UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2025-037

Faculty of Science, Technology and Medicine

# DISSERTATION

Defence held on 13 March 2025 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES DE L'INGÉNIEUR

by

**Meisam KABIRI**

Born on 28 January 1991 in Arak (Iran (Islamic Republic of))

# 5G-ENHANCED INDOOR UAV LOCALIZATION AND SLAM THROUGH SENSOR FUSION

## Dissertation Defence Committee:

Dr. Holger VOOS, Dissertation supervisor
*Professor, Université du Luxembourg*

Dr. Miguel Angel OLIVARES MENDEZ, Chairman
*Associate Professor, Université du Luxembourg*

Dr. Jose Luis SANCHEZ LOPEZ, Vice Chairman
*Research scientist, Université du Luxembourg*

Dr. Sébastien LAMBOT
*Professor, Université Catholique de Louvain*

Dr. Maria Rita PALATTELLA
*Senior R&T Associate, Luxembourg Institute of Science And Technology (LIST)*

# Abstract

Indoor localization and navigation for Unmanned Aerial Vehicles (UAVs) remain challenging due to GPS denial and the limitations of traditional visual-inertial systems. The emergence of 5G networks offers new opportunities for precise indoor positioning, but their integration with existing UAV navigation systems remains unexplored. This thesis systematically investigates the feasibility of integrating 5G Time-of-Arrival (TOA) measurements with advanced sensor fusion techniques to enhance indoor localization and SLAM (Simultaneous Localization and Mapping). Two approaches are proposed for localization: a real-time Error State Kalman Filter (ESKF) framework and a Pose Graph Optimization (PGO) method. The study leverages the EuRoC MAV dataset, augmented with simulated 5G TOA measurements, to evaluate system performance across diverse indoor scenarios and 5G base station densities. Using just IMU and TOA measurements as a minimal sensor setup, both proposed methods demonstrate significant improvements in pose estimation accuracy and drift reduction, with the PGO-based approach achieving superior results, reaching accuracies up to 13 cm with five base stations. A unified SLAM framework is developed that incorporates 5G TOA measurements alongside visual-inertial data, providing global localization capabilities and resolving scale ambiguity in monocular configurations. For local SLAM, the system operates effectively even with unknown base station positions and intermittent connectivity patterns, demonstrating an average 4.40% improvement in local accuracy while maintaining reliable scale estimation. Furthermore, TOA integration serves as an effective alternative to loop closure, improving accuracy by up to 29.6% in scenarios where traditional loop closure is unavailable. Comparative analysis with state-of-the-art approaches confirms the robustness of the proposed methods even under relaxed operational constraints. This research bridges the gap between visual-inertial and 5G radio-frequency-based approaches, establishing realistic baselines for understanding the practical impact of 5G technology on robotic localization and navigation in complex indoor environments.

# Index

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Introduction to Drone Localization and SLAM

Unmanned aerial vehicles (UAVs) have been found to be extensively used in indoor environments for surveillance, monitoring, and intralogistics. Their successful deployment depends on accurate localization - the crucial capability to determine position and orientation in real-time. Without precise localization, drones cannot navigate effectively, maintain stable flight paths, or reliably avoid obstacles. Various localization approaches exist, from traditional methods using external reference systems to more advanced techniques integrating multiple onboard sensors. While basic localization works well in controlled environments, complex scenarios may benefit from additional capabilities like Simultaneous Localization and Mapping (SLAM), which enables drones to build environmental maps while tracking their position. These dual capabilities form the foundation for autonomous operations and their critical functionalities such as:

- Obstacle Avoidance: Allowing drones to navigate safely around objects and structures.
- Path Planning: Enabling efficient route selection in complex environments.
- Multi-robot Coordination: Facilitating collaborative tasks among multiple drones or with ground-based robots.

- Data Georeferencing: Accurately assigning precise geographic coordinates (latitude, longitude, altitude) to collected data (images, sensor readings, measurements) to enable accurate spatial positioning and analysis.

As drone technology rapidly evolves, the development of robust localization and SLAM algorithms has become critical for expanding autonomous aerial operations. While GNSS provides reliable outdoor positioning, its performance significantly degrades in indoor or urban environments. The integration of 5G technology, with its high bandwidth and widespread infrastructure, offers a promising solution to these challenges. The fusion of 5G measurements with existing sensor modalities shows promising potential for enhancing localization system resilience across diverse environments, potentially expanding the capabilities of autonomous aerial systems.

## 1.2 Evolving Beyond Conventional Sensors: The 5G Paradigm for UAV Localization and SLAM

Drones, Unlike ground-based, operate within three-dimensional space, often in dynamic and unstructured environments. This lead is accompanied by additional complexities such as rapid changes in altitude. These factors make precise localization and mapping a formidable challenge, which is essential for a wide range of applications, from search and rescue operations to infrastructure inspection and environmental monitoring.

Traditionally, UAV localization has relied heavily on Global Navigation Satellite Systems (GNSS), particularly the Global Positioning System (GPS). While GNSS provides global coverage and meter-level accuracy in optimal conditions, it faces significant limitations in various scenarios crucial for UAV operations:

- **Indoor and Urban Environments:** GNSS signals are severely attenuated or blocked by buildings, rendering them unreliable or unavailable in indoor spaces and urban canyons.
- **Precision Requirements:** Many UAV applications demand centimeter-level accuracy, which standard GNSS solutions cannot consistently deliver.

- **Multipath Effects:** In complex environments, GNSS signals can reflect off surfaces, leading to multipath errors that degrade accuracy.
- **Vulnerability to Interference:** GNSS is susceptible to both unintentional interference and deliberate jamming or spoofing attacks.

To address the limitations of GNSS-based systems, the robotics community has turned to complementary sensing modalities, each offering unique advantages and challenges. Inertial Measurement Units (IMUs), visual sensors, and Light Detection and Ranging (LiDAR) systems have emerged as prominent alternatives, particularly for indoor and GPS-denied environments.

Inertial Navigation Systems (INS), leveraging IMU data, provide high-frequency motion estimates. These systems offer valuable short-term motion estimation and attitude determination, often serving as a crucial component in multi-sensor fusion schemes. However, INS are susceptible to accumulating errors over time due to sensor noise and bias. This inherent drift leads to significant deviations in positional accuracy if left uncorrected, limiting their effectiveness as a standalone solution for long-term localization.

Recent advancements in sensor technology have positioned LiDAR and image-based methods as promising alternatives for indoor localization of Micro Aerial Vehicles (MAVs), thanks to their high accuracy and robustness. However, several challenges hinder their broader adoption. These include the significant computational complexity and the need for advanced algorithms, as well as the relatively high cost, substantial size, and weight of LiDAR sensors. Additionally, image-based methods struggle in environments with poor texture or inadequate lighting, limiting their effectiveness in certain indoor settings.

Indoor localization alternatives based on wireless communication, including WLAN, Bluetooth, and Ultra-Wideband (UWB), exhibit limitations in accuracy, scalability, energy efficiency, and cost, as highlighted by various studies [1, 2]. For instance, WLAN is notably susceptible to noise interference, Bluetooth faces constraints related to range and precision, and UWB has seen slow progress in standardization. Furthermore, with its focus on low-power communication and constrained range, Zigbee has limited potential for accurate indoor positioning.

These difficulties highlight the need for other UAV indoor positioning technologies to offer sufficient precision and dependability without relying on GNSS signals.

These limitations have driven research toward innovative localization solutions, with particular emphasis on Radio Frequency (RF) signals from emerging 5G networks and advanced sensor fusion techniques. The exceptional characteristics of 5G technology - high bandwidth, low latency, and dense network infrastructure - present transformative opportunities for UAV localization in GPS-denied environments. The convergence of 5G capabilities with sophisticated sensor fusion approaches opens new possibilities for robust, scalable localization solutions, potentially revolutionizing autonomous operations in challenging indoor and urban environments.

## 1.3 Why 5G

The emergence of 5G networks presents new opportunities for complementing drone localization and simultaneous localization and mapping (SLAM) systems. Current localization methods face inherent challenges including accuracy limitations, environmental sensitivity, and dependence on infrastructure like GPS, which may be unreliable or unavailable in certain scenarios. 5G infrastructure offers potential advantages that could help address some of these constraints. While existing wireless technologies like Wi-Fi and UWB provide indoor localization capabilities, 5G's characteristics such as increased bandwidth, higher frequency bands, and low latency suggest promising applications for real-time UAV localization. This emerging technology trend motivates investigation into how 5G measurements can complement existing localization methods.

### 1.3.1 Benefits of 5G for Localization and SLAM

5G networks provide several key features that are particularly beneficial for drone localization and SLAM:

- **High Bandwidth and Carrier Frequency:** Operating at higher carrier frequencies and offering greater bandwidth, 5G networks enable finer resolution in time-based

measurements. This results in more precise estimation of parameters critical for localization, such as Time of Arrival (TOA).

- **Massive MIMO Technology:** The implementation of massive Multiple Input Multiple Output (MIMO) antennas enhances angular resolution, which aids in accurate positioning and orientation estimation of drones.
- **Wide Area Coverage:** Extensive deployment of 5G infrastructure provides broad and reliable coverage, facilitating drone operations over large areas without the need for additional localization infrastructure.
- **Enhanced Positioning Signals:** Introduction of Positioning Reference Signals (PRS) in the downlink allows for the extraction of precise TOA measurements, improving localization capabilities.

### 1.3.2 Why Indoor 5G? Key Advantages for UAV Localization

While 5G technology can be deployed both indoors and outdoors, this research focuses on indoor environments due to their unique advantages for high-precision UAV localization. Indoor settings provide superior control over environmental factors, predictable signal propagation, and alignment with industrial applications requiring robust and accurate navigation. Unlike outdoor environments, which rely on widely-spaced macro-cells optimized for coverage rather than precision, indoor networks enable ultra-dense configurations of small cells and private 5G deployments, offering significant benefits for localization tasks.

**Network Architecture (Small Cells and Ultra-Dense Deployments):** Outdoor 5G networks typically use macro-cells with large coverage areas, prioritizing broad connectivity over positioning accuracy. In contrast, indoor deployments leverage small cells and low-power base stations, strategically placed to maintain shorter distances between transmitters and UAVs. These reduced propagation paths mitigate signal attenuation and improve timing measurements, enabling higher accuracy. The ultra-dense deployment of indoor networks also increases the number of reference points, supporting advanced positioning algorithms

and achieving near-decimeter precision.

**Private 5G Networks(Control and Customization) :** Private 5G networks in industrial settings offer complete control over frequency allocation, power management, and infrastructure layout. This allows for optimization tailored to specific operational requirements, such as warehouse automation or factory monitoring. Base stations can be mounted on ceilings or walls, creating optimal geometric configurations for triangulation while benefiting from readily available power and connectivity. This level of control is rarely feasible in outdoor environments, where network design is constrained by regulatory and logistical challenges.

**High-Frequency Bands (Practicality Indoors):** The shorter propagation distances in indoor environments make high-frequency bands (e.g., mmWave) particularly advantageous. While these frequencies suffer from severe path loss over long outdoor distances, their expanded bandwidth and superior timing resolution are practical indoors, enabling centimeter-level accuracy.

**Alignment with Industrial Applications:** Many critical UAV applications requiring high-precision localization—such as automated inventory management, confined-space inspections, and collaborative robotics—primarily occur in indoor industrial settings. This focus aligns with real-world use cases, making indoor 5G localization particularly relevant for practical applications.

**Practical Experimental Validation:** A critical factor favoring indoor testing is the ability to leverage controlled experimental conditions. In our lab, we utilized a high-precision OptiTrack motion capture system to provide accurate ground truth data for validation. Conducting similar experiments outdoors would have introduced significant challenges, including obtaining flight permissions, ensuring safety, and acquiring reliable ground truth measurements. These constraints make indoor testing not only more feasible but also more rigorous for evaluating the performance of our visual-inertial dataset and 5G-assisted localization algorithms.

### 1.3.3 Time of Arrival (TOA) Measurements

Among the various positioning techniques available in 5G networks, Time of Arrival (TOA) measurements offer distinct advantages for indoor localization applications. While Received Signal Strength (RSS) methods are simpler to implement, they suffer from significant accuracy limitations due to their sensitivity to shadowing, multipath, and environmental variations in indoor settings. Similarly, while Angle of Arrival (AoA) techniques are available, they typically require complex antenna arrays and are more susceptible to multipath effects. TOA measurements, in contrast, benefit directly from 5G's high bandwidth and carrier frequencies, which inherently reduce timing estimation errors.

The theoretical bounds for TOA estimation precision can be analyzed through the Cramér-Rao Lower Bound (CRLB) [3]:

$$\mathrm{var}(\mathsf{TOA}) \geq \frac{1}{8\pi^2 B T_s F_c^2 \mathrm{SNR}}, \tag{1.1}$$

where:

- $B$ is the signal bandwidth,
- $T_s$ is the signal duration,
- $F_c$ is the carrier frequency,
- $\mathrm{SNR}$ is the signal-to-noise ratio.

This expression indicates that increasing the bandwidth $B$ and the carrier frequency $F_c$ lowers the theoretical bound on the variance of TOA estimation, thereby improving localization accuracy. The inherent ability of 5G to operate with wider bandwidths and higher carrier frequencies directly contributes to this potential enhancement.

However, while the theoretical bound highlights the promise of improved TOA estimation under such conditions, real-world deployments face multiple challenges:

- **Multipath Propagation:** Indoor environments often introduce reflections and scattering, distorting TOA measurements.

- **Signal Attenuation:** Higher carrier frequencies experience greater attenuation and penetration loss.
- **Hardware and Synchronization Constraints:** The achievable precision depends on equipment quality and timing synchronization protocols.
- **Non-Line-of-Sight (NLOS) Conditions:** Obstructions may cause additional errors that require specialized mitigation strategies.

Consequently, the actual performance of TOA-based localization depends on specific deployment conditions, the network architecture (e.g., standalone vs. non-standalone), and the chosen frequency band (sub-6 GHz or mmWave). These practical considerations ultimately define the level of accuracy achievable for indoor localization applications.

### 1.3.4 Application to UAV Localization and SLAM

Integrating 5G TOA measurements into drone localization and SLAM algorithms can provide additional positioning information in environments where conventional sensors are limited or GPS signals are unavailable. This integration may strengthen the robustness and reliability of drone navigation systems, enabling more autonomous and complex operations.

As illustrated in Figure 1.1, warehouse environments represent an ideal application for 5G ToA-integrated SLAM. Modern warehouses present several challenging conditions for traditional SLAM systems: large indoor spaces with repetitive visual features (identical shelving units), limited opportunities for loop closure in long aisles, dynamic obstacles, and requirements for precise global positioning for inventory management.

A drone performing inventory verification in this environment must navigate narrow aisles, maintain precise positioning relative to inventory locations, and operate without returning to previous locations. The experimental results in this research demonstrate that 5G ToA integration can enable global positioning accuracy of 12.6cm with 78 GHz measurements, allowing drones to maintain precise positioning for inventory management tasks. The system's ability to operate with sequential base station coverage ensures robust performance even when moving between coverage zones. Most importantly, ToA integration's effectiveness as

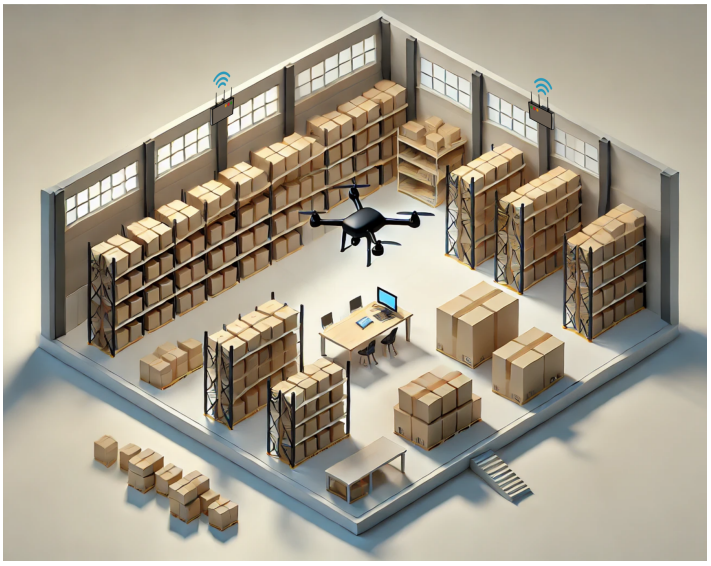


Figure 1.1: Illustration of a warehouse environment with a drone navigating using 5G base stations for localization. The two 5G base stations provide Time of Arrival (TOA) measurements for increased positioning accuracy within the environment.

an alternative to loop closure addresses a critical limitation in warehouse operations, where drones typically navigate long linear trajectories without returning to starting positions.

This warehouse example validates key research hypotheses: 5G ToA integration enables robust global positioning, higher frequency measurements provide superior performance, and ToA integration effectively compensates for limited loop closure opportunities in real-world applications. A detailed discussion of 5G's capabilities and state-of-the-art methods is presented in Section 2.1.

## 1.4 Research Objectives and Contributions

The central question of this research is: How can 5G measurements be integrated into UAV localization and SLAM systems to potentially enhance system performance in indoor environments? The study examines the feasibility of combining 5G Time of Arrival (TOA) measurements with Inertial Measurement Unit (IMU) data and visual information for UAV pose estimation during indoor flights. The integration of these sensing modalities presents opportunities for addressing challenges in localization accuracy, scalability, adaptability, and

real-time performance within sensor fusion frameworks. While existing studies have largely focused on the communication aspects of 5G technology, this research explores its potential application in drone localization, aiming to contribute to the development of indoor drone positioning techniques.

This research addresses fundamental challenges in UAV localization and SLAM systems by exploring the integration potential of 5G TOA measurements with diverse sensor modalities, including monocular cameras, RGB-D and stereo data, and inertial units. The investigation provides comprehensive insights into how 5G can enhance robotic localization under varying sensor configurations, network setups, and base station densities. The work establishes baseline metrics and practical guidelines for future 5G-enabled systems through systematic evaluation across multiple conditions.

Two complementary approaches for state estimation are introduced. First, an Error State Kalman Filter (ESKF) framework efficiently fuses high-frequency IMU data with intermittent 5G TOA signals. Second, a Pose Graph Optimization (PGO) method leverages factor graphs with newly formulated TOA-based constraints, offering superior accuracy through batch optimization. Both approaches demonstrate significant drift reduction and enhanced robustness across various deployment scenarios in real time. Using the minimal sensor setup of IMU and 5G TOA measurements, both approaches show remarkable performance: with five base stations, the PGO method achieves average accuracies of 16.6 cm, outperforming the ESKF approach which reaches around 30 cm accuracy under similar conditions. Even with reduced infrastructure of three base stations, both methods maintain sub-meter accuracy, with the PGO approach consistently demonstrating superior performance. These approaches show particular effectiveness when fusing IMU data - a sensor commonly available on commercial drones - with 5G TOA measurements, providing a practical solution that balances accuracy with minimal hardware requirements.

Then, a novel unified SLAM framework is proposed that deeply and seamlessly integrates 5G TOA data with visual-inertial sensing. This framework significantly extends the capabilities of ORB-SLAM3 through several key contributions: First, the integration of 5G TOA measurements enables globally consistent map creation and localization across extended trajectories

and multiple sessions, facilitating precise loop closure detection and map merging in different environments. Second, the novel sensor fusion approach addresses the fundamental challenge of scale ambiguity in monocular SLAM systems.

A significant advancement of this research is the successful implementation of SLAM with unknown base station positions and relaxed connectivity conditions. The system functions reliably with sequential base station connections, where only one base station maintains connection during specific, non-overlapping time intervals (e.g., BS1: 10-40s, BS2: 50-70s, BS3: 80-100s). These relaxed conditions show consistent improvement across datasets, with an average improvement of 4.40% in local accuracy for monocular configurations while maintaining reliable scale estimation (average error of 1.30%). The research also demonstrates that TOA measurements can effectively serve as an alternative to loop closure in rich sensor configurations. When loop closure is disabled (simulating scenarios like linear trajectories or exploration of new areas), TOA integration from three sequential base stations improves accuracy by 18.6% compared to the no-loop-closure baseline, with continuous TOA measurements achieving even greater enhancements (29.6%). Extensive experimental validation, utilizing both the EuRoC MAV and Aerolab datasets, underscores the advantages of 5G-based fusion across various operational conditions. Comparative analysis with state-of-the-art approaches, including UWB-VO implementations [4], confirms the robustness of the proposed methods even under relaxed operational constraints. Despite using a lower TOA measurement frequency (5 Hz versus 30 Hz in UWB-VO), our approach achieves comparable performance (0.202m mean RMSE) in challenging scenarios. The research establishes a solid foundation for integrating 5G with visual-inertial sensor fusion, demonstrating real-time performance and robust SLAM capabilities in challenging indoor settings, even with limited infrastructure knowledge and intermittent connectivity.

## 1.5 Publications

The research carried out throughout this thesis has led to several peer-reviewed publications. A comprehensive survey of radio frequency-based localization systems, with a focus on 5G

applications for aerial and ground robots, was published in *Sensors* in 2022. This work is part of the background material discussed in Chapter 2.

The core contributions on pose estimation and sensor fusion have been published in multiple venues. The initial approach, which uses Pose Graph Optimization (PGO) for UAV indoor localization by combining 5G NR Time-of-Arrival measurements with IMU data, was presented at the 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN) in 2023. This research was expanded into a journal article in *Journal of Intelligent & Robotic Systems* in 2024, providing a comparative analysis between graph-based methods and Error State Kalman Filter (ESKF)-based approaches to fuse 5G and inertial data. These contributions form the foundation of Chapter 4, which delves into sensor fusion strategies for UAV pose estimation.

The latest work, focusing on Global SLAM in Visual-Inertial Systems with 5G Time-of-Arrival Integration, has been submitted to *IEEE Access* and represents the culmination of the research in Chapter 5, which advances sensor fusion for robust indoor localization.

**Publications:**

- Kabiri, M., et al. "A review of radio frequency-based localization for aerial and ground robots with 5G future perspectives." *Sensors* 23.1 (2022): 188. *arXiv:2212.05126 [cs.RO]*. URL: https://www.mdpi.com/1424-8220/23/1/188.
- Kabiri, M., et al. "Pose Graph Optimization for a MAV Indoor Localization Fusing 5GNR TOA with an IMU." 2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE, 2023. *arXiv:2306.09826 [cs.RO]*. URL: https://ieeexplore.ieee.org/abstract/document/10332506.
- Kabiri, M., et al. "Graph-Based vs. Error State Kalman Filter-Based Fusion of 5G and Inertial Data for MAV Indoor Pose Estimation." *Journal of Intelligent & Robotic Systems* 110.2 (2024): 87. *arXiv:2404.00691 [cs.RO]*. URL: https://link.springer.com/article/10.1007/s10846-024-02111-5.

