WiSion: Bolstering MAV 3D Indoor State Estimation by Embracing Multipath of WiFi

Fei Xiao, Shengkai Zhang, *Member, IEEE*, Sheyang Tang, *Student Member, IEEE*, Shaojie Shen, *Member, IEEE*, Huixin Dong, *Student Member, IEEE*, Yi Zhong*, *Member, IEEE*

***Abstract*—Recent years have witnessed a remarkable growth of micro aerial vehicle (MAV) technologies, which are desirable for many applications, *e.g.*, warehouse inventory management and home entertainment. These indoor needs significantly improve work efficiency while posing fundamental challenges in the design of MAV state estimation. The state includes a vehicle's position, orientation, and velocity, which are fundamental to guide the motor control and adjust that vehicle's actions in autonomous flight. Existing vision-based solutions only work in well-lit texture-rich environments, while laser-based solutions are limited to MAVs' payload and budget. This paper presents WiSion, a robust and low-cost state estimator that leverages ubiquitous WiFi to estimate six-degree-of-freedom states for MAVs. Our observation is that the multipath of WiFi conceals a wealth of information about a vehicle's state, which helps combat the temporal drift of inertial sensors for smooth state estimation. We realize WiSion by an absolute-relative WiFi sensing module and a WiFi-inertial state estimation module. It works without knowing access points' (APs') positions. We implement the prototype with off-the-shelf products and conduct experiments in indoor venues. Our results show that WiSion achieves the position error of 35.25 cm and the attitude error of 2.6° with a maximum linear velocity of 1.74 m/s. Moreover, WiSion can recover APs' positions and is robust to indoor hindrances such as obstacles and multipath.**

***Index Terms*—State estimation, autonomy, wireless localization, multi-sensor fusion**

I. Introduction

Micro aerial vehicle (MAV) technology has undergone an impressive expansion in its applications over recent years, as this once previously exclusive technology for military purposes has found a thriving purpose in civilian scenarios such as warehouse inventory management [1]–[3] and home entertainment [4], [5]. Specifically, it has been speculated that MAV automation in warehouses will grow by 29 billion by 2027, with an annual growth rate of almost 20% [6]. Meanwhile, home entertainment is one of the major segments in the drone service market, projected to reach USD 63.6 billion by 2025 [7]. The autonomy of MAVs is the critical demand for their intelligent applications. MAVs that inspect items along aisles in a warehouse are usually blocked by stocks, making human control inefficient and requiring autonomous operations. Kids' drones [8] are required to be cheap and intelligent to interact with humans autonomously. State estimation [9], [10] is fundamental to these autonomous operations. The state, including the vehicle's position, orientation, and velocity, is the key information to the flight control module, adjusting the rotary speed of rotors to achieve desired actions.

Conventional methods for MAV state estimation include global navigation satellite system (GNSS) based, vision-based, and laser scanner based approaches. Since GNSS signals are blocked in indoor applications, the mainstream uses cameras or laser scanners for state estimation [11]–[17]. Laser-based approaches can provide highly accurate results, but they are not suitable to be installed on MAVs (power < 100 W) [18] due to their expensiveness and heavyweight. A laser scanner will substantially increase the production cost and reduce the battery life. On the other hand, although vision-based approaches are low-cost and lightweight, they are limited to well-lit and texture-rich environments as they require enough visual features to recover camera poses via projective geometry. Applications for warehouse management and home entertainment cannot guarantee that MAVs always fly in vision-friendly environments. Typically, warehouses lack well lighting due to their high ceilings, stacks shading, and energy conservation [19]. Home, there can be textureless objects such as white walls and solid color floors [20].

To ease the problem of visual sensing, recent progress in radio frequency (RF)-based state estimation demonstrates that RF signals, *e.g.*, LoRa [21], Ultra-wideband (UWB) [22], [23], and WiFi [24], [25], bring new opportunities to the MAV state estimation that are highly resilient to visual limitations. However, none of the existing works has made it into the domestic use-cases of MAVs that requires: 1) a **low-cost** sensing module that fits the compactness and helps MAVs go into a mass-market; 2) an **instantly deployable** system that works in any unknown venues without any endeavor to setup; 3) a **6-DoF** (six degrees of freedom) state estimator that fully supports the MAV mobility in 3D space. Specifically, UWB and LoRa based approaches [22], [23], [26] require specialized hardware, violating the first and second requirements. The state-of-the-art WiFi-based state estimator [25] only supports

F. Xiao is with Wuhan University of Technology and Huazhong University of Science and Technology, Wuhan, China (e-mail: feixiao@hust.edu.cn).

S. Zhang and Y. Zhong are with Wuhan University of Technology, Wuhan, China (e-mail: {shengkai, zhongyi}@whut.edu.cn).

S. Tang is with University of Waterloo, Waterloo, Canada (e-mail: s99tang@uwaterloo.ca).

S. Shen is with Hong Kong University of Science and Technology, Hong Kong, China (e-mail: eeshaojie@ust.hk).

H. Dong is with Huazhong University of Science and Technology, Wuhan, China (e-mail: huixin@hust.edu.cn).

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant WUT: 223109003, the Research Project of Wuhan University of Technology Chongqing Research Institute under Grant YF2021-06, the National Natural Science Foundation of China under Grant 62002104.

*The corresponding author is Yi Zhong (zhongyi@whut.edu.cn).

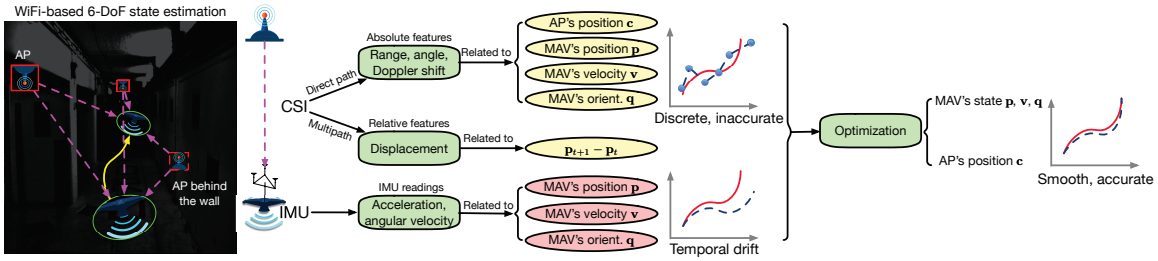


Fig. 1. Commodity WiFi assists a MAV to autonomously fly through a dark corridor where visual sensing is crippled.

the MAV mobility in 2D space and requires that WiFi access points (APs) are installed at the same altitude as the MAV, failing to meet the last requirement.

This paper proposes WiSion, a new state estimator that satisfies all the above requirements. WiSion allows autonomous navigation in any environment where few WiFi APs are available. The system uses a WiFi card, which is indispensable to a MAV for communications. It does not need prior knowledge of APs' positions in the venue and supports 6-DoF state estimation for 3D mobility.

The intuition of using WiFi signals for state estimation is that we can view connected APs as environmental features, holding the cues of MAV positions and attitudes like the visual features in vision-based approaches. Compared to numerous but volatile visual features from camera images, APs are few but persistent. APs can even be connected out of sight (behind walls), while visual features are limited in the field of camera view. We thoroughly explore the encoded geometric features from the direct path and the multipath reflections of WiFi signals and find that these features are all related to 6-DoF MAV state and APs' positions, as illustrated in Fig. 1. Although only WiFi signals may not provide smooth and accurate state estimates required by the MAV control, such an exteroceptive sensing modality is promising to achieve this goal by fusing with an inertial measurement unit (IMU). The sensor fusion aims to use WiFi sensing to correct the IMU's temporal drift, mimicking the working principle of VIO (Visual-Inertial Odometry) [12]–[14].

Realizing WiSion poses two challenges, thereby designing two components to address them:

Absolute-relative WiFi sensing: The most recent WiFi tracking technique, mD-Track [27], has shown that WiFi signals can estimate rich geometrical parameters for localization. Besides multipath in indoor venues, another challenge for MAVs is that the *absolute* features, which connect the current state with the sensor output, such as range, angle-of-arrival (AoA), and Doppler shift, are corrupted by MAV rotations. We, therefore, propose an inertial-based algorithm to extract the absolute features correctly. Also, due to the noise of absolute features, we propose an algorithm that extracts the *relative* feature, *i.e.*, displacement, which connects the current and past states with the sensor output. The relative feature reaches centimeter-level accuracy and imposes additional geometrical constraints to the vehicle's state, thus producing more accurate state estimation.

WiFi-inertial state estimation: Without knowing APs' positions and the noise of the above WiFi sensing measurements make state estimation still challenging. Recent simultaneous localization and mapping (SLAM) [28], [29] and multi-dimensional scaling (MDS) [30] based approaches require visual assistance or fail to deliver accurate results in real-time. We leverage the MAV's onboard IMU that measures the accelerations and the angular rate of its motions to estimate the APs' positions and the 6-DoF state simultaneously in real-time. Our WiFi-inertial state estimation employs a nonlinear optimization framework to take the WiFi sensing information to correct the IMU's temporal drift without cumbersome initialization.

Results. We implement WiSion on a DJI M100 equipped with an IMU and an Intel 5300 WiFi card on an Intel Next Unit of Computing (NUC). We use the 802.11 channel state information (CSI) tool [31] to obtain wireless channel information for absolute-relative WiFi sensing. The experiments are conducted in a 12×8 m^2 underground drone test site to validate individual system modules and overall performance. The results show that WiSion achieves the MAV's position error of 35.25 cm and the attitude error of 2.6° with a maximum flying velocity of 1.74 m/s. The indoor performance outperforms outdoor commercialized GPS-based state estimation systems [32]. Moreover, WiSion can recover the APs' positions with an accuracy of 39.7 cm, and it is robust to indoor hindrances such as obstacles and multipath.

Contributions. WiSion is a novel state estimator that leverages ubiquitous WiFi to deliver accurate 6-DoF MAV states in real-time. It fuses the absolute-relative WiFi measurements, including range, AoA, Doppler shift, and displacement, with IMU measurements. WiSion consists of a comprehensive WiFi sensing module and a new WiFi-inertial state estimator to simultaneously estimate the vehicle's state and surrounding APs' positions without cumbersome initialization. We implement WiSion on commodity devices and experimentally validate the system in indoor venues.

The remainder of this paper is organized as follows. Section II summarizes related works. Then we introduce the system overview in Section III. Section IV elaborates on the WiFi sensing module that extracts the absolution and relative features. The WiFi-inertial state estimation is presented in Section V. Section VI evaluates the performance of individual modules as well as the integrated system. Finally, we conclude this work in Section VII.

II. Related Work

State estimation in GNSS-denied environments has been well studied in the robotics community. The mainstream uses laser scanners [15]–[17] or cameras [12]–[14], [33]. However, a laser scanner, *e.g.*, LiDAR, is typically not suitable for MAVs as its heavyweight and high cost will substantially reduce the battery life and increase the manufacturing cost. On the other hand, monocular cameras are more acceptable to MAVs as the vision-based approaches are lightweight and accurate. Nevertheless, they demand well lighting and rich texture, which cannot be guaranteed in a general use case, *e.g.*, a house with white walls or a dim warehouse. In summary, LiDAR and cameras use light as a medium of perception, sensing the world in a line-of-sight (LOS) fashion. Exploring non-LOS (NLOS) sensing modalities, whose sensing medium can traverse through visual blockages, may overcome the limitations of these optical-based solutions.

Recent years have witnessed advances in versatile applications using RF sensing, *e.g.*, WiFi imaging [34]–[36], WiFi localization and tracking [27], [37]–[40], health monitoring [41], [42], and object detection [43]–[45]. These bring the opportunities to make RF signals a powerful sensing modality that complements optical-based sensing. RF signals can traverse through obstacles or penetrate walls due to their large wavelength, being highly resilient to the visual/laser limitations.

