wi-Fi-based pose tracking accuracy, we introduce a factor-graph-based *joint optimization and fusion* (JOF) module to fuse the EPE and PMT module. As mentioned above, the rationale behind the sufficient fusion is that EPE and PMT leverage two distinct wireless features and treat CSI samples as *exteroceptive* and *proprioceptive* measurements, respectively. Consequently, the two modules benefit from their individual yet complementary advantages, and hence a joint optimization of them would boost the overall performance, getting a 6-DoF trajectory with both low bias and low cumulative drift.

As illustrated in Fig. 6, JOF includes both short-term (inter-frame tracking, Fig. 6a) and long-term (local pose tracking, Fig. 6b) optimizations that work hand in hand to improve the pose tracking
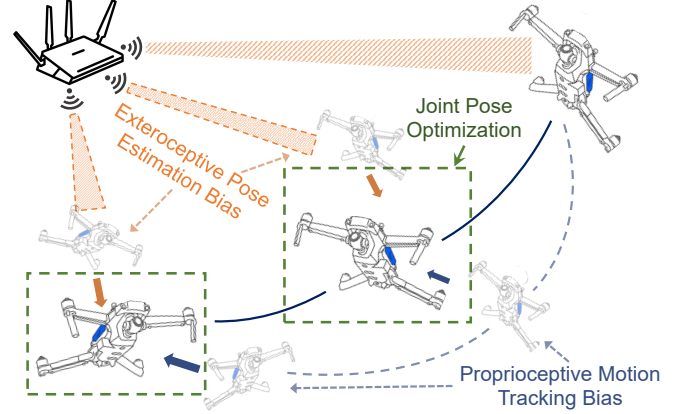

Figure 5: Illustration of EPE, PMT, and JOF. JOF jointly optimizes the EPE and PMT module to fuse their derived 6-DoF poses, which overcomes the drawbacks of each module and reports the drone's accurate trajectory.

performance continuously. Besides EPE and PMT, JOF also integrates prior knowledge such as drone flight characteristics (e.g., the continuity of the velocity) to optimize a more smooth and accurate trajectory.

5.1 Factor Graph

A factor graph consists of two types of nodes: the variable nodes which indicate the states to be optimized (e.g., $\boldsymbol{x}_i$), and the factor nodes which represent the probability of certain states given a measurement result. In Wi-Drone, these measurements come from the EPE module [1] (denoted as $\boldsymbol{v}^i$) and PMT module (denoted as $\Delta^2 \Phi_i$).

In order to estimate the values of a certain set of variable nodes $\mathcal{X} = \{\boldsymbol{x}_i | i \in \mathcal{T}\}$ given all relevant measurements $\mathcal{Z} = \{\Delta^2 \Phi_i, \boldsymbol{v}_i | i \in \mathcal{T}\}$, Wi-Drone optimizes all the factor nodes connected with them based on maximum a posteriori (MAP) estimation:

$$\begin{aligned} \hat{\mathcal{X}} &= \underset{\mathcal{X}}{\arg\max} \, p(\mathcal{X}|\mathcal{Z}) = \underset{\mathcal{X}}{\arg\max} \, p(\mathcal{X}) p(\mathcal{Z}|\mathcal{X}) \\ &= \underset{\mathcal{X}}{\arg\max} \, p(\mathcal{X}) \prod_{i \in \mathcal{T}} p(\Delta^2 \Phi_i | \boldsymbol{x}_i) p(\boldsymbol{v}_i | \boldsymbol{x}_i), \end{aligned} \tag{18}$$

where the $p(\boldsymbol{v}_i | \boldsymbol{x}_i)$ and $p(\Delta^2 \Phi_i | \boldsymbol{x}_i)$ are the likelihood of the EPE and PMT measurements respectively, and $p(\mathcal{X})$ is the prior information of trajectory, which can be inferred from the drone flight characteristics.

5.2 Probabilistic Representation

To infer the drone's 6-DoF pose based on MAP, both the prior term and the likelihood term in Eqn. 18 are needed.

Prior term. The prior term $p(\boldsymbol{x}_i)$ indicates the probability distribution of the drone's location and orientation at time i without knowing any measurement result. Based on the kinetic characteristics of the drone, the constant velocity model, which has been widely used in both flight control [40] and poses estimation [25], is adopted to derive the prior term. Specifically, the drone is assumed to move and rotate at an approximately constant speed during a

---

[1]Denote the direct path parameters in EPE module as $\boldsymbol{v} = \{\boldsymbol{n}_\mathrm{A}, \boldsymbol{n}_\mathrm{D}, \tau_\mathrm{A}, \tau_\mathrm{D}\}$.
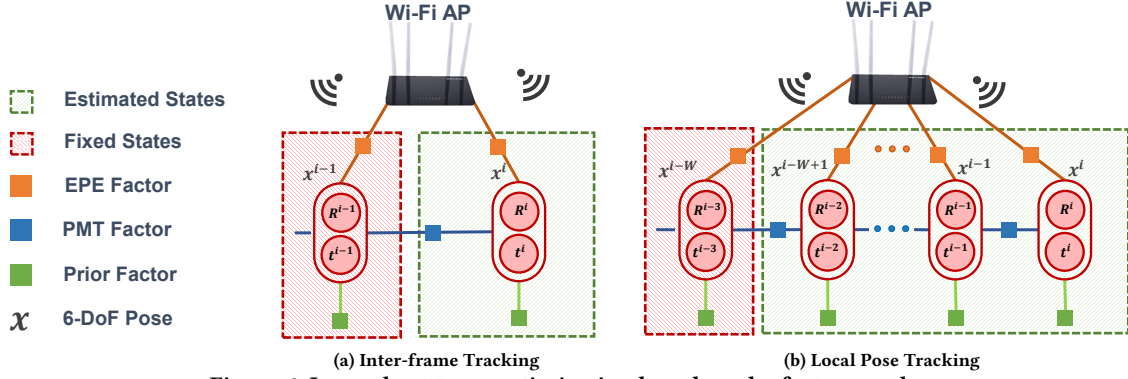
(a) Inter-frame Tracking

(b) Local Pose Tracking

Figure 6: Long-short term optimization based on the factor graph.

short period of time. On this basis, the prior pose $\overline{x}_i = \{\overline{R}_i, \overline{t}_i\}$ can be inferred from the following equations:

$$\overline{R}_i = R_{i-1}(R_{i-2})^{\mathsf{T}} R_{i-1}, \quad \overline{t}_i = 2t_{i-1} - t_{i-2}. \tag{19}$$

Suppose the prediction error follows a zero-mean Gaussian distribution, we have $p(x_i|x_{i-1}, x_{i-2}) \sim \mathcal{N}(\overline{x}_i, \sigma_x)$, where the σ_x is an empirical value set according to the flying mode of drone.