## 1.6 Thesis Structure

Chapter 2 provides background on relevant work in the field, covering various approaches to SLAM, including visual, visual-inertial, and radio-frequency-based methods. This chapter also surveys RF-based localization techniques, focusing on the potential of 5G for robotics and its promises for localization. Chapter 3 examines the fundamentals of 5G signal structure and TOA estimation techniques. This chapter begins with a comprehensive overview of 5G signal architecture and Positioning Reference Signals (PRS), followed by a detailed explanation of TOA estimation methodologies in 5G networks. The chapter then presents the QuaDRiGa channel model and simulation framework used to generate realistic ToA measurements. Finally, an in-depth statistical analysis of the simulated TOA error distributions is conducted, with particular focus on evaluating their Gaussian characteristics and implications for localization algorithms. Chapter 4 introduces the fusion of 5G and inertial data using graph-based approaches and Error-State Kalman Filters (ESKF) for indoor UAV pose estimation. Chapter 5 focuses on integrating 5G TOA into visual-inertial SLAM frameworks, particularly using ORB-SLAM3 to improve localization accuracy. Finally, Chapter 6 provides a summary of the research findings, discusses the contributions to the field of UAV localization, and suggests potential avenues for future research.

# Chapter 2

# Background and Literature Review

## 2.1 Fundamentals of RF-Based Localization

Radio-frequency (RF) technology underlies a broad range of indoor localization methods. This section covers essential RF measurements—such as RSS, TOA, TDoA, AoA, and CSI—and surveys the primary algorithms used to convert these measurements into position estimates. The discussion also addresses key architectural considerations like distributed vs. centralized processing, cooperative vs. non-cooperative designs, and differences between anchor-based and anchor-free systems. Understanding these fundamentals lays the groundwork for subsequent sections, where RF techniques are integrated into broader robot localization and SLAM frameworks.

### 2.1.1 Key Radio Frequency Features in Localization

The received radio frequency (RF) signal encodes information about the relative position of the transmitter and receiver, making it a relevant and widely used feature for localization purposes. In the following sections introduce the most important RF features commonly used in localization.

**Received Signal Strength (RSS):** Received Signal Strength (RSS) is widely used for localization due to its low hardware requirements and straightforward implementation compared

to other techniques. RSS represents the received power at the receiver over a specific bandwidth. The fundamental principle behind using RSS for localization is that the signal's power attenuation from the transmitter to the receiver correlates with the distance between them. However, extracting the exact relationship between RSS and distance in real-world environments is challenging due to the complexity of the channel model. As a result, most literature relies on simplified models to map RSS to distance or range. The most commonly used model is the log-distance one-slope propagation model:

$$\begin{aligned} \mathcal{P}_d[dB] =& \mathcal{P}_0[dB] - 10\alpha log_{10}(d/d_0) + \boldsymbol{\eta}[dB] + \mathbf{b}[dB]\,, \\ d =& ||\mathbf{p} - \mathbf{L}||\,, \end{aligned} \tag{2.1}$$

where:

- $\mathcal{P}_0$: power at the reference distance $d_0$ from the transmitter (usually 1m),
- $\mathcal{P}_d$: received power at distance $d$ from the transmitter,
- $\boldsymbol{\eta}$: shadowing effect (typically modeled as Gaussian noise),
- $\alpha$: Path Loss Exponent (PLE), indicating the rate at which power decreases over distance,
- $\mathbf{b}$: bias error.
- $\mathbf{p}$: target (receiver position)
- $\mathbf{L}$: base station (landmark) position

More complex propagation models have been proposed, including two-slope, third-order, and higher-order models up to sixth-order [5, 6]. Lee et al. [7] proposed a method using a Genetic algorithm to find the best model for each transmitter-receiver pair, searching among multi-state path loss models with varying states. It should be noted that these models assume fixed transmission power. When power control is applied at the transmitter, as is often the

case in 5G networks, range inference based on RSS alone becomes infeasible. Transmission power and PLE are usually obtained through a pre-test process, where data is collected from the environment, and suitable values are matched to the model. Algorithms capable of handling unknown PLE and transmission power offer the advantage of removing the need for this intensive pre-test phase.

RSS-based localization also suffers from instability and variability in RSS readings, which are influenced by factors such as antenna orientation, obstacles, and beacon density, as identified by Lin et al. [8]. These factors significantly affect localization accuracy.

**Time-Of-Arrival (TOA):** The Time-of-Arrival (TOA) technique measures the time taken for an RF signal to travel from the transmitter to the receiver. By multiplying this time by the speed of propagation (typically the speed of light), the distance between the transmitter and receiver can be calculated. TOA generally offers higher accuracy than RSS, especially in Line-of-Sight (LOS) conditions, provided that precise clock synchronization between the transmitter and receiver is maintained.

However, two major challenges arise in TOA estimation: the requirement for LOS conditions and stringent clock synchronization. When LOS is obstructed, multipath components arrive, each traveling a longer distance than the direct path, leading to inaccuracies. Additionally, achieving tight synchronization requires complex and expensive hardware.

**Time-Difference-Of-Arrival (TDOA):** Time-Difference-Of-Arrival (TDOA) measures the difference in arrival times of signals from two transmitters to the same receiver. The conventional approach relies on cross-correlation to extract this value by calculating the delay that maximizes the cross-correlation function. A key advantage of TDOA over TOA is that synchronization is only required among the transmitters. However, TDOA suffers from similar challenges as TOA, including hardware imperfections and LOS blockage.

**Angle-Of-Arrival (AOA):** The Angle-Of-Arrival (AOA) technique uses the direction from which an RF signal arrives at the receiver to estimate the position of the transmitter. While less commonly used compared to other techniques, AOA has gained attention with the

advent of directional antennas, multi-element arrays for MIMO, and mmWave technology, particularly in 5G networks. Like TOA, AOA also suffers from LOS blockage. Furthermore, Non-Line-of-Sight (NLoS) multipath components can provide misleading information about the signal's direction. As a result, AOA is typically used in conjunction with other localization data to improve accuracy.

**Channel State Information (CSI):** Compared to the other RF features discussed, Channel State Information (CSI) provides much richer information about the communication channel, such as fading, scattering, and power decay, as well as how the signal is affected at a specific carrier frequency along different paths. CSI comprises a matrix of Channel Frequency Responses (CFRs), which describe the amplitude and phase of the received signal across different frequencies. While other methods like RSS, TOA, and TDOA provide single scalar values per measurement, CSI offers detailed raw data that is dependent on the transmitter's and receiver's positions and the surrounding environment.

Because CSI data is more complex and less intuitive, machine learning algorithms and fingerprinting techniques are often used to process and interpret the information for localization purposes. Although CSI provides significant advantages, its complexity presents challenges regarding data handling and analysis.

### 2.1.2 Overview of RF-Based Localization Techniques

Broadly speaking, RF-based localization can be divided into two categories: range-based and fingerprinting methods.

- **Range-Based Techniques:** These techniques estimate the distance or angle between the target and some reference nodes based on various signal measurements. Common methods include Time-Of-Arrival (TOA), Time-Difference-Of-Arrival (TDOA), and Received Signal Strength (RSS) for range measurements, while Angle-Of-Arrival (AOA) provides angular (bearing) measurements. In practice, combining two or more of these techniques in a sensor fusion framework can improve localization accuracy. After extracting the range or bearing measurements, various mathematical tools such as

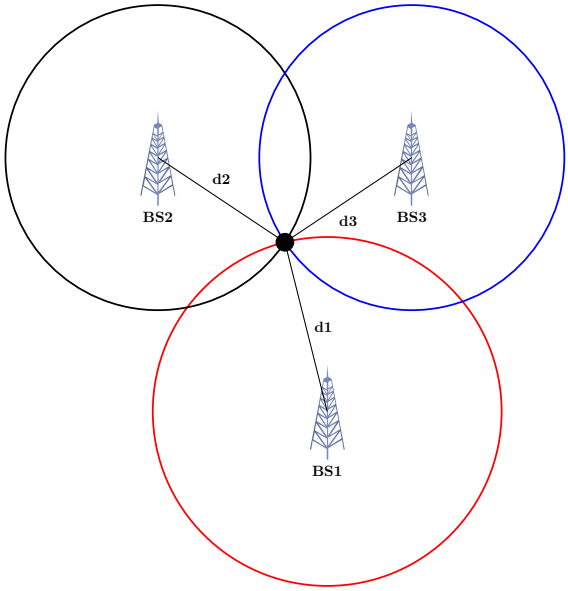



Figure 2.1: Trilateration: localization based on the range from 3 anchors using TOA and RSS measurements.

Maximum Likelihood (ML), Least Squares (LS), Bayesian models, or filtering techniques like Kalman Filter (KF), Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Particle Filter (PF) are used to estimate the target's location. In the following section, the methods used for range-based localization will be elaborated upon

- **Range-Free or Fingerprinting Techniques:** Instead of calculating distances or angles, these methods rely on an environmental survey to create a database of "fingerprints," such as location-RSS pairs. During the online phase, localization is performed by matching new measurements to the closest entry in the database. Generally, fingerprinting consists of two stages: mapping and matching. Compared to range-based techniques, fingerprinting methods are often more accurate but also more complex, as they require a comprehensive pre-survey to generate the database. Fingerprinting techniques also vary in how the dataset is generated, updated, and how the matching is performed. Despite the additional complexity, fingerprinting is widely used, particularly for CSI and RSS-based localization.

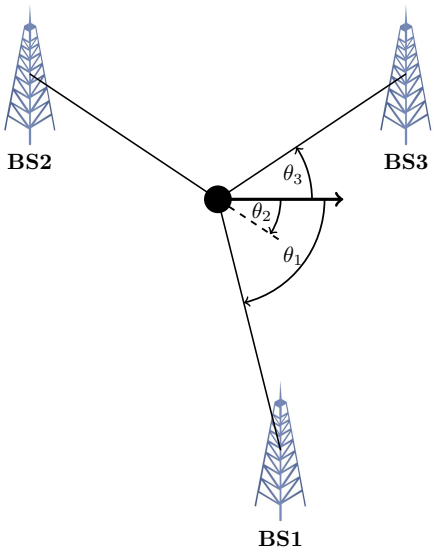



Figure 2.2: Triangulation: localization based on AOA from 3 BS.

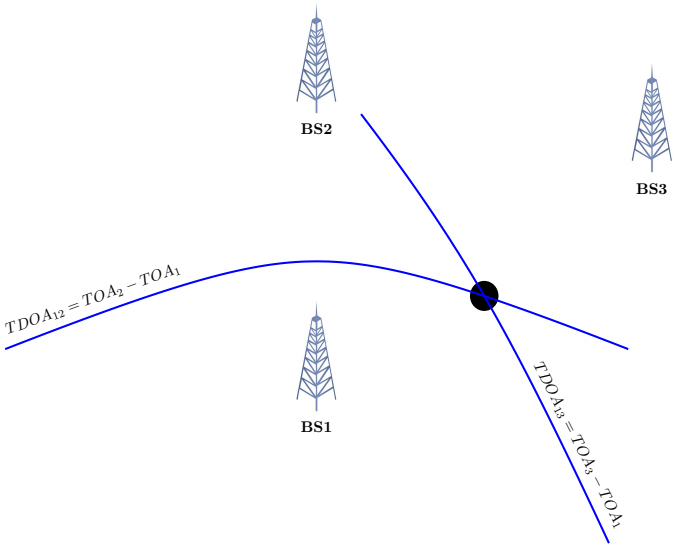



Figure 2.3: Localization based on TDOA, the intersection of the hyperbolas.

### 2.1.3 Range-based Algorithms

Having extracted ranges or bearing from the RF signal, intuitive mathematical or geometrical approaches are then leveraged to infer the receiver's position. The most important methods include Multi-Lateration/Triangulation, Min-Max, Multidimensional scaling (MDS), Least Squares (LS), Maximum Likelihood (ML), Bayesian inference method, and Bayesian Filters. Table 2.1 gives a comparison of range-based methods.

**Multi-Lateration/Triangulation:** Based on the ranges, angles, or range differences a set of equations can be constructed which usually results in an over-determined set of non-linear equations. For N number of BS position at $\mathbf{L}_i = [\mathbf{L}_{ix}, \mathbf{L}_{i_y}]^T$ and the target at position $\mathbf{p} = [\mathbf{p}_x, \mathbf{p}_y]^T$, one can write

$$d_i = ||\mathbf{p} - \mathbf{L}_i|| + \boldsymbol{\eta}, \quad i = 1, 2, .., N\,, \tag{2.2}$$

$$\delta d_i = ||\mathbf{p} - \mathbf{L}_i - \mathbf{L}_1|| + \boldsymbol{\eta}, \quad i = 2, ..., N\,, \tag{2.3}$$

$$a_i = \arctan(\frac{\mathbf{p}_y - \mathbf{L}_{iy}}{\mathbf{p}_x - \mathbf{L}_{ix}}) + \boldsymbol{\eta}, \quad i = 1, ..., N\,, \tag{2.4}$$

Where $\boldsymbol{\eta}$ represents the measurement noise, $d_i$, $\delta d_i$, and $a_i$ denote the range, range-difference, and angle measurements, respectively. Geometrically, in the simplest scenario of 2D localization using range measurements from a target to three Base Stations (BS), the target's position is determined by the intersection of three circles, each centered at a BS with a radius corresponding to the measured range (see Figure 2.1). In 3D, the solution corresponds to the intersection of four spheres. For TDOA, hyperbolas are constructed, where the foci represent the positions of two BS (see Figure 2.3). In the AOA framework, measured angles, along with the geometric properties of triangles, are used to determine the target's position (see Figure 2.2). In an ideal scenario, these methods would intersect at a single point. However, in real-world applications, this is rarely the case due to measurement errors, leading to several possible scenarios [9].

The set of equations can then be solved either in an approximate closed-form or iterative

way. Closed-form formulation leads to an easy, low-complexity solution. For example, in [10], a closed-form algebraic solution for a target localization for both trilateration (three reference points) and multi-lateration (more than three reference points) is solved. The uncertainty of each information piece is considered by adding a variance matrix to the equations. However, it assumes a known covariance matrix for the measurement noise, which can be difficult to estimate accurately in real-world scenarios. In an extended version of this method [11], after applying the standard multi-lateration procedure, if the found solution lies within the reference node positions, it is considered the final solution. Otherwise, the algorithm keeps searching for the solution inside the zones that are determined within the reference nodes based on the strength of the RSS level (the zone closer to the base station from which the strongest RSS level is read has the highest probability of containing the solution). In the search process for a solution inside a zone, virtual positions are defined inside zones, and then the final target location is selected based on the distance between the first estimation and the virtual positions.

**Min-Max:** Min-Max is a simple, intuitive, and geometrical-based technique when an easy implementation is desired [12, 13, 14]. According to ranges, squares are formed that circumscribe the circles around each BS with radius $d_i$ as the distance between the BS and target. Then the vertices of a rectangle called the area of interest are found as shown in Figure 2.4. In the simplest version, the centroid of this rectangle is selected as the estimated position. Among the variants of Min-Max, Extended Min-Max [15] improves upon this by using a weighted centroid rather than the geometric centroid, enhancing accuracy. Additionally, Yang et al. [13] propose an advanced approach that involves partitioning the area to refine the position estimation further.

**Multidimensional scaling (MDS):** Multidimensional Scaling (MDS) is a computationally efficient technique for node localization in cooperative frameworks where high-dimensional search spaces make optimization challenging. As a visualization method, MDS maps pairwise distances to lower-dimensional Cartesian coordinates that can be graphically represented. The method's key advantage lies in its analytical closed-form solution, which significantly

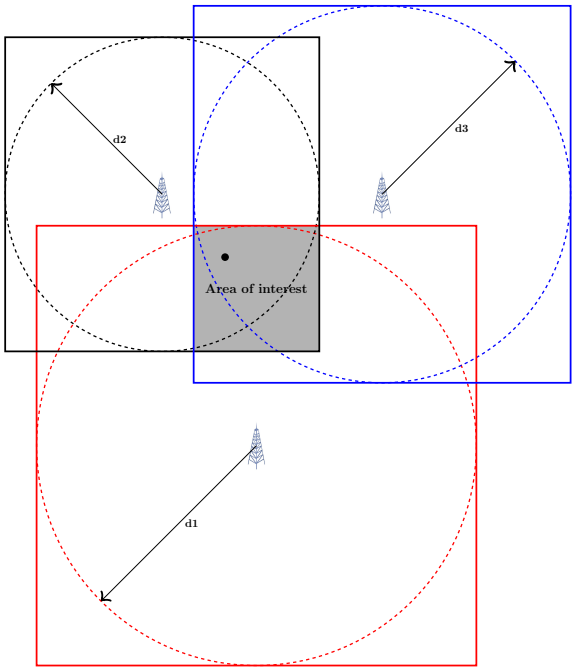



Figure 2.4: Min-Max algorithm.

reduces computational burden while enhancing efficiency and ease of implementation [16, 17].

**Least squares (LS):** Least Squares (LS) is a fundamental method for parameter estimation that minimizes the sum of squared errors. In localization contexts, for single target localization it can be formulated as:

$$\arg\min_{\mathbf{p}} \sum_{i=1}^{N} \left(RSS_i - \left(\mathcal{P}_0^i - 10\alpha log_{10}(||\mathbf{p} - \mathcal{L}_i||)\right)\right)^2, \tag{2.5}$$

$$\arg\min_{\mathbf{p}} \sum_{i=1}^{N} \left(TOA_i c - ||\mathbf{p} - \mathcal{L}_i||\right)^2, \tag{2.6}$$

$$\arg\min_{\mathbf{p}} \sum_{i=1}^{N} \left(TDOA_i c - ||\mathbf{p} - \mathcal{L}_i|| + ||\mathbf{p} - \mathcal{L}_1||\right)^2, \tag{2.7}$$

$$\arg\min_{\mathbf{p}} \sum_{i=1}^{N} \left(AOA_i - \arctan(\frac{\mathbf{p}_y - \mathcal{L}_{yi}}{\mathbf{p}_x - \mathcal{L}_{xi}})\right)^2, \tag{2.8}$$

where c is the speed of the light, $RSS_i$, $TOA_i$, $TDOA_i$, and $AOA_i$ are the measurements with respect to the ith BS. For the cooperative case where there are M targets to be localized, it can be written as :

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M} \left(RSS_{ij} - \left(\mathcal{P}_0^i - 10\alpha log_{10}(||X_j - \mathcal{L}_i||)\right)\right)^2 , \tag{2.9}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M} \left(TOA_{ij}c - ||\mathbf{p}_j - \mathcal{L}_i||\right)^2 , \tag{2.10}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M} \left(TDOA_{ij}c - ||\mathbf{p}_j - \mathcal{L}_i|| + ||\mathbf{p}_j - \mathcal{L}_1||\right)^2 , \tag{2.11}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M} \left(AOA_{ij} - \arctan(\frac{\mathbf{p}_{yi} - \mathcal{L}_{yi}}{\mathbf{p}_{xi} - \mathcal{L}_{xi}})\right)^2 . \tag{2.12}$$

The optimization problems presented in the equations above exhibit significant non-linearity. A straightforward solution to address this non-linearity is linearization, typically employing Taylor expansion [18]. Building upon this foundation, the recursive Least Squares method offers improved fidelity and accuracy, though at the cost of increased computational complexity and burden [18]. The Weighted Least Squares (WLS) method presents a more efficient alternative by applying different weights to measurements. In its most intuitive implementation, weights are selected based on distance, but the optimal approach exploits the covariance of the measurement noise. When such noise information is not readily available, researchers have developed methods to estimate it effectively [19].

Beyond position estimation alone, Least Squares methods can simultaneously estimate additional unknown parameters, such as the Path Loss Exponent (PLE) [20] and transmission power in RSS-based localization scenarios [21]. Recent work [21] has expanded this concept further by considering both unknown weight matrices and Non-Line-of-Sight (NLOS) impact, modeled by introducing random variables to the propagation model. While this approach enhances accuracy, it significantly increases problem complexity. To address this challenge, the semi-definite relaxation (SDR) technique serves as an effective approximation, transforming

the non-convex problem into convex semi-definite programming.

In pursuit of even greater efficiency, some researchers have proposed using relative error instead of absolute estimation error, an approach known as Least Squares Relative Error (LSRE) [22, 23]. LSRE defines the relative error as the ratio of the absolute error to the measured value, addressing a critical limitation of traditional LS methods where all observations are treated equally regardless of their actual precision. This approach proves particularly valuable in real-world scenarios where measurement quality can vary significantly across different nodes or environmental conditions.

**Maximum Likelihood (ML):** Maximum Likelihood (ML) is one of the most widely used approaches for localization, providing a statistically sound framework for estimating a target's position. However, it typically leads to a non-convex, non-linear optimization problem, which can be challenging to solve. The objective of ML is to maximize the likelihood function:

$$\arg\max_{\mathbf{p}} \quad p(\mathbf{p}|z), \tag{2.13}$$

where $p(\mathbf{p} \mid z)$ represents the probability of the target's position given the set of observations $z$.

In practice, the ML solution is often obtained by minimizing the Least Squares (LS) errors, where the observation errors are normalized by their respective measurement variances. This approach leads to the following optimization problems:

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M}\left(\frac{RSS_{ij}-\left(\mathcal{P}_0^i-10\alpha log_{10}(||\mathbf{p}_j-\mathcal{L}_i||)\right)}{\sigma_{ij}}\right)^2, \tag{2.14}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M}\left(\frac{TOA_{ij}c-||\mathbf{p}_j-\mathcal{L}_i||}{\sigma_{ij}}\right)^2, \tag{2.15}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M}\left(\frac{TDOA_{ij}c-||\mathbf{p}_j-\mathcal{L}_i||+||\mathbf{p}_j-\mathcal{L}_1||}{\sigma_{ij}}\right)^2, \tag{2.16}$$

$$\arg\min_{\mathbf{p}} \sum_{j=1}^{N}\sum_{i=1}^{M}\left(\frac{AOA_{ij}-\arctan(\frac{\mathbf{p}_{yi}-\mathcal{L}_{yi}}{\mathbf{p}_{xi}-\mathcal{L}_{xi}})}{\sigma_{ij}}\right)^2. \tag{2.17}$$

Here, $\sigma_{ij}$ represents the standard deviation of the measurement noise for the $i$-th Base Station (BS) and $j$-th target. It reflects the uncertainty or variance in the measurements, helping to weigh the contributions of each observation in the LS minimization. The greater the variance (or noise), the smaller the contribution of that measurement to the overall error minimization.

Solving these optimization problems is not straightforward and imposes significant computational costs due to their non-convex nature. To address this complexity, various relaxation techniques are used, such as Semi-Definite Programming (SDP) relaxation and Second-Order Cone Programming (SOCP) relaxation, which approximate the original problem. Additionally, numerical methods like Newton-Raphson are often employed to iteratively solve the optimization.