Researchers have proven the feasibility of using RF signals, such as WiFi [24], [25], [46], LoRa [21], and UWB [47]–[49], to assist the MAV state estimation. H. Xu *et al.* [49] leveraged the omnidirectional sensing capability of UWB to overcome the field of view limitation of cameras for aerial swarm state estimation. S. Zhang *et al.* [21] exploited the high sensitivity in signal decoding of LoRa signals to achieve long-range/through-wall MAV state estimation with a single controller. Nevertheless, they require dedicated signal sources such as UWB nodes and LoRa transceivers to support the system. B. Li *et al.* [24] proposed CWISE, a WiFi-based state estimator that uses the phases of WiFi signals to correct the IMU drift, being the first feasible MAV state estimator using commodity WiFi. However, its primitive signal processing method cannot resolve the indoor multipath. S. Zhang *et al.* [25] proposed WINS, an indoor WiFi-based pose estimation that resolves multipath and allows MAVs to fly in indoors. However, WINS only supports state estimation in 2D space due to the limitation of a linear antenna array and the exponential computation complexity of their parameter estimation. R. Venkatnarayan *et al.* [46] proposed WIO (WiFi-Inertial Odometry), which is the closest to our work. It leverages WiFi CSI to estimate the moving distance and then fuses with IMU by Kalman filter. Although it provides promising results in tracking positions, it only addresses 2D localization and does not consider the effect of rotations. Moreover, the Kalman filter can only combat the Gaussian noise for linear systems. Due to the presence of rotation, MAV dynamic systems are highly nonlinear. Even the extended Kalman filter fails to solve our problem because it is susceptible to initialization. Early fixing linearization points leads to suboptimal results.

Since ubiquitous WiFi is promising to make MAV into domestic scenarios, we dig deeper into WiFi sensing and propose WiSion, an extraordinarily lightweight and low-cost system that achieves accurate 3D 6-DoF state estimation. WiSion designs new algorithms to 1) efficiently extract multi-dimensional parameters to enable robust MAV 3D localization and orientation estimation; 2) automatically initialize the system; 3) effectively solve the nonlinearity of 6-DoF state estimation to improve the accuracy over a long-term run. Dive into WiSion the closest to the WiFi sensing module are mD-Track [27] and WiCapture [38]. The fundamental difference in our context is that a MAV can rotate and change its heading during the flight. Such a rotation adds rotational Doppler shifts and corrupts the extracted feature referring to the MAV state. Therefore, we formulate new models that incorporate MAV rotations and design new algorithms that utilize the onboard IMU to solve the problem.

III. System Overview

WiSion allows low-cost and plug-and-play 3D navigation for MAVs in indoors by 6-DoF state estimation using commodity WiFi. Fig. 2 shows WiSion's system block diagram. Two sensors on the MAV are functional in WiSion, a WiFi network card and an IMU. The network card receives WiFi signals from multiple ($\geq$ 1) APs, whose positions are initially unknown, to extract the absolute and relative measurements, including range, angle, Doppler shift, and displacement. Then WiSion takes them with IMU measurements to produce smooth and accurate MAV states. Note that before running the WiFi sensing module, we use a Vector Network Analyzer to conduct a one-time calibration of transceiver responses of WiFi chips. The calibration can substantially reduce the frequency and phase offset of signals from radio chains [38].

WiSion consists of two components.

The first component is **WiFi sensing**. This module takes the calibrated CSI to estimate the parameters of signal paths related to the absolute and relative features of MAV states. The absolute features include range, AoA, and Doppler shift, encoding the vehicle's absolute positions and velocities. The relative feature is displacement, imposing spatial constraints of positions over time. These features provide comprehensive cues to determine a MAV's position, velocity, and orientation.

Sensor fusion is the second component. This module fuses the WiFi sensing with the IMU measurements, including accelerations and angular velocities, transmitting through a Universal Asynchronous Receiver-Transmitter (UART). Its design purpose is to use the drift-free WiFi measurements to correct the IMU's temporal drift. It employs a factor graph and formulates a nonlinear optimization problem that simultaneously estimates 6-DoF MAV states and surrounding APs' positions. Solving a nonlinear optimization problem requires a good initialization point to bootstrap iterative algorithms, *e.g.*, the Gauss-Newton solver. We propose an automatic method to find a good initialization point efficiently to ease the initialization.

The proposed algorithms are run in the airborne computer, *i.e.*, the Intel NUC. Finally, we visualize the state estimate, including the MAV's position, velocity, orientation, and con-

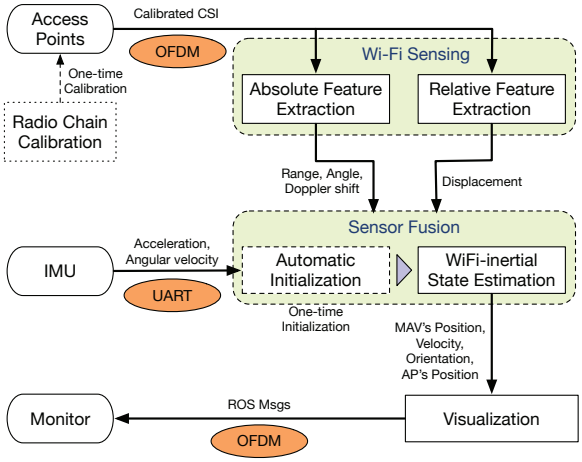


Fig. 2. System block diagram.

nected APs' positions, by sending these data to the user's computer using WiFi.

IV. Absolute-relative WiFi Sensing

This section describes how WiSion obtains the absolute motional features, *i.e.*, range, AoA, and Doppler shift, as well as the relative motional feature, *i.e.*, displacement, between consecutive packets.

TABLE I
Mathematical notation in the WiFi sensing.

Symbol	Description
$\hat{(\cdot)}$	the quantity that can be measured/estimated
$\mathbf{x}_t/\mathbf{x}(t)$	the MAV's state at time t
$\mathbf{n}_t/\mathbf{n}(t)$	random noise at time t
$\mathbf{Q}_t$	the covariance of measurements
θ/θ_k	the azimuth angle of the direct path/k^{th}path
ϕ/ϕ_k	the elevation angle of the direct path/k^{th}path
ρ/ρ_k	the range of the direct path/k^{th}path
η/η_k	the Doppler shift along the direct path/k^{th}path
r_k	the signal amplitude of the k^{th}path
c	the light speed in the air
d	the uniform spacing between antennas
λ	the signal wavelength
$\boldsymbol{\omega}_t$	the angular velocity at time t
Δt	the time interval between consecutive AoAs
$\mathbf{R}$	the rotation matrix from time k to current t
η_{r_i}	the rotational Doppler shift w.r.t. i^{th}antenna
$\mathbf{l}$	the displacement between consecutive packets
$\mathbf{\Gamma}$	channel state information matrix
Δf	time-variant frequency offset
$\mathbf{H}$	steering matrix
$\mathbf{A}$	the matrix of complex attenuations

Table I describes the mathematical notation used in the WiFi sensing, listed in the order they appear in the text.

A. Absolute Feature Extraction

Absolute features can be modeled in the following form:

$$\mathbf{z}_t = f_{\text{abs}}\left(\mathbf{x}_t, \mathbf{n}_t\right), \tag{1}$$

where $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ is the measurement noise and $\mathbf{Q}_t$ denotes the covariance of measurements. Variable $\mathbf{x}_t$ denotes the current state, $\mathbf{z}_t$ the sensor output. Function $f_{\text{abs}}(\cdot)$ is a general nonlinear function. An absolute feature connects the current state with the sensor output. In particular, we are interested in ranges, angles, and Doppler shifts. A range and an angle with respect to an AP can determine the MAV's position. And the Doppler shift is related to the velocity.

The range and angle that we need are concealed from multipath refections. Suppose there are K paths of the signal propagation. The received signal can be expressed by multi-dimensional parameters.

$$\mathbf{y}(t) = \sum_{k=1}^{K} r_k e^{j2\pi\eta_k t}\mathbf{h}(\theta_k, \phi_k)x(t-\rho_k/c) + \mathbf{n}(t), \tag{2}$$

where $x(t)$ denotes the transmitted signal at time t, $\mathbf{n}(t)$ 3-dimensional Gaussian noise, c the light speed. Variable $r_k, \mathbf{h}(\theta_k, \phi_k), \eta_k, \theta_k, \phi_k, \rho_k$ denote the amplitude, steering vector, Doppler frequency shift, azimuth angle, elevation angle, and range of k^{th}path, respectively. The absolute features that contain cues of a MAV's state are the direct-path parameters, denoted by η, θ, ϕ, ρ, as shown in Fig. 3. Since the signal amplitude r_k plays no role in our design, we omit this variable in the following text for brevity.

There have been many approaches that address the multipath challenge to find the direct-path component for localization [27], [38], [39]. The critical issue for MAVs is that the vehicle's rotation corrupts the antenna observed AoA and causes rotational Doppler shifts that degenerate the accuracy of the parameter estimation. Therefore, we employ a joint parameter estimation methodology [27] and design a new algorithm to address this issue.

AoA estimation. A MAV requires azimuth and elevation AoAs to determine 3D positions. Thus, we equip the receiver with a circular array of 3 antennas, having a uniform spacing of d between each other. Considering the signal received along k^{th}path, the steering vector $\mathbf{h}(\theta_k, \phi_k)$ of the circular array is

$$\mathbf{h}(\theta_k, \phi_k) = [1, \Phi(\theta_k, \phi_k), \Psi(\theta_k, \phi_k)]^\top, \tag{3}$$

where

$$\Phi(\theta_k, \phi_k) = e^{-j2\pi d \sin(\phi_k)\cos(\theta_k + \frac{\pi}{3})/\lambda},$$
$$\Psi(\theta_k, \phi_k) = e^{-j2\pi d \sin(\phi_k)\cos(\theta_k)/\lambda}.$$

λ is the wavelength.

The AoA estimation steers the steering vector by scanning the incident angle (θ_k, ϕ_k) and removes the phase differences at the antennas of an array due to different propagation lengths. The incident angle that aligns signals across all antennas in time and produces the maximum signal strength is the AoA estimate. However, the obtained AoA cannot localize the MAV due to its rotations. Since the circular array is fixed on the vehicle, its rotations change the AoAs without any translations. To obtain the AoA that can localize the vehicle, we have to compensate for the rotation upon the AoA estimate.

We integrate the angular velocity $\hat{\boldsymbol{\omega}}_t \in \mathbb{R}^3$ measured by the onboard IMU to compute the rotation between two consecutive AoAs. Since the IMU data rate ($\approx$ 100 Hz) is much higher

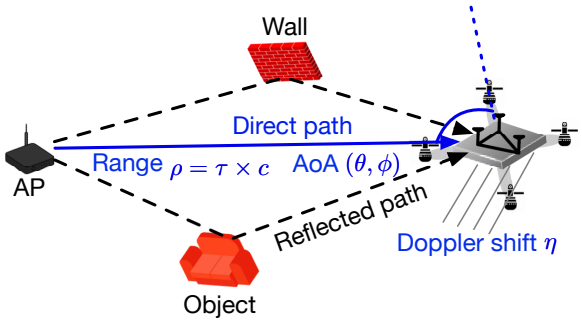


Fig. 3. The absolute features of signal paths related to the MAV state. Variable τ denotes the signal's traveling time between the AP and the vehicle.

than the AoA rate ($\approx$ 10 Hz), there are multiple IMU readings between two consecutive AoAs. Then the rotation of the vehicle between AoA i and $i+1$ can be computed by

$$\hat{\mathbf{R}}^i_{i+1} = \int_{\Delta t} \lfloor \hat{\omega}_t \times \rfloor \mathrm{d}t, \tag{4}$$

where $\lfloor \hat{\omega}_t \times \rfloor$ is the skew-symmetric matrix from the angular velocity $\hat{\omega}_t$. This solution is effective because the time interval between consecutive AoAs Δt is typically very short (less than 100 ms). The temporal drift of short-term IMU integration is negligible [50]. Then we convert the rotation matrix $\hat{\mathbf{R}}^i_{i+1}$ to angles and correct the AoA estimates.

Note that due to the noisy CSI from WiFi cards, the three-antenna circular array has limited multipath resolvability. Therefore, there may exist multiple candidates that meet the geometrical constraints in the AoA estimation. Nevertheless, the following joint parameter estimation can help eliminate false candidates as they will cause geometrical contradictions among the parameters such as Doppler shift and range estimates.