EPE likelihood. The likelihood of the EPE module $p(v_i|x_i)$ indicates the distribution of the measured ToF and AoA at a given pose. According to Eqn. 5, the following relationship can be derived:

$$\overline{n}_A = -t_{AD}/\|t_{AD}\|_2, \quad \overline{n}_D = R_{AD}^{\mathsf{T}} t_{AD}/\|R_{AD}^{\mathsf{T}} t_{AD}\|_2,$$
$$\overline{\tau}_A = \overline{\tau}_D = \|t_{AD}\|_2/c. \tag{20}$$

Similar as most existing wireless tracking systems, Wi-Drone assume the measured AoA and ToF suffers from a Gaussian noise:

$$p(n_A|x_i) \sim \mathcal{N}(\overline{n}_A, \sigma_{AoA}), \quad p(n_D|x_i) \sim \mathcal{N}(\overline{n}_D, \sigma_{AoA}),$$
$$p(\tau_A|x_i) \sim \mathcal{N}(\overline{\tau}_A, \sigma_{ToF}), \quad p(\tau_D|x_i) \sim \mathcal{N}(\overline{\tau}_D, \sigma_{ToF}), \tag{21}$$

where the σ_{AoA} and σ_{ToF} are the standard deviation of AoA and ToF measurements respectively.

PMT likelihood. The likelihood of the PMT module $p(\Delta^2\Phi_i|x_i)$ indicates the distribution of the measured phase difference at a given pose. According to Eqn. 17, the following relationship can be derived:

$$\overline{\Delta^2\Phi_i} = 2\pi N \left((\Delta R_i - I)\Omega + \Delta t_i\right)/\lambda, \tag{22}$$

where $\Delta R_i = R_i(R_{i-1})^{\mathsf{T}}$, and $\Delta t_i = t_i - R_i(R_{i-1})^{\mathsf{T}} t_{i-1}$. Suppose the CSI phase noise follows a Gaussian distribution of $p(\Delta^2\Phi_i|x_i) \sim \mathcal{N}\left(\overline{\Delta^2\Phi_i}, \sigma_\Phi\right)$, where σ_Φ denotes the standard deviation of the phase.

5.3 Fusion-based Tracking

As shown in Fig. 6, two types of fusion schemes are adopted in Wi-Drone. Specifically, the inter-frame tracking infers the drone's instant 6-DoF pose in real-time. In contrast, the local pose tracking focuses on the overall accuracy of the flight trajectory over a period of time.

Inter-frame tracking. As shown in Fig. 6a, once a new Wi-Fi packet i arrives, the prior factor, the EPE factor and the PMT factor

are formulated as follows:

$$E_i^{\text{prior}} = -\log p(x_i) \propto \|x_i - \overline{x}_i\|_{\sigma_x}^2,$$
$$E_i^{\text{EPE}} = -\log p(n_A, n_D, \tau_A, \tau_D | x_i)$$
$$\propto \|n_A - \overline{n}_A\|_{\sigma_{AoA}}^2 + \|n_D - \overline{n}_D\|_{\sigma_{AoA}}^2 \tag{23}$$
$$+ \|\tau_A - \overline{\tau}_A\|_{\sigma_{ToF}}^2 + \|\overline{\tau}_D - \overline{\tau}_D\|_{\sigma_{ToF}}^2,$$
$$E_i^{\text{PMT}} = -\log p(\Delta^2\Phi_i|x_i) \propto \|\Delta^2\Phi_i - \overline{\Delta^2\Phi_i}\|_{\sigma_\Phi}^2.$$

On this basis, the inter-frame optimization is performed to give an instant pose tracking result. According to Eqn. 18:

$$\hat{x}_i = \arg\max_{x_i} p(x_i|x_{i-1}, x_{i-2}) p(\Delta^2\Phi_i|x_i) p(v_i|x_i)$$
$$= \arg\min_{x_i} -\log\left(p(x_i|x_{i-1}, x_{i-2}) p(\Delta^2\Phi_i|x_i) p(v_i|x_i)\right) \tag{24}$$
$$= \arg\min_{x_i}\left(E_i^{\text{prior}} + E_i^{\text{EPE}} + E_i^{\text{PMT}}\right).$$

Local pose tracking. For every few seconds, the local pose tracking is triggered to correct the cumulative drift. As shown in Fig. 6b, local pose tracking takes all frames based on a sliding window, and jointly optimizes their poses:

$$\hat{\mathcal{X}} = \arg\max_{\mathcal{X}} p(\mathcal{X}) \prod_{i\in\mathcal{T}} p(\Delta^2\Phi_i|x_i) p(v_i|x_i)$$
$$= \arg\min_{\mathcal{X}} \sum_{i\in\mathcal{T}}\left(E_i^{\text{prior}} + E_i^{\text{EPE}} + E_i^{\text{PMT}}\right). \tag{25}$$

Loop detection and global optimization. To correct the drone's trajectory during the flight, we exploit the locations it has visited before as constraints for long-term cumulative error elimination. To detect the loop, we adopt the time-reversal resonating strength (TRRS) proposed in [43] as a loop closure index, indicating the similarity between the current CSI measurement H_i and a previous one H_{i-T}. A TRRS higher than the threshold (set to 0.7 in our implementation) indicates the existence of a closed loop, and the global optimization is performed. The global optimization is a special case of local pose tracking, where all the frames between $i-T$ and i are optimized, under a constraint that $x_{i-T} = x_i$.

6 PLATFORM IMPLEMENTATION

As shown in Fig. 7, we implement Wi-Drone using commercial Wi-Fi NIC AR9580 operating in 5.8 GHz with 40 MHz bandwidth. To realize multi-NIC integration, all the NICs are set on a Pericom PCI-E extension card attached to an Intel NUC with Core i7-1165G7 CPU and 8GB RAM. Wi-Drone is integrated with ArduPilot [1], a widely-used open source drone flight control framework. The output 6-DoF poses are streamed to the ArduPilot Mega (APM) controller through a Micro-USB port, making real-time flight control available. The main part of Wi-Drone runs on Unbuntu 20.04 LTS. The NICs are driven by the PicoScenes [17]. All algorithms in Wi-Drone are implemented by MATLAB/C++ hybrid programming. Several implementation techniques are introduced to make our system practical and push the limit of tracking accuracy.