In joint-ML schemes, both the location and channel parameters are estimated simultaneously [24, 25]. For example, [25] explores a hybrid approach combining multi-lateration with ML to balance performance and computational complexity. In this method, multi-lateration is used to estimate the position, while ML is applied to estimate the channel parameters. The paper compares this hybrid method to the traditional joint-ML approach.

The impact of faults and various types of noise is considered in [26, 27]. [26] specifically addresses Byzantine faults and Non-Line-Of-Sight (NLOS) effects. Byzantine faults are modeled as non-Gaussian interference noise corrupting the transmission data, while the

NLOS effect is accounted for by introducing a bias term into the propagation model. A first-order Taylor series expansion is employed to simplify the model, transforming the problem into a Generalized Trust-Region Sub-problem (GTRS). Additionally, [27] uses an Expectation-Maximization framework to handle additive and multiplicative noise, taking advantage of the iterative nature of ML for noise mitigation.

It should be noted that the key distinction between ML and Least Squares lies in their fundamental principles and assumptions. While LS minimizes the sum of squared errors between measurements and theoretical values, ML maximizes the probability of obtaining the observed measurements. Under the assumption of Gaussian noise with constant variance, the ML estimator coincides with the LS estimator. However, when measurement noise follows different distributions or has varying variance, ML can provide more accurate estimates by explicitly incorporating the noise characteristics into the estimation process. For instance, in RSS-based localization, measurement noise often follows a log-normal distribution, making ML particularly suitable. The ML framework naturally accommodates various noise models and can incorporate prior knowledge about the statistical properties of measurements, offering greater flexibility than traditional LS methods.

**Bayesian Inference Method:**

Building upon the Maximum Likelihood approach, the Bayesian method provides a more informative result by yielding a probability distribution over the estimated parameters, rather than a single point estimate like in Least Squares (LS). In contrast to ML, where parameters are assumed to be fixed, the Bayesian approach treats them as random variables with known prior distributions [28]. This distinction allows Bayesian methods to incorporate prior knowledge into the estimation process, updating the posterior distribution of unknown parameters based on new observations according to Bayes' theorem. For example, in [29], PLE and position are treated as mutually independent random variables from which the posterior distributions are derived. To do so, a message passing algorithm, called belief propagation [30], is used on the factor graph, which allows for efficient computation of marginal distributions and dealing with the problem's intractability. The cooperative localization

scenario is also dealt with using the Bayesian method in [29]. In this paper, in addition to PLE, transmission power is estimated.

**Bayesian Filters:** While the Bayesian inference method is highly effective for estimating static parameters, Bayesian filters are more suitable when dealing with dynamic states, which is often the case in robotics and tracking applications. These filters operate through two main steps: prediction and update. A prime example is the Kalman Filter, which serves as the optimal estimator for a linear system with Gaussian noise. The Kalman Filter is widely used for sensor fusion, allowing the integration of observations from multiple sensors to improve localization accuracy.

In robotics, where the system dynamics often follow a constant velocity or constant acceleration model, Kalman Filters typically perform well in practice. However, while the process model is usually linear, the observation model is often non-linear. In such cases, sub-optimal non-linear filters are employed, including the Extended Kalman Filter (EKF) [31], Particle Filter (PF) [32], and Unscented Kalman Filter (UKF) [33].

The EKF approximates non-linearities using a first-order Taylor series expansion, which is effective in many applications but can introduce errors in posterior distribution estimation due to its linearization step [34]. In contrast, the UKF and PF outperform the EKF by avoiding linearization altogether. The Unscented Kalman Filter (UKF) utilizes an unscented transform, which generates a set of sigma points to capture the mean and covariance of the state distribution. These sigma points are then used in both prediction and update steps, providing better accuracy in non-linear systems. The Particle Filter (PF), on the other hand, employs sequential importance sampling, where particles are drawn from the probability density, and each particle is assigned a weight based on its likelihood. This approach can handle non-Gaussian noise and complex non-linearities effectively.

Table 2.1: Comparison between range-based methods.

| Range-based methods | Scenario | Advantages | Disadvantages |
|---|---|---|---|
| Multi-lateration<br>Triangulation | fast and rough estimation scenarios | simple calculation | limited accuracy,<br>sensitive to measurement error |
| Min-Max | fast and rough estimation scenarios | low complexity,<br>easy implementation | limited accuracy |
| Multidimensional Scaling (MDS) | cooperative localization | reduce the complexity | difficult to include the knowledge<br>about unequal measurement error |
| Least Square (LS) | high accuracy | easier implementation and less<br>demanding than ML and Bayesian,<br>gives estimation uncertainty | computationally demanding,<br>less optimal compared to ML and<br>Bayesian |
| Maximum Likelihood (ML) | high accuracy,<br>inaccurate prior information<br>(outperform Bayesian) | gives estimation uncertainty | computationally demanding |
| Bayesian Inference | higher accuracy,<br>sparse observations | gives estimation uncertainty | computationally demanding (more<br>demanding than LS and ML) |
| Extended Kalman Filter (EKF) | real-time dynamic state estimation<br>Easy implementation for real-time | simpler multi-sensor fusion,<br>suitable for mobile targets,<br>easy implementation,<br>gives estimation uncertainty | not useful for non-gaussian noise,<br>less optimal compared to UKF and PF |
| Unscented Kalman filter (UKF) | real-time dynamic state estimation,<br>better accuracy compared to EKF | simpler multi-sensor fusion,<br>suitable for mobile target,<br>gives estimation uncertainty | not useful for non-gaussian noise |
| Particle Filter (PF) | high accuracy dynamic state<br>estimation | handling non-gaussian noise,<br>gives estimation uncertainty | computationally demanding,<br>difficult implementation |

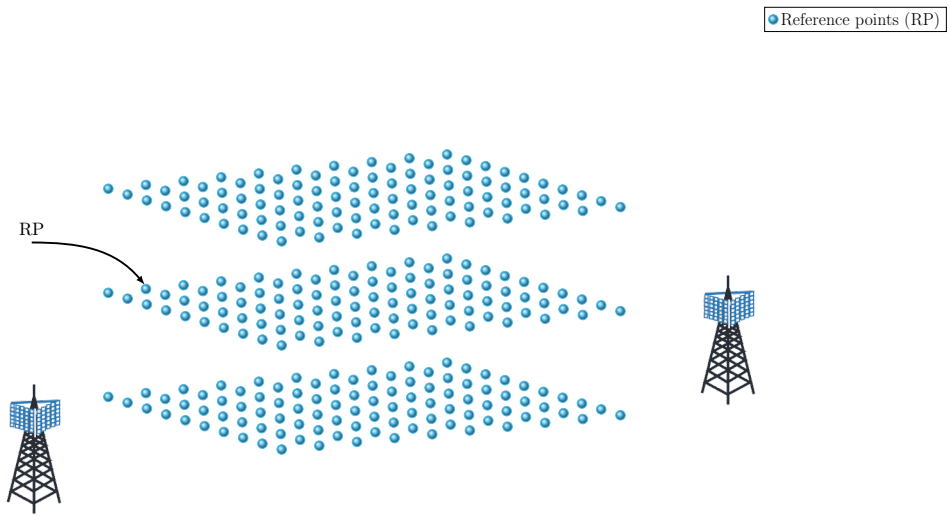



Figure 2.5: Reference points in 3D.

### 2.1.4 Fingerprinting

As an alternative to range-based approaches, fingerprinting methods infer the position based on collected data rather than relying solely on model assumptions. This technique can more effectively mitigate errors introduced by modeling inaccuracies and noise. While fingerprinting is predominantly used in RSS and CSI-based localization, it has also been applied to other RF-based measurements [35, 36, 37, 38, 39], as well as in hybrid schemes [40, 41, 42].

This method usually results in more accurate estimation and comes at the cost of the laborious step of data collection. In general, fingerprinting involves two stages: an offline phase and an online phase. During the offline phase, a dataset of recorded measurements or RF features, paired with ground truth positions at reference points (RPs) (see Figure 2.5), is collected to form the "fingerprints." In the online phase, position estimation is performed by matching real-time measurements with these pre-recorded fingerprints (see Figure 2.6).

**Offline step:** One of the main challenges in the fingerprinting approach is the creation and

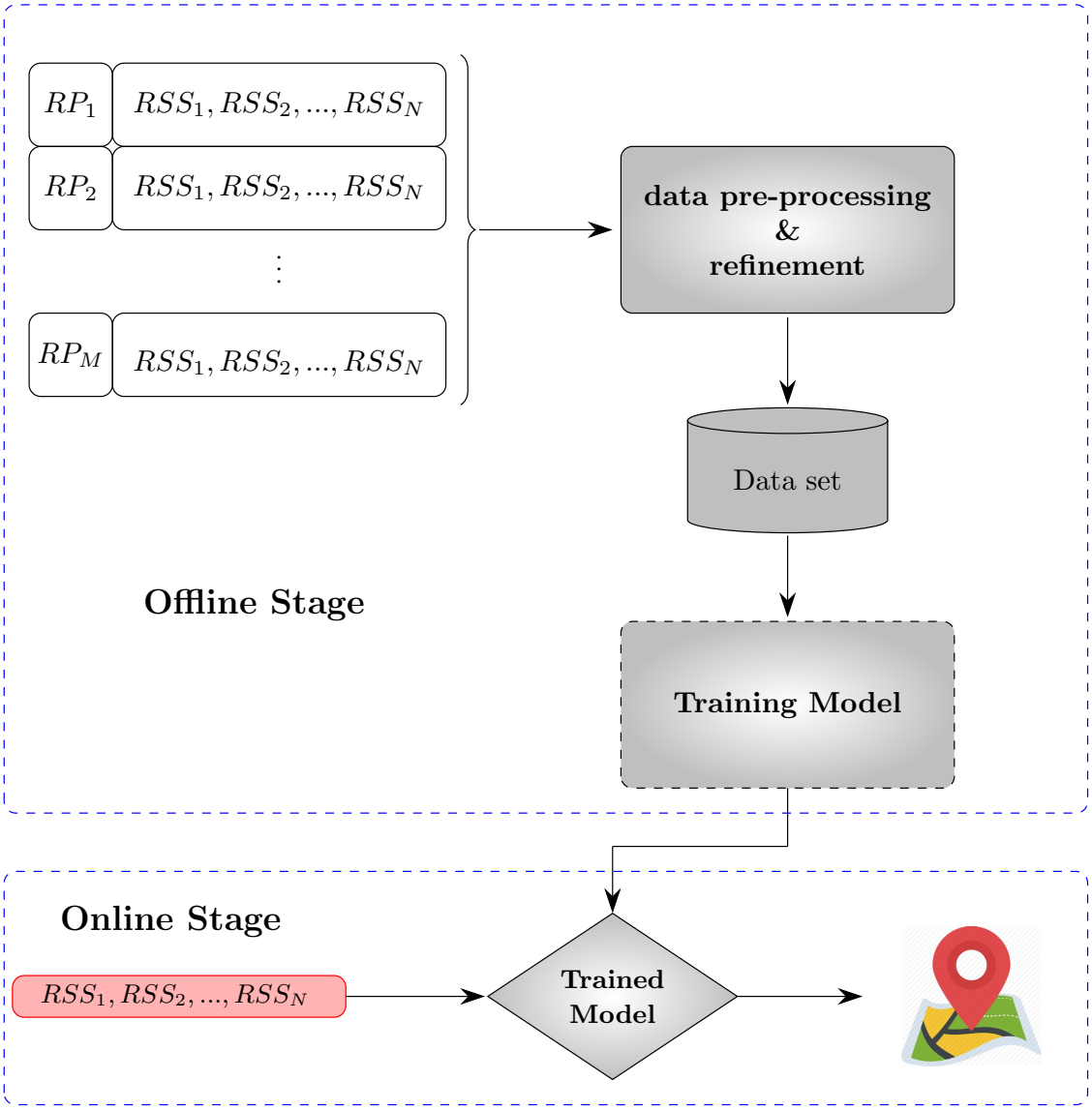



Figure 2.6: Fingerprinting block diagram.

maintenance of an accurate fingerprint database. The simplest way to generate the radio map or dataset is by measuring and recording fingerprints across all areas. However, this is often impractical due to the large coverage areas involved, making detailed measurements and site surveys both time-consuming and costly. Additionally, the fingerprint map can quickly become outdated due to changes in radio signal characteristics and dynamic environmental factors. This issue is discussed in [43], which explores the impact of incomplete datasets and suggests the use of interpolation and extrapolation techniques to recover missing data.

Efficient radio map generation, which involves both the collection of raw signal data and the processing techniques to enhance or fill gaps in the dataset, is essential to mitigate the challenges of large-scale data collection. In [44], a deterministic adaptive path loss model is employed to generate radio maps faster and reduce construction time. This method extracts each access point's (AP) RSS path loss model using available reference points, least squares (LS) estimation, and interpolation. Another common approach is Gaussian Process Regression (GPR), as explored in [45], where GPR is used to predict the spatial distribution of RSS based on available datasets. Instead of using a basic zero-mean function with a single squared exponential kernel, compound GPR kernel functions are applied to enhance the accuracy of the predictions. Additional information on various radio map generation techniques and real-world experimental comparisons can be found in [46].

It is important to note that most fingerprinting localization methods rely on raw or scaled data collected from RF signals. However, this approach can be inefficient in heterogeneous environments, where the transmitter and receiver devices may have varying hardware characteristics, such as differing antenna gains. This hardware heterogeneity can result in inconsistent signal measurements even in identical situations. To address this issue, alternative approaches based on relative measurements have been proposed. For example, [47] uses differential signal strengths to account for both device heterogeneity and temporal variations in RSS. Similarly, [48] employs a differential RSS-based localization approach. Another solution is hyperbolic location fingerprinting, as presented in [49], where fingerprints are recorded as RSS log ratios between pairs of base stations rather than absolute RSS values.

**Online phase:** In the online phase, machine learning algorithms play a crucial role in matching and position estimation. Commonly used algorithms include KK Nearest Neighbor (KNN), Weighted K Nearest Neighbor (WKNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Random Forest. These methods are employed to analyze the real-time data and match it to the pre-constructed radio map, enabling accurate localization.

K-Nearest Neighbor (KNN) is the first and simple candidate to select K points in the data set based on similarity. Considering the computation complexity, the achieved performance for this algorithm is acceptable. In a WKNN, different weights are assigned to measurements based on criteria such as the distance to the AP. For example, in [50], WKNN is employed where based on the inverse of RSS distance, weights are assigned to the APs. This criterion, however, is not consistent with the logarithmic nature of path loss which decays logarithmically based on distance (the reader is referred to [51] for more details). To resolve this inconsistency, [51] introduces a new weighting scheme and distance measure that considers both RSS similarity and spatial position.

A limitation of traditional KNN is the use of a fixed number of neighbors (K). To address this, [52] proposes a Weighted Adaptive KNN algorithm that selects a variable number of reference points (RPs) based on an enhanced combination of RSS similarity and positional proximity. Density-based spatial clustering of applications with noise (DBSCAN) and Affinity Propagation Clustering (APC) are also two approaches that do not require the number of clusters to be pre-determined and are suitable for handling large databases. The authors in [53] show that clustering could reduce noise's impact in large data sets. Hence, the DBSCAN method is proposed for localization, where the fingerprints are divided around a center point based on the density. Similarly, [54] applies APC offline until the algorithm converges, using message-passing techniques to determine clusters (attraction and attribution messages).

Random Forest, a classification and regression method based on ensembles of decision trees, is another strong candidate for handling large, high-dimensional datasets. It offers fast training and prediction, making it effective for localization tasks [39, 55]. Support Vector Machines (SVM), a robust classification engine, can also be applied to regression problems

in localization. For instance, multi-class SVM has been utilized for localization in [56, 57]. To enhance separability in fingerprint data, [56] employs a fuzzy kernel that maps the data into a higher-dimensional space, improving accuracy.

Neural Networks (NN) are particularly well-suited for fingerprinting due to their ability to model complex input-output relationships. Notably, NNs can directly map fingerprints to Cartesian coordinates, as demonstrated in [58], where an Artificial Neural Network (ANN) takes RSS measurements as input and outputs the estimated position in 2D space. Deep learning techniques, including NNs, offer more power and flexibility than traditional machine learning methods, especially when dealing with large, noisy, and multi-dimensional data [59]. A significant advantage of deep learning is its ability to harness the parallelism of GPU architectures, leading to faster run-times.

However, there are challenges associated with deep learning models. First, they require access to a rich and diverse dataset to learn meaningful patterns and relationships between input and output effectively. Additionally, there is a lack of tools to interpret these models, making it difficult to understand which parts of the input space have been accurately learned and where more data is needed. This issue ties into overfitting, where the model struggles to generalize to unseen data and instead makes predictions that are too closely aligned with the training data. Finally, determining appropriate training hyperparameters is often a time-consuming and iterative process.

One of the promising deep learning-based methods for localization is Convolutional Neural Networks (CNN), which have been widely deployed and successfully used for image classification. CNNs can perform very fast in the online phase, but their training process is meticulous and time-consuming, which still requires further investigation in the context of fingerprinting localization. For instance, in [60], a six-layer CNN classifier is used to learn and predict over 74 classes. Similarly, in [61], a Stacked Auto-encoder (SAE) is incorporated to reduce data dimensionality before feeding it into a CNN classifier for multi-floor, multi-building localization.

In a 5G Internet of Things (IoT) test-bed, [62] combines two CNNs for localization. The first CNN acts as a regressor to estimate the position in 2D, and its output is used as input

to the second CNN, which functions as a classifier to estimate the 3D location. To address challenges such as temporal RSS variations, device heterogeneity, and sparse RSS data in a lightweight algorithm, a Siamese network architecture consisting of two sub-networks (convolutional neural network with shared weights) with offline fine-tuning is proposed in [55]. During training, an embedding function is learned that uniquely translates RSS-distance pairs into embeddings. The training is executed so that each sub-network can take an RSS vector as input and estimate the relative distance of the locations corresponding to the RSS pairs. In the online step, the location is calculated as the weighting average of the RPs, based on the probabilities assigned for the embeddings of online RSS using Random Forest.

It is worth noting that these methods can also be combined in a multi-stage process for more efficient and accurate localization. In practice, an initial rough estimation is typically performed to identify a sub-area candidate for the target position. A more precise estimate is then obtained by searching within this selected sub-area. For example, in [63], the data in the database is clustered based on signal levels, including RSS and directional antenna gain. After an initial rough estimation, the solution is refined through up-sampling until the desired accuracy is achieved. A similar two-step process is used in [36] for UAV positioning in 3D based on TOA. Coarse localization is followed by a neural network fitting step to refine the estimation. Additionally, [41] employs a hybrid approach where rough localization is first performed using TDOA. Then, a deep neural network (DNN) refines the position estimate within the identified sub-area using RSS data. These multi-stage methods help balance the trade-off between speed and accuracy, allowing for an initial quick estimate and a subsequent more detailed search for precise localization.

### 2.1.5 Extended Classification and System Aspects

Aside from the techniques based on which the works fall under two broad main categories, there are other points of view that RF-based localization can be taken into account.

**Distributed vs Centralized :** Two approaches can be distinguished based on how computation and processing are handled: centralized and distributed. In centralized systems,

all data is transmitted to a central node or station where computations are performed. This approach typically requires simpler algorithms but demands a powerful processor to handle the workload. In contrast, distributed systems share the computation across all nodes and subsystems, with each node contributing to the overall result. While this distributed approach requires more sophisticated and complex algorithms, it becomes increasingly attractive as the number of nodes in a Wireless Sensor Network (WSN) grows.

**Cooperative vs Non-cooperative:** In cooperative localization, nodes communicate with one another, sharing information with neighboring nodes to improve accuracy. The advantages of cooperation, such as improved accuracy in terms of the Cramer-Rao Lower Bound (CRLB), are well documented [64]. Non-cooperative methods, on the other hand, focus on minimizing energy consumption and reducing algorithmic complexity, making them more suitable for resource-constrained environments.

**Anchor-based vs Anchor-free:** Anchor-based localization relies on the known positions of certain nodes or base stations, referred to as anchors. These anchors, which are typically determined via GPS or manual deployment, serve as fixed reference points for the localization process, providing more accurate results. In contrast, anchor-free approaches do not assume prior knowledge of any node's position, offering greater scalability and flexibility. While anchor-free methods are less accurate, they eliminate the need for deploying and maintaining anchor nodes, as discussed in [65, 66].

**Static vs Mobile:** Most RF-based localization algorithms are designed for static targets. Extending these methods to mobile targets, such as those commonly found in robotic applications, poses additional challenges, especially in terms of real-time feasibility. Mobile localization introduces new complexities, such as continuous tracking and maintaining accuracy while accounting for dynamic changes in the environment. The next section explores mobile robot localization in greater detail.

**Technologies:** A variety of technologies can be employed for RF-based localization, depending on the infrastructure, desired accuracy, cost, and environmental factors. Popular options

include WiFi, Bluetooth, Ultra-Wideband (UWB), Zigbee, Radio Frequency Identification (RFID), cellular networks, and Long Range (LoRa) radio. Each technology has its strengths and weaknesses. For instance, UWB offers centimeter-level accuracy over short ranges, making it ideal for indoor environments, while WiFi leverages existing infrastructure at the cost of reduced accuracy. A comprehensive survey of these technologies can be found in [67].

**2D vs 3D:** While much of the existing research focuses on 2D localization, extending these methods to 3D introduces additional challenges. Although it is theoretically possible to apply most 2D techniques to 3D localization, the vertical (z-axis) component often results in higher error margins than the horizontal (x-y) plane. This issue is particularly significant in applications such as drone navigation, where accurate 3D positioning is crucial. The challenges and solutions for 3D localization are discussed in greater depth in [68, 25].

**Performance Parameters:** Evaluating the performance of localization algorithms requires a multi-faceted approach. Several criteria need to be considered, including accuracy, precision, computational complexity, scalability, security, reliability, cost, and stability [69]. Balancing these parameters is essential to selecting or designing a localization method that meets the specific requirements of an application.

## 2.2 Robot Localization and SLAM Approaches

This section reviews the literature under three categories: visual SLAM, visual-inertial SLAM, and radio-based SLAM summarizing key advancements and challenges in each domain.

### 2.2.1 Visual SLAM

Visual SLAM systems, which rely primarily on camera sensors, have undergone significant evolutionary stages in algorithmic development. Feature-based methods pioneered early advancements by extracting keypoints and descriptors from images to estimate camera motion and 3D structure. MonoSLAM [70] introduced real-time monocular SLAM using

an Extended Kalman Filter (EKF) to track sparse feature points, laying groundwork for subsequent approaches. Parallel Tracking and Mapping (PTAM) [71] revolutionized the field by separating tracking and mapping into parallel threads, enabling more efficient Bundle Adjustment (BA) for improved pose estimation.