Doppler shift and range estimation. A MAV is capable of rotating and changing its heading directions during its flight. Such motions make the onboard antennas experience the Doppler frequency shift from translations and rotations, as shown in Fig. 4. The Doppler shift for each antenna combines translational shift η and rotational shift η_r. The rigid body makes all antennas on the vehicle experience the same translational velocity but the rotational velocities with different directions. Thus, the i^{th}antenna has a frequency shift

$$\eta_i = \eta_t + \eta_{r_i},\ i = \{1, 2, 3\}. \tag{5}$$

To eliminate the effect of such a shift, we need to multiply the received signal from i^{th}antenna by $e^{-j2\pi(\eta+\eta_{r_i})t}$. There are 4 parameters for our three-antenna circular array, making the estimation intractable. Fortunately, based on the geometry of the rotation, the sum of the rotational velocity is zero, *i.e.*, $\sum_i \mathbf{v}_{r_i} = \mathbf{0}$ and thus $\sum_i \eta_{r_i} = 0$. Taking the cubic root of the product of the received signals over all the antennas can eliminate the effect of the rotational shifts as

$$\begin{aligned} s(t,\eta) = \left(\prod_{i=1}^{3} y_i(t,\eta,\eta_{r_i})\right)^{\frac{1}{3}} &= \left(\prod_{i=1}^{3} e^{j2\pi(\eta+\eta_{r_i})t} x'(t)\right)^{\frac{1}{3}} \\ &= e^{j2\pi\eta t} x'(t), \end{aligned} \tag{6}$$

where $x'(t)$ absorbs $\mathbf{h}(\theta,\phi)$ in (2). Therefore, we can cancel the Doppler shift by multiplying $s(t,\eta)$ by $e^{-j2\pi\eta t}$:

$$y'(t,\eta) = e^{-j2\pi\eta t} s(t,\eta). \tag{7}$$

We take this corrected signal to estimate the Doppler shift and the range jointly.

The range is computed by multiplying the time-of-flight (ToF) of RF signals with the light speed in the air, thus the ToF is ρ/c. We take an autocorrelation between the corrected signal and the delay reversed signal, then find the parameter configuration that results in the maximum correlation. Specifically,

$$z(\eta,\rho) = \int_T y'(t,\eta) x^* \left(t - \frac{\rho}{c}\right) \mathrm{d}t, \tag{8}$$

where T denotes the signal duration. We can find a peak in $|z|$ when the shift and the delay are corrected.

Put it together. We combine all the above parameter estimation into a comprehensive estimator as follows.

$$z(\theta,\phi,\eta,\rho) = \int_T e^{-j2\pi\eta t} \mathbf{h}^H(\theta,\phi) s(t,\eta) x^* \left(t - \frac{\rho}{c}\right) \mathrm{d}t. \tag{9}$$

Then the parameter estimation can be obtained by

$$(\hat{\theta},\hat{\phi},\hat{\eta},\hat{\rho}) = \underset{(\theta,\phi,\eta,\rho)}{\arg\max} \left| z(\theta,\phi,\eta,\rho) \right|. \tag{10}$$

We instead employ a sequential estimation strategy to avoid the exponential computation complexity in the combinatorial parameter estimation. In practice, however, it works as fine as the exhaustive parameter search because these parameters are partially decoupled. Precisely, the AoA estimation uses the phase difference among antennas. On the other hand, the Doppler shift and range have equal impacts on the signal phases across all antennas, independent of the AoA estimation. Therefore, we first perform the AoA estimation, then fix the AoA and estimate the Doppler shift and range.

The above models only consider one path in signal propagation. In indoors, the received signal is the superposition of signals from multiple reflections, as shown in Fig. 3. We use the multipath resolving technique in [27] to find the absolute features of the direct path.

B. Relative Feature Extraction

Although the above absolute features have already encoded 3D position, the accuracy is inferior due to the narrow bandwidth of WiFi signals. We need to improve positioning accuracy, which will substantially benefit the state estimation as the velocity is the first derivative of position. To this end, we propose an algorithm that extracts a relative position-related feature, *i.e.*, displacement, to impose spatial constraints over time. Unlike the absolute features, the relative feature connects the current and past states with sensor outputs. It can be modeled as:

$$\mathbf{z}_t = f_{\text{rel}}(\mathbf{x}_t, \mathbf{x}_{t-k}, \mathbf{n}_t), \tag{11}$$

where $k > 0$, $\mathbf{x}_{t-k}$ denotes the past state at time $t-k$ before the current time t. Function $f_{\text{rel}}(\cdot)$ is a general nonlinear function that represents the mapping between the current state, a past state, sensor noise, and sensor outputs. In particular,

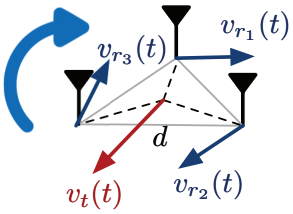


Fig. 4. Translational and rotational Doppler shifts.

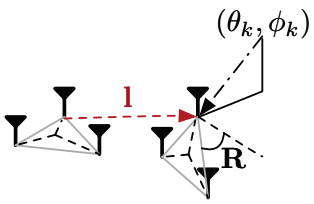


Fig. 5. Displacement with rotations.

the displacement is a special relative feature where $k = 1$. It measures the incremental change between two consecutive positions.

We first investigate the channel change from a small displacement $\mathbf{l} \in \mathbb{R}^3$ between two consecutive packets. Suppose there are K paths the signal propagates to the receiver. Consider one antenna of the array as shown in Fig. 5, the additional traveling distance of k^{th}path signal is $(\mathbf{R}\mathbf{r}_k)^\top \mathbf{l}$, where $\mathbf{R} \in \text{SO}(3)$[1] is the rotation matrix between two consecutive packets which can be obtained by (4), $\mathbf{r}_k$ the unit vector along the direction of the signal path (θ_k, ϕ_k), $\mathbf{r}_k = [\sin\theta_k \cos\phi_k, \cos\theta_k \cos\phi_k, \sin\phi_k]^\top$. This shifts the phase of the second packet by $2\pi(\mathbf{R}\mathbf{r}_k)^\top \mathbf{l}/\lambda$. Extending (3) to multipath has the channel model

$$\hat{\mathbf{\Gamma}} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]\mathbf{A} = \mathbf{H}\mathbf{A}, \tag{12}$$

where K is the number of paths, $\mathbf{h}_k$ the abbreviation of $\mathbf{h}(\theta_k, \phi_k)$, $\mathbf{H} \in \mathbb{C}^{3\times K}$ the steering matrix, $\mathbf{A} \in \mathbb{C}^{K\times 1}$ the matrix of complex attenuations. Matrix $\hat{\mathbf{\Gamma}} \in \mathbb{C}^{3\times 1}$ is the channel state information (CSI) matrix, which is reported by WiFi card. From (12), we have the channel of the second packet

$$\hat{\mathbf{\Gamma}}' = \mathbf{H}'\mathbf{A}' = \mathbf{H}'\mathbf{L}\mathbf{A}e^{j\Delta f} = \mathbf{H}'\mathbf{L}^f\mathbf{A}, \tag{13}$$

where $\mathbf{A}'$ is the matrix of attenuations of the second packet, $\mathbf{L} \in \mathbb{C}^{K\times K}$ a diagonal matrix with $e^{-j2\pi(\mathbf{R}\mathbf{r}_1)^\top \mathbf{l}/\lambda}$, $\cdots$, $e^{-j2\pi(\mathbf{R}\mathbf{r}_K)^\top \mathbf{l}/\lambda}$. Variable Δf is the time-variant frequency offset due to the clock drift between the transmitter and receiver since they are not synchronized [27], [52]. Variable $\mathbf{L}^f$ absorbs $e^{j\Delta f}$ into $\mathbf{L}$. Notice that in MAV scenarios, even a small displacement can result different steering matrices of two consecutive packets due to the vehicle's rotation.

In this model, λ is known, the MAV rotation $\mathbf{R}$ and AoAs of multiple paths $\mathbf{r}_1, \cdots, \mathbf{r}_K$ can be obtained by § IV-A. Then, the steering matrix $\mathbf{H}$ and $\mathbf{H}'$ can be computed from $\mathbf{R}$ and AoAs.

To estimate the displacement $\mathbf{l}$, we need to cancel the time-variant offset Δf. The offset Δf changes over different packets. However, it imposes the same effect on different paths when receiving a packet. Specifically, i^{th}diagonal entry of $\mathbf{L}^f$ is $L_i e^{-j2\pi(\mathbf{R}\mathbf{r}_i)^\top \mathbf{l}/\lambda + \Delta f}$ where L_i is the magnitude. Thus, the following operation on any two diagonal entries, $\mathbf{L}_{i,i}$ and $\mathbf{L}_{j,j}$, $i \neq j$, cancels Δf:

$$\begin{aligned}\frac{\mathbf{L}_{i,i}}{\mathbf{L}_{j,j}} &= \frac{L_i}{L_j} e^{-j\frac{2\pi}{\lambda}(\mathbf{R}\mathbf{r}_i)^\top \mathbf{l} + \Delta f - (-j\frac{2\pi}{\lambda}(\mathbf{R}\mathbf{r}_j)^\top \mathbf{l} + \Delta f)} \\ &= \frac{L_i}{L_j} e^{-j\frac{2\pi}{\lambda}\left(\mathbf{R}(\mathbf{r}_i - \mathbf{r}_j)\right)^\top \mathbf{l}}.\end{aligned} \tag{14}$$

Thus, we take the phase of both sides of (14) and have

$$s_{i,j} = \text{Phase}\left(\frac{\mathbf{L}_{i,i}}{\mathbf{L}_{j,j}}\right) = -\frac{2\pi}{\lambda}\left(\mathbf{R}(\mathbf{r}_i - \mathbf{r}_j)\right)^\top \mathbf{l}. \tag{15}$$

From (12), (13), and (15), we consider K paths of the signal propagation and construct the following problem:

$$\begin{aligned}&\arg\min_{\mathbf{L}^f, \mathbf{l}} \|\mathbf{A}' - \mathbf{L}^f\mathbf{A}\| + \|\mathbf{Q}\mathbf{R}^\top \mathbf{l} - \mathbf{s}\| \\ &\text{subject to } \mathbf{L}^f \text{ is diagonal,}\end{aligned} \tag{16}$$

where $\mathbf{s} \in \mathbb{R}^{{}^K C_2}$, ${}^K C_2 = \frac{K!}{2!(K-2)!}$, whose entries are from the combination of $s_{i,j}$, and $\mathbf{Q} \in \mathbb{R}^{{}^K C_2 \times 3}$, whose rows are from the combination of $-\frac{2\pi}{\lambda}(\mathbf{r}_i - \mathbf{r}_j)^\top$. When there are multiple APs, matrix $\mathbf{Q}$ and vector $\mathbf{s}$ obtained from multiple access points are concatenated vertically. The system (16) is linear so that it can be efficiently solved [53] to obtain displacement $\mathbf{l}$ between consecutive packets.

V. WiFi-inertial State Estimation

So far, we have discussed how WiSion measures absolute-relative features related to a MAV state via WiFi. The features, including AoA, range, Doppler shift, and displacement, are sufficient to obtain the MAV state. However, the result is dissatisfactory in accuracy and smoothness due to the noise of WiFi CSI. This section leverages the onboard IMU that measures 3D acceleration and angular velocity to achieve accurate and smooth state estimation.

Our key idea is to take the drift-free features to combat the IMU drift for better accuracy over a long-term run. Although IMU provides reliable odometry measurements, the measured accelerations and angular velocities induce the temporal drift of state estimation as the error can propagate and accumulate with integration. On the other hand, the WiFi-based absolute-relative features are less accurate, but they are drift-free in that the position, orientation, and velocity can be inferred by the features measured at the current time, with no error propagation from past data. We will fuse WiFi and IMU measurements to improve the accuracy of state estimation and combat the IMU's temporal drift.

A. Problem Formulation

Table II describes the mathematical notation used in the WiFi-inertial state estimation, listed in the order they appear in the text.

Typical sensor fusion techniques can be categorized by filtering-based approaches and graph-based approaches. Although filtering-based approaches, *e.g.*, extended Kalman filter, particle filter, and unscented Kalman filter, have the advantages of fast processing thanks to its continuous marginalization of past states, their performance can be suboptimal due to

[1] SO(3) is the 3D rotation group, a.k.a., special orthogonal (SO) group. It is the group of all rotations about the origin of 3D Euclidean space $\mathbb{R}^3$. Please refer to Chapter 7 of [51] for more information.