6.1 Flight Controller Integration

It takes a lot of effort to integrate Wi-Drone with ArduPilot. To realize the data transmission from the Intel NUC to the APM flight controller, we dive into the transmission protocol used by ArduPilot called MAVLink and implement a communication middleware to forward the 6-DoF pose to APM in real-time. To evaluate the flight control performance of Wi-Drone, both the GPS module and IMU module of the controller are manually disabled. The workflow of the controller is also modified based on our system flow design, and the corresponding flight control firmware is also recompiled and burned in the APM controller.

6.2 NIC Calibration

Most commercial NICs suffer from imperfect hardware implementations. Therefore, they generally need to be before performing wireless sensing tasks.

Radio chain phase offset. The signals transmitted or received by different antennas experience different phase offsets, which incurs inaccurate AoA measurements [46]. Such offset is constant across different packets and frequency bands. Therefore, we connect each transmitter (Tx) and receiver (Rx) via coaxial cable in turn, and then record the phase offset of each Tx-Rx pair. The recorded offsets are canceled from the raw CSI before any processing steps.

Amplitude and phase nonlinearity. Due to non-ideal analogue filter design, both the CSI amplitude and phase suffer from nonlinear error, and it has been proved that such nonlinearity is constant for one specific NIC [44]. Our basic idea is to connect the Tx and Rx with coaxial cables to record the nonlinear pattern, and then subtract the pattern from the measured CSI amplitude and phase respectively.

6.3 Push the Limit of Tracking Accuracy

The following techniques are used in our implementation to fully unleash the potential of Wi-Drone.

Antenna array setup. As shown in Fig. 7, 12 antennas on 4 NICs forms into a 3×4 square-shaped array. To improve the accuracy of AoA measurement, the arrangement of each antenna is carefully designed. As the previous research [41] points out, for AoA measurement, there's a trade-off between improving accuracy and removing ambiguity: an antenna space larger than $\lambda/2$ (half-wavelength) improves the angle resolution, while causing the
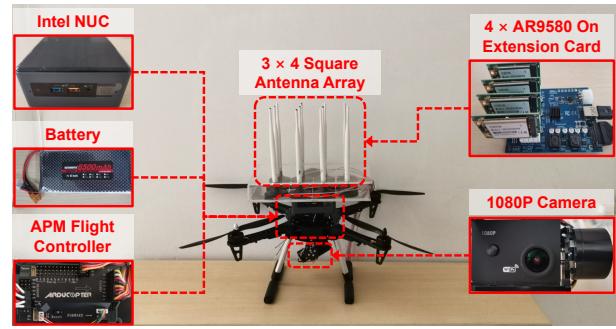

Figure 7: Hardware setup of Wi-Drone platform.

ambiguity. Therefore, 2 NICs are designed with an antenna space $\lambda/2$ for global search to avoid ambiguity, while the other 2 NICs are designed with an antenna space λ responsible for AoA refinement. By comprehensively using the AoA measurement results from multiple NICs, the AoA accuracy generally has 3× improvements.

Erroneous packet rejection. Multipath parameter estimation lies in the vital part of Wi-Drone. Though only taking up a small portion, those packets in which multipath cannot be resolved correctly still incur severe drift in practice. Therefore, Wi-Drone adopts a straightforward yet effective erroneous frame rejection strategy. Concretely, Wi-Drone assumes that all frames received within a short time window W_T (set to 20 ms in our implementation) experience similar multipath conditions. Therefore, Wi-Drone computes the mean and covariance matrix of all multipath parameters of these packets within the W_T, and removes the "outliers".

Multi-AP collaboration. Wi-Drone achieves satisfactory tracking accuracy with only one Wi-Fi AP. Nevertheless, since Wi-Fi APs have been densely deployed in many indoor spaces, we expect our system perform better in realistic scenarios through multi-AP collaboration. Specifically, by adding constraints of the APs' relative poses and CSI timestamps to Eqn. 24 and Eqn. 25, Wi-Drone leverages CSI data collected from multiple APs, and further improves tracking accuracy. Since the link quality (e.g., signal strength, LoS condition) between the drone and APs are different, during the optimization process, EPE and PMT factors corresponding to each AP are weighted by the attenuation η in Eqn. 13.

Extra spatial stream. According to the 802.11 protocol [3], multiple consecutive high throughput long training fields (HT-LTFs) are allowed to be inserted into the physical layer protocol data unit (PPDU), which means more than one CSI measurement can be collected from a single packet. Wi-Drone collects 3 independent CSI measurements from one packet by adding 2 extra spatial streams. By averaging 3 CSI measurements in the same packet, the background noise can be eliminated effectively.