The ORB-SLAM series [72, 73, 74] marked a significant milestone by introducing robust feature extraction using Oriented FAST and Rotated BRIEF (ORB) descriptors [75]. These implementations progressively expanded SLAM capabilities, with ORB-SLAM2 supporting monocular, stereo, and RGB-D cameras, and ORB-SLAM3 integrating visual-inertial capabilities and multi-map SLAM techniques.

Direct methods represent another critical approach, operating directly on pixel intensities without explicit feature extraction. Dense Tracking and Mapping (DTAM) [76] was among the first to use an inverse-depth representation for dense map construction. Large-Scale Direct SLAM (LSD-SLAM) [77] introduced a semi-dense approach focusing on high-gradient pixels, while Direct Sparse Odometry (DSO) [78] minimized photometric error over selected pixels.

Hybrid methods emerged to balance the strengths of feature-based and direct approaches. Semi-Direct Visual Odometry (SVO) [79] utilized direct methods for rotation estimation and feature-based methods for translation, achieving high-speed performance in dynamic environments. Direct Sparse Mapping (DSM) [80] further advanced this paradigm by introducing photometric Bundle Adjustment for global optimization.

Despite significant advancements, visual SLAM continues to face critical challenges: scale ambiguity in monocular systems, sensitivity to environmental conditions like lighting and texture variations, and the persistent need to balance computational complexity with real-time performance.

### 2.2.2 Visual-Inertial SLAM

Integrating Inertial Measurement Units (IMUs) with cameras addresses some limitations of visual SLAM by providing complementary motion information. IMUs offer high-frequency localization and orientation data, compensating for visual sensor shortcomings in fast motion or low-light conditions. Conversely, visual sensors help mitigate the cumulative drift commonly

associated with IMUs, leading to more accurate and robust localization and mapping.

Filter-based approaches, such as Multi-State Constraint Kalman Filter (MSCKF) [81], MSCKF 2.0 [82], and Robust Visual-Inertial Odometry (ROVIO) [83], utilize recursive estimation techniques to fuse visual and inertial data. These methods primarily use Extended Kalman Filters (EKFs) to track system states and estimate uncertainties. MSCKF introduced a novel feature marginalization technique to decrease computational complexity, showing robustness during aggressive movements and brief feature losses. ROVIO extended this approach by incorporating direct photometric error minimization, particularly improving performance in low-texture environments. Optimization-based methods, including Open Keyframe-Based Visual-Inertial SLAM (OKVIS) [84], Visual-Inertial Navigation System (VINS-Mono) [85], VINS-Fusion [86], and ORB-SLAM3 [74], solve for system states by minimizing cost functions over data windows. These approaches employ sophisticated non-linear optimization techniques like bundle adjustment to achieve global consistency. VINS implementations, for instance, utilized sliding window optimization frameworks that demonstrated remarkable robustness across diverse real-world datasets, while ORB-SLAM3 comprehensively integrated IMU data across multiple camera modalities.

Challenges in visual-inertial SLAM include accurate intrinsic and extrinsic calibration, robust initialization that involves scale and gravity estimation, handling dynamic environments, and maintaining computational efficiency.

### 2.2.3 Radio-Based SLAM

While visual and visual-inertial SLAM methods have shown remarkable performance, they often struggle in feature-sparse environments, under poor lighting conditions. These methods can also suffer from accumulated drift over time and the lack of absolute reference frames, particularly in GPS-denied or indoor environments.

Radio-based SLAM has emerged as a valuable complement to traditional methods. It leverages Radio Frequency (RF) signals—such as Ultra-Wideband (UWB), Wi-Fi, millimeter-wave (mmWave), and 5G technologies—for localization and mapping.

Recent advancements in localization techniques have explored innovative approaches

for positioning mobile receivers in challenging environments [87]. Gentner et al. introduced Channel-SLAM, an algorithm leveraging multipath signals for positioning by treating multipath components as virtual transmitters and employing recursive Bayesian filtering with a Rao-Blackwellized particle filter. This approach demonstrates the potential of exploiting reflections and scattering without prior environmental knowledge, achieving accurate positioning in both line-of-sight and non-line-of-sight conditions. However, the method is computationally intensive and limited to multipath signal processing without incorporating visual or inertial sensors.

In a graph-based framework in radio SLAM, researchers have explored belief propagation and factor graphs to jointly localize mobile agents and map environments using multipath components [88, 89]. Specular reflections are modeled as virtual anchors (VAs), which are mirror images of physical anchors (PAs), allowing the system to simultaneously estimate the positions of VAs, PAs, and the mobile agent. Such techniques enhance localization accuracy by incorporating advanced signal parameters like the Angle of Arrival (AoA) but often face challenges in computational complexity. Similarly, Chu et al. [90] and Mendrzik et al. [91] have investigated joint localization and radio mapping, using multipath information to estimate vehicle positions and environmental features simultaneously via factor graphs. However, their work lacked network simulations and fusion schemes necessary for practical implementations.

Complementary approaches have utilized Wi-Fi technologies, such as 60 GHz IEEE 802.11ad, to provide indoor localization. Bielsa et al. [92] have developed a real-time system that achieves sub-meter accuracy in 70% of cases. While promising, these Wi-Fi-based methods may be limited by the quality of the wireless signal.

The emergence of 5G technology has further expanded localization possibilities for localization [93, 94]. Researchers developed frameworks like NR5G-SAM, presented in [93], combined Time of Arrival (TOA) and Received Signal Strength Indicator (RSSI) measurements with inertial sensing, employing factor graphs for trajectory estimation and Radio Environmental Map (REM) creation. This approach showed promise, especially in GNSS-denied and rural areas. However, it faced several limitations. The reliance on RSSI-based

mapping reduced precision, particularly in vertical positioning. Additionally, the absence of loop closure mechanisms and dependence on multilateration increased computational demands. Del Peral-Rosado et al. [95] and Saleh et al.[96] explored 5G-based positioning techniques, achieving accuracies of 20-25 cm and sub-meter level, respectively. Other 5G-based localization methods have explored various techniques, including fingerprinting, machine learning, and signal processing. Talvitie et al. [97] utilized 5G synchronization signals to achieve sub-meter accuracy for high-speed train tracking. Zhang et al. [98] employed a deep neural network to improve positioning accuracy using 5G AoA and amplitude information, even in non-line-of-sight environments. Shamaei and Kassas [99] proposed an opportunistic TOA estimation approach using 5G synchronization signals and PBCH, achieving a ranging error standard deviation of 1.19 m. All of these studies primarily focused on localization and did not address the simultaneous mapping aspect of SLAM.

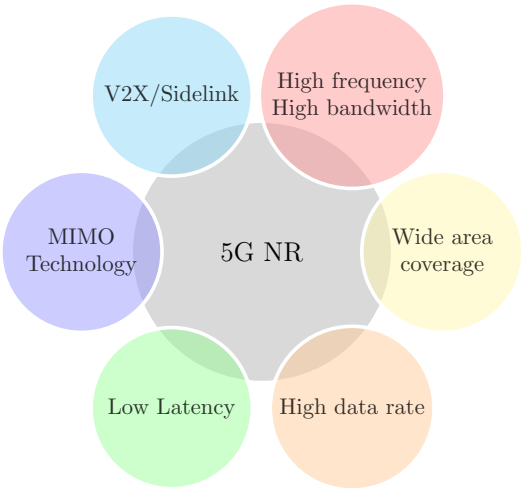
Despite recent progress in radio-based localization methods, especially those leveraging 5G technologies, they typically focus solely on localization without addressing the simultaneous mapping aspect essential for SLAM applications. Moreover, existing radio SLAM approaches often rely on complex multipath signal processing without integrating other sensor data, limiting their robustness and applicability.

## 2.3 5G potentials and promises for robot applications

Rolling out of the 5G New Radio (NR) technology provides great potential to boost the localization of robots and UAVs in terms of accuracy, robustness, cost, and coverage. The promising features of 5G NR for robot applications include the following:

- Wide area coverage
- MIMO technology
- High carrier frequency
- High bandwidth

Figure 2.7: 5G NR enabler for improved robot localization




- Vehicle-to-Everything (V2X)
- Low latency
- High throughput

### 2.3.1 Wide area coverage and in-expensive localization systems

Compared to other technologies like WiFi, UWB, etc., 5G will be available almost anywhere, indoors or outdoors, since the cellular infrastructure is widely deployed in the cities. Using Vehicle-to-Everything (V2X) also makes it feasible to take advantage of 5G, where there is no full coverage. For example, in a collaborative scheme, one or some part of the device(s) or vehicle(s) can play the role of anchor or Pseudo BS for others. 5G is also considered an in-expensive solution because there is no specific equipment to set up as long as operating under the coverage of BS. Accordingly, taking advantage of the available infrastructure of 5G, in some cases, might remove the need for costly and also energy-consuming GPS devices.

In addition, 5G confers robustness to the localization system. Robustness is an essential feature in highly mobile scenarios, in the case where safety is of great concern. For example, consider a vehicle or a group operating in a wide area where some part of it might be

GPS-denied. In such a situation, relying only on GPS information may cause failure.

### 2.3.2 RF Measurements with more resolution

The 3GPP 5G New Radio (NR) is envisioned to pave the way for achieving higher accuracy and robustness of localization in GPS-denied environments like indoors and improving localization outdoors combined with GPS. 5G NR provides improved measurements for localization like time-based, angular-based, and energy-based measurements. The measurements include TOA, TDOA, AOA/AOD, and multi-cell round-trip time. For robot localization, 5G in the downlink defines a new reference signal called positioning reference signal (PRS) based on which these measurements can be extracted. (the readers are referred to [100] for further information ).

Owing to the high carrier frequency, high bandwidth, and MIMO technology accurate measurements will be delivered. 5G NR operates at high-frequency bands: Frequency Range 1 (FR1) (450 MHz to 6 GHz) and Frequency Range 2 (FR2) ( frequency bands from 24.25 GHz to 52.6 GHz). Relying on the Cramer-Rao Lower Bound (CRLB) in (1.1), it can be observed that higher frequency and bandwidth contribute to higher TOA estimation accuracy.

The probability of LOS also increases due to the strong path loss in higher frequencies (mmWave). This feature and the large transmission bandwidth make distinguishing between LOS and NLOS measurements in the multipath effect feasible by applying proper analysis over the received signals, resulting in a highly precise time-based and angular-based localization.

The Massive MIMO technology is one of the most noticeable enhancements and relevant features for localization offered by 5G NR. It allows for the implementation of ultra-massive antenna arrays consisting of hundreds or thousands of antennas in a single base station leading to the finer angular resolution (azimuth and elevation of the beam) of even less than one degree which can contribute to accurate localization.

### 2.3.3 Vehicle-to-Everything Standard

The Third Generation Partnership Project (3GPP) deploys the Vehicle-to-Everything (V2X) based on Dedicated Short Range Communications which includes Vehicle-to-vehicle (V2V),

Vehicle-to-Network (V2N) or Vehicle-to-Infrastructure (V2I). The cellular V2X standard based on the 5G air interface is a fascinating feature for cooperative robotic missions, specifically through introducing the sidelink (SL), which permits vehicles to directly exchange information without other parts of the network being involved. This will play a vital role in cooperative tasks and localization, whether the operation is within the coverage area or without BS coverage. Furthermore, the V2X feature allows for not only communication of the vehicles with each other but with the infrastructure and even the internet. This increased connectivity improves the efficiency of the cooperative systems, enhances localization accuracy, and makes some missions feasible which were not before. For example, in an environment with adverse NLOS effects from the BS, which results in erroneous measurements, one or some parts of the vehicles can play the role of BSs or act as a pseudo BS.

### 2.3.4 Low Latency

Ultra-reliable low latency communication (URLL) offered in 5G new radio allows for future applications which is on-demand for aggressive latency for quick reaction. The existing 4G cellular network is not appropriate for this purpose.

Low latency means a slight delay between sending and receiving information indispensable for autonomous driving and flying. For example, 1 ms minimum latency (average of 10 ms) is expected to be brought by 5G, which is a substantial breakthrough compared to the 200 ms latency typical of 4G. The central station that might do the localization, planning, or control must receive and send back data and command fast enough to the vehicles operating at high speed. The time for transmitting and receiving data is vital for autonomous control, where the car moves rapidly. For example, for a drone to be controlled, both localization and control commands need to be processed and sent back to the UAV with an acceptable slight delay, resulting in a fast reaction of the UAV.

### 2.3.5 High throughput

Significantly for the uplink transmission, a high data rate is needed when offloading computations. Many algorithms try to balance computation power and accuracy mostly because

onboard computers are usually not equipped with a powerful processing unit. This concern would be prevented with offloading, which allows for implementing complex algorithms with high accuracy in real-time on a powerful server using edge computing. In addition to low latency, a high data rate is crucial for the real-time transfer of extensive sensor data -such as high-resolution images or LIDAR data. In [101], the role of edge computing and the impact of data rate in the uplink for vision-based drone navigation in the 5G network is explored. Three scenarios are tested and compared: no offloading, partial, and complete offloading. It shows how offloading can be advantageous in the network capable of providing a fast uplink rate.

### 2.3.6 Localization based on 5G

As the 5G roll-out is in the early stages, not many works have investigated 5G localization, especially for robotics applications and in a sensor fusion framework. Besides, most are just based on simulation results, neglecting many practical aspects. For example, the impact of synchronization error in time-based positioning or simplification of channel models is considered. The current state-of-the-art 5G-based localization is reviewed below.

In 5G NR, pilot signals are included for positioning purposes, Positioning Reference Signal (PRS) in down-link and Sounded Reference Signal (SRS) in up-link [100]. In addition to network centre frequency and bandwidth, PRS and SRS configuration also play roles in localization accuracy. In [94], for different combinations of centre frequency, sub-carrier spacing, and PRS comb-size, localization accuracy for simple scenarios is compared in terms of Root Mean Square Error (RMSE) using simulation. The authors in [102] simulate the roadside 5G network implementation for assisted driving, showing accuracy below 20-25 cm for 50-100 MHz bandwidth. Localization is performed based on TOA extracted as the first correlation peak between PRS and the received signal. The channel is modelled based on path loss and the TDL channel model. The impact of the geometrical placement of roadside 5G base stations on the localization based on EKF and how the distance from BS affects EKF linearization error is investigated in [103]. In this paper, EKF for location estimation is also presented where the covariance matrix is tuned dynamically, and improvement is shown through simulation. There are several attempts for localization in the 5G network based on

CIR and CFR. The work of [104] addresses localization in the up-link side based on 5G SRS information, in which, based on the received signal, the channel frequency response (CFR) is estimated. TOA and Direction of Arrival (DoA) are then evaluated using the well-known 2D multiple signal classification (MUSIC) algorithms. Localization is finally performed, and an indoor experimental setup achieves an accuracy of less than 1m. The authors in [98] generated a fingerprint dataset of AOA and its corresponding amplitude based on the CSI matrix in a 5G network. Deep Neural network (DNN) is trained and used as a regressor for online estimation. Quasi Deterministic Radio channel Generator (QuaDRiGa) [105] is exploited for channel modelling. Approximately 1-meter accuracy for NLOS and 0.1 m for LOS are reported. Based on the SRS symbol, in up-link, CFR is estimated for each base station in [106], and subsequently, TOA and AOA are jointly estimated for localization. Accuracy below 1m is acquired. Localization under the fingerprinting framework is explored in [107] based on CSI, where the transfer learning concept is leveraged to reduce the real-world training effort. QuaDRiGa is leveraged for obtaining synthetic CSI to pre-train the CNN model. In [108], angle-based fingerprint localization is conducted. The fingerprints include the angles (zenith and azimuth) along with their corresponding power for all observed paths. To validate the results, simulation is done by recreating 3D outdoor environments, including building geometry. AOA-based position estimation in 5G network experimentally reported in [109], where EKF is used at edge cloud for localization. The research of [110, 111] is related to receiver localization harnessing AOA under 5G MIMO System and beam-formed RSS, respectively.

While TOA, TDOA, AOA, and RSS of the LOS path could be directly related to the relative positions of the transmitter and receiver, there is no explicit connection between the NLOS path and close distance. Thus, the localization performance will be degraded noticeably only relying on those measurements. In literature, NLOS error mitigation techniques and ray tracing-based approaches are carried out to compensate for the NLOS ( see [112] and the references therein). Under the 5G network, this issue is addressed in [113]. The effect of NLOS in an unknown environment is dealt with in a fusion framework. Localization and navigation are accomplished by fusing TDOA and Pedestrian Dead Reckoning (PDR). TDOA

from LOS base stations is combined with TDOA from virtual Base stations placed in an unknown area whose locations are determined based on the NLOS base stations. Simulation and experiments are performed. The work of [91] goes beyond just localization by mapping the radio environment simultaneously, taking advantage of NLOS-rich information about transmitter and receiver positions and environmental obstacles. This paper proposes joint position and orientation estimation for a mobile target and the position estimation of reflectors and scatters relying on NLOS paths. Leveraging AOA, AOD, and TOA for each NLOS path, the receiver's location is determined only based on the received signal from one base station in 2D if there are at least 3 NLOS paths. It is shown in [90], in NLOS situation, cooperation among vehicles improves situational awareness and localization performance as several cars operate in the same environment where they might share one or more scatters, which results in a correlated multipath structure that can contribute to the improved localization.

### 2.3.7 Future Research Directions and Challenges

Most of the research on the use of 5G for localization approaches the problem from the pure communication point of view, while its use in various robotic applications is still in the infancy stage. In this section, future opportunities, research gaps, and challenges provided by 5G for robot applications are highlighted.

**Fingerprinting and Deep Learning applied to CSI:** Compared to range-based methods, fingerprinting has more potential to deliver higher localization precision. On the one hand, range-based methods are limited by the accuracy of the model they apply to extract ranges or angles. On the other hand, fingerprinting may utilize the signal information to its full extent. Mainly, CSI data contains latent knowledge that can be captured by the complex AI approaches linked with fingerprinting. Hence, appropriate Deep learning models may be the key to exploiting the CSI large matrix structure effectively to hallucinate necessary abstract features for localization. For example, deep neural networks may be able to decode these high-dimensional matrices to localize the obstacles around the receiver from the NLOS path.

Furthermore, if multi-array antennae are deployed at the receiver and transmitter, CSIs

contain even richer information since different versions of the same signal would be available as separate fingerprints. Therefore, effectively profiting from CSI information may prevent the need for several BSs for localization. It is shown in [91] that just one BS might suffice. Lastly, in this method, the main disadvantage of the traditional RF-based method, i.e., NLOS situations, is circumvented, under which these approaches become erroneous and unreliable.

There are challenges in fingerprinting that are worth much more attention. The availability of data and the collection of enough data is not trivial. The collected data set is not entirely reliable as the environment constantly evolves. Transfer learning as one possible solution is suggested to exploit synthetic data or already available data-set. In this area, there still seems to be much room to investigate the integration of deep learning with CSI fingerprints.

**Fusion of RF with other sensor data:** The mobility and the issue of receiving diverse information with different frequencies is a challenge that is missing in the literature. Localization algorithms are expected to be implemented in real-time, with each measurement coming in its own time. In fingerprinting, most existing works rely on fixed target localization and data sets consisting of RF features. However, in either the online or offline step, other sensor data like IMU and images can be used effectively with radio-based data sets. They can either narrow down the searching area in the online phase or directly be fused with fingerprints in the offline step. From a theoretical point of view, there is still room for localization improvement in accuracy and robustness in a data fusion scheme. For instance, to the author's best knowledge, there is no study on fusing promising measurements provided by 5G with images, LIDAR, etc. Several combinations of the sensor data and their performance under different situations and parameters need to be researched.

**Combination of Multiple Estimators:** Combining multiple estimators for range-based or fingerprinting could compensate for each method's disadvantages, improve accuracy, and introduce resiliency to failure cases. This technique is well-established in the general machine learning task and is known as ensemble learning [114]. The ensemble method combines different estimators and often yields much more accurate results than individual methods into four main paradigms Bayesian averaging, error-correcting output coding, Bagging, and

boosting [115].

**Cooperative Localization:** With V2X technology, many heterogeneous multi-robot applications comprising UGVs and UAVs can be envisioned. This enables more efficient use of robots for diverse scenarios like coverage, formation, task distribution, etc. For example, in a search and rescue operation, a group of heterogeneous robots may collaborate while some act as an anchor in the absence of a signal for some or all team members. On the other hand, sharing information by robots operating in the same area would endow robustness and accuracy to the localization system, as each robot can benefit from the knowledge and estimation of its neighbour(s).

**Orientation Estimation:** Deployment of a multi-array antenna system paves the way for orientation estimation. However, attitude estimation would be challenging, especially in 3D space while the UAV is flying. Therefore, integrating RF signals with other data to improve attitude estimation in 3D seems to be an attractive research topic left untouched.

**Experimental setup and Realistic Simulation:** Real experiments and setups for the use of 5G with robots are lacking in the state-of-the-art. This is critical as the actual setup's outcome does not necessarily match the predictions and expectations. Two examples are the offloading and handover in the 5G network. In the uplink, which is crucial for offloading, the data rate in 5G NR is supposed to be significantly improved, but [116] recorded the maximum of 67 Mbit/s in the uplink for a flying drone showing no improvement compared to 4G. The impact of handover is also worth further researching, especially for high mobility scenarios. [116] shows that the handover rate between LTE and 5G is too high, which is unacceptable.

**Off-loading:** With offloading, more computation power will be available. This calls for new complex strategies and algorithms capable of being run and worked in parallel. For example, the onboard computer and edge server might work collaboratively and separately. While onboard computers process some part of the data to execute a rough localization, heavy algorithms are to be implemented on the edge server separately when there is no communication between the two processors. Both algorithms then need to be merged to

output the refined localization result. At the same time, new strategies might be developed with the processing power on edge devices to focus more on accuracy and robustness instead of the computational burden. Most state-of-the-art localization tries to balance accuracy, complexity, and computation power.

**Simultaneous Localization and Mapping:** A critical task in robotics is Simultaneous Localization and Mapping (SLAM). As the robot moves in an unfamiliar environment, it needs to construct a map and localize itself on the map while navigating. The use of multi robots instead of a single one, while each shares its maps with others, is a fascinating topic that adds efficiency to the SLAM. Multi-robot SLAM (MRSLAM) is implemented in two ways. Using a central station to which all robots disseminate the collected data to do all the processing, construct the global map, and transmit this data back to each robot. In a decentralized system, each robot performs locally, and whenever they visit, other robots share their local map, based on which they update their map. Accordingly, one big challenge in centralized and distributed MRSLAM is the constraint on the communication bandwidth, limited computation power, and memory. 5G NR will circumvent these limitations by facilitating edge computing [117, 118], providing accurate RF-based measurements, and enabling relative localization for each vehicle in the team via sidelink and V2V technologies.

**Vertical Localization Accuracy:** As discussed earlier, the localization accuracy in the z-direction is always less than in the x-y. Notably, the multi-path effect and the small offset of anchors in the z-direction yield poor vertical accuracy relying on conventional methods [119]. MIMO technology and wide-band mmWave system of 5G provide promising solutions to improve the 3D positioning, like taking full advantage of CFR in all frequencies or analyzing antenna radiation pattern [120].