TABLE II
MATHEMATICAL NOTATION IN THE WIFI-INERTIAL STATE ESTIMATION.

Symbol	Description
$(\cdot)^b$	IMU body frame
$(\cdot)^a$	antenna frame
$(\cdot)^Y_X$	the frame transformation from Y to X
$\boldsymbol{\mathcal{X}}$	the full state vector
$\mathbf{x}_k$	the MAV state when obtaining k^{th}WiFi features
$\mathbf{c}_i$	the i^{th}AP's position
$\mathbf{p}$	MAV position
$\mathbf{v}$	MAV velocity
$\mathbf{q}$	MAV orientation (quaternion representation)
n	the number of frames in the bundle
m	the number of APs
$\mathbf{x}^b_a$	antenna-IMU extrinsic parameter
f_c	the signal's central frequency
$\mathbf{a}_t$	the acceleration at time t
$\mathbf{g}^{b_k}$	the earth's gravity in the body frame of k^{th}state
$\mathcal{I}$	the set of IMU measurements
$\mathcal{A}$	the set of absolute features
$\mathcal{R}$	the set of relative features
$\mathbf{P}$	covariance matrix
$\boldsymbol{\mathcal{X}}_0$	the initial state vector
$\delta(\cdot)$	the error state representation of a variable
$\mathbf{\Lambda}$	the Jacobian matrix of a measurement error term
$\mathbb{I}$	the 3×3 identity matrix

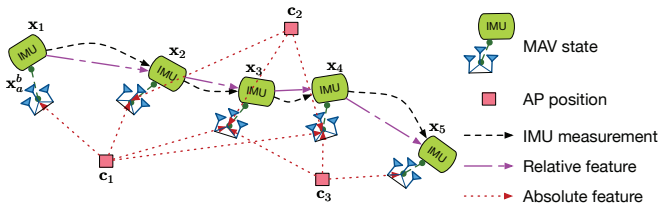


Fig. 6. Graph-based WiFi-inertial state estimation.

early fix of linearization points. In contrast, graph-based approaches, *e.g.*, pose graph optimization, benefit from iterative re-linearization of states but it requires more computation power. With proper marginalization [54], we can obtain a constant complexity sliding window graph-based approach.

Therefore, we employ a graph-based optimization framework to formulate the sensor fusion problem via multi-view constraints [21], [55], as illustrated in Fig. 6. Several WiFi and IMU measurements are kept in a bundle. The bundle size is limited to bound computational complexity. When the bundle is full, the oldest state and the corresponding measurements are marginalized to accept a new state. Here we use the marginalization technique proposed in our previous work [54]. In this formulation, we aim to find MAV states and APs' positions that best match the measurement constraints in the pose graph.

Notice that different sensors provide measurements in their local reference frames. Precisely, the antenna array measures AoAs of the array to APs, and the onboard IMU measures the accelerations and angular velocities of the MAV body. Thus, we define $(\cdot)^b$ as the body frame, which is aligned with the IMU body, $(\cdot)^a$ as the antenna frame, which is aligned with the antenna array. Superscript b_k and a_k are the body frame and the antenna frame when obtaining k^{th}WiFi measurement. The system takes the first body frame $(\cdot)^{b_0}$ as the global frame. Then we define the full state vector in the bundle as:

$$\begin{aligned}\boldsymbol{\mathcal{X}} &= [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m]^\top \\ \mathbf{x}_k &= [\mathbf{p}^{b_0}_{b_k}, \mathbf{v}^{b_k}_{b_k}, \mathbf{q}^{b_0}_{b_k}]^\top, \quad k \in [0, n],\end{aligned} \tag{17}$$

where $\mathbf{x}_k$ is the state when obtaining k^{th}WiFi measurement, which contains position $\mathbf{p}^{b_0}_{b_k}$, velocity $\mathbf{v}^{b_k}_{b_k}$, and rotation $\mathbf{q}^{b_0}_{b_k}$. And $\mathbf{p}^{b_0}_{b_0} = [0,0,0]^\top$, $\mathbf{q}^{b_0}_{b_0} = [0,0,0,1]^\top$. We use Hamilton quaternions [56] $\mathbf{q} \in \mathbb{R}^4$ to represent rotation in order to avoid singularity problems. Notation $(\cdot)^X_Y$ indicates the frame transformation from Y to X, *e.g.*, $\mathbf{p}^{b_0}_{b_k}$ is the position from k^{th}body frame to the first body frame. Variable n is the number of frames in the bundle, m the number of APs, $\mathbf{c}_i$ the i^{th}AP's position. Notice that there is a constant antenna-IMU extrinsic parameter $\mathbf{x}^b_a = \left[[\mathbf{p}^b_a; 0], \mathbf{q}^b_a\right]^\top$ in Fig. 6. Since the antenna array and IMU are fixed on the platform, $\mathbf{x}^b_a$, consisting of the relative rotation and translation between the two frames, can be manually measured.

Our goal is to estimate the state vector using the WiFi absolute-relative features to correct the IMU drift. To achieve this, we first derive the error models of absolute-relative features and IMU measurements. Then we solve an optimization problem that adjusts the state to minimize the errors.

Absolute-relative WiFi sensing error. From the absolute features to i^{th}AP, including j^{th}corrected AoA[2]

$$\hat{\mathbf{r}}^{a_j}_i = \left[\sin \hat{\theta}_i{}^{a_j} \cos \hat{\phi}^{a_j}_i, \cos \hat{\theta}_i{}^{a_j} \cos \hat{\phi}^{a_j}_i, \sin \hat{\phi}^{a_k}_i\right]^\top, \tag{18}$$

range $\hat{\rho}^{a_j}_i$, and Doppler shift $\hat{\eta}^{a_j}_i$, the measurement error term can be expressed as

$$e_{(i,j)}(\mathcal{X}) = \begin{bmatrix} \hat{\rho}^{a_j}_i \hat{\mathbf{r}}^{a_j}_i - \left(\hat{\mathbf{q}}^a_b \left(\mathbf{q}^{b_j}_{b_0}(\mathbf{c}_i - \mathbf{p}^{b_0}_{b_j})\right) + \hat{\mathbf{p}}^a_b\right) \\ \mathbf{v}^{b_j}_{b_j} - \left(\hat{\mathbf{r}}^{a_j}_i \hat{\mathbf{r}}^{a_j\top}_i\right)^{-1} \hat{\mathbf{r}}^{a_j}_i \frac{c\hat{\eta}^{a_j}_i}{f_c} \end{bmatrix}, \tag{19}$$

where f_c is the signal's central frequency, $[\hat{\mathbf{p}}^a_b, \hat{\mathbf{q}}^a_b]$ is the inverse of $\mathbf{x}^b_a$ in Fig. 6, which can be manually measured. The second block row is derived from the Doppler shift equation $\hat{\eta}^{a_j}_i = \frac{f_c}{c}\left(\hat{\mathbf{r}}^{a_j}_i\right)^\top \mathbf{v}^{a_j}_{a_j}$ and $\mathbf{v}^{a_j}_{a_j} = \mathbf{v}^{b_j}_{b_j}$ since the vehicle is a rigid body.

From the relative feature, *i.e.*, j^{th}displacement $\hat{\mathbf{l}}^{a_j}$, obtained by multiple APs, the error model is linear as it only involves the most recent two positions in b_0 frame:

$$e_j(\mathcal{X}) = \mathbf{p}^{b_0}_{b_{j+1}} - \mathbf{p}^{b_0}_{b_j} - \hat{\mathbf{q}}^b_a \left(\hat{\mathbf{l}}^{a_j} - \hat{\mathbf{p}}^a_b\right). \tag{20}$$

IMU measurement error. Since the IMU data rate is higher than the WiFi feature rate, multiple IMU measurements exist between two consecutive WiFi measurements. We need to pre-integrate IMU data between two WiFi measurements to construct the odometry constraint that aligns with the WiFi

[2]We mark the results obtained or trivially computed from the outputs of WiFi sensing or IMU as $\hat{(\cdot)}$.

data. The IMU preintegration technique has been studied by [13]. We summarize its usage within WiSion.

The raw measurements of IMU at time t are acceleration $\hat{\mathbf{a}}_t$ and angular velocity $\hat{\boldsymbol{\omega}}_t$. Given two time instants $[k, k+1]$ that correspond to two consecutive WiFi measurements, we can write the IMU preintegration as follows [13]:

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{b_k}_{b_{k+1}} &= \iint_{t\in[k,k+1]} \hat{\boldsymbol{\gamma}}^{b_k}_t \hat{\mathbf{a}}_t \,\mathrm{d}t^2, \quad \hat{\boldsymbol{\beta}}^{b_k}_{b_{k+1}} = \int_{t\in[k,k+1]} \hat{\boldsymbol{\gamma}}^{b_k}_t \hat{\mathbf{a}}_t \,\mathrm{d}t,\\ \hat{\boldsymbol{\gamma}}^{b_k}_{b_{k+1}} &= \int_{t\in[k,k+1]} \hat{\boldsymbol{\gamma}}^{b_k}_t \otimes \begin{bmatrix} 0 \\ \frac{1}{2}\hat{\boldsymbol{\omega}}_t \end{bmatrix} \mathrm{d}t,\end{aligned} \tag{21}$$

where $\otimes$ is the quaternion multiplication operation. This preintegration summarizes multiple IMU measurements into a standalone measurement term. Then, based on the kinematics theory, the IMU measurement error term can be defined as [21]

$$e_k(\boldsymbol{\mathcal{X}}) = \begin{bmatrix} \mathbf{q}^{b_k}_{b_0}\left(\mathbf{p}^{b_0}_{b_{k+1}} - \mathbf{p}^{b_0}_{b_k} + \mathbf{g}^{b_0}\frac{\Delta t_k^2}{2}\right) - \mathbf{v}^{b_k}_{b_k}\Delta t_k - \hat{\boldsymbol{\alpha}}^{b_k}_{b_{k+1}} \\ \mathbf{q}^{b_k}_{b_0}\left(\mathbf{q}^{b_0}_{b_{k+1}}\mathbf{v}^{b_{k+1}}_{b_{k+1}} + \mathbf{g}^{b_0}\Delta t_k\right) - \mathbf{v}^{b_k}_{b_k} - \hat{\boldsymbol{\beta}}^{b_k}_{b_{k+1}} \\ 2\left[\left(\hat{\boldsymbol{\gamma}}^{b_k}_{b_{k+1}}\right)^{-1} \otimes \left(\mathbf{q}^{b_0}_{b_k}\right)^{-1} \otimes \mathbf{q}^{b_0}_{b_{k+1}}\right]_{xyz} \end{bmatrix}, \tag{22}$$

where Δt_k denotes the time interval between two consecutive states. Variable $\mathbf{g}^{b_0}$ is the initial earth's gravity vector in the IMU body frame, which will be initialized in § V-B. Operation $[\mathbf{q}]_{xyz}$ extracts the vector part of a quaternion.

Our design mimics the idea of bundle adjustment [13] that aims to find a configuration of the state parameters that minimizes the sum of the Mahalanobis norm of all measurement errors:

$$\begin{aligned}\min_{\boldsymbol{\mathcal{X}}} \Bigg\{ &\sum_{k\in\mathcal{I}} \left\|e_k(\boldsymbol{\mathcal{X}})\right\|^2_{\mathbf{P}^{b_k}_{b_{k+1}}} + \sum_{(i,j)\in\mathcal{A}} \left\|e_{(i,j)}(\boldsymbol{\mathcal{X}})\right\|^2_{\mathbf{P}^{a_j}_i} \\ &+ \sum_{j\in\mathcal{R}} \left\|e_j(\boldsymbol{\mathcal{X}})\right\|^2_{\mathbf{P}^{a_j}} \Bigg\},\end{aligned} \tag{23}$$

where $\mathcal{I}$, $\mathcal{A}$, and $\mathcal{R}$ denote the set of IMU measurements, absolute features, and relative features, respectively. Matrix $\mathbf{P}^{b_k}_{b_{k+1}}$, $\mathbf{P}^{a_j}_i$, $\mathbf{P}^{a_j}$ are the corresponding covariance matrices. We use the Mahalanobis norm to compute the cost because the norm rescales the multi-dimensional variables by their correlations, which is key to achieve a high-precision autonomous system [13], [21].