7 EVALUATION

7.1 Experimental Methodology

Field studies. As illustrated in Fig. 8, we integrate Wi-Drone into a drone platform and conduct extensive experiments in an office building with several typical indoor scenarios. As shown in Fig. 9, various 3D experimental trajectories are elaborately set with diverse rotation angles, pose transfer speeds, and path lengths ranging from 5 m to 30 m. Three types of drone rotation modes (i.e., roll, pitch,
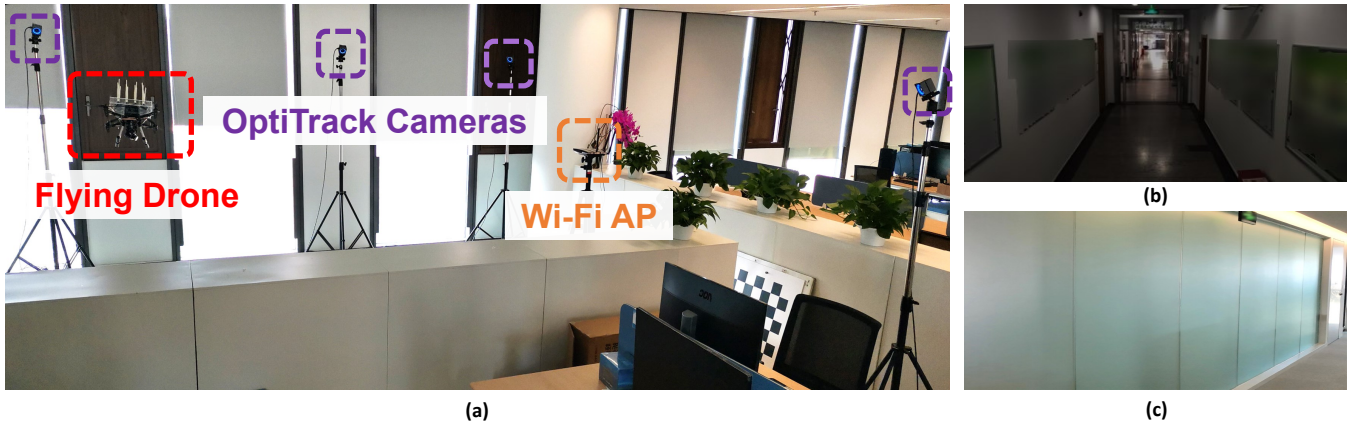
Figure 8: Experimental setup and scenarios of Wi-Drone. (a) An experimental scenario with satisfactory light and texture conditions. (b)-(c) A low-light-level scenario and a texture-less scenario from the drone's perspective.
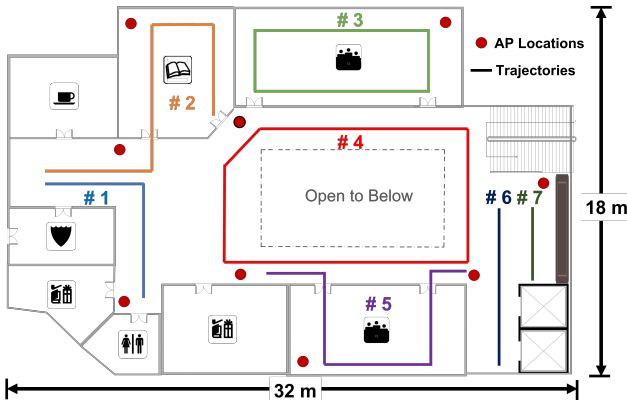

Figure 9: Floorplan of the experimental scenario with AP deployment and planned trajectories.

and yaw) are included in the designed trajectories. During the flight, drastic speed fluctuations are avoided, and the flight speed is controlled within a given interval.

Ground truth acquisition. Multiple *OptiTrack* [27] cameras working at 240 FPS are deployed along the experimental trajectories to obtain sub-millimeter-level ground truth.

Evaluation metrics. Wi-Drone reports the drone's 6-DoF pose (3D location and 3D rotation) in real-time. We compare the reported pose with the ground truth for all trajectories to calculate the average location error (a.k.a. Absolute Trajectory Error, ATE) in meters and the average rotation error in degrees. We also evaluate the latency for reporting the drone's instant pose. Moreover, the CPU workload and RAM usage are recorded as the runtime resource consumption.

Comparative methods. We compare Wi-Drone with several related systems, which can be divided into three categories:

- *Visual-inertial-based pose tracking system*, **VINS-Mono** [29], which leverages both the monocular camera and IMU for accurate drone 6-DoF pose tracking and has already been embedded in today's drone flight control system.
- *Wi-Fi-based location tracking solutions*, **mD-Track** [45] and **SpotFi** [18], both of which estimate the device's location based on the resolved AoA and ToF parameters.

- *Wi-Fi-based rotation tracking solutions*, **RIM** [43], which tracks the device's rotation in the 2D plane by applying multiple antenna alignment algorithm, and **MonoLoco** [35], which estimates the device's azimuth (a.k.a. the rotation in the 2D plane) based on the plane geometry.

We first compare Wi-Drone with VINS-Mono to demonstrate its capability to support indoor drone flight control. Then we compare Wi-Drone with the latter two types of Wi-Fi-based tracking systems to present our system's superiority in terms of both dimensionality and accuracy. For VINS-Mono, the video is recorded by a monocular 1080P RGB camera at 60 FPS, and a built-in IMU working at 200 Hz is leveraged for inertial measurements. Additionally, we provide VINS-Mono and RIM with accurate initial location and orientation to transform their relative pose changes into absolute ones.

7.2 Overall Performance

7.2.1 Accuracy Comparison with the On-drone System. We first compare the 6-DoF pose tracking accuracy of Wi-Drone with VINS-Mono, one of the most prevalent 6-DoF tracking solutions that has been widely adopted for drone flight control. As shown in Fig. 10, Wi-Drone outperforms VINS-Mono under low-light-level scenarios (VINS-Mono-LL) and texture-less scenarios (VINS-Mono-TL) and achieves comparable performance with VINS-Mono under satisfactory light and texture conditions. The average location accuracy of Wi-Drone is 26.1 cm, which exceeds VINS-Mono-LL by 63.0% and VINS-Mono-TL by 70.2%. As for the rotation, the average accuracy of Wi-Drone is 3.8°, outperforming VINS-Mono-LL and VINS-Mono-TL by 42.7% and 58.9% respectively. Besides, the 95th percentile accuracy of Wi-Drone outperforms VINS-Mono in all scenarios, improving location accuracy and rotation accuracy by 17.6% and 12.7%, respectively. The rationale behind the delightful performance is two-fold: 1) Compared with visual clues, Wi-Fi is agnostic to either frequent illumination change or insufficient texture, which are common cases in complex indoor environments. 2) From the perspective of the proposed tracking algorithm, the EPE module in Wi-Drone leverages AP as the anchor point to provide the drone with an absolute reference, bounding the tracking bias within 0.7 m. In contrast, visual-inertial odometry inevitably suffers from severe cumulative drift as the path length increases. In general,
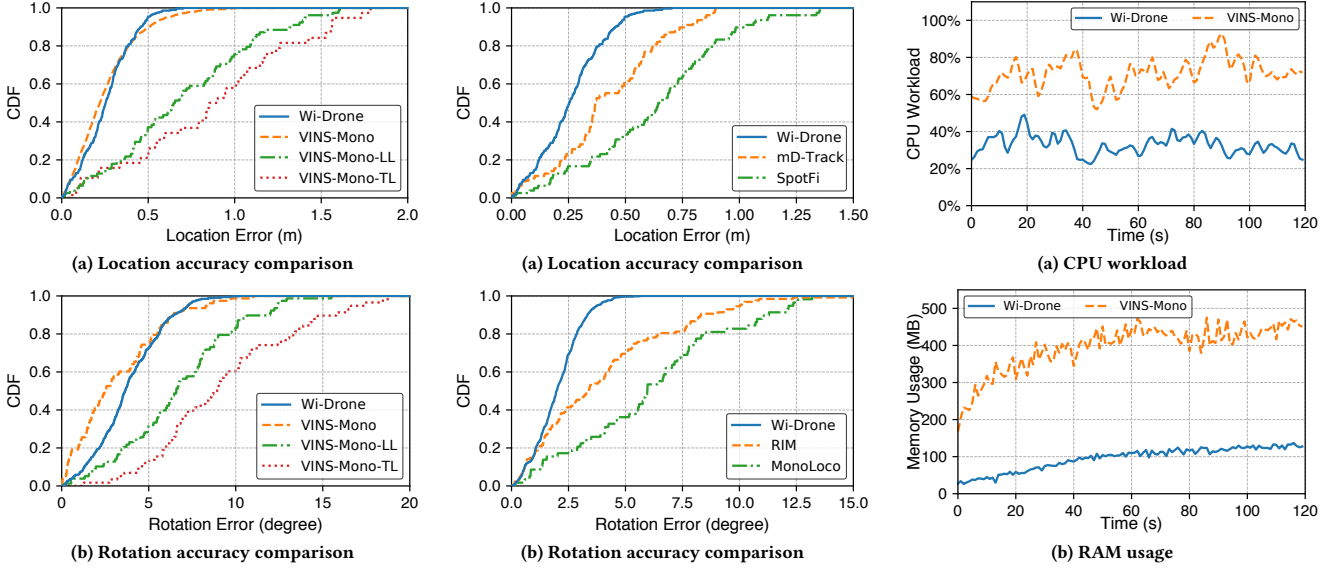
(a) Location accuracy comparison