**Safety:** Safety is a major concern in robotics-related tasks. In general, there are two types of communication that apply to UAVs: Control and Non-Payload Communication (CNPC) and Payload Communication (PC). CNPC refers to control commands, way-points and navigation, usually in the order of several Kbps. Instead, the PC includes data transmission to the edge

server, ground or aerial centre for processing. This information ranges from small-sized data, e.g., IMU, GPS, and RSS, to large-size high-resolution aerial images or LIDAR scans. While communication link reliability might not be vital for PC, avoiding communication interruptions is critical regarding CNPC. This leads to the concern of interruption of the CNPC link when there is no LOS path. Remarkably, using mmWave and beam-forming features of 5G call for additional care because they come with the disadvantage of very high propagation pass loss [121]. Therefore, the design of trajectory and path-planning algorithms that guarantee LOS communication would be an interesting research topic. Another concern for CNPC links is cyber-physical attacks that might corrupt data transmission or cause the misaction of the UAVs or UGVs [26].

## 2.4 Research Gaps and Contributions in 5G-Based UAV SLAM

As 5G technology continues to advance, several critical research gaps remain unresolved. This systematic analysis identifies these knowledge gaps and technical limitations to guide future research priorities and drive innovation in 5G-enabled UAV positioning systems.

- **Limited scope:**
  - Many existing methods target static targets or vehicles with constant velocity, while UAVs exhibit dynamic behaviors and complex motion patterns
  - Prior studies typically focus on idealized scenarios that do not reflect the complexities of real-world indoor environments, particularly in terms of signal propagation and multipath effects
  - There is a notable scarcity of comprehensive sensor fusion frameworks specifically designed to integrate 5G measurements with existing UAV sensor suites for robust indoor operation
- **Methodological and Integration Limitations:**
  - Many existing studies rely on basic multilateration or simplified sensor fusion techniques, lacking the sophistication needed for robust UAV operation

  - Current visual SLAM systems have not been adequately adapted to leverage the potential of 5G measurements for enhanced global consistency and drift correction
  - The integration of 5G measurements with visual-inertial systems remains largely unexplored, particularly in addressing fundamental challenges like scale ambiguity and global reference frame alignment
  - Existing approaches often fail to consider the temporal characteristics of different sensor modalities, leading to suboptimal fusion strategies
  - Some approaches emphasize communication aspects, such as signal analysis and estimation of angles/times of arrival, without fully addressing the challenges of accurate localization

- **Accuracy and Robustness Limitations:**
  - A number of methods achieve only sub-meter accuracy, which may be insufficient for precise indoor UAV operations
  - There is limited research targeting the mitigation of drift from inertial measurement units (IMUs), which is crucial for reliable UAV localization
  - Existing approaches lack robust solutions for maintaining accurate localization in visually degraded or feature-poor environments
  - Current methods struggle to maintain consistent performance under challenging signal conditions or partial base station visibility

- **Evaluation and Benchmarking Deficiencies:**
  - The field lacks standardized evaluation frameworks that incorporate realistic 5G network configurations and channel models
  - Existing studies often use proprietary or limited datasets, making direct performance comparisons challenging
  - There is insufficient analysis of system performance across different sensor combinations and environmental conditions

  - The impact of varying base station densities and network configurations on localization accuracy remains inadequately quantified
  - Few works provide comprehensive comparisons across different localization approaches in consistent experimental settings

## Research Gap Connection

The contributions presented in this thesis directly address the identified research gaps as follows:

1. **Limited Scope and Dynamic Challenges:**

- Developing comprehensive sensor fusion frameworks (Ch. 4 & Ch. 5) that handle dynamic UAV trajectories, achieving up to 13cm accuracy with PGO and consistent sub-meter performance across diverse indoor environments
- Successfully integrating 5G measurements with existing UAV sensor suites, demonstrating significant improvements across various SLAM configurations

2. **Methodological and Integration Solutions:**

- Introducing sophisticated ESKF- and PGO-based algorithms (Ch. 4) that achieve 34 cm and 13 cm accuracy respectively with just IMU and 5G TOA measurements
- Enhancing ORB-SLAM3 with 5G integration (Ch. 5), enabling globally consistent map creation and localization while addressing key limitations in traditional visual SLAM
- Implementing novel approaches for unknown base station positions and relaxed connectivity conditions, maintaining effective operation with sequential connections to base stations at different time intervals
- Developing optimal fusion strategies that maintain real-time performance while handling multi-rate sensor data from IMU, vision, and 5G measurements

3. **Enhanced Accuracy and Robustness:**

- Demonstrating effective SLAM operation with sequential base station connectivity, where only one base station maintains line-of-sight during specific non-overlapping time intervals
- Providing TOA measurements as an effective alternative to loop closure, improving accuracy by up to 29.6% in scenarios where traditional loop closure is unavailable
- Maintaining reliable performance even with minimal infrastructure of three base stations while achieving optimal accuracy with five base stations

4. **Comprehensive Evaluation Framework:**

- Augmenting EuRoC MAV and Aerolab datasets with realistic 5G simulations using QuaDRiGa and MATLAB
- Conducting systematic performance analysis across multiple indoor scenarios with varying base station configurations and connectivity patterns
- Providing detailed comparative analysis with state-of-the-art approaches, including UWB-VO implementations, demonstrating robust performance even with lower TOA measurement frequency (5 Hz versus 30 Hz)
- Evaluating system performance across different sensor combinations, base station configurations, and connectivity scenarios

These contributions bridge the gap between visual-inertial and radio-based methods for UAV localization and SLAM by integrating 5G measurements with existing sensor modalities. The developed frameworks demonstrate practical solutions for indoor environments that can operate effectively without prior knowledge of base station positions and with intermittent connectivity. The comprehensive evaluation across different scenarios and configurations provides valuable insights for implementing 5G-enabled UAV navigation systems in real-world applications where infrastructure knowledge may be limited or connectivity patterns irregular.

# Chapter 3

# Fundamentals of 5G Signal Structure and Time of Arrival Simulation

The evaluation of 5G-based localization systems presents significant challenges, particularly in terms of infrastructure requirements. Although we had access to a 5G base station installed on the rooftop of a building for outdoor experiment, this setup was insufficient for conducting comprehensive outdoor TOA measurements. Additionally, hardware limitations and a lack of specialized equipment for indoor measurements further restricted our ability to perform thorough real-world experiments. Additionally, real-world testing would have constrained the systematic evaluation of various network configurations and signal parameters. To address these challenges, a simulation-based approach using MATLAB and QuaDRiGa was adopted. This validated framework generates realistic TOA measurements while ensuring experimental control and reproducibility, enabling a thorough investigation of 5G-based localization system performance under diverse conditions.

This chapter details the key concepts behind 5G signal structures, including Positioning Reference Signals (PRS), establishing the fundamental knowledge essential for understanding TOA measurement simulation. In addition, the chapter introduces QuaDRiGa as the channel simulator and describes how its features are leveraged to produce realistic 5G TOA data.

## 3.1 5G Signal Structure and PRS Fundamentals for Localization

This section explains the underlying signal structure of 5G networks and its role in localization using PRS. The section covers the essential parameters—from OFDM modulation to resource grid configuration—that enable TOA estimation.

### 3.1.1 OFDM and Subcarrier Spacing in 5G

5G employs *Orthogonal Frequency Division Multiplexing* (OFDM) along with a cyclic prefix, a guard interval added at the start of each OFDM symbol to mitigate inter-symbol interference. In the time domain, OFDM breaks down a high-speed stream of digital bits into multiple parallel slower-moving streams. In the frequency domain, OFDM divides a wideband signal into multiple narrowband subcarriers, each of which carries a small portion of the data.

*Subcarrier spacing*, referring to the separation or interval between adjacent subcarriers within the frequency domain, is a crucial parameter in 5G localization accuracy. 5G offers specific subcarrier spacing options of 15, 30, 60, 120, and 240 kHz, following a $2^n 15$ kHz pattern. The choice of subcarrier spacing influences not only the number of subcarriers but also the symbol duration and cyclic prefix length. Larger subcarrier spacing reduces symbol duration, enabling shorter slot lengths for low-latency applications.

### 3.1.2 5G Frame Structure and Resource Grid

In 5G's downlink transmission, a frame has a duration of 10 milliseconds, which is further divided into ten subframes, each lasting 1 millisecond. Figure 3.1 shows the time-domain structure of the 5G frame, illustrating its reliance on the subcarrier spacing. In the time domain, the choice of subcarrier spacing determines the number of slots in a subframe, each slot containing 14 OFDM symbols. As illustrated in this figure, subframes are subdivided into 1, 2, 4, 8, or 16 slots, depending on the chosen subcarrier spacing, all utilizing a standard cyclic prefix.

The *resource grid* is a two-dimensional structure where time is represented along one axis, and frequency is represented along the other (matrix of subcarriers and OFDM symbols).

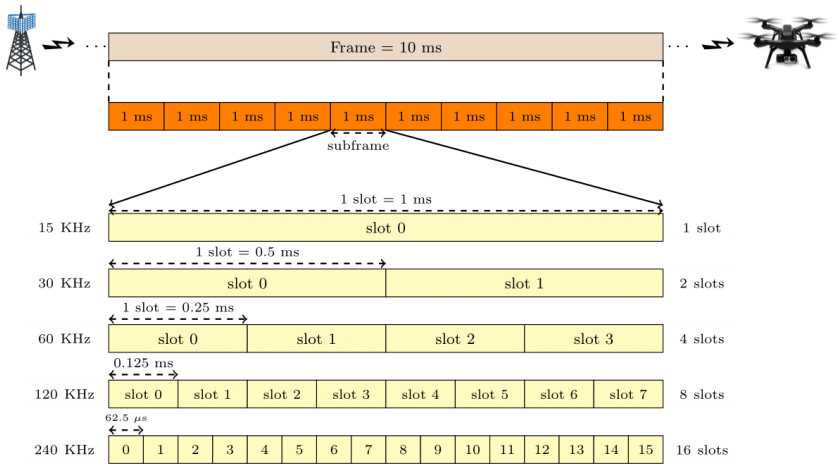



Figure 3.1: 5G Frame Structure

A resource grid is characterized by one subframe in the time domain and full carrier bandwidth in the frequency domain. This grid is used to allocate resources for communication between the base station (gNodeB) and user devices (User Equipment, UE).

Resource grids are further divided into *Resource Blocks* (RBs). Each RB is composed of a group of subcarriers. The number of subcarriers in an RB depends on the subcarrier spacing. For example, with 15 kHz subcarrier spacing, an RB contains 12 subcarriers. A *Physical Resource Block* (PRB) is the smallest unit of resource that can be allocated to a user in 5G. A PRB is defined as a group of consecutive subcarriers in the frequency domain and a group of consecutive OFDM symbols in the time domain. Finally, a *Resource Element* (RE) is the smallest unit of resource allocation in the 5G resource grid. Each RE contains one subcarrier and one OFDM symbol. Figure 3.2 shows the resource grid, PRBs, RBs, REs, as well as the distribution of the reference signal PRS, which will be explained in more detail in the next section.

### 3.1.3 Positioning Reference Signals (PRS) in 5G

For precise positioning, 5G supports *Positioning Reference Signals* (PRS) in the downlink. PRS are generated using Gold sequences, which are a type of pseudo-random code, enabling easy reproduction of the PRS at the receiver and resistance to interference. Gold sequences

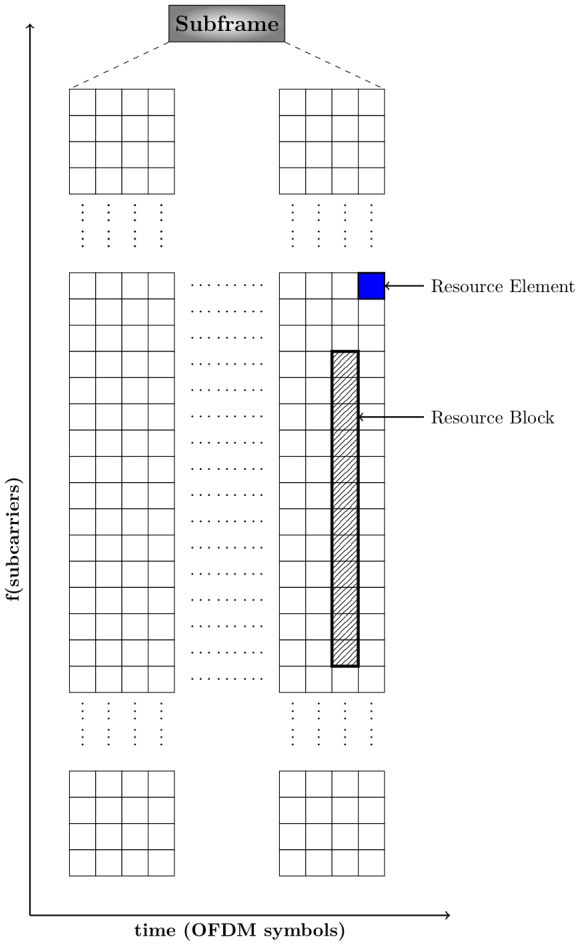



Figure 3.2: Visualization of the 5G Resource Grid Structure, highlighting Resource Blocks (RBs) and Resource Elements (REs), showcasing the allocation of resources in the Resource Grid

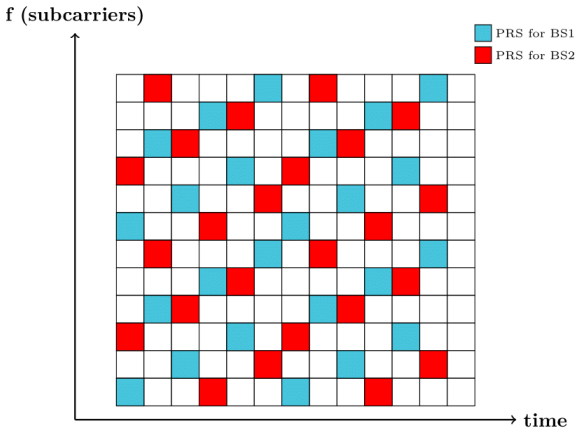



Figure 3.3: PRS distribution in a Physical Resource Block (PRB) in 5G NR with two BSs using a comb-6 structure

also have good cross-correlation properties, meaning that they have a low correlation with all other sequences in the same set of Gold sequences.

PRS symbols are generated using *Quadrature Phase Shift Keying* (QPSK) modulation and distributed across time and frequency in the resource grid. The PRS pattern can be configured with parameters like starting resource element, periodicity, and density, depending on the specific requirements of the positioning application. One critical configuration that significantly impacts the TOA estimation accuracy is the *PRS bandwidth*. A wider PRS bandwidth will result in more accurate and robust localization but at the cost of increased interference.

Downlink PRS utilizes a comb-like frequency structure by transmitting PRS within a subset of the sub-bands derived from the divided PRS bandwidth. The comb size N in 5G PRS dictates that every Nth subcarrier is used by PRS symbols, with configurable values of N (2, 4, 6, or 12). This effectively diminishes interference from neighboring base stations. It also helps to improve the correlation properties of the PRS signal, which leads to enhanced accuracy and robustness of TOA estimation. *PRS muting* is another feature in 5G PRS which can also decrease the interference of PRS from neighboring BSs. Figure 3.3 illustrates a Physical Resource Block (PRB) featuring PRS symbols configured with a comb-6 structure for two BSs.

Before transmission, the PRS and data signals are OFDM modulated, and a cyclic prefix is added to the beginning of each OFDM symbol.

### 3.1.4 Time of Arrival Estimation at the Receiver

At the receiver, the UE processes the received signal by first eliminating the cyclic prefix and applying a Fast Fourier Transform (FFT) to reconstruct all OFDM symbols in the frame, including PRSs. While perfect synchronization between the receiver and base station is ideal, practical implementations must account for timing offset errors. These timing errors are estimated and compensated for during the correlation process, where the UE correlates the received PRS signal with the known PRS pattern to identify the TOA. The UE identifies the peak of the correlation function, which corresponds to the TOA of the signal from the base station. This process is repeated for all BSs from which signals are received.

### 3.1.5 Line-of-Sight Constraints and Research Focus

This study focuses exclusively on Line-of-Sight (LOS) conditions, as leveraging TOA measurements effectively in Non-Line-of-Sight (NLOS) scenarios typically requires intensive signal processing. Accurately modeling of complex multipath propagation, estimating virtual anchor locations, and maintaining real-time performance under dynamic flight conditions pose significant challenges, particularly for resource-constrained UAV platforms.

From a practical standpoint, 5G signals—especially at millimeter-wave (mmWave) frequencies—are highly directional and have short propagation distances. This characteristic increases the likelihood of achieving LOS in many indoor environments when base stations are optimally placed. The use of beamforming and MIMO (Multiple Input Multiple Output) technologies further enhances 5G's effectiveness in LOS environments, making it particularly well-suited for such scenarios. While NLOS propagation does occur, the substantial attenuation of 5G signals in such conditions makes multipath components less reliable for precise localization.

Indoor drone applications impose stringent real-time computational constraints that often exceed the capabilities of typical UAV hardware. Traditional NLOS channel modeling and

multipath mitigation algorithms require high-dimensional filtering and dense environmental maps, which are computationally demanding. These requirements are particularly challenging for agile drones.

Researchers has developed theoretical frameworks for NLOS localization, but these are often based on idealized simulations with basic geometric reflectors and perfect channel estimations [90, 91]. While these studies provide valuable insights into fundamental NLOS principles, their practical implementation for UAV navigation in real-world indoor environments would require substantial additional development. Such efforts would significantly broaden the scope of this thesis beyond its robotics-oriented objectives.

The research prioritizes robotic systems, sensor fusion, and SLAM, with a deliberate focus on challenges directly relevant to UAV navigation and localization. Communication-specific challenges—such as advanced NLOS signal processing—are avoided to maintain computational feasibility and alignment with robotics-oriented objectives. Instead, the study prioritizes LOS scenarios, where the inherent strengths of 5G technology (e.g., beamforming, MIMO, and mmWave directional propagation) synergize with the computational capabilities of UAV platforms.

Furthermore, preliminary results from this study indicate that even in LOS conditions, TOA measurements provide only marginal improvements when fused with rich sensor data such as visual features. This observation suggests that in NLOS scenarios—where TOA measurements are less reliable and require significant additional processing—the cost-benefit ratio of implementing complex NLOS mitigation strategies becomes even less favorable. This empirical finding further justifies the focus on LOS conditions, where TOA can deliver meaningful benefits without excessive computational overhead.

Given these considerations, this study maintains its focus on LOS TOA localization as a practical, robotics-centric solution. By leveraging LOS scenarios—often achievable with carefully planned 5G infrastructure—the system can achieve meaningful gains in accuracy and robustness while maintaining low computational overhead. The goal is to demonstrate how TOA integration with drone sensor fusion can improve state estimation in real-world indoor environments, even with limited infrastructure. This work provides a strong foundation

for robotics applications in LOS conditions and outlines clear pathways for extending these methods to NLOS scenarios through future interdisciplinary collaboration.

## 3.2 QuaDRiGa for 5G Time-of-Arrival Simulations

Accurate wireless channel modeling is a cornerstone of high-fidelity TOA simulation in 5G networks, particularly for UAV applications. In this research, we utilize the *quasi-deterministic radio channel generator* (QuaDRiGa) [122] to generate realistic channel impulse responses (CIRs) that enable 5G TOA simulation. QuaDRiGa's ability to blend geometric consistency with stochastic multipath modeling makes it uniquely suited to indoor or mixed environments where UAVs often operate.

### 3.2.1 Overview of QuaDRiGa

QuaDRiGa (QUAsi Deterministic RadIo channel GenerAtor) evolved from the WINNER family of channel models and implements the 3GPP TR 38.901 standard for 3D wireless propagation. It is widely adopted in academia and industry for simulating 5G-compliant channel characteristics, including supports for carrier frequencies ranging from 450 MHz to 100 GHz—covering both FR1 and FR2 5G bands—with bandwidths up to 1 GHz. This wide frequency range, combined with its geometric approach, allows researchers to simulate diverse propagation scenarios accurately, from indoor office layouts to large outdoor spaces.

*: Statistical Ray-Tracing Methodology Unlike purely deterministic ray-tracing that relies on detailed environmental geometry, QuaDRiGa implements a *statistical ray-tracing* approach. As depicted in Figure 3.4, the model stochastically distributes scattering clusters in 3D space to represent potential reflection and diffraction points (e.g., walls, furniture, or other obstacles). For each path, QuaDRiGa computes:

- **Angles of departure/arrival:** Relating the transmitter, receiver, and cluster locations
- **Path lengths and delays:** Translating geometric distance into signal delay (key for TOA)

- **Path power contributions:** Accounting for both line-of-sight (LOS) and non-LOS (NLOS) components

This hybrid approach balances geometric consistency with stochastic realism, resulting in an accurate yet efficient framework for simulating UAV flight paths in changing environments.

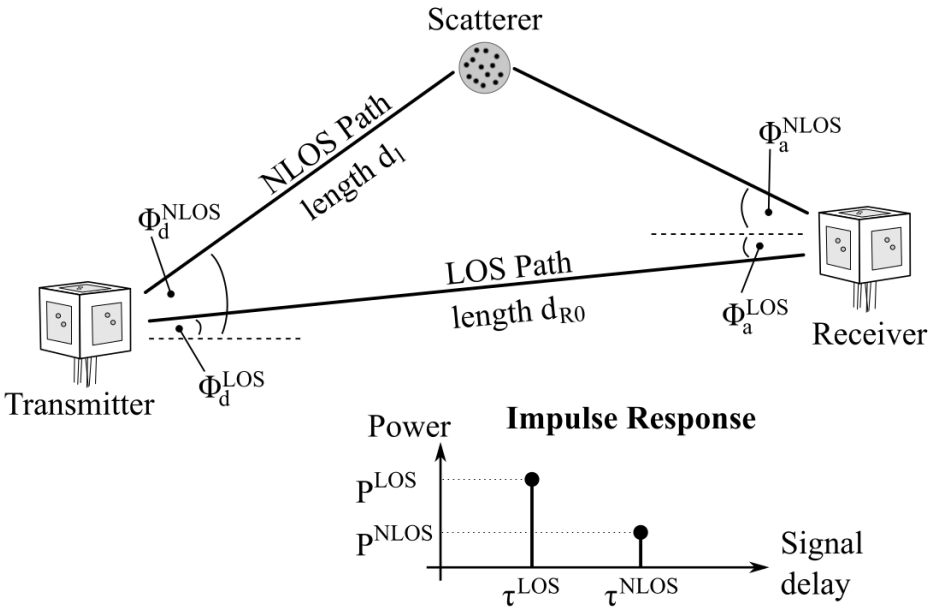



Figure 3.4: QuaDRiGa's statistical ray-tracing approach showing LOS (direct) and NLOS (scattered) paths. Angles of departure ($\phi_d$) and arrival ($\phi_a$) are computed relative to randomly placed clusters. Adapted from [122].