From (19) and (22), we can see that the system (23) is nonlinear due to the unknown orientation. To solve such a nonlinear optimization problem, we need to find a good initialization to bootstrap an iterative solution, *e.g.*, the Gauss-Newton algorithm.

B. Automatic Initialization

The initialization requires a user to hold the MAV and move around by hand with enough accelerations for several seconds. It aims to determine the initial MAV state as well as surrounding APs' positions. The initial state includes MAV position, velocity, and orientation. Notice that the initial orientation is represented by the earth's gravity in the body frame $\mathbf{g}^{b_0}$. If the initial orientation is aligned with the earth's gravity, then $\mathbf{g}^{b_0} = [0, 0, 9.8]^\top$. Unfortunately, in practice, we cannot guarantee the ground/platform where a MAV takes off is perfectly horizontal. Thus $\mathbf{g}^{b_0}$ depends on the initial orientation and needs to be recovered.

We define the initial state vector as

$$\begin{aligned}\boldsymbol{\mathcal{X}}_0 &= \left[\mathbf{x}'_0, \mathbf{x}'_1, \cdots, \mathbf{x}'_n, \mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_m\right]^\top \\ \mathbf{x}'_k &= [\mathbf{p}^{b_0}_{b_k}, \mathbf{v}^{b_k}_{b_k}, \mathbf{g}^{b_k}]^\top,\ k\in[0,n], \quad \mathbf{p}^{b_0}_{b_0} = [0,0,0]^\top.\end{aligned} \tag{24}$$

Compare to (17), the initial state vector considers $\mathbf{g}^{b_k}$ instead of orientation $\mathbf{q}^{b_0}_{b_k}$. Quaternion $\mathbf{q}^{b_0}_{b_k}$ at this stage can be calculated by the integration of angular velocities in that during the short-term initialization the drift of such an integration is very limited. The measured $\hat{\mathbf{q}}^{b_0}_{b_k}$ makes a key difference that the system (23) becomes linear.

The error model of absolute features can be rewritten as:

$$\begin{aligned}\hat{\mathbf{z}}^{a_j}_i &= \begin{bmatrix} \hat{\rho}^{a_j}_i \hat{\mathbf{r}}^{a_j}_i - \hat{\mathbf{p}}^a_b \\ \left(\hat{\mathbf{r}}^{a_j}_i \hat{\mathbf{r}}^{a_j\top}_i\right)^{-1} \hat{\mathbf{r}}^{a_j}_i \frac{c\hat{\eta}^{a_j}_i}{f_c} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{q}}^a_b\left(\hat{\mathbf{q}}^{b_j}_{b_0}(\mathbf{c}_i - \mathbf{p}^{b_0}_{b_j})\right) \\ \mathbf{v}^{b_j}_{b_j} \end{bmatrix} \\ &= \hat{\mathbf{H}}^{a_j}_i \boldsymbol{\mathcal{X}}_0 + \mathbf{n}^{a_j}_i,\end{aligned} \tag{25}$$

where $\mathbf{n}^{a_j}_i$ is the additional Gaussian noise for absolute features. The error model of relative features is:

$$\hat{\mathbf{z}}^{a_j} = \hat{\mathbf{q}}^b_a\left(\hat{\mathbf{l}}^{a_j} - \hat{\mathbf{p}}^a_b\right) = \mathbf{p}^{b_0}_{b_j} - \mathbf{p}^{b_0}_{b_{j-1}} = \hat{\mathbf{H}}^{a_j}\boldsymbol{\mathcal{X}}_0 + \mathbf{n}^{a_j}, \tag{26}$$

where $\mathbf{n}^{a_j}_i$ is the additional Gaussian noise for relative features. The error model of IMU measurements can be expressed as:

$$\begin{aligned}\hat{\mathbf{z}}^{b_k}_{b_{k+1}} &= \begin{bmatrix} \hat{\boldsymbol{\alpha}}^{b_k}_{b_{k+1}} \\ \hat{\boldsymbol{\beta}}^{b_k}_{b_{k+1}} \\ \hat{\mathbf{0}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{q}}^{b_k}_{b_0}\left(\mathbf{p}^{b_0}_{b_{k+1}} - \mathbf{p}^{b_0}_{b_k} + \mathbf{g}^{b_0}\frac{\Delta t_k^2}{2}\right) - \mathbf{v}^{b_k}_{b_k}\Delta t_k \\ \hat{\mathbf{q}}^{b_k}_{b_0}\left(\hat{\mathbf{q}}^{b_0}_{b_{k+1}}\mathbf{v}^{b_{k+1}}_{b_{k+1}} + \mathbf{g}^{b_0}\Delta t_k\right) - \mathbf{v}^{b_k}_{b_k} \\ \hat{\mathbf{q}}^{b_k}_{b_{k+1}}\mathbf{g}^{b_{k+1}} - \mathbf{g}^{b_k} \end{bmatrix} \\ &= \hat{\mathbf{H}}^{b_k}_{b_{k+1}}\boldsymbol{\mathcal{X}}_0 + \mathbf{n}^{b_k}_{b_{k+1}},\end{aligned} \tag{27}$$

where $\mathbf{n}^{b_k}_{b_{k+1}}$ is the additional Gaussian noise for IMU measurements. We can then recover the initial state and APs' positions by solving the following least-square problem:

$$\begin{aligned}\min_{\boldsymbol{\mathcal{X}}_0} \Bigg\{ &\sum_{k\in\mathcal{I}} \left\|\hat{\mathbf{z}}^{b_k}_{b_{k+1}} - \hat{\mathbf{H}}^{b_k}_{b_{k+1}}\boldsymbol{\mathcal{X}}_0\right\| + \sum_{(i,j)\in\mathcal{A}} \left\|\hat{\mathbf{z}}^{a_j}_i - \hat{\mathbf{H}}^{a_j}_i\boldsymbol{\mathcal{X}}_0\right\| \\ &+ \sum_{j\in\mathcal{R}} \left\|\hat{\mathbf{z}}^{a_j} - \hat{\mathbf{H}}^{a_j}\boldsymbol{\mathcal{X}}_0\right\| \Bigg\}.\end{aligned} \tag{28}$$

At this stage, the initialization is completed, and the initial values will bootstrap the solution of our nonlinear WiFi-inertial state estimation.

C. State Estimation

After the initialization, we take the initial values to bootstrap the Gaussian-Newton algorithm that solves the nonlinear system (23). We use Ceres Solver [57], which is an open-source C++ library for solving complicated optimization problems. To this end, we need to derive the linearized model of system (23).

Based on the full state vector (17), the residuals for the position, velocity, orientation, and APs' positions can be

defined:

$$\begin{aligned} &\mathbf{p} = \hat{\mathbf{p}} + \delta\mathbf{p}, \quad \mathbf{v} = \hat{\mathbf{v}} + \delta\mathbf{v}, \quad \mathbf{c} = \hat{\mathbf{c}} + \delta\mathbf{c}, \\ &\mathbf{q} = \hat{\mathbf{q}} \otimes \delta\mathbf{q}, \quad \delta\mathbf{q} \approx \begin{bmatrix} \frac{1}{2}\delta\boldsymbol{\theta} \\ 1 \end{bmatrix}. \end{aligned} \tag{29}$$

Following this definition, we operate on the error state representation:

$$\begin{aligned} \delta\boldsymbol{\mathcal{X}} &= [\delta\mathbf{x}_0, \delta\mathbf{x}_1, \ldots, \delta\mathbf{x}_n, \delta\mathbf{c}_1, \delta\mathbf{c}_2, \ldots, \delta\mathbf{c}_m]^\top \\ \delta\mathbf{x}_k &= \left[\delta\mathbf{p}_{b_k}^{b_0}, \delta\mathbf{v}_{b_k}^{b_k}, \delta\boldsymbol{\theta}_{\boldsymbol{b_k}}^{\boldsymbol{b_0}}\right]. \end{aligned} \tag{30}$$

Then we linearize the cost function (23) with respect to $\delta\boldsymbol{\mathcal{X}}$. Given the current best state estimates $\hat{\boldsymbol{\mathcal{X}}}$, we have:

$$\begin{aligned} \min_{\delta\boldsymbol{\mathcal{X}}} \Bigg\{ &\sum_{k\in\mathcal{I}} \left\| e_k\left(\hat{\boldsymbol{\mathcal{X}}}\right) + \boldsymbol{\Lambda}_{b_{k+1}}^{b_k}\delta\boldsymbol{\mathcal{X}} \right\|^2_{\mathbf{P}_{b_{k+1}}^{b_k}} + \sum_{j\in\mathcal{R}} \left\| e_j\left(\hat{\boldsymbol{\mathcal{X}}}\right) + \boldsymbol{\Lambda}^{a_j}\delta\boldsymbol{\mathcal{X}} \right\|^2_{\mathbf{P}^{a_j}} \\ &+ \sum_{(i,j)\in\mathcal{A}} \left\| e_{(i,j)}\left(\hat{\boldsymbol{\mathcal{X}}}\right) + \boldsymbol{\Lambda}_i^{a_j}\delta\boldsymbol{\mathcal{X}} \right\|^2_{\mathbf{P}_i^{a_j}} \Bigg\}, \end{aligned} \tag{31}$$

where $\boldsymbol{\Lambda}_{b_{k+1}}^{b_k}$, $\boldsymbol{\Lambda}_i^{a_j}$, and $\boldsymbol{\Lambda}^{a_j}$ are the Jacobians of $e_k(\boldsymbol{\mathcal{X}})$, $e_{(i,j)}(\boldsymbol{\mathcal{X}})$, and $e_j(\boldsymbol{\mathcal{X}})$ at $\hat{\boldsymbol{\mathcal{X}}}$. We take the Jacobians to rearrange system (31) as

$$\left(\boldsymbol{\Gamma}_{\mathcal{I}} + \boldsymbol{\Gamma}_{\mathcal{R}} + \boldsymbol{\Gamma}_{\mathcal{A}}\right)\delta\boldsymbol{\mathcal{X}} = \mathbf{b}_{\mathcal{I}} + \mathbf{b}_{\mathcal{R}} + \mathbf{b}_{\mathcal{A}}, \tag{32}$$

where $\boldsymbol{\Gamma}_{\mathcal{I}}, \boldsymbol{\Gamma}_{\mathcal{R}}, \boldsymbol{\Gamma}_{\mathcal{A}}$ and $\mathbf{b}_{\mathcal{I}}, \mathbf{b}_{\mathcal{R}}, \mathbf{b}_{\mathcal{A}}$ are information matrices and vectors that derived from the Jacobians $\boldsymbol{\Lambda}_{b_{k+1}}^{b_k}, \boldsymbol{\Lambda}^{a_j}, \boldsymbol{\Lambda}_i^{a_j}$ and the covariance matrices $\mathbf{P}_{b_{k+1}}^{b_k}, \mathbf{P}^{a_j}, \mathbf{P}_i^{a_j}$. Solving this system updates the state estimates as

$$\hat{\boldsymbol{\mathcal{X}}} = \hat{\boldsymbol{\mathcal{X}}} \oplus \delta\boldsymbol{\mathcal{X}}, \tag{33}$$

where $\oplus$ is the compound operator that adds positions, velocities, and APs' positions but multiplies quaternions for orientations as in (29). The above procedure indicates that the Jacobians and covariance matrices in (31) are crucial to solve the problem.

Absolute WiFi Measurement Terms. We take the derivative of the residual $e_{(i,j)}(\boldsymbol{\mathcal{X}})$ with respect to error state $\delta\mathbf{x}_j$, $\delta\mathbf{c}_i$ and have the Jacobian

$$\begin{aligned} \boldsymbol{\Lambda}_i^{a_j} &= \begin{bmatrix} \frac{\partial e_{(i,j)}}{\partial\delta\mathbf{x}_j} & \frac{\partial e_{(i,j)}}{\partial\delta\mathbf{c}_i} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{R}}_b^a\mathbf{R}_{b_0}^{b_j} & \mathbf{0} & -\hat{\mathbf{R}}_b^a\left\lfloor\mathbf{R}_{b_0}^{b_j}\left(\mathbf{c}_i - \mathbf{p}_{b_j}^{b_0}\right)\times\right\rfloor & -\hat{\mathbf{R}}_b^a\mathbf{R}_{b_0}^{b_j} \\ \mathbf{0} & \mathbb{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \end{aligned} \tag{34}$$

where $\hat{\mathbf{R}}_b^a$ is the rotation matrix converted from quaternion $\hat{\mathbf{q}}_b^a$, $\mathbb{I}$ denotes the identity matrix of size 3. The absolute feature covariance $\mathbf{P}_i^{a_j}$ can be estimated by statistically analyzing the absolute features.