(b) Rotation accuracy comparison

Figure 10: Accuracy comparison with VINS-Mono.

(a) Location accuracy comparison

(b) Rotation accuracy comparison

Figure 11: Accuracy comparison with related Wi-Fi-based solutions.

(a) CPU workload

(b) RAM usage

Figure 12: Resource consumption with VINS-Mono.

(a) Impact of flight speed   (b) Impact of path length   (c) Impact of AP number   (d) Impact of antenna number
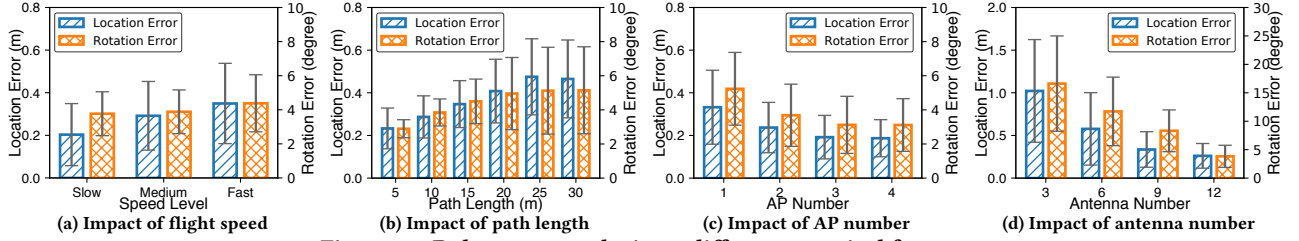
Figure 13: Robustness analysis on different practical factors.

compared with existing visual-inertial-based method, Wi-Drone is a competitive solution for indoor drone pose tracking.

7.2.2 Accuracy Comparison with Wi-Fi-based Solutions. We further demonstrate the superiority of our system compared to the state-of-the-art Wi-Fi-based tracking solutions. In this part, we compare Wi-Drone's location tracking performance with mD-Track and SpotFi and its azimuth tracking performance with RIM and MonoLoco. Specifically, we extend mD-Track and SpotFi to 3D space by applying multiple antennas and then compare their 3D location tracking results with Wi-Drone. Limited by the inherent drawbacks of the algorithm, RIM and MonoLoco are constrained to the 2D plane, so we focus on comparing the 2D azimuth tracking performance. As depicted in Fig. 11, Wi-Drone's average location accuracy is 26.1 cm, exceeding mD-Track and SpotFi by 40.9% and 59.0% respectively. Besides, Wi-Drone achieves an average 2.1° azimuth tracking accuracy, outperforming RIM and MonoLoco by 49.5% and 66.8%. The results demonstrate that Wi-Drone outperforms existing Wi-Fi-based tracking solutions in terms of both dimensionality and accuracy. Being the first Wi-Fi-based work capable of tracking an object's 6-DoF pose, Wi-Drone fully exploits the spatial-temporal CSI to infer the drone's states and optimize the trajectory, providing finer-grained pose estimations continuously.

7.2.3 Resource Consumption. Compared with the widely used visual-inertial 6-DoF tracking method, Wi-Drone takes fewer computing resources and thus is more suitable for devices with limited hardware on board. As shown in Fig. 12a, the CPU workload for running Wi-Drone is only 32.8%, outperforming VINS-Mono by about 40%. Our system also requires less memory, Fig. 12b depicts that Wi-Drone takes only 100 MB memory on average, achieving more than 3× memory saving compared with VINS-Mono.

7.3 Robustness Analysis

7.3.1 Impact of Flight Speed. To evaluate the robustness of Wi-Drone against different velocity, we conduct experiments on the same trajectories with various flight speeds. The flight speed is divided into three levels corresponding to different types of indoor drone applications. In Fig. 13a, the slow, medium, and fast indicate a flight speed of $v \le 2$ m/s, $2 < v \le 3$ m/s, and $3 < v \le 5$ m/s respectively. As shown, at three levels of flight speed, the average location error is 20.3 cm, 29.2 cm, and 34.9 cm, respectively. Similarly, the corresponding average rotation error is 3.8°, 3.9° and 4.4°. Although the accuracy of tracking performance slightly degrades as the flight speed increases, the tracking performance under a high flight speed (34.9 cm location accuracy with 4.4° rotation accuracy) still meets the needs of most indoor drone applications. Wi-Drone's
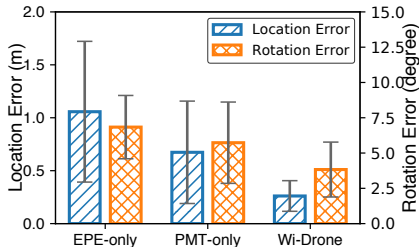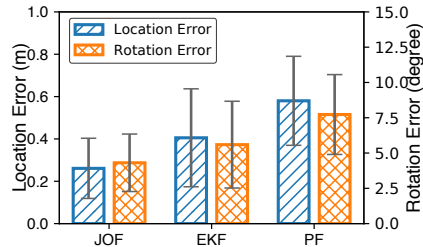
Figure 14: Component study.
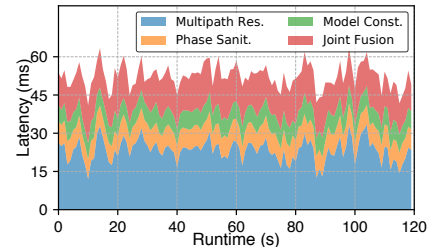

Figure 15: Fusion frameworks.