### 3.2.2 Channel Modeling Framework in QuaDRiGa

***:** Scattering Cluster Architecture QuaDRiGa groups propagation paths into clusters, each containing multiple sub-paths. These sub-paths share similar arrival and departure angles but differ slightly in phase and delay. A typical scenario may feature 8–20 clusters, each consisting of 20 sub-paths, thereby capturing rich multipath conditions without exhaustive environmental modeling. Though individual sub-paths within a cluster exhibit narrow angular spreads (1–6°), the cumulative effect of multiple clusters distributed across the environment results in wide overall angular spreads. This behavior accurately models real-world indoor

or mixed indoor-outdoor UAV scenarios, where signals reflect off walls, ceilings, and other obstacles from diverse directions.

***:** Dynamic Channel Evolution A key advantage of QuaDRiGa for UAV-based localization is its capability to handle mobility and dynamic channel conditions:

- **Continuous Parameter Updates:** As the UAV moves, path delays, angles, and fading parameters adjust over time to reflect the changing transmitter–receiver geometry.
- **Variable Speed Support:** The model accommodates acceleration/deceleration profiles and can smoothly transition between different propagation environments.
- **Spatial Consistency:** Large-scale and small-scale fading parameters remain correlated over space, an essential factor for realistic simulation of UAV flight trajectories.

***:** Predefined Scenarios and Validation QuaDRiGa provides a library of predefined parameter sets for diverse scenarios, including indoor environments, which are the focus of this thesis. These parameter sets are aligned with 3GPP standards [123] and real-world measurement campaigns [124, 125], ensuring accurate modeling of channel characteristics relevant to UAV localization. Importantly, QuaDRiGa's modeling principles and scenario definitions have been validated extensively through peer-reviewed research and real-world measurement campaigns, such as those conducted in industrial indoor environments[125]. This comprehensive validation ensures that simulated channel characteristics—such as multipath delay profiles and angular spreads—closely mirror those observed in practice. As a result, the generated TOA measurements are highly reliable, enabling precise localization in indoor UAV applications.

### Summary

QuaDRiGa's blend of geometric detail and statistical modeling makes it an ideal choice for simulating 5G TOA measurements in UAV localization systems. Its flexible framework, wide frequency support, and dynamic channel evolution features enable high-fidelity emulation

of multipath effects and realistic signal delays. By integrating QuaDRiGa with imaging and IMU sensor data, this work provides a robust platform to evaluate and refine next-generation 5G-based localization and SLAM algorithms for drone operations.

## 3.3 Simulation Pipeline and Integration

The simulated 5G TOA data is integrated with real vision and inertial data from the UAV's sensor suite, forming a comprehensive multi-modal dataset. This fusion is critical for evaluating Localization and SLAM algorithms such as the ESKF, PGO and visual-inertial SLAM Framework. The process of TOA simulation includes several sequential steps:

- **Signal Generation:** Generates the 5G signals using MATLAB 5G Toolbox with standard-compliant configurations. This includes defining carrier frequencies in both FR1 and FR2 bands, system bandwidth, PRS configurations (comb size, PRB allocation, periodicity), subcarrier spacing, and appropriate SNR levels for indoor environments. The specific network configuration parameters are detailed in the respective experimental chapters.
- **QuaDRiGa Channel Simulation:** Simulates the propagation of 5G signals through the environment. A typical simulation run requires:
  - Setting transmitter positions in the global coordinate system
  - Defining antenna properties for transmitters and receiver
  - Configuring user trajectory based on real UAV flights dataset (EuRoC MAV or AeroLab datasets)
  - Defining states/segments along the trajectory
  - Assigning appropriate propagation environments
- **Channel Response Application:** The channel impulse responses generated by QuaDRiGa are applied to the transmitted signals to obtain the received signals at the UAV's position, accounting for all propagation effects modeled by QuaDRiGa.

- **TOA Estimation:** Processes the received signals through signal processing techniques including correlation with the transmitted PRS and peak detection to extract precise timing information. These time measurements are then converted to distance estimates.
- **Dataset Augmentation:** The simulated 5G TOA data is applied to two datasets for comprehensive evaluation of Localization and SLAM algorithms:
  - **EuRoC MAV Dataset:** In **Chapter 4**, the widely used EuRoC MAV dataset—comprising vision and inertial data from indoor drone flights—is augmented with simulated TOA measurements for benchmarking localization algorithms.
  - **AeroLab Dataset:** In **Chapter 5**, our AeroLab dataset (collected from a custom indoor experimental setup) is similarly enhanced with simulated 5G TOA data to assess the robustness of pose estimation in varying environments.
- **Localization and SLAM Implementation:** The augmented datasets are used to evaluate different pose estimation approaches:
  - Error State Kalman Filter (ESKF) as detailed in Chapter 4
  - Pose Graph Optimization (PGO) as presented in Chapter 4
  - Our developed SLAM framework presented in Chapter 5

## 3.4 Analysis of Time-of-Arrival (TOA) Error Distribution

This section analyzes the distribution of TOA errors to justify the Gaussian noise assumption commonly used in sensor fusion frameworks. The analysis is based on two comprehensive datasets: (1) 17,430 simualted TOA measurements on the EuRoC MAV dataset at 28 GHz frequency, and (2) 33,500 TOA simulated measurements on the Aerolab dataset at 78 GHz frequency. The goal is to determine whether the TOA errors can be reasonably modeled as Gaussian, or if a more complex error model is required.

### 3.4.1 Methodology

To analyze the TOA error distribution, the following steps are employed:

1. **Data Preprocessing**: The raw TOA errors were computed as the range difference between the measured TOA and the ground truth. The dataset was then analyzed for skewness and kurtosis to assess deviations from Gaussianity.

2. **Winsorization**: Extreme values in the dataset were truncated by clipping the data to the 1st and 99th percentiles. This step reduces the influence of outliers while preserving the central distribution.

3. **Z-score Filtering**: Data points with absolute z-scores greater than 2.5 were removed to filter out extreme outliers. This step ensures that the analysis focuses on the central portion of the error distribution.

4. **Central 95% Analysis**: The middle 95% of the data was isolated to analyze the core distribution, excluding the tails that may contain non-Gaussian behavior.

5. **Gaussian Mixture Model (GMM)**: A GMM was fitted to the data to determine if the error distribution could be better represented as a mixture of multiple Gaussian components. The Bayesian Information Criterion (BIC) was used to select the optimal number of components.

6. **Normality Tests**: The Shapiro-Wilk, D'Agostino's $K^2$, and Kolmogorov-Smirnov tests were applied to assess Gaussianity. These tests were performed on both the full dataset and the central 95% of the data.

### 3.4.2 Results for EuRoC MAV Dataset

The dataset analysis of 17,430 TOA errors revealed a distribution with a mean $\mu = -0.007141$ and a standard deviation $\sigma = 0.332053$, demonstrating excellent symmetry with a skewness of approximately $-0.003$ and an excess kurtosis of $0.095$, suggesting a distribution very close to Gaussian. When applying Winsorization by truncating the extreme $1\%$ of data, the

standard deviation decreased slightly to $\sigma = 0.327804$, indicating minimal impact from outliers. Z-score filtering with $|z| \leq 2.5$ retained $99.93\%$ of the original data (17,417 data points), further confirming few extreme outliers. The central $95\%$ of the data maintained a similar mean $\mu = -0.007404$ with a reduced standard deviation $\sigma = 0.308018$. BIC analysis for the Gaussian Mixture Model suggested that the error distribution is best modeled as a mixture of 4 Gaussian components:

- Component 1: Weight $= 0.2424$, $\mu = -0.440011$, $\sigma = 0.103571$
- Component 2: Weight $= 0.2016$, $\mu = 0.000007$, $\sigma = 0.001031$
- Component 3: Weight $= 0.1347$, $\mu = -0.143382$, $\sigma = 0.100800$
- Component 4: Weight $= 0.4214$, $\mu = 0.282012$, $\sigma = 0.234313$

Despite the near-zero skewness and minimal excess kurtosis suggesting a distribution remarkably close to Gaussian, all statistical tests (Shapiro-Wilk, D'Agostino's $K^2$, and Kolmogorov-Smirnov) rejected the null hypothesis of Gaussianity for both the full dataset and the central $95\%$ ($p < 0.05$). The observed percentages of data within $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ were $57.36\%$, $99.86\%$, and $99.93\%$, respectively, showing some deviation at the $\pm 1\sigma$ level from the expected $68.27\%$, but excellent conformity at higher thresholds. As illustrated in Figure 3.5, the error distribution exhibits a multi-modal nature with a pronounced central spike, while maintaining excellent agreement with theoretical Gaussian properties at higher sigma thresholds.

### 3.4.3 Results for Aerolab Dataset

The dataset analysis of 33,500 TOA errors revealed a distribution with a mean $\mu = -0.010457$ and a standard deviation $\sigma = 0.151413$, demonstrating good symmetry with skewness of $0.042$, while the excess kurtosis of $-0.554$ suggested a slightly flatter distribution than a perfect Gaussian. When applying Winsorization by truncating the extreme $1\%$ of data, the standard deviation decreased negligibly to $\sigma = 0.151237$, indicating minimal impact from outliers. Z-score filtering with $|z| \leq 2.5$ retained $100.00\%$ of the original data (33,500 data points), suggesting no extreme outliers beyond this threshold. The central $95\%$ of the data

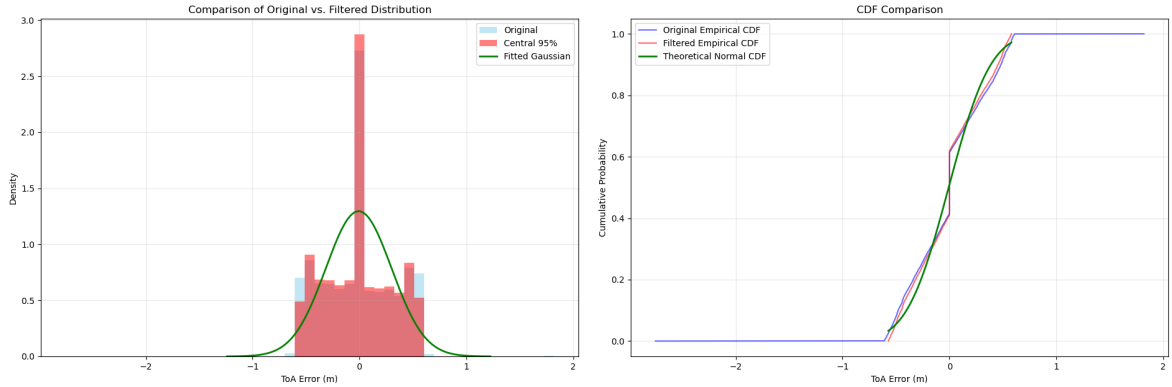



Figure 3.5: Visual analysis of EuRoC MAV TOA error distribution (28 GHz). (left) Histogram of error distribution with fitted Gaussian curve, illustrating the multi-modal nature of the data with a pronounced central spike (Component 2: weight=0.2016, $\mu \approx 0$) and distinct shoulders corresponding to other mixture components. Despite the multi-modal structure, the distribution exhibits near-perfect symmetry (skewness=-0.003) and kurtosis (0.095) remarkably close to Gaussian. (right) CDF comparison showing excellent agreement between empirical and theoretical Gaussian distributions at higher sigma thresholds critical for robust estimation: 99.86% at $\pm 2\sigma$ (vs. expected 95.45%) and 99.93% at $\pm 3\sigma$ (vs. expected 99.73%). These results support the practical validity of Gaussian modeling for TOA errors in sensor fusion applications despite statistical tests rejecting perfect Gaussianity.

maintained a similar mean $\mu = -0.010992$ with a reduced standard deviation $\sigma = 0.139882$. BIC analysis for the Gaussian Mixture Model suggested that the error distribution is best modeled as a mixture of 4 Gaussian components:

- Component 1: Weight $= 0.1509$, $\mu = -0.241996$, $\sigma = 0.039834$
- Component 2: Weight $= 0.2195$, $\mu = -0.000271$, $\sigma = 0.001477$
- Component 3: Weight $= 0.3985$, $\mu = 0.119797$, $\sigma = 0.109785$
- Component 4: Weight $= 0.2310$, $\mu = -0.093571$, $\sigma = 0.078189$

Despite the low skewness and negative excess kurtosis, all statistical tests (Shapiro-Wilk, D'Agostino's $K^2$, and Kolmogorov-Smirnov) rejected the null hypothesis of Gaussianity for both the full dataset and the central $95\%$ ($p < 0.05$). The observed percentages of data within $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ were $59.94\%$, $94.01\%$, and $100.00\%$, respectively, showing some deviation at the $\pm 1\sigma$ level from the expected $68.27\%$, and a small deviation at the $\pm 2\sigma$ level from the

expected $95.45\%$, but excellent conformity at the $\pm 3\sigma$ threshold. The low skewness $(0.042)$ indicates good symmetry in the error distribution. The negative excess kurtosis $(-0.554)$ suggests a distribution with slightly lighter tails than a perfect Gaussian, which is favorable for sensor fusion applications as it indicates fewer extreme values. Figure 3.6 visualizes this distribution, showing both the multi-modal characteristics and the good agreement with theoretical Gaussian properties at the critical thresholds.

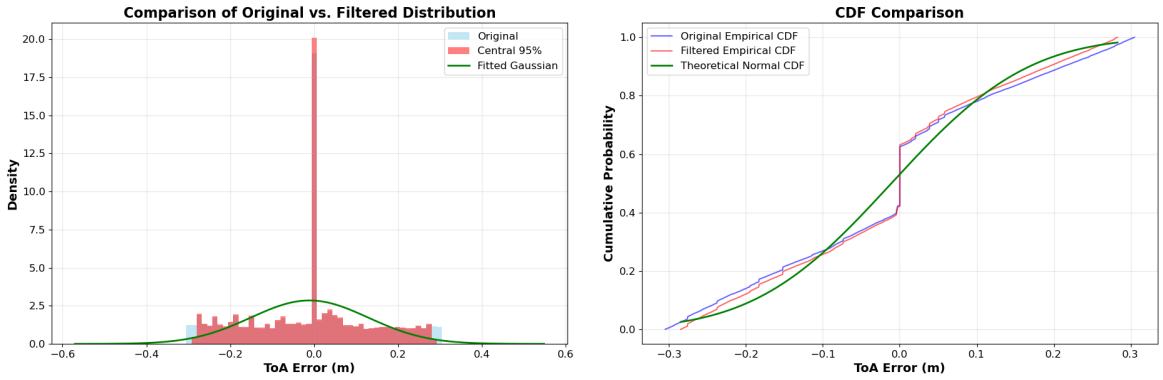



Figure 3.6: Visual analysis of Aerolab TOA error distribution (78 GHz). (left) Comparison of original vs. filtered distribution with fitted Gaussian curve, showing the multi-modal nature of the data with a pronounced central spike (Component 2: weight=0.2195, $\mu \approx 0$) and distinct shoulders from the mixture components. The distribution exhibits good symmetry (skewness=0.042) and a slightly flatter distribution (excess kurtosis=-0.554) than a perfect Gaussian. (right) CDF comparison showing good agreement between empirical and theoretical Gaussian distributions at the critical thresholds: 94.01% at $\pm 2\sigma$ (vs. expected 95.45%) and 100.00% at $\pm 3\sigma$ (vs. expected 99.73%). These results support the practical validity of Gaussian modeling for TOA errors at higher frequencies as well.

### 3.4.4 Comparative Analysis and Practical Validation

The summary of analyses of both datasets is shown in Table 3.1. Both datasets exhibit the following key characteristics:

- **Multi-modal Distribution**: Both datasets are best characterized by a 4-component Gaussian mixture model, suggesting inherent complexity in the TOA error behavior.
- **Near-zero Mean**: Both distributions have means very close to zero, indicating minimal

Table 3.1: Comparison of TOA Error Distributions Across Datasets

| **Metric** | **EuRoC MAV** | **Aerolab** |
|---|---|---|
| Sample size | 17,430 | 33,500 |
| Mean ($\mu$) | -0.007141 | -0.010457 |
| Standard deviation ($\sigma$) | 0.332053 | 0.151413 |
| Skewness | -0.003 | 0.042 |
| Excess kurtosis | 0.095 | -0.554 |
| GMM components | 4 | 4 |
| $\pm 1\sigma$ coverage | 57.36% | 59.94% |
| $\pm 2\sigma$ coverage | 99.86% | 94.01% |
| $\pm 3\sigma$ coverage | 99.93% | 100.00% |

systematic bias in the TOA measurements.

- **Similar Sigma Coverage Pattern**: Both datasets show a similar pattern in their sigma coverage, with lower-than-expected coverage at $\pm 1\sigma$ but good conformity at higher sigma ranges.

### 3.4.5 Normality Tests and Practical Validation

Comparing the empirical distributions with theoretical Gaussian properties:

Table 3.2: Comparison of Empirical and Theoretical Gaussian Properties

| **Sigma Range** | **Expected %** | **EuRoC MAV (28 GHz)** | | **Aerolab (78 GHz)** | |
|---|---|---|---|---|---|
| | | **Observed %** | **Difference** | **Observed %** | **Difference** |
| $\pm 1\sigma$ | 68.27 | 57.36 | -10.91 | 59.94 | -8.33 |
| $\pm 2\sigma$ | 95.45 | 99.86 | +4.41 | 94.01 | -1.44 |
| $\pm 3\sigma$ | 99.73 | 99.93 | +0.20 | 100.00 | +0.27 |

These results indicate that while there are some deviations at the $1\sigma$ level, both distributions show good conformity at the critical $2\sigma$ and excellent conformity at $3\sigma$ thresholds that are most relevant for robust estimation. The high coverage at these higher sigma levels indicates that the tails of both distributions are well-bounded, which is a favorable characteristic for sensor fusion applications where outlier rejection is important.

The analysis of TOA error distributions across two different datasets (EuRoC MAV at 28

GHz and Aerolab at 78 GHz) leads to several key findings:

- Both datasets exhibit multi-modal distributions best characterized by 4-component Gaussian mixture models, revealing the complex nature of TOA error behavior.
- Despite the multi-modal structure, both distributions show remarkably Gaussian-like properties in terms of symmetry (near-zero skewness) and tail behavior (excellent conformity at $3\sigma$ thresholds).
- While formal statistical tests reject perfect Gaussianity (common with large datasets), the practical validation through sigma coverage analysis reveals that both distributions behave sufficiently Gaussian-like at the critical thresholds relevant for robust estimation.

The findings support the practical validity of using Gaussian error models for TOA measurements in sensor fusion frameworks, particularly when robust estimators capable of handling moderate deviations from perfect Gaussianity are employed.

For most practical sensor fusion applications, a Gaussian approximation with the following parameters is justified:

- **EuRoC MAV (28 GHz)**: $\mu \approx -0.007141$, $\sigma \approx 0.332053$
- **Aerolab (78 GHz)**: $\mu \approx -0.010457$, $\sigma \approx 0.151413$

### 3.4.6 Additional theoretical justification for using Gaussian models

While the analysis demonstrates that 5G TOA errors cannot be perfectly modeled by a single Gaussian distribution, the following section examines why Gaussian models remain appropriate and effective for practical multi-sensor fusion applications. The following section provides theoretical justification and empirical evidence for this modeling choice, considering the effects of sensor fusion, computational requirements, and robustness mechanisms.

**Central Limit Theorem in Multi-Sensor Fusion:** The Central Limit Theorem (CLT) suggests that the *aggregated residuals* of a multi-sensor system will approximate a Gaussian distribution, even if individual sensor errors are non-Gaussian. Our framework fuses TOA

measurements with inertial (IMU) and visual (ORB-SLAM3) data, where the combined residual vector $\mathbf{r}_k$ at time $k$ is:

$$\mathbf{r}_k = \mathbf{r}_{\text{TOA},k} + \mathbf{r}_{\text{IMU},k} + \mathbf{r}_{\text{Vision},k} \tag{3.1}$$

Assuming independence between sensors, the summation of these residuals converges toward Gaussianity as the number of measurements increases.

**Algorithmic Tractability and Consistency:** Gaussian noise models enable computationally efficient solutions critical for real-time operation:

- **Kalman Filter Compatibility**: Our IMU pre-integration relies on the Kalman filter, which assumes Gaussian noise for optimality. Deviating from this would require computationally expensive alternatives (e.g., particle filters).
- **ORB-SLAM3 Consistency**: ORB-SLAM3 assumes Gaussian reprojection errors in its bundle adjustment. Adopting a non-Gaussian TOA model would create inconsistency in the joint optimization.
- **Least-Squares Simplification**: The Gaussian assumption allows formulating the maximum likelihood estimator (MLE) as a least-squares problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_i \|\mathbf{z}_i - h_i(\mathbf{x})\|^2_{\Sigma_i^{-1}} \tag{3.2}$$

  where $\Sigma_i$ is the covariance matrix. Non-Gaussian models would require iterative reweighting or sampling, increasing complexity by $\mathcal{O}(n^2)$.

**Robustness to Non-Gaussianity Through Fusion:** The fusion framework mitigates non-Gaussian TOA errors through:

- **Huber Loss for Robust Optimization**: The pipeline uses Huber loss in nonlinear optimization, which applies quadratic (L2) penalties to small residuals (Gaussian-like

regime) and linear (L1) penalties to large residuals. This reduces sensitivity to outliers without assuming heavy-tailed noise [126].

- **Fusion-Driven Redundancy**: Cross-validation between sensors (e.g., visual, inertial, TOA) dilutes the impact of non-Gaussian errors from any single source. The error-state Kalman filter naturally suppresses outliers by weighting measurements proportionally to their covariance.

### 3.4.7 Discussion

The comprehensive analysis of TOA errors across both 28 GHz (EuRoC MAV dataset) and 78 GHz (Aerolab dataset) frequencies reveals a consistent pattern that supports the Gaussian approximation for sensor fusion applications, despite formal statistical tests rejecting perfect Gaussianity. At 78 GHz frequency, the error distribution exhibits improved precision ($\sigma \approx 0.151413$) compared to 28 GHz ($\sigma \approx 0.332053$), demonstrating the potential benefit of higher frequency systems for localization tasks.

Both datasets show remarkable Gaussian-like properties in three critical aspects: (1) near-zero mean errors ($\mu \approx -0.010457$ at 78 GHz), indicating minimal systematic bias; (2) good symmetry (skewness of 0.042 at 78 GHz); and (3) excellent conformity at higher sigma thresholds—particularly at $\pm 3\sigma$ where coverage reaches 100% (vs. expected 99.73%). The multi-modal nature revealed by GMM analysis (optimally fitted with 4 components) explains the lower-than-expected coverage at $\pm 1\sigma$ (59.94% vs. expected 68.27%), but this deviation is effectively mitigated in practical applications through robust estimation techniques.