Relative WiFi Measurement Terms. Similarly, taking the derivative of the residual $e_j(\boldsymbol{\mathcal{X}})$ with respect to error state $\delta\mathbf{x}_j$ and $\delta\mathbf{x}_{j+1}$ gives the Jacobian

$$\boldsymbol{\Lambda}^{a_j} = \begin{bmatrix} \frac{\partial e_j}{\partial\delta\mathbf{x}_j} & \frac{\partial e_j}{\partial\delta\mathbf{x}_{j+1}} \end{bmatrix} = \begin{bmatrix} -\mathbb{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{35}$$

where $\mathbb{I}$ is the identity matrix of size 3. The covariance $\mathbf{P}^{a_j}$ can also be estimated by analyzing the relative features. Notice that the measurement model of relative feature is linear as in (20). Therefore, the entries in the Jacobian $\boldsymbol{\Lambda}^{a_j}$ are constants.

The IMU measurement terms, including the Jacobian $\boldsymbol{\Lambda}_{b_{k+1}}^{b_k}$ and the covariance matrix $\mathbf{P}_{b_{k+1}}^{b_k}$, have been studied and derived in inertial-based fusion systems [13], [21]. We omit the details here.

Finally, we summarize the steps of our WiFi-inertial state estimation in Algorithm 1. The state estimation requires one-time initialization by solving linear system (28) to obtain the initial state $\boldsymbol{\mathcal{X}}_0$. Then we bootstrap the Gauss-Newton algorithm implemented by Ceres-Solver [57] with $\boldsymbol{\mathcal{X}}_0$ to solve nonlinear system (31) and obtain an incremental update of the state $\delta\boldsymbol{\mathcal{X}}$. Finally, the state vector is updated using (33). The nonlinear optimization process is repeated as long as a new set of WiFi-based features is coming. The bundle will marginalize the oldest state and the corresponding measurements and accept the new ones to maintain a fixed size and bound the computational complexity. Please refer the details of such a marginalization technique to [54].

Algorithm 1 WiFi-inertial State Estimation

Goal: Output the state of the MAV over time
Input: The WiFi module provides AoA $\hat{\mathbf{r}}_i^{a_j}$, range $\hat{\rho}_i^{a_j}$, Doppler shift $\hat{\eta}_i^{a_j}$ to i^{th}AP at time j, as well as the displacement $\hat{\mathbf{l}}^{a_j}$ from time j to $j+1$. The IMU provides acceleration $\hat{\mathbf{a}}_t$ and angular velocity $\hat{\boldsymbol{\omega}}_t$ at time t
Keep receiving the WiFi and IMU measurements until the bundle is full
Initialization:
$\left\{\hat{\mathbf{z}}_i^{a_j}, \hat{\mathbf{H}}_i^{a_j}\right\} \leftarrow$ (25); $\left\{\hat{\mathbf{z}}^{a_j}, \hat{\mathbf{H}}^{a_j}\right\} \leftarrow$ (26)
$\left\{\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \hat{\mathbf{H}}_{b_{k+1}}^{b_k}\right\} \leftarrow$ (27)
Solving linear system (28) to obtain the initial state $\boldsymbol{\mathcal{X}}_0$
Non-linear optimization:
while True **do**
  if Obtaining a new set of WiFi-based features **then**
    Remove the oldest state and the corresponding measurements and accept the new measurements in the bundle
    $\left\{\boldsymbol{\Lambda}_{b_{k+1}}^{b_k}, \mathbf{P}_{b_{k+1}}^{b_k}\right\} \leftarrow$ [13]; $\left\{\boldsymbol{\Lambda}_i^{a_j}, \mathbf{P}_i^{a_j}\right\} \leftarrow$ (34) and absolute feature analysis; $\{\boldsymbol{\Lambda}^{a_j}, \mathbf{P}^{a_j}\} \leftarrow$ (35) and relative feature analysis
    Bootstrapping the Gauss-Newton algorithm [57] with $\boldsymbol{\mathcal{X}}_0$ to solve nonlinear system (31) and obtain the incremental update $\delta\boldsymbol{\mathcal{X}}$
    Update the state vector using (33)
  end if
end while

VI. System Evaluation

A. Implementation and Experimental Setup

We implement WiSion on an Intel NUC with a 1.8 GHz Core i5 processor with four cores, an 8 GB RAM, and a 120 GB SSD, running Ubuntu 14.04 LTS equipped with Intel 5300 wireless NICs. A LORD MicroStrain 3DM-GX4-45 IMU is attached to the NUC. We fix the NUC on a DJI M100 platform, and the battery of the vehicle powers the NUC. We use the

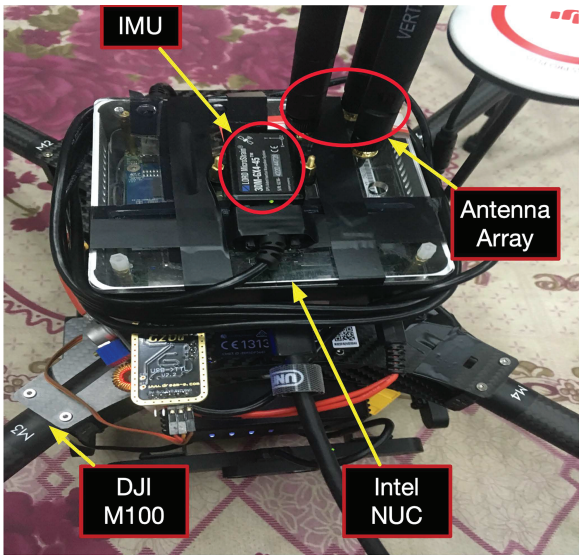


Fig. 7. The prototype implemented on a DJI M100.

Linux 802.11 CSI tool [31] to obtain the wireless channel information for each packet and implement the state estimator in C++. Fig. 7 shows the prototype.

To improve the accuracy of absolute and relative WiFi-based measurements, we use a Vector Network Analyzer to conduct a one-time calibration of transceiver responses of WiFi chips [38]. Utilizing ROS as the communication middleware, we combine WiFi and IMU measurements to estimate the MAV state and APs' positions. In our experiments, we use multiple $(1-3)$ APs. The WiFi NIC operates in 5.32 GHz center frequency on a 40 MHz band. The APs operate in monitor mode and send packets at 300 Hz. We conduct experiments in the 12×8 square meters of the MAV test site in our laboratory, a typical indoor setting.

B. Absolute-relative WiFi Measurements

Absolute measurement accuracy. We evaluate the performance in both stationary and mobile scenarios. We place three APs in different corners of the MAV test site. Both the transmitters and the receiver are configured to transmit packets with 3 antennas. Thus, we obtain a $3 \times 3 \times 30$ CSI matrix for each packet.

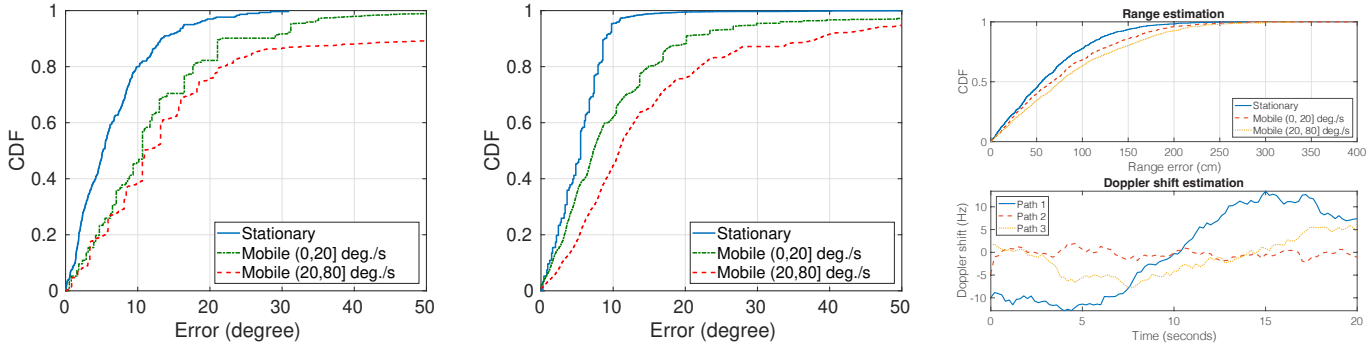
In stationary experiments, we use a digital protractor to provide the ground truth of AoAs. We manually set the orientation of the antenna array and fix the ground-truth azimuth AoA from 0 to 360 degrees in a step size of 10 degrees, fix the ground-truth elevation AoA from 0 to 90 degrees in the same step size. At each step, each AP transmits 1000 packets to the NUC. In mobile experiments, we experiment by placing the NUC on a rotating table driven by a stepper motor and repeat experiments in different angular velocities 50 times. The controller of the table provides the ground truth of AoAs. We control the rotational speed from 0 to 80 degree/s and consider slow rotations in $(0, 20]$ degree/s and fast rotations in $(20, 80]$ degree/s for evaluations.

Fig. 8 (a)-(b) plot the cumulative distribution functions (CDFs) of azimuth and elevation AoA errors in stationary, mobile in angular velocity less than 20 degree/s, and mobile in $(20, 80]$ degree/s. As expected, the stationary result is the best. 80 percentile error is 10 degrees in azimuth AoAs and 8 degrees in elevation AoAs. The AoA accuracy degenerates with the increasing angular velocity. Nevertheless, the error does not increase indefinitely. The CDF shows the accuracy is slightly worse when it goes from 20 to 80 degree/s. The 80 percentile error is 21 and 22 degrees in azimuth and elevation AoAs. These results demonstrate the reliability of AoAs in assisting MAV state estimation. The upper part of Fig. 8 (c) shows that the range and Doppler shift accuracy. We can see a median error of 56.7 cm, 65.9 cm, and 72.3 cm for range estimation in stationary and mobile scenarios. The lower part of Fig. 8 (c) depicts the estimated Doppler shift introduced by the rotation. It shows the Doppler shifts of three paths, demonstrating the capability of Doppler shift estimation.

Relative measurement accuracy. The baseline we compare with is a WiFi tracking algorithm named WiCapture [38]. Its model assumes the target of interest has a fixed heading, which is not valid for agile MAVs, and thus it cannot cope with rotational operations. We conduct experiments in two scenarios to highlight our proposal's effectiveness, *i.e.*, with and without rotations. We first put the NUC on a focusing rail and precisely move the NUC 100 mm along the rail without rotation. Fig. 9 shows that the performance of our method is similar to WiCapture as expected. The 80 percentile of displacement errors is 17.9 mm for WiCapture and 20.2 mm for WiSion. Then we move the NUC while rotating the NUC at a speed of 10 degree/s by controlling the stepper motor. The results show that WiSion significantly outperforms WiCapture. The 80 percentile is 32.7 mm for WiSion while 81.3 mm for WiCapture, demonstrating that our model effectively incorporates a vehicle's rotation.

Computational cost. Real-time processing is key to MAV state estimation due to the fast dynamics. WiSion has two computationally intensive components that keep running on the fly, the joint search of multi-dimensional parameters for WiFi sensing and the nonlinear optimization for state estimation. Note that although the automatic initialization is computationally intensive, it only needs to be run once. We enable different WiFi parameter estimations to see their computation time. To further reduce the computation time, we resort to the IMU to reduce the parameter searching range [25]. Also, we are interested in the computation time with various bundle sizes in the optimization.