Figure 16: Latency analysis.

outstanding tracking performance at high-speed conditions benefits from its quick response mechanism due to the following facts: 1) the Wi-Fi NICs could transmit and receive up to 1000 packets per second; 2) the acceleration techniques significantly reduce the computational complexity of both EPE module and PMT module.

7.3.2 Impact of Path Length. Fig. 13b depicts that as the path length increases, the location tracking error and the rotation tracking error gradually converge to an upper bound of about 47 cm and 5.1° respectively. There is no significant performance degradation in both location accuracy and rotation accuracy when the path length increases from 25 m to 30 m. Such a good long-path tracking performance benefits from the advantages of the EPE module and the JOF module. Specifically, by fusing the result of the EPE module, Wi-Drone corrects the trajectory in real-time, preventing the location error from accumulating.

7.3.3 Impact of AP Deployment Density. By changing the number of APs monitoring the drone via the direct path from 1 to 4, we evaluate the performance of Wi-Drone with different AP deployment densities. Fig. 13c shows that as the AP number increases, the location error and the rotation error gradually converge to 18.7 cm and 3.1° respectively. The result also shows that even with only one AP, Wi-Drone achieves a satisfactory result of 32.2 cm location accuracy and 5.2° rotation accuracy, which ensures the robustness of the system being used in real-world scenarios.

7.3.4 Impact of Antenna Number. We further evaluate how the number of antennas affects the system accuracy. As shown in Fig. 13d, a larger number of antennas brings a significant accuracy improvement. The result shows that, as the antenna number increases from 3 to 12, the average location error reduces from 102.2 cm to 26.1 cm, and the average rotation error also reduces from 8.4° to 3.8°. The number of antennas improves the Wi-Drone accuracy in two aspects: 1) More antennas come with better AoA estimation results; 2) Independent measurements of multiple antennas form into the overdetermined equations (Eqn. 17) in the PMT module, which means a larger number of antennas helps solve the relative 6-DoF pose more accurately.

7.4 Micro Benchmarks

7.4.1 Effectiveness of Each Component. In this part, we evaluate the performance of EPE, PMT and further demonstrate the improvement brought by the joint optimization. As illustrated in Fig. 14, the performance of a fusion-based Wi-Drone far exceeds that of the EPE-only strategy and the PMT-only strategy in terms of location accuracy and rotation accuracy. Specifically, the location tracking accuracy for Wi-Drone, EPE, and PMT are 26.1 cm, 105.7 cm, and

67.4 cm, respectively, which means Wi-Drone reduces 75.3% and 61.5% location error compared with EPE and PMT. As for the rotation tracking, fusion-based Wi-Drone also outperforms EPE and PMT by 43.9% and 33.1%, respectively. The results indicate that Wi-Drone effectively incorporates the complementary characteristics of the two modules: the EPE module with AP location reference doesn't incur any cumulative drift, while the PMT module based on fine-grained phase has extremely high short-term accuracy.

7.4.2 Effectiveness of the Fusion Framework. To demonstrate the effectiveness of the Wi-Drone's fusion performance, we compare the proposed factor-graph-based fusion framework, JOF, with another two classic fusion approaches, extended Kalman filter (EKF) and particle filter (PF), both of which are widely used in previous works [16, 48]. As shown in Fig. 15, Wi-Drone based on the factor graph achieves enhanced location tracking accuracy for more than 35.6% and 55.0% compared with EKF and PF. This performance gain becomes 22.9% and 41.3% when tracking the rotation. The outstanding performance benefits from leveraging the factor graph in a nonlinear and tight-coupling manner. Specifically, both the PF and EKF are based on the approximation with local linearity, and thus falling short in complex nonlinear problem [8]. In contrast, the factor graph realizes efficient nonlinear optimization and fusion based on the maximum posterior principle, and estimates the final result from the error distribution of the initial measurements (e.g., CSI phase).

7.4.3 Latency Analysis. Wi-Drone is designed to be an effective solution for indoor drones' flight control, which makes its real-time performance critical. Fig. 16 illustrates the end-to-end latency of Wi-Drone through the whole tracking process, including multipath resolution, phase sanitization, EPE and PMT model construction, and JOF. As seen, the average end-to-end latency of Wi-Drone is 52.34 ms, with an average of 23.35 ms for multipath resolution and 13.91 ms for fusion. The evaluation results illustrate that the intensive nonlinear optimizations are still the main cause of system latency, even after various acceleration strategies. To conclude, the latency evaluation indicates Wi-Drone can run up to 20 FPS.

7.4.4 Loop Detection Analysis. To fully understand the necessity of the TRRS-based loop detection function, we evaluate Wi-Drone on the looped trajectories (i.e., #3 and #4 in Fig. 9) with/without loop detection. Our evaluation result shows that, with loop detection disabled, the location accuracy and rotation accuracy degrades by over 40% and 30%, respectively, showing the effect of the loop detection function. With a loop detection function, the global optimization will be triggered on a closed loop, bringing about global-consistent 6-DoF trajectories.

8 DISCUSSIONS AND FUTURE WORK

Wi-Drone is the first attempt towards ubiquitous and accurate Wi-Fi-based 6-DoF tracking, and there is room for continued research in various perspectives.

Antenna array. The current prototype of Wi-Drone employs several NICs to support multiple antennas. As Wi-Fi technology evolves, the latest 802.11be standard supports up to 16 antennas, and the size of the PCB antenna array will be smaller enough to be embedded into various mobile devices (e.g., VR/AR headsets, smart watches) and further enable a broad spectrum of applications. For the current prototype, Wi-Drone can already be integrated into the COTS drone for 6-DoF flight control.