The justification for Gaussian approximation extends beyond empirical validation to include theoretical and practical considerations: (1) the Central Limit Theorem's effect when aggregating measurements across multiple anchors; (2) the inherent robustness of the fusion framework which employs outlier rejection mechanisms; (3) the computational tractability requirements for real-time SLAM applications; and (4) the consistency with established practices in both visual-inertial SLAM systems and 5G localization literature.

# Chapter 4

# Graph-Based and ESKF Fusion of 5G and Inertial Data for UAV Indoor Pose Estimation

This chapter investigates the integration of 5G TOA measurements with IMU data to achieve robust UAV localization in GPS-denied indoor environments. While IMUs provide high-frequency motion estimates, their inherent drift poses significant challenges for extended operations. Similarly, while 5G TOA measurements offer drift-free range information, their lower update rates and potential measurement sparsity limit standalone usage. This research addresses these limitations by developing and analyzing two complementary fusion approaches that leverage the strengths of both sensing modalities:

- An **Error State Kalman Filter (ESKF)** framework that efficiently fuses high-frequency IMU predictions with intermittent 5G TOA updates, providing real-time pose estimates while maintaining computational efficiency. The filter's error-state formulation enables robust handling of large uncertainties and nonlinear error accumulation inherent in IMU integration.
- A **Pose Graph Optimization (PGO)** approach that maintains a factor graph structure incorporating IMU preintegration factors and 5G TOA range constraints. This global

optimization framework achieves superior accuracy by jointly optimizing the full trajectory, demonstrating remarkable precision of up to 13 cm with five base stations while remaining computationally tractable for real-time operation.

Both methodologies are extensively evaluated using the EuRoC MAV dataset, augmented with carefully simulated 5G TOA measurements that incorporate realistic channel models from the QuaDRiGa simulator. The comprehensive evaluation examines critical aspects including:

- **Base Station Configuration:** Localization accuracy across varying base station configurations (2-5 stations), providing insights into minimum infrastructure requirements and performance scaling
- **Frequency Performance:** Detailed assessment under different frequency bands (5 GHz, 28 GHz, and 78 GHz) and bandwidths (100, 200, and 400 MHz), evaluating the impact of network configurations on system accuracy
- **Real-time Capability:** Comprehensive analysis of computational efficiency, demonstrating feasibility for real-time operations on standard hardware platforms
- **Reliability Testing in Diverse Indoor Scenarios:** Systematic testing under diverse indoor scenarios with limited infrastructure, validating system reliability in practical deployment conditions

Through rigorous experimental validation, this chapter provides fundamental insights into the capabilities and limitations of minimal sensor fusion approaches for UAV localization. The results demonstrate that both methods achieve consistent sub-meter accuracy even with just three base stations, establishing a practical foundation for infrastructure-light indoor navigation solutions.

## 4.1 Methodology for UAV Localization

### 4.1.1 Problem Definition

In this section, the UAV localization problem in 5G networks is formulated by combining IMU data, i.e., angular velocity and linear acceleration, with TOA measurements, which provide the radial distance from multiple BSs. These measurements can then be fused to track the UAV motion using the estimation approaches described in the following sections.

It is worth noting that while recent breakthroughs in machine learning, like Variational Autoencoders (VAEs) and other learning-based approaches, hold promise for localization tasks, this study concentrates on model-based estimation techniques, specifically the ESKF and PGO, for a few key reasons:

- **Data Availability, Offline Learning**
  - Learning-based methods often require large datasets for effective learning, which can be challenging to obtain, especially for unique environments or specific 5G deployments. Although transfer learning could mitigate this issue, it falls outside the scope of our current study due to the additional data needed for fine-tuning and the potential computational burden on resource-constrained drones. As such, alternative approaches better suited to our real-time and resource-constrained scenario were prioritized.
- **Interpretability and Safety:**
  - In safety-critical applications like drone navigation, understanding the decision-making process behind localization is paramount. Model-based methods such as ESKF and PGO provide greater transparency in their calculations compared to learning-based methods, which can be black boxes and difficult to interpret. This transparency aligns better with the safety requirements of our application.

**Sensing and State Representation for UAV Localization:** The goal is to determine the

3D location and orientation of the UAV's body center, which is aligned with the IMU frame B, in the world-fixed frame W. The set of state variables $\mathcal{X}$ contains the pose and velocity of the UAV. Moreover, one needs to estimate the time-variant biases of the IMU's gyroscope $\mathbf{b}^g \in \mathbb{R}^3$ and accelerometer $\mathbf{b}^a \in \mathbb{R}^3$ to account for the IMU noise drift.

The state vector $\mathcal{X}$ can then be represented as:

$$\mathcal{X} = \begin{bmatrix} \mathbf{q}^T & (\mathbf{b}^g)^T & \mathbf{v}^T & (\mathbf{b}^a)^T & \mathbf{p}^T \end{bmatrix}^T. \tag{4.1}$$

Here, each transformation $\mathtt{T} = (\mathtt{R}, \mathbf{p}) \in \mathrm{SE}(3)$ is composed of a rotation $\mathtt{R} \in \mathrm{SO}(3)$ and a translation $\mathbf{p} \in \mathbb{R}^3$. This transformation matrix $\mathtt{T}$ effectively transforms the body frame B to the world frame W, in which the base stations (BSs) are positioned. Additionally, the quaternion form of rotation, denoted as $\mathbf{q}$, is related to the rotation matrix $\mathtt{R}$ in the transformation $\mathtt{T}$.

**IMU Measurements:** IMU measurements are provided w.r.t. the B frame. Our approach involves a 6-axis IMU that measures the linear acceleration ${}_{\mathtt{B}}\bar{\mathbf{a}}$ and angular velocity ${}_{\mathtt{B}}\bar{\boldsymbol{\omega}}$. The IMU real motion state $\{{}_{\mathtt{B}}\mathbf{a}, {}_{\mathtt{B}}\boldsymbol{\omega}\}$ is altered by additive Gaussian white noise $\{\boldsymbol{\eta}^a, \boldsymbol{\eta}^g\}$ and slowly time-varying biases $\{\mathbf{b}^a, \mathbf{b}^g\}$ affecting respectively the accelerometer and gyroscope as defined by the following IMU measurement model [127, 128]:

$$\begin{aligned} {}_{\mathtt{B}}\bar{\boldsymbol{\omega}} &= {}_{\mathtt{B}}\boldsymbol{\omega} + \mathtt{R}^T\boldsymbol{\omega}^G + \mathbf{b}^g + \boldsymbol{\eta}^g, \\ {}_{\mathtt{B}}\bar{\mathbf{a}} &= \mathtt{R}^{\mathsf{T}}\left({}_{\mathtt{W}}\mathbf{a} - \mathbf{g} + 2\boldsymbol{\omega}^G_{\times}\mathbf{v} + [\boldsymbol{\omega}^G_{\times}]^2\mathbf{p}\right) \\ &\quad + \mathbf{b}^a + \boldsymbol{\eta}^a \end{aligned} \tag{4.2}$$

where ${}_{\mathtt{B}}\bar{\boldsymbol{\omega}}$ and ${}_{\mathtt{B}}\bar{\mathbf{a}}$ denote the gyro and accelerometer measurements in the body frame, and ${}_{\mathtt{W}}\mathbf{a}$ is the acceleration of the body frame expressed in the world frame. The terms $\boldsymbol{\eta}^a$ and $\boldsymbol{\eta}^g$ represent the stochastic noise affecting accelerometer and gyroscopic measurement readings, respectively.

The term $\boldsymbol{\omega}^G$ incorporates the effects of the earth's rotation, and the accelerometer measurements include the gravitational acceleration $\mathbf{g}$. The operator $\boldsymbol{\omega}_{\times}$ denotes a skew-symmetric matrix, and for $\boldsymbol{\omega} = (\boldsymbol{\omega}_x, \boldsymbol{\omega}_y, \boldsymbol{\omega}_z)$, it is defined by:

$$\boldsymbol{\omega}_{\times} = \begin{bmatrix} 0 & -\boldsymbol{\omega}_z & \boldsymbol{\omega}_y \\ \boldsymbol{\omega}_z & 0 & -\boldsymbol{\omega}_x \\ -\boldsymbol{\omega}_y & \boldsymbol{\omega}_x & 0 \end{bmatrix}.$$

**Time of Arrival (TOA) Measurements :** In this context, TOA measurements denoted as $\bar{\rho} = (\bar{\rho}_1, \ldots, \bar{\rho}_K)$ are leveraged, where $K$ is the total number of 5G BSs. By multiplying the TOA values with the speed of light $c$, one can deduce the distances to each BS, respectively. Let represent these distances using the vector:

$$\bar{\mathbf{d}} = (\bar{\mathbf{d}}_1, \ldots, \bar{\mathbf{d}}_K)^T.$$

This distance is determined by both the dynamic position of the drone and the fixed positions of the 5G BSs (landmarks), defined by the equation:

$$\bar{\mathbf{d}}_k = \mathbf{d}_k + \boldsymbol{\eta}_k^{\text{Dist}} = \|\mathbf{p} - \mathbf{L}_k\|_2 + \boldsymbol{\eta}_k^{\text{Dist}}, \quad k \in \{1, ..., K\}. \tag{4.3}$$

Here $\bar{\mathbf{d}}_k$ denotes the distance measurement between the drone and the $k$-th BS and $\mathbf{L}_k \in \mathbb{R}^3$ denotes the 3D position of the $k$-th 5G BS.

Please note that unless explicitly stated, the explicit time index is omitted for brevity and all variables will be interpreted as values at time $t$.

**Dynamic UAV model:** Assuming the UAV has a rigid frame, the dynamics of a rigid body are applied to characterize its motion. The UAV's attached IMU allows us to use IMU measurements as inputs to the dynamic model. This results in the following dynamic model governing IMU states:

$$
\begin{aligned}
\dot{\mathbf{q}} &= \frac{1}{2}\Omega\left(\boldsymbol{\omega}(t)\right)\mathbf{q}(t), \\
\dot{\mathbf{b}}^g &= \boldsymbol{\eta}^{wg}(t), \\
\dot{\mathbf{v}} &= {}_{\mathrm{W}}\mathbf{a}, \\
\dot{\mathbf{b}}^a &= \boldsymbol{\eta}^{wa}(t), \\
\dot{\mathbf{p}} &= \mathbf{v},
\end{aligned}
\tag{4.4}
$$

where $\boldsymbol{\eta}^{wa}$ and $\boldsymbol{\eta}^{wg}$ are random walk noise of IMU accelerometer and gyroscope biases, and

$$
\Omega(\boldsymbol{\omega}) = \begin{bmatrix} -\boldsymbol{\omega}_{\times} & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}. \tag{4.5}
$$

Taking into account the IMU measurement model described in Eq. (4.2) and applying the expectation operator $\hat{}$, one obtains the following set of dynamic equations:

$$
\begin{aligned}
\dot{\hat{\mathbf{q}}} &= \frac{1}{2}\Omega\left(\hat{\boldsymbol{\omega}}(t)\right)\hat{\mathbf{q}}(t), \\
\dot{\hat{\mathbf{b}}}^g &= \mathbf{0}, \\
\dot{\hat{\mathbf{v}}} &= \hat{\mathrm{R}}^T\hat{\mathbf{a}} + \mathbf{g} - 2[\boldsymbol{\omega}^G{}_{\times}]\hat{\mathbf{v}} - [\boldsymbol{\omega}^G_{\times}]^2\hat{\mathbf{p}}, \\
\dot{\hat{\mathbf{b}}}^a &= \mathbf{0}, \\
\dot{\hat{\mathbf{p}}} &= \hat{\mathbf{v}},
\end{aligned}
\tag{4.6}
$$

with $\hat{\mathbf{a}} = {}_{\mathrm{B}}\bar{\mathbf{a}} - \hat{\mathbf{b}}^a$ and $\hat{\boldsymbol{\omega}} = {}_{\mathrm{B}}\bar{\boldsymbol{\omega}} - \hat{\mathbf{b}}_g - \hat{\mathrm{R}}^T\boldsymbol{\omega}^G$. Note that W is considered as an inertial frame of reference, neglecting the effect of the Earth's rotation, i.e., $\boldsymbol{\omega}^G = \mathbf{0}$.

### 4.1.2 Error State Kalman Filter for UAV Localization (Indirect Method)

The Extended Kalman filter (EKF) is a widely employed technique for estimating and tracking system states, particularly for non-linear system dynamics or measurement models. The EKF operates by linearizing the system dynamics and the observation (measurement) model

around the current state estimate. Although the EKF performs accurately enough in many situations, its efficacy may diminish in highly nonlinear systems.

One effective approach is to utilize the error-state Kalman filter (ESKF) method for enhanced precision and confidence in estimations, as elucidated by [129]. ESKF entails approximating the error state, representing the disparity between the true and estimated states, rather than directly estimating the states themselves. The rationale behind adopting an ESKF, also known as the indirect method, is that errors are typically smaller and exhibit more linearity than the states, making them well-suited for linear Gaussian filtering. This enables the ESKF to refine estimations with greater precision. This process unfolds in two main steps, namely prediction and update.

During the prediction stage, the system state is forecasted using integration, disregarding minor disturbances and noise. However, this straightforward projection can result in the gradual accumulation of errors due to noise and disturbances. The ESKF simultaneously calculates a Gaussian estimate of the error state's distribution while integrating the nominal state to address these inaccuracies. This dual process refines our real-time understanding of the evolving errors. By identifying small, subtle signals within the system's behavior, the ESKF enhances state estimation accuracy.

The update step is triggered by the reception of the new measurements, a less frequent occurrence than prediction. These measurements serve to expose accumulated errors. Exploiting this new information, the ESKF refines the error state estimate, enhancing the comprehension of its distribution and contributing to higher accuracy. The error state is then augmented with the estimated state, enabling necessary adjustments. While the error state is reset to zero, its uncertainty is updated.

Table 4.1 presents the key differences between the EKF and ESKF, highlighting the trade-off between accuracy and computational complexity. In summary, while EKF directly estimates states and linearizes around the current estimate, leading to lower computational complexity, its accuracy may be compromised in highly non-linear systems. The ESKF, on the other hand, tackles this challenge by estimating the error between true and estimated states and linearizing around this error state. This approach achieves higher accuracy in non-linear

Table 4.1: Comparison of Extended Kalman Filter (EKF) and Error State Kalman Filter (ESKF)

| Feature | EKF | ESKF |
|---|---|---|
| State Representation | Actual system states | Error between true and estimated states |
| Linearization Point | Current state estimate | Error state |
| Suitability for Non-linearity | Less accurate | More accurate |
| Computational Complexity | Lower | Slightly higher |
| Prediction Stage | Predicts system state based on linearized dynamics | Predicts error state evolution & nominal state |
| Update Stage | Updates state estimate directly | Updates error state estimate & refines state through error augmentation |

scenarios, although at a slightly increased computational cost due to additional calculations and memory requirements. In both filters, the prediction stage utilizes linearized dynamics to forecast the state (ESKF predicts both error and nominal state), while the update stage refines the estimate using new measurements (ESKF updates error before refining the full state).

**Prediction Step:** Having the states $\mathcal{X}$ defined in Eq. (4.1), the error state is defined as per [128]:

$$\delta\mathcal{X} = \begin{bmatrix} \delta\theta^T & (\delta\mathbf{b}^g)^T & \delta\mathbf{v}^T & (\delta\mathbf{b}^a)^T & \delta\mathbf{p}^T \end{bmatrix}^T. \tag{4.7}$$

Herein, $\delta x = x - \hat{x}$, where $x$ represents the true value, and $\hat{x}$ is the estimated value. However, the quaternion error is defined as $\mathbf{q} = \delta\mathbf{q} \otimes \hat{\mathbf{q}}$, with $\mathbf{q}$ being a real value and $\hat{\mathbf{q}}$ as the estimated one. To effectively represent slight rotational errors, the following approximation is adopted:

$$\delta\mathbf{q} \simeq \begin{pmatrix} \frac{1}{2}\delta\theta^T & 1 \end{pmatrix}^T. \tag{4.8}$$

This approximation serves to simplify the representation of minor rotational errors within quaternion-based state estimation.

Note that the IMU bias states are explicitly estimated within the overall state vector ($\delta\mathcal{X}$). This allows the filter to learn and compensate for these constant offsets over time. In the error state equation, bias state dynamics would be modeled as: $\delta\dot{\mathbf{b}}^g \approx 0$ and $\delta\dot{\mathbf{b}}^a \approx 0$ (assuming random walk noise is negligible).

Employing the minimal representation provided by $\delta\theta$, it reduces computational complexity while simultaneously maintaining accuracy.

Based on Eq. (4.6) and Eq. (4.7), the error state equation can be expressed in the following form:

$$\delta\dot{\mathcal{X}} = \mathbf{F}\delta\mathcal{X} + \mathbf{G}\boldsymbol{\eta}^{\text{IMU}}, \tag{4.9}$$

where $\boldsymbol{\eta}^{\text{IMU}} = \begin{bmatrix} (\boldsymbol{\eta}^g)^T & (\boldsymbol{\eta}^{wg})^T & (\boldsymbol{\eta}^a)^T & (\boldsymbol{\eta}^{wa})^T \end{bmatrix}^T$, $\mathbf{F}$, and $\mathbf{G}$ are given as follows:

$$\mathbf{F} = \begin{bmatrix} -[\hat{\boldsymbol{\omega}}_\times] & -I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ -\hat{\mathrm{R}}^T[\hat{\mathbf{a}}_\times] & 0_{3\times3} & -2[\boldsymbol{\omega}^G_\times] & -\hat{\mathrm{R}}^T & -[\boldsymbol{\omega}^G_\times]^2 \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix},$$

and

$$\mathbf{G} = \begin{bmatrix} -I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -\hat{\mathrm{R}}^T & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_3 \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix},$$

where $I_3 \in \mathbb{R}^{3\times3}$ denotes the Identity matrix.

In the discrete implementation, the IMU captures measurements at discrete time intervals, namely $\bar{\boldsymbol{\omega}}_t$ and $\bar{\mathbf{a}}_t$, which are then employed to propagate the state and covariance matrix. The ESKF is initialized with the initial state estimate and covariance matrix. Upon the arrival of a fresh IMU measurement, the IMU-derived state estimate undergoes propagation through a numerical integration, effectively accounting for the dynamic model of the system as specified by Eq. (4.6). The covariance matrix is also propagated using a numerical integration of the Lyapunov equation:

$$\dot{\mathbf{P}} = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T + \mathbf{G}\mathbf{Q}^{\mathrm{IMU}}\mathbf{G}^T. \tag{4.10}$$

Note that the IMU noise characteristics are represented by a noise covariance matrix ($\mathbf{Q}^{IMU}$) tailored to the specific sensor used. This matrix captures the variances and correlations of different noise sources (gyroscope noise, accelerometer noise, etc.). Accurate estimation of $\mathbf{Q}^{IMU}$, whether through empirical measurements or manufacturer specifications, is crucial to precise noise modeling. The $\mathbf{G}$ matrix in the error state equation (4.9) acts as a gain matrix, mapping IMU noise directly into the error state dynamics. Each non-zero entry in $\mathbf{G}$ indicates how a specific noise component influences a particular state error. For example, gyroscope noise primarily affects the attitude error, while accelerometer noise impacts both attitude and velocity errors. Furthermore, in the propagation of the covariance matrix, the process noise term ($\mathbf{G}\mathbf{Q}^{IMU}\mathbf{G}^T$) in the Lyapunov equation (4.10) drives the growth of the state covariance matrix ($\mathbf{P}$). As the covariance matrix increases, it reflects the increasing

uncertainty in the state estimates due to IMU noise accumulation

**Update Step:** Upon receiving Time of Arrival (TOA) measurements, the update stage is proceeded with. Given Eq. (4.3), the measurement function $h(\mathcal{X})$ is defined as follows:

$$h(\mathcal{X}) = \begin{bmatrix} \|\mathbf{p} - \mathbf{L}_1\|_2 \\ \vdots \\ \|\mathbf{p} - \mathbf{L}_K\|_2 \end{bmatrix} \in \mathbb{R}^{K\times 1}.$$

Now the residual $\mathbf{r}^\rho \in \mathbb{R}^{K\times 1}$ is formulated as:

$$\begin{aligned} \mathbf{r}^\rho &= \bar{\mathbf{d}} - h(\hat{\mathcal{X}}) = h(\mathcal{X}) - h(\hat{\mathcal{X}}) + \boldsymbol{\eta}^{\text{Dist}} \\ &= h(\hat{\mathcal{X}} + \delta\mathcal{X}) - h(\hat{\mathcal{X}}) + \boldsymbol{\eta}^{\text{Dist}} \\ &\approx \mathbf{H}\delta\mathcal{X} + \boldsymbol{\eta}^{\text{Dist}}, \end{aligned} \tag{4.11}$$

where $\boldsymbol{\eta}^{\text{Dist}}$ represents the vector of noise associated with the distance measurements and the measurement Jacobian matrix $\mathbf{H}$ is defined by:

$$\begin{aligned} \mathbf{H} =& \frac{\partial \mathbf{r}^\rho}{\partial(\delta\mathcal{X})}|_{\hat{\mathcal{X}}} \\ =& \begin{bmatrix} 0_{1\times 12} & \frac{\hat{\mathbf{p}}_x - L_{1x}}{d_1} & \frac{\hat{\mathbf{p}}_y - L_{1y}}{d_1} & \frac{\hat{\mathbf{p}}_z - L_{1z}}{d_1} \\ \vdots & \vdots & \vdots & \vdots \\ 0_{1\times 12} & \frac{\hat{\mathbf{p}}_x - L_{Kx}}{d_K} & \frac{\hat{\mathbf{p}}_y - L_{Ky}}{d_K} & \frac{\hat{\mathbf{p}}_z - L_{Kz}}{d_K} \end{bmatrix}_{k\times 15} \end{aligned} \tag{4.12}$$

Here $L_k = (L_{kx}, L_{ky}, L_{kz})$, for $k \in 1, ..., K$ represent the position of the k$th$ base station. The update rules at time $t$ will be:

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T \left(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{4.13}$$

$$\delta\hat{\mathcal{X}}_{t|t} = \mathbf{K}_t\mathbf{r}_t^\rho, \tag{4.14}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1}. \tag{4.15}$$

In these equations, $\mathbf{R}$ represents the covariance matrix of the distance measurement noise. The process involves calculating the Kalman Gain $\mathbf{K}_t$, updating the error state estimate $\delta\hat{\mathcal{X}}_{t|t}$ and the error covariance matrix. Finally, $\delta\hat{\mathcal{X}}_{t|t}$ is appropriately incorporated into the predicted state $\hat{\mathcal{X}}_{t|t-1}$ based on the specific calculations for each component of the state vector. This step refines the estimated state using TOA measurements and their associated uncertainties.