We summarize the computation cost in Table III. It shows that the multi-dimensional parameter estimation has a linear increase in computation time to the number of parameters. Using the IMU cuts the computation time in half. We also tune the bundle size in the optimization and see that the accuracy improves marginally with the increasing bundle size while the computation time nearly increases exponentially. Therefore, we choose $n = 20$ in our implementation, as it is a balancing point between computation cost and accuracy. WiSion is implemented with three threads to process the packets from three APs, achieving an output rate of 10 Hz of the WiFi sensing module. Overall, WiSion provides a state update rate of 10 Hz, ensuring real-time processing.



(a) Azimuth AoA error (b) Elevation AoA error (c) Range and Doppler shift error

Fig. 8. The accuracy of absolute WiFi measurements.

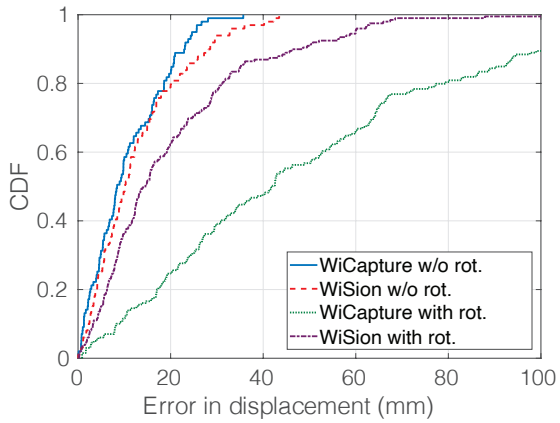


Fig. 9. The accuracy of relative WiFi measurements.

TABLE III
THE COMPUTATION COST OF WIFI SENSING AND OPTIMIZATION IN WISION.

Module Configurations	Mean Comp. Time (ms)	STD (ms)	Accuracy (cm)
Azimuth-elevation AoA	84.8	22.2	N/A
AoA + Range	112.2	27.3	N/A
AoA + Range + Doppler shift	162.5	25.4	N/A
IMU-AoA	22.4	41.3	N/A
IMU-(AoA + Range)	54.9	54.1	N/A
IMU-(AoA + Range + Doppler shift)	77.9	52.8	N/A
Nonlinear optimization (n = 8)	8.9	0.6	55.57
Nonlinear optimization (n = 10)	12.6	1.1	52.24
Nonlinear optimization (n = 15)	15.4	0.9	43.66
Nonlinear optimization (n = 20)	20.5	2.3	36.56
Nonlinear optimization (n = 30)	59.9	4.6	38.11
Nonlinear optimization (n = 40)	102.4	9.7	32.29
WiSion (n = 20)	105.6	29.2	36.56

C. System-level Performance

We program the MAV to fly in a rectangular pattern in the test site and configure a wireless network by ROS to build a connection between a laptop and the airborne NUC. The ROS connection allows us to run and recompile the airborne code while the drone is flying. We place three APs at the corners of the test site. OptiTrack [58] provides the ground truth.

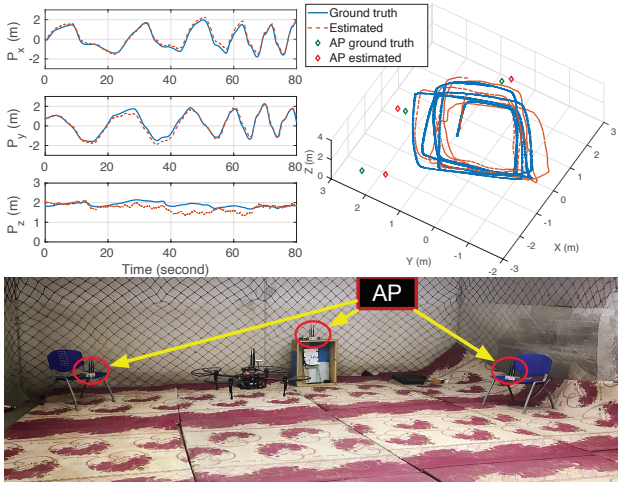


Fig. 10. The overall performance of WiSion.

1) LOS setting: We first evaluate WiSion under line-of-sight (LOS) conditions. There is no obstacle between the vehicle and the APs. Fig. 10 shows the experimental setup and the overall performance of WiSion in terms of trajectory tracking. The maximum velocity during the flight is 1.74 m/s. We can see that the positioning error in z axis is larger than other axes. This is because that the z-axis component of the position is partially observable when the vehicle does not change its height too much. The final positioning and orientation estimation errors are 35.25 cm and 2.6°, respectively. The AP's average position error is 39.23 cm.

To further highlight the effectiveness of WiSion's modules, we evaluate the performance of state estimation by incorporating different sensor measurements: absolute-WiFi only, relative-WiFi only, and WiSion that takes all measurements. In this experiment, the MAV flies 10 times to collect the data for performance analysis. Fig. 11 shows the impacts on the positioning. The mean position error is 88.23 cm with the absolute WiFi measurements due to their noise, while the mean position error with relative measurements is 109.25 cm. Although the relative measurements are very accurate, it does not provide drift-free positions and rotations, making the

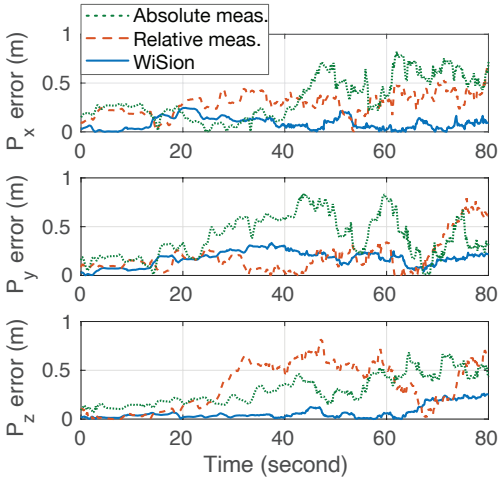


Fig. 11. Impacts of WiFi measurements.

system state not completely observable. Finally, when we take both the absolute and relative WiFi measurements (WiSion), the mean error is significantly reduced to 34.51 cm.

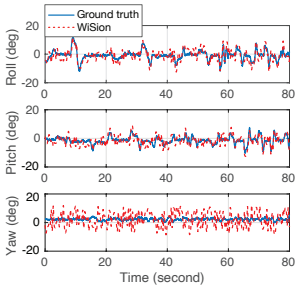


Fig. 12. LOS rotation estimation.

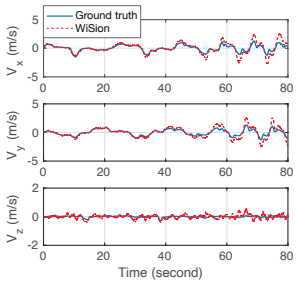


Fig. 13. LOS velocity estimation.

Fig. 12 shows the rotation estimation error of WiSion. Since an IMU only drifts in the yaw direction, we can see that the roll and pitch angles are highly accurate, while the average error of the yaw angle is 2.3°. The yaw angle is drift-free with the correction of WiFi AoAs. In addition, Fig. 13 shows that WiSion can accurately track the velocity. The maximum velocity in this experiment is 1.738 m/s.

2) NLOS setting: Now, we evaluate WiSion under NLOS settings where we put wooden plates in front of the APs so that two or fewer APs have a decent direct path from the MAV due to blockages. Fig. 14 and 15 show the details of position and velocity. As expected, the errors are larger than LOS deployments. The direct-path AoAs have larger uncertainties, as WiSion estimates them by the signals with lower magnitudes after passing through obstacles. In this experiment, the mean error of the positions over the trajectory is 96.19 cm with the maximum linear velocity of 1.72 m/s.

3) Accuracy of APs' Positions: WiSion employs a SLAM-style formulation. It estimates the MAV state and the APs' positions simultaneously. We repeat 20 experiments in LOS and NLOS settings to evaluate the performance of positioning APs. Fig. 16 is the box plot of AP's position error. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25thand 75thpercentile, respectively. The median error in the LOS setting is 39.7 cm, while

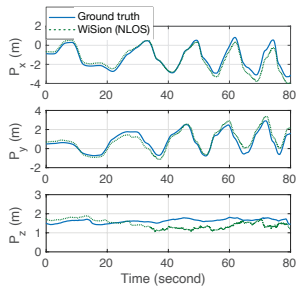


Fig. 14. NLOS position tracking.

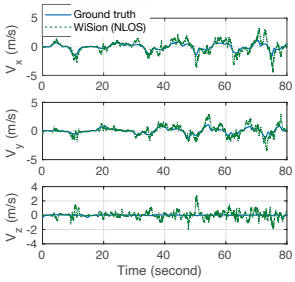


Fig. 15. NLOS velocity estimation.

the median in the NLOS setting is 91.1 cm. The worsening accuracy in the NLOS setting is because lower signal strengths in this scenario cause larger errors in absolute measurements, which are highly correlated with APs' positions.

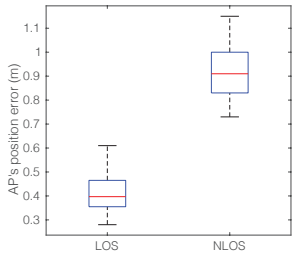


Fig. 16. The accuracy of AP's position.

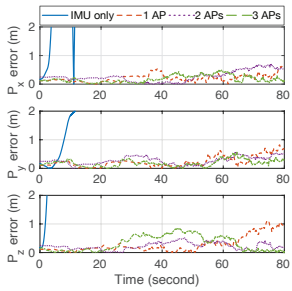


Fig. 17. Impacts of AP density.

4) Impacts of AP density: We repeat the experiments using different numbers (1 – 3) of WiFi APs to demonstrate the robustness of WiSion. The MAV flies five times for each deployment and collects the data for performance analysis. Fig. 17 plots the position errors to different numbers of APs. As expected, the accuracy decreases with fewer APs. However, the errors when using 2 and 3 APs are comparable, and the final errors are 76.91 cm and 32.36 cm, respectively. However, the final error of using 1 AP is much larger, 135.29 cm. The reason is that using 1 AP cannot estimate the displacement. Only absolute WiFi measurements are available with 1 AP. We also plot the result of IMU integration without any correction. It drifts so fast that the final error becomes meaningless, indicating that the WiFi assistance with only 1 AP is still functional to mitigate the IMU's drift. The accuracy is still applicable to spacious places where do not require precise control, *e.g.*, flying in a living room or a stadium.

5) Comparison with other related approaches: To highlight the effectiveness of our design, we compare WiSion with WINS [25] in 3D space using three APs. WINS is an indoor WiFi-based state estimation that resolves multipath and allows MAVs to fly in indoors. However, it only supports state estimation in 2D space due to the limitation of its linear antenna array. Thus, WINS can no longer support the MAV control where the heights of APs and the MAV are different. In this experiment, we use the ground truth from OptiTrack to guide the MAV control, and run WINS to record data for performance analysis.

Fig. 18 shows the statistical results of positioning and orientation estimation. Since WINS cannot cope with motions

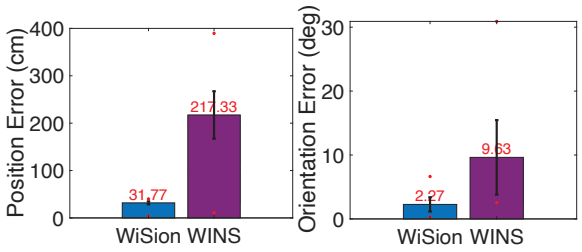


Fig. 18. Comparison with the state-of-the-art WiFi based state estimation. The red asterisks indicate the maximum and minimum of the errors. The error bars denote the 95% confidence interval.

along z-axis with its linear antenna array, its mean positioning error goes up to 217.33 cm. The maximum error even reaches 389.67 cm. Meanwhile, the poor positioning performance indicates that the WiFi measurements of WINS do not deliver reliable state information. Thus, its AoAs fail to suppress the gyroscope's drift in orientation. Although the mean error of the orientation estimation in WINS is 9.63°, its maximum error reaches 30.9°, and the 95% confidence interval is very large. In contrast, WiSion adequately addresses the 3D challenges with new designs and a circular array. The mean positioning and orientation error of WiSion are 31.77 cm and 2.27° with mild maximum errors and 95% confidence intervals.

D. Discussion

WiSion has proven the feasibility of enabling autonomous navigation for MAVs using PHY information of Wi-Fi in indoor environments. Nevertheless, there are still limitations that we need to address to improve the robustness in practice.