LoS dependency. Like most Wi-Fi-based systems [5, 45], Wi-Drone relies on the direct path to reach the best tracking performance. Under NLoS conditions, Wi-Drone still works based on the PMT module and prior information, however, suffers severe performance degradation. In a nutshell, though occasional LoS blockage during the flight is tolerable for Wi-Drone, accurate Wi-Fi-based 6-DoF tracking under long-term NLoS conditions remains open for future research.

Multimodal fusion. Wi-Drone has a great potential to incorporate with existing modalities to achieve complementary benefits. As a Wi-Fi-standalone solution, Wi-Drone doesn't aim to completely replace existing 6-DoF tracking systems (e.g., VIO, LIO, etc.). Instead, Wi-Drone provides a novel pervasive modality for 6-DoF tracking, especially for those light-weight and resource-constraint applications. In application scenarios with sufficient computing resources and ideal lighting conditions, Wi-Drone can also be incorporate with other sensors (e.g., cameras and radars) with little efforts to achieve better performance.

9 RELATED WORK

We briefly review the related works in the following.

6-DoF pose estimation. As a key enabler of autonomous driving and AR/MR applications, numerous 6-DoF pose estimation solutions have been proposed over the past decade. Most existing solutions leverage various onboard sensors (e.g., IMU, LiDAR, mmWave radar, RGB, and RGB-D camera) to infer the device's 6-DoF pose change. Benefiting from low cost and ubiquity, IMU-based solutions [7, 11, 34], which infer the real-time rotation and motion from the built-in gyroscope and accelerometer, are widely used in various mobile platforms, however suffer from severe cumulative drift. Vision-based solutions have been proposed using RGB [9, 12] or RGB-D [13] cameras to track the device's relative pose by feature matching. Some recent works integrate the visual odometry and IMU into VIO [26, 29] and achieve better performance. However, vision-based solutions require significant computing resources and frequently lose tracking in low-light-level and visual-texture-less scenarios, limiting their scope of application, especially in indoor scenarios. Previous researches also introduce sophisticated sensors and infrastructures including LiDAR [33], mmWave radar [21], infrared ray [27, 38], and their fusions [39, 49]. However, both the price and resource overhead prevent them from ubiquitous deployment.

In contrast, Wi-Drone provides an accurate and light-weight 6-DoF tracking solution and is competent in any adverse lighting and texture scenarios, shedding light on pervasive 6-DoF tracking applications.

Wireless localization and tracking. As a pervasive and ubiquitous approach, wireless localization and tracking have gained great interest over the past decade. Existing works exploit a variety of channel parameters, such as AoA [18, 46], ToF [32, 37], and their fusion [4, 45, 51], to accomplish a variety of wireless sensing tasks. Most of them resort to extra hardware [15], frequency hopping [37], and protocol modifications [47] to achieve meter-level or decimeter-level precision. By further leveraging plane geometry [5], some pioneer works can even infer a device's 2D location and azimuth simultaneously [35]. Lacking motion information, the methods based on channel parameters can only form discrete locations instead of a continuous trajectory, which lead to severe performance degradation when tracking high speed moving devices. In contrast, a recent work, RIM [43], directly estimates the device's heading direction and moving distance by leveraging multiple antennas. Another work [19] leverages the relative ToF and attenuation change of the received signals to generate the device's fine-grained trajectory. These approaches achieve centimeter-level tracking accuracy over a relatively short distance. However, they tend to accumulate drift error, thus falling short in long-distance tracking scenarios.

The emerging UWB [10] and Bluetooth 5.1 standards provide high range resolution and AoA measurement capability, respectively. Some recent RFID-based tracking systems [22] achieve delightful performance in 3D space. Both of them, however, still lack the ability for simultaneous rotation and location tracking.

Compared with the aforementioned works that are either 2-DoF (e.g., 2D location) or 3-DoF (e.g., 3D location, 2D location plus azimuth), Wi-Drone pushes the limit of wireless device tracking dimensions by simultaneously estimating the device's rotation and location in 3D space, realizing 6-DoF tracking. Besides, by proposing the EPE, PMT, and JOF modules, Wi-Drone enables accurate tracking over both short and long trajectories, and thus are competent for more challenging scenarios.

10 CONCLUSION

This paper proposes the design and implementation of Wi-Drone, the first Wi-Fi-based 6-DoF device tracking system capable of indoor drone flight control. Wi-Drone exploits the idea of rigid-body coordinate transformation and further derives two different 6-DoF tracking algorithms by jointly utilizing the information of multipath and multiple antennas. Wi-Drone also adopts a novel paradigm of fusing two complementary Wi-Fi tracking modules to push the limit of the tracking accuracy. By doing so, Wi-Drone achieves accurate and robust 6-DoF pose tracking performance. Our practical experience reveals that Wi-Drone's accuracy is sufficient to support most drone-based applications in typical indoor scenarios.

11 ACKNOWLEDGMENTS

REFERENCES

[1] ArduPilot. 2016. https://ardupilot.org/ardupilot/

[2] K Somani Arun, Thomas S Huang, and Steven D Blostein. 1987. Least-squares fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1987).

[3] IEEE Standard Association. 2016. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *802.11-2016 - IEEE Standard for Information Technology* (2016). https://doi.org/10.1109/IEEESTD.2016.7786995

[4] Roshan Ayyalasomayajula et al. 2020. Deep learning based wireless localization for indoor navigation. In *Proceedings of the ACM MobiCom*.

[5] Roshan Ayyalasomayajula, Aditya Arun, et al. 2020. LocAP: Autonomous millimeter accurate mapping of WiFi infrastructure. In *Proceedings of the USENIX NSDI*.

[6] Marius Beul et al. 2018. Fast autonomous flight in warehouses for inventory applications. *IEEE Robotics and Automation Letters* (2018).

[7] Martin Brossard, Axel Barrau, and Silvère Bonnabel. 2020. AI-IMU dead-reckoning. *IEEE Transactions on Intelligent Vehicles* (2020).

[8] Cesar Cadena et al. 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* (2016).

[9] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics* (2021).

[10] Yifeng Cao et al. 2021. ITrackU: tracking a pen-like instrument via UWB-IMU fusion. In *Proceedings of the ACM MobiSys*.