The ESKF algorithm's core steps are outlined in the pseudocode presented in algorithm 1, serving as an overview of the ESKF implementation, helping to grasp the algorithm's logical flow and key components.

**Algorithm 1:** ESKF-based Localization

**Input:** IMU and TOA measurements
**Result:** state estimate $\hat{\mathcal{X}}$
Initialize: $\hat{\mathcal{X}}_0$, $\mathbf{P}$ ;
**while** *new IMU measurement received* **do**
  **Prediction Stage:**
  Propagate state using Runge-Kutta integration;
  Propagate covariance matrix using Runge-Kutta integration;
  **if** *new TOA measurements received* **then**
    **Update Stage:**
    Obtain distance measurements $\bar{\mathbf{d}}_t$;
    Calculate residual $\mathbf{r}_t^{\rho} \leftarrow \bar{\mathbf{d}}_t - h(\hat{\mathcal{X}}_{t|t-1})$;
    Calculate Jacobian matrix $\mathbf{H}_t$ for measurement model;
    Calculate Kalman Gain $\mathbf{K}_t$;
    Update error state estimate $\delta\hat{\mathcal{X}}_{t|t} \leftarrow \mathbf{K}_t\mathbf{r}_t^{\rho}$;
    Update covariance matrix $\mathbf{P}_{t|t} \leftarrow (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}$;
    Update state estimate $\hat{\mathcal{X}}_{t|t} \leftarrow \hat{\mathcal{X}}_{t|t-1} \oplus \delta\hat{\mathcal{X}}_{t|t}$;
  **end**
**end**

### 4.1.3 Pose Graph Optimization (PGO)

As a second alternative method to estimate the drone's 6DoF pose, graph-based optimization techniques are also leveraged. This method models the relationships between the pose variables based on sensor measurements and then performs the estimation using least

squares optimization. For that purpose, a factor graph model [130] is created to abstract the problem. Factor graphs are bipartite graphs with two types of nodes, namely variables and factors, where factors represent functions on subsets of the variables. Edges in the factor graph between a factor and a set of variables indicate that the factor depends only on those variables. In our case, the variables are the state variables that should be estimated at certain instants and the factors correspond to the likelihood of the adjacent state variable nodes given the related measurements, see the structure of the resulting factor graph in Fig. 4.1. The figure illustrates a factor graph used for optimizing variables, with circles representing variables and squares representing factors. Nodes $T_t$ represent pose variables, $v_t$ represent velocity, and $b_t$ represent bias. IMU pre-integration factors connect these nodes, while TOA measurements create range factors. Prior factors constrain initial values.

Therefore, the factor graph can be used to specify the joint posterior probability density $p(\mathcal{X}|\mathcal{Z})$ of the whole set of state variables $\mathcal{X}$ and the whole set of measurements $\mathcal{Z}$ as the product of all factors $f_i(\mathcal{X}_i)$ in the graph

$$p(\mathcal{X}|\mathcal{Z}) \propto \prod_{\forall f_i \in \mathcal{F}} f_i(\mathcal{X}_i)\,, \tag{4.16}$$

where $\mathcal{X}_i$ is the set of state variable nodes adjacent to the factor $f_i(\mathcal{X}_i)$ and $\mathcal{F}$ is the set of all factors in the graph. Herein, the measurements are no longer explicitly represented but simply become a fixed parameter in the corresponding factor. This factorization will be exploited for pose estimation as described in the following.

**IMU Factor:** Due to the IMU's higher sampling frequency than other sensors, it typically captures multiple measurements between two TOA instances. The IMU factor is constructed utilizing a *preintegrated measurement* [131] constraining the relative motion increments. Especially, the condensed measurements $\Delta\bar{\mathrm{R}}_{ij}$ of rotation, $\Delta\bar{\mathbf{p}}_{ij}$ of position, and $\Delta\bar{\mathbf{v}}_{ij}$ of velocity are obtained by integrating multiple IMU readings $\{{}_{\mathrm{B}}\bar{\mathbf{a}}_t, {}_{\mathrm{B}}\bar{\boldsymbol{\omega}}_t : \forall t \in [t_i, t_j]\}$. So, the

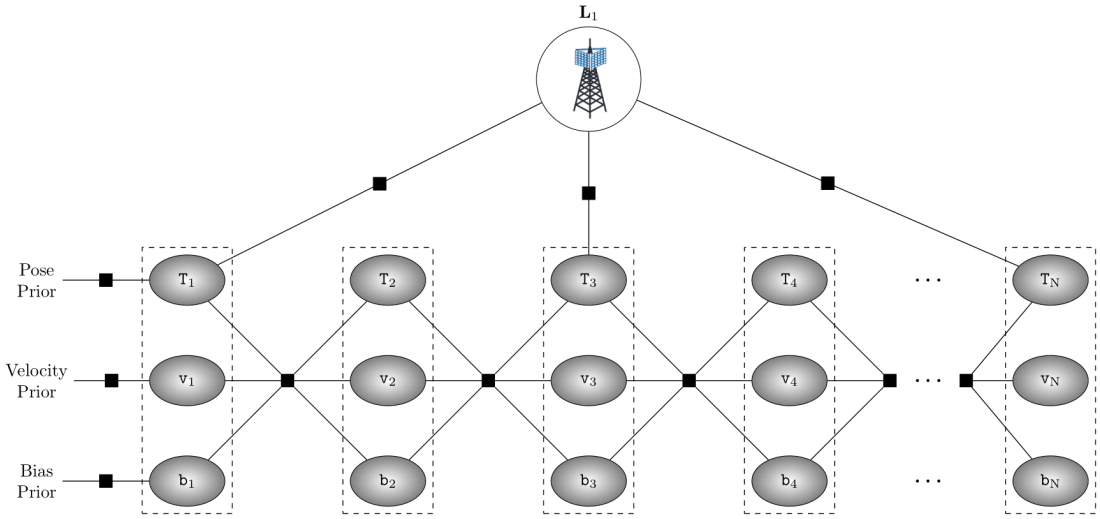



Figure 4.1: The figure visualizes the structure of the factor graph used to optimize the variables, represented by circles, by relating them through factors, represented by squares. The nodes $\mathtt{T}_t$ incorporate the 6DoF pose variables, $\mathtt{v}_t$ nodes encapsulate the velocity variables, and $\mathtt{b}_t$ nodes denote the bias variables, encompassing biases from both gyroscopes and accelerometers. IMU pre-integration factors connect all of these nodes. TOA measurements create range factors between robot pose nodes and BSs position nodes, with a single instance, $\mathbf{L}_1$, visualized here to enhance graph clarity. Prior factors, namely prior pose, prior velocity, and prior bias, are connected to the respective nodes $\mathtt{T}_1$, $\mathtt{v}_1$, and $\mathtt{b}_1$ to constrain them with their initial values in the trajectory.

residual terms $\mathbf{r}$ for the states $\{\mathtt{R}_{ij}, \mathbf{p}_{ij}, \mathbf{v}_{ij}\}$ can be defined as:

$$\mathbf{r}_{ij}^{\mathtt{R}} = \mathrm{Log}\left(\Delta\bar{\mathtt{R}}_{ij}^{\mathsf{T}}\, \mathtt{R}_i^{\mathsf{T}}\mathtt{R}_j\right) , \tag{4.17}$$

$$\mathbf{r}_{ij}^{\mathbf{p}} = \mathtt{R}_i^{\mathsf{T}}\left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{ij} - \tfrac{1}{2}\mathbf{g}\Delta t_{ij}^2\right) - \Delta\bar{\mathbf{p}}_{ij} , \tag{4.18}$$

$$\mathbf{r}_{ij}^{\mathbf{v}} = \mathtt{R}_i^{\mathsf{T}}\left(\mathbf{v}_j - \mathbf{v}_i - \mathbf{g}\Delta t_{ij}\right) - \Delta\bar{\mathbf{v}}_{ij} , \tag{4.19}$$

where $\Delta t_{ij} = t_j - t_i$ is the total time interval with $t_i < t_j$. Also, $\mathrm{Log} : \mathrm{SO}(3) \to \mathbb{R}^3$ defines the logarithm map that associates elements of the rotation manifold $\mathrm{SO}(3)$ to vectors on the Euclidean tangent space $\mathbb{R}^3$ representing rotation increments. Regarding the biases, the total residual $\mathbf{r}_{ij}^{\mathbf{b}}$ between time $t_i$ and $t_j$ is set as follows:

$$\mathbf{r}_{ij}^{\mathbf{b}} = \ \mathbf{b}_j^g - \mathbf{b}_i^g + \mathbf{b}_j^a - \mathbf{b}_i^a . \tag{4.20}$$

**TOA Range Factor:** By multiplying the estimated TOA values $\rho_{sk}$ by the speed of light $c$, *i.e.,* $d_{sk} = \rho_{sk} \cdot c$, one obtains $\mathrm{K}$ metric distance measurements $d_{sk} \in \mathbb{R}$ of the drone to the $k$-th BS at location $\mathbf{L}_k$ at time $t_s$. Notably, the possibility of having fewer TOA measurements than the number of tracked poses is explicitly expressed. The residual $\mathbf{r}_{sk}^{\rho}$ of the TOA factor at time $t_s$ with the BS $\mathbf{L}_k$ is defined as:

$$\mathbf{r}_{sk}^{\rho} = d_{sk} - \|\mathbf{p}_s - \mathbf{L}_k\|_2 \,. \tag{4.21}$$

**Optimization:** The pose graph optimization problem is formulated as Maximum a Posteriori (MAP) estimation that involves finding the state $\mathcal{X}^*$ that maximizes the posterior:

$$\mathcal{X}^* = \arg\max_{\mathcal{X}} p(\mathcal{X}|\mathcal{Z}). \tag{4.22}$$

Considering the proportional relationship in Eq. (4.16), Eq. (4.22) is equivalent to the maximization of the product of all factors in the factor graph:

$$\mathcal{X}^* = \arg\max_{\mathcal{X}} \prod_{\forall f_i \in \mathcal{F}} f_i(\mathcal{X}_i) \,. \tag{4.23}$$

In our application, the factor graph evolves with time, where $t = \{1, \ldots, N\}$ represents instants and $\mathcal{X}_t$ is the state at instant $t$. Therefore, the overall set of states is $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_N\}$ starting with a given initial state $\mathcal{X}_0$. The factors are likelihoods derived from the respective previously described measurements, assumed to be corrupted by zero-mean, normally distributed noise. Now $f_{ti}$ denotes the factors between state $\mathcal{X}_t$ and the state $\mathcal{X}_{t+1}$ and, if TOA measurements are available, the factors between state $\mathcal{X}_t$ and the respective BSs. Taking the negative log of Eq. (4.23) finally leads to the minimization of the sum of the respective residuals in the following form:

$$\mathcal{X}^* = \arg\min_{\mathcal{X}} \|\mathbf{r}_0\|^2_{\mathbf{\Sigma}_0} + \sum_{t=1}^{\mathrm{N}-1} \sum_{\forall f_{ti} \in \mathcal{F}} \|\mathbf{r}_{f_{ti}}\|^2_{\mathbf{\Sigma}_{f_{ti}}} \,, \tag{4.24}$$

where $\|\mathbf{r}\|^2_{\mathbf{\Sigma}} = \mathbf{r}^{\mathsf{T}} \mathbf{\Sigma}^{-1} \mathbf{r}$ is the squared Mahalanobis norm, and $\mathbf{r}_{f_{ti}}$ are the residual functions

related to the aforementioned factors $f_{ti}$ with covariance matrix $\mathbf{\Sigma}_{f_{ti}}$. $\mathbf{r}_0$ denotes the residual derived from the prior on the initial pose, with $\mathbf{\Sigma}_0$ being its covariance matrix.

To efficiently solve the MAP optimization problem, the iSAM2 (Incremental Smoothing and Mapping 2) iterative optimization algorithm [132] implemented in GTSAM [133] is utilized. This algorithm can automatically identify the variables that require linearization at each step, and it enables us to keep our graph solution updated while adding new nodes without experiencing memory overload.

iSAM2 effectively leverages a Bayesian tree structure to incorporate historical data during optimization, but its performance can be impacted by the growing tree depth for extended temporal horizons. In scenarios where both long-term memory retention and real-time performance are critical, [134] offers a promising solution that is based on a spanning tree-based hierarchy. This method leverages a simplified spanning tree representation of the pose graph, reducing complexity. It also employs a coarse-to-fine optimization strategy, achieving faster convergence by first optimizing on supernodes formed by connected components and then refining within each supernode using the original measurements. While this approach introduces a slight accuracy approximation, its reduced complexity and potential for real-time performance make it a valuable option for specific applications.

Within algorithm 2, the essence of graph-based state estimation is encapsulated in a streamlined manner.

## 4.2 Evaluation and Results

The evaluation of the two derived approaches for pose estimation requires application to real experimental data wherever possible. For this purpose, the EuRoC MAV dataset [135] was selected, which comprises vision data and inertial sensor measurements as well as the related ground truth pose data collected during multiple UAV indoor flight sequences. The dataset is widely used in the literature as a benchmark for the evaluation of visual-inertial odometry or SLAM solutions. While it provides the necessary inertial measurements for evaluation, it does not yet include the required 5G TOA data. Therefore, the EuRoC MAV

**Algorithm 2:** Graph-Based Localization

**Input:** IMU and TOA measurements
**Result:** Optimized state estimates $\mathcal{X}^*$
Initialize: $\mathcal{X}_0^* = \mathcal{X}_0$;
**while** *true* **do**
  **for** *each fixed time interval* **do**
    Accumulate IMU measurements;
    preintegrate accumulated IMU measurements;
    Add a new node;
    Create an IMU factor linking the new node to the previous one;
    **if** *New TOA measurement arrives* **then**
      Determine the temporally nearest node to the TOA measurement.;
      Incorporate a range factor between the node and the relevant landmark;
    **end**
    Perform optimization using iSAM2;
    Update the state estimates $\mathcal{X}^*$;
  **end**
**end**

dataset needed augmentation with suitable data that could be generated either experimentally or via simulations. However, experimental generation of these 5G TOA data would require not only an exact repetition of the flights with identical equipment and configurations but also the setup of a configurable 5G indoor network and an on-board 5G receiver with TOA measurements. While such 5G networks (e.g., 5G femtocells) are foreseen for the near future by many mobile communications equipment suppliers, acquiring and installing them in an academic lab environment currently remains a major challenge.

Therefore, the missing 5G TOA data was generated via a suitable simulation environment. The following sections describe in detail the approach to generate such simulated yet highly realistic data and present the assessment based on the augmented EuRoC MAV dataset.

### 4.2.1 Augmenting the EuRoC MAV dataset with simulated 5G TOA data

Fig. 4.2 shows the overall structure of the simulation environment with its core components and their interaction for the augmentation of the EuRoC MAV dataset. In addition, it shows the connection of the resulting dataset with the pose estimators for the assessment. Herein,

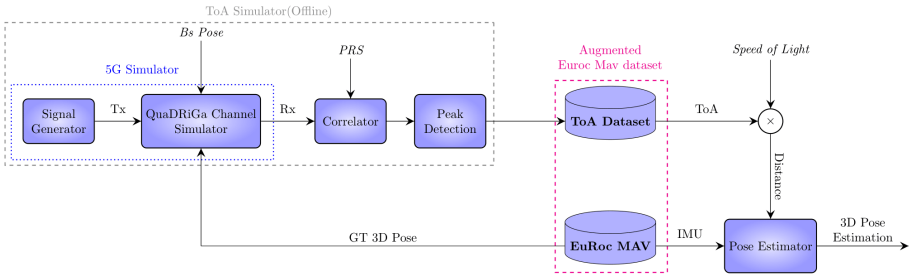



Figure 4.2: System Architecture for Augmentation of EuRoC MAV Dataset with 5G Simulation.

the main components are:

- **EuRoC MAV dataset:** provides inertial measurements and ground truth pose data from real MAV flights.
- **Signal generation:** Generates the 5G signals as emitted from a number of simulated BSs placed at defined locations.
- **QuaDRiGa channel simulation:** Simulates the propagation of the 5G signals from the BSs with their respective fixed poses through the environment to a supposed flying UAV with its current pose provided by the EuRoC MAV ground truth data.
- **TOA Calculation:** Calculates the current TOA measurements as provided by a supposed UAV on-board 5G receiver by analyzing the received 5G signal, and stores them as a dataset.
- **Pose estimator:** Estimates the pose of the UAV using the stored simulated TOA measurements and the inertial measurement from the EuRoC MAV dataset using one of the two pose estimators (i.e., ESKF or PGO) as derived in Sections 4.1.3 and 4.1.2, respectively.

The EuRoC MAV dataset [135] was collected by an indoor flying MAV equipped with a stereo-camera module capturing images at 20 Hz and a calibrated IMU providing inertial measurements at a rate of 200 Hz. The dataset also contains the UAV's position and orientation data obtained through the Vicon motion capture system as ground truth, recording

the full 6DoF at approximately 100 Hz. The full set of calibrated rigid transformations between sensors and the Vicon system is also given. The EuRoC MAV dataset comprises multiple flight sequences while this study examines all six sequences from the Vicon Room 1 and Vicon Room 2 datasets.

The distance estimation to 5G BSs requires generation of 5G signals, including PRS and Physical Downlink Shared Channel (PDSCH) resources transmitted by each BS. The MATLAB 5G Toolbox generates resource blocks for these signals, with transmit power set to 0 dBm (1 mW) and SNR set to 10 dB, which represents a conservative value, as real-world scenarios often have higher SNR levels.

The QuaDRiGa (quasi-deterministic radio channel generator) channel simulator [122] creates an impulse response that emulates the wireless channel characteristics based on specific network configurations, given receiver and transmitter positions, orientations, velocities, and the chosen indoor environment. QuaDRiGa is a realistic channel simulator that has been validated through extensive measurements. The UAV trajectories utilize the available 6DoF ground truth poses of the UAV provided by the EuRoC MAV dataset, from which the required velocity is computed considering the translation vectors between two time-consecutive poses. Two up to five fictitious BSs are virtually placed in the room where the trajectory is recorded. The positions of the BSs in the EuRoC MAV Vicon system's coordinate frame are $\mathrm{BS}_1 = (-10, -7, 2)$, $\mathrm{BS}_2 = (7, 13, 3)$, $\mathrm{BS}_3 = (25, -35, 4)$, $\mathrm{BS}_4 = (-6, 9, 5)$, $\mathrm{BS}_5 = (-4, -14, 6)$. These values are used to initialize the corresponding state variables of the optimization problem with a small covariance.

The simulation incorporates three QuaDRiGa channel simulator scenarios, each operating at different frequencies, to simulate the wireless communication environment realistically: QuaDRiGa-Industrial-LOS for 5 GHz, 3GPP-38.901-Indoor-LOS for 28 GHz, and mmMAGIC-Indoor-LOS for 78 GHz.

- **QuaDRiGa_Industrial_LOS [125]**: This scenario is designed to replicate a line-of-sight (LOS) environment for industrial applications. The simulation is optimized for frequencies ranging from 2 to 6 GHz and aims to capture the radio propagation behavior typically observed in automation industry halls. The scenario is validated through mea-

surements conducted across five factory halls within Siemens' Nuremberg operational premises.

- **3GPP_38.901_Indoor_LOS [136]**: This scenario simulates an indoor environment with a 0.5-100 GHz frequency in LOS conditions. It aims to encompass various indoor deployment scenarios, such as those commonly found in office buildings and shopping centers. These indoor settings typically consist of open cubicle spaces, private enclosed offices, common areas, corridors, etc.
- **mmMAGIC_Indoor_LOS [137]**: This is designed specifically for frequencies in the range of 6-100 GHz and indoor scenarios with LOS like traditional enclosed offices, semi-closed offices (cubicle areas), and open offices.

It is also assumed that both the receiver and all transmitters use omnidirectional antennas. Other configurations are detailed in Table 4.2. It includes information such as frequency band, bandwidth, subcarrier spacing, number of resource blocks (RBs), comb size, signal-to-noise ratio (SNR), and cyclic prefix type for each 5G scenario.

Table 4.2: 5G system configurations.

| 5G Sim. Scenario | Freq (GHz) | Bw (MHz) | Subcarrier Spacing (KHz) | number of RBs | comb size | SNR | cyclic prefix |
|---|---|---|---|---|---|---|---|
| QuaDRiGa-Industrial-LOS | 5 (FR1) | 100 | 30 | 275 | 6 | 10 dB | normal |
| 3GPP-38.901-Indoor-LOS | 28(FR2) | 200 | 60 | 275 | 6 | 10 dB | normal |
| mmMAGIC-Indoor-LOS | 78(FR2) | 400 | 120 | 275 | 6 | 10 dB | normal |

The transmitted signal is convolved with the impulse response to replicate the effects of the transmission environment, generating the received signal. The received signal is generated at the receiver every 0.2 seconds, enabling TOA calculation at a frequency of 5 Hz. The received signal is then correlated with the transmitter's PRS pattern, and the delay is calculated by analyzing the correlation profile. Typically, the initial or highest peak is considered as the response. However, this approach can be compromised by noise or when the LOS coefficient is weaker than the multipath coefficient. To address this, a threshold is set to eliminate values below it, and the first peak above the threshold is chosen as the

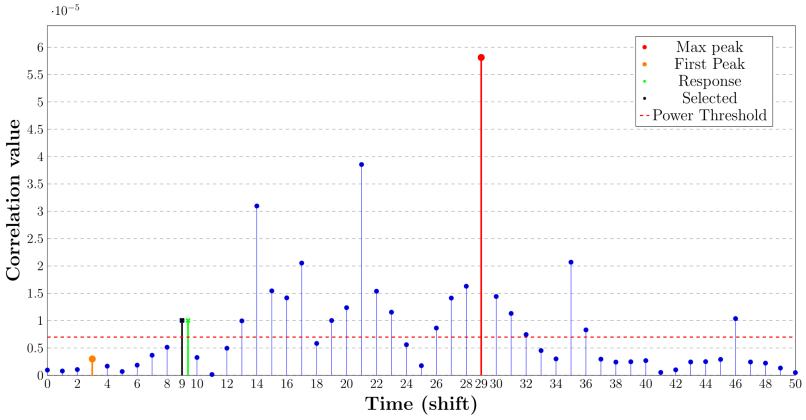



Figure 4.3: PRS Correlation Profile

response. Through empirical testing, a threshold of 0.2 was determined to be optimal for TOA estimation with the generated data. An example of a correlation profile in the simulation is displayed in Fig.4.3. The response shown in green is neither the first nor the maximum peak. Still, a suitable threshold allowed the selection of the first peak as the response. The statistics of the error in the resulting estimated distance to each base station are provided in Appendix A.1, Table A.1.

### 4.2.2 Evaluation Metrics

For the evaluation, two popular metrics in Simultaneous Localization and Mapping (SLAM) are used: Absolute Trajectory Error (ATE) and Relative Pose