First, WiSion requires at least one direct path between the MAV and an AP to obtain reliable WiFi measurements. The unreliable measurements from weak signals should be rejected. One workaround that identifies and rejects unreliable measurements is to use the RSSI. Intuitively, if the direct-path component is completely blocked or severely attenuated, the RSSI should be much lower. Through our experiments, we observe that the WiFi measurements are very erroneous when the RSSI is lower than −40 dBm. The MAV receives packets with such a low RSSI when the AP is blocked by thick walls, metal obstacles, or too far away from the MAV due to severe signal degradations. To extend WiSion to be robust in the absence of direct-path signals, fusing IMU data with the NLOS WiFi localization technique [59] can be promising.

Second, the moderate accuracy makes WiSion fail to support precise MAV control. A promising method is to fuse additional sensors, *e.g.*, optical flow module, to improve the system's robustness and accuracy. However, the sensor fusion is non-trivial to analyze the measurement model and ensure the real-time processing to support the fast dynamics of MAVs.

In summary, WiSion is the key step towards realizing the low-cost and instantly deployed state estimation system using ubiquitous Wi-Fi. In future, we will work on the NLOS signal rejection and more advanced sensor fusion methods to improve the robustness of Wi-Fi based state estimators in practice. Furthermore, with the development of Wi-Fi technology, *e.g.*, Wi-Fi 7 [60], Wi-Fi signals have larger signal bandwidth and more accurate timestamps to enhance the ability of both communication and sensing. Thus, WiSion holds great potential to achieve centimeter-level state estimation soon.

VII. Conclusion

This paper presents WiSion, the first 6-DoF state estimation system that fully leverages ubiquitous commodity WiFi. WiSion allows a low-cost solution of MAV navigation to enter the massive domestic market. It consists of two components: the absolute-relative WiFi sensing module that estimates multi-dimensional geometric parameters with MAV rotations, and the WiFi-inertial state estimation that fuses the WiFi sensing with the onboard IMU in a nonlinear optimization framework. We implement WiSion on an Intel NUC attached to a DJI Matrice100 platform. The Intel NUC equips with an IMU and a three-antenna circular array. The experiments demonstrate the ability of 6-DoF state estimation in indoor venues, even with hindrances between the vehicle and APs. We believe that WiSion can be more accurate and robust by incorporating other sensing modalities, *e.g.*, ultrasonic sensors.

Acknowledgement

The authors would like to take this opportunity to thank all the anonymous reviewers for all their insightful comments and feedback and thank Prof. Yang Yang for his help with prototype implementations.

References

[1] M. Beul, D. Droeschel, M. Nieuwenhuisen, J. Quenzel, S. Houben, and S. Behnke, "Fast autonomous flight in warehouses for inventory applications," *Robotics and Automation Letters*, 2018.

[2] Y. Ma, N. Selby, and F. Adib, "Drone relays for battery-free networks," in *Proc. ACM SIGCOMM*, 2017.

[3] G. Raja, S. Suresh, S. Anbalagan, A. Ganapathisubramaniyan, and N. Kumar, "Pfin: An efficient particle filter-based indoor navigation framework for uavs," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4984–4992, 2021.

[4] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, "Indoor follow me drone," in *Proc. ACM MobiSys*, 2017.

[5] G. Quiroz and S. J. Kim, "A confetti drone: Exploring drone entertainment," in *Proc. IEEE ICCE*, 2017.

[6] L. Wawrla, O. Maghazei, and T. Netland, "Applications of drones in warehouse operations," Department of Management, Technology and Economics, ETH Zurich, Tech. Rep., 2019.

[7] "Drone service market," https://www.marketsandmarkets.com/Market-Reports/drone-services-market-80726041.html, Online; Accessed: 20 July, 2021.

[8] "Best kids drones 2020: perform stunts and take aerial selfies with the top kids drones," https://www.t3.com/features/best-kids-drones, Online; Accessed: 20 July, 2021.

[9] C. Wang, J. Wang, Y. Shen, and X. Zhang, "Autonomous navigation of uavs in large-scale complex environments: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2124–2136, 2019.

[10] K. Tanaka, M. Tanaka, Y. Takahashi, A. Iwase, and H. O. Wang, "3-d flight path tracking control for unmanned aerial vehicles under wind environments," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 11 621–11 634, 2019.

[11] A. R. Vetrella, R. Opromolla, G. Fasano, D. Accardo, and M. Grassi, "Autonomous flight in gps-challenging environments exploiting multi-uav cooperation and vision-aided navigation," in *AIAA Information Systems-AIAA Infotech@ Aerospace*, 2017.

[12] Y. Lin, F. Gao, T. Qin, W. Gao, T. Liu, W. Wu, Z. Yang, and S. Shen, "Autonomous aerial navigation using monocular visual-inertial fusion," *J. Field Robot.*, vol. 35, no. 1, pp. 23–51, 2018.

[13] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[14] X. Wang, M. Christie, and E. Marchand, "Tt-slam: Dense monocular slam for planar environments," in *Proc. IEEE ICRA*, 2021.

[15] R. Dubé, A. Gawel, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, "An online multi-robot SLAM system for 3D LiDARs," in *Proc. IEEE IROS*, 2017.

[16] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LiDAR SLAM," in *Proc. IEEE ICRA*, 2016.

[17] T.-M. Nguyen, M. Cao, S. Yuan, Y. Lyu, T. H. Nguyen, and L. Xie, "Liro: Tightly coupled lidar-inertia-ranging odometry," in *Proc. IEEE ICRA*, 2021.

[18] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, "A 64mw dnn-based visual navigation engine for autonomous nano-drones," *IEEE Internet of Things Journal*, 2019.

[19] J. Fichtinger, J. M. Ries, E. H. Grosse, and P. Baker, "Assessing the environmental impact of integrated inventory and warehouse management," *Int. J. Prod. Econ.*, vol. 170, pp. 717–729, 2015.

[20] G. d. Croon and C. D. Wagter, "Challenges of autonomous flight in indoor environments," in *Proc. IEEE IROS*, 2018.

[21] S. Zhang, W. Wang, N. Zhang, and T. Jiang, "Rf backscatter-based state estimation for micro aerial vehicles," in *Proc. IEEE INFOCOM*, 2020.

[22] Y. Liu, Y. Wang, J. Wang, and Y. Shen, "Distributed 3d relative localization of uavs," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 756–11 770, 2020.

[23] K. Guo, Z. Qiu, W. Meng, L. Xie, and R. S. H. Teo, "Ultra-wideband based cooperative relative localization algorithm and experiments for multiple unmanned aerial vehicles in gps denied environments," *Int. J. Micro Air Veh.*, vol. 9, no. 3, pp. 169–186, 2017.

[24] B. Li, S. Zhang, and S. Shen, "CSI-based WiFi-inertial state estimation," in *Proc. IEEE MFI*, 2016.

[25] S. Zhang, W. Wang, and T. Jiang, "Wi-fi-inertial indoor pose estimation for microaerial vehicles," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4331–4340, 2021.

[26] J. Tiemann, F. Schweikowski, and C. Wietfeld, "Design of an uwb indoor-positioning system for uav navigation in gnss-denied environments," in *Proc. IEEE IPIN*, 2015.

[27] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "md-track: Leveraging multi-dimensionality in passive indoor Wi-Fi tracking," in *Proc. ACM MobiCom*, 2019.

[28] R. Ayyalasomayajula, A. Arun, C. Wu, S. Sharma, A. R. Sethi, D. Vasisht, and D. Bharadia, "Deep learning based wireless localization for indoor navigation," in *Proc. ACM MOBICOM*, 2020, pp. 1–14.

[29] K. Huo, Y. Cao, S. H. Yoon, Z. Xu, G. Chen, and K. Ramani, "Scenariot: Spatially mapping smart things within augmented reality scenes," in *Proc. ACM CHI*, 2018, pp. 1–13.

[30] C. Di Franco, A. Prorok, N. Atanasov, B. Kempke, P. Dutta, V. Kumar, and G. J. Pappas, "Calibration-free network localization using non-line-of-sight ultra-wideband measurements," in *Proc. ACM/IEEE IPSN*, 2017, pp. 235–246.

[31] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, 2011.

[32] R. van der Merwe, E. Wan, and S. Julier, "Sigma-point Kalman filters for nonlinear estimation and sensor-fusion: Applications to integrated navigation," in *Proc. AIAA Guidance, Navigation, and Control Conference and Exhibit*. American Institute of Aeronautics and Astronautics, 2004.

[33] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs," in *Proc. IEEE ICRA*, 2015.

[34] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 1–13, 2015.

[35] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *Proc. USENIX NSDI*, 2014.

[36] C. R. Karanam and Y. Mostofi, "3D Through-Wall Imaging with Unmanned Aerial Vehicles Using WiFi," in *Proc. ACM/IEEE IPSN*, 2017.

[37] R. Akter, V.-S. Doan, T. Huynh-The, and D.-S. Kim, "Rfdoa-net: An efficient convnet for rf-based doa estimation in uav surveillance systems," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.

[38] M. Kotaru and S. Katti, "Position tracking for virtual reality using commodity wifi," in *Proc. IEEE CVPR*, 2017.

[39] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using WiFi," in *Proc. ACM SIGCOMM*, 2015.

[40] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in *Proc. USENIX NSDI*, 2016.

[41] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. ACM MobiCom*, 2016.

[42] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proc. ICML*, 2018.

[43] C. Wang, J. Liu, Y. Chen, H. Liu, L. Xie, W. Wang, B. He, and S. Lu, "Multi-touch in the air: Device-free finger tracking and gesture recognition via COTS RFID," in *Proc. IEEE INFOCOM*, 2018.

[44] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang, "Towards in-baggage suspicious object detection using commodity WiFi," in *Proc. IEEE CNS*, 2018.

[45] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Design and implementation of a csi-based ubiquitous smoking detection system," *IEEE/ACM Transactions on Networking*, 2017.

[46] R. H. Venkatnarayan and M. Shahzad, "Enhancing indoor inertial odometry with WiFi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–27, 2019.

[47] R. Liu, C. Yuen, T.-N. Do, D. Jiao, X. Liu, and U.-X. Tan, "Cooperative relative positioning of mobile users by fusing imu inertial and uwb ranging information," in *Proc. IEEE ICRA*, 2017.

[48] A. Ledergerber, M. Hamer, and R. D'Andrea, "A robot self-localization system using one-way ultra-wideband communication," in *Proc. IEEE IROS*, 2015.

[49] H. Xu, L. Wang, Y. Zhang, K. Qiu, and S. Shen, "Decentralized visual-inertial-uwb fusion for relative state estimation of aerial swarm," in *Proc. IEEE ICRA*, 2020.

[50] H. Nyqvist and F. Gustafsson, "A high-performance tracking system based on camera and imu," in *Proc. IEEE FUSION*, 2013.

[51] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.

[52] D. R. Brown, R. Mudumbai, and S. Dasgupta, "Fundamental limits on phase and frequency tracking and estimation in drifting oscillators," in *Proc. IEEE ICASSP*, 2012.

[53] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[54] S. Zhang, W. Wang, S. Tang, S. Jin, and T. Jiang, "Robot-assisted backscatter localization for IoT applications," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5807–5818, 2020.

[55] C. X. Lu, Y. Li, P. Zhao, C. Chen, L. Xie, H. Wen, R. Tan, and N. Trigoni, "Simultaneous localization and mapping with power network electromagnetic field," in *Proc. ACM MobiCom*, 2018.

[56] J. B. Kuipers *et al.*, *Quaternions and rotation sequences*. Princeton university press Princeton, 1999, vol. 66.

[57] S. Agarwal, K. Mierle, and Others, "Ceres solver," http://ceres-solver.org, Online; Accessed: 22 July, 2021.

[58] "Optitrack – motion capture systems," https://optitrack.com/, Online; Accessed: 20 July, 2021.

[59] X. Zhang, L. Chen, M. Feng, and T. Jiang, "Toward reliable non-line-of-sight localization using multipath reflections," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 1, 2022.

[60] E. Khorov, I. Levitsky, and I. F. Akyildiz, "Current status and directions of ieee 802.11be, the future Wi-Fi 7," *IEEE Access*, vol. 8, pp. 88 664–88 688, 2020.