[11] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. 2018. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[12] Guoxuan Chi, Jingao Xu, Jialin Zhang, Qian Zhang, Qiang Ma, and Zheng Yang. 2021. Locate, Tell, and Guide: Enabling Public Cameras to Navigate the Public. *IEEE Transactions on Mobile Computing* (2021).

[13] Weichen Dai, Yu Zhang, Ping Li, Zheng Fang, and Sebastian Scherer. 2020. RGB-D SLAM in dynamic environments using point correlations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[14] Jeffrey A Fessler et al. 1994. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing* (1994).

[15] Jon Gjengset, Jie Xiong, Graeme McPhillips, and Kyle Jamieson. 2014. Phaser: Enabling phased array signal processing on commodity WiFi access points. In *Proceedings of the ACM MobiCom*.

[16] Sejong Heo, Jaehyuck Cha, and Chan Gook Park. 2018. EKF-based visual inertial navigation using sliding window nonlinear optimization. *IEEE Transactions on Intelligent Transportation Systems* (2018).

[17] Zhiping Jiang, Tom H Luan, Xincheng Ren, Dongtao Lv, Han Hao, Jing Wang, Kun Zhao, Wei Xi, Yueshen Xu, and Rui Li. 2021. Eliminating the Barriers: Demystifying Wi-Fi Baseband Design and Introducing the PicoScenes Wi-Fi Sensing Platform. *IEEE Internet of Things Journal* (2021).

[18] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. Spotfi: Decimeter level localization using wifi. In *Proceedings of the ACM SIGCOMM*.

[19] Manikanta Kotaru and Sachin Katti. 2017. Position tracking for virtual reality using commodity WiFi. In *Proceedings of the IEEE/CVF CVPR*.

[20] Ying-Chih Lai and Shau-Shiun Jan. 2011. Attitude estimation based on fusion of gyroscopes and single antenna GPS for small UAVs under the influence of vibration. *GPS solutions* (2011).

[21] Chris Xiaoxuan Lu et al. 2020. milliEgo: single-chip mmWave radar aided ego-motion estimation via deep sensor fusion. In *Proceedings of the ACM SenSys*.

[22] Zhihong Luo et al. 2019. 3D backscatter localization for fine-grained robotics. In *Proceedings of the USENIX NSDI*.

[23] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. 2017. Indoor follow me drone. In *Proceedings of the ACM MobiSys*.

[24] Victor RF Miranda et al. 2021. Autonomous Navigation System for a Delivery Drone. *arXiv preprint arXiv:2106.08878* (2021).

[25] Raul Mur-Artal and Juan D Tardós. 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics* 33, 5 (2017), 1255–1262.

[26] Raúl Mur-Artal and Juan D Tardós. 2017. Visual-inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters* (2017).

[27] Opti-Track. 2016. https://optitrack.com/

[28] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2. 0: Passive human tracking with a single Wi-Fi link. In *Proceedings of the ACM MobiSys*.

[29] Tong Qin, Peiliang Li, and Shaojie Shen. 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* (2018).

[30] Amazon Ring. 2020. https://spectrum.ieee.org/ring-indoor-security-drone

[31] Rodrigues' rotation formula. 2016. https://en.wikipedia.org/wiki/Rodrigues%27_rotation_formula

[32] Souvik Sen, Jeongkeun Lee, Kyu-Han Kim, and Paul Congdon. 2013. Avoiding multipath to revive inbuilding WiFi localization. In *Proceeding of the ACM MobiSys*.

[33] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus. 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *2020 IEEE/RSJ IROS*. 5135–5142.

[34] Sheng Shen et al. 2018. Closing the gaps in inertial motion tracking. In *Proceedings of the ACM MobiCom*.

[35] Elahe Soltanaghaei et al. 2018. Multipath triangulation: Decimeter-level wifi localization and orientation with a single unaided receiver. In *Proceedings of the ACM MobiSys*.

[36] James Svacha, James Paulos, Giuseppe Loianno, and Vijay Kumar. 2020. Imu-based inertia estimation for a quadrotor using newton-euler dynamics. *IEEE Robotics and Automation Letters* (2020).

[37] Deepak Vasisht, Swarun Kumar, and Dina Katabi. 2016. Decimeter-level localization with a single WiFi access point. In *Proceedings of the USENIX NSDI*.

[38] VICON. 2016. https://www.vicon.com/

[39] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. 2018. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. *IEEE Robotics and Automation Letters* 3, 2 (2018), 994–1001.

[40] Csaba Virágh et al. 2016. Self-organized UAV traffic in realistic environments. In *2016 IEEE/RSJ IROS*.

[41] Jue Wang, Deepak Vasisht, and Dina Katabi. 2014. RF-IDraw: Virtual touch screen in the air using RF signals. *ACM SIGCOMM Computer Communication Review* (2014).

[42] Lukas Wawrla, Omid Maghazei, and Torbjørn Netland. 2019. Applications of drones in warehouse operations. *Whitepaper. ETH Zurich, D-MTEC* (2019).

[43] Chenshu Wu, Feng Zhang, Yusen Fan, and KJ Ray Liu. 2019. RF-based inertial measurement. In *Proceedings of the ACM SIGCOMM*.

[44] Yaxiong Xie et al. 2018. Precise power delay profiling with commodity Wi-Fi. *IEEE Transactions on Mobile Computing* (2018).

[45] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *Proceedings of the ACM MobiCom*.

[46] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: a fine-grained indoor location system. In *Proceedings of the USENIX NSDI*.

[47] Jie Xiong, Karthikeyan Sundaresan, and Kyle Jamieson. 2015. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proceedings of ACM MobiCom*.

[48] Jingao Xu, Hengjie Chen, Kun Qian, Erqun Dong, Min Sun, Chenshu Wu, Li Zhang, and Zheng Yang. 2019. ivr: Integrated vision and radio localization with zero human effort. *PACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2019).

[49] Jingao Xu, Guoxuan Chi, Zheng Yang, Danyang Li, Qian Zhang, Qiang Ma, and Xin Miao. 2021. FollowUpAR: enabling follow-up effects in mobile AR applications. In *Proceedings of the ACM MobiSys*.

[50] Milos Zefran, Vijay Kumar, and Christopher Croke. 1999. Metrics and connections for rigid-body kinematics. *International Journal of Robotics Research* (1999).

[51] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the ACM MobiSys*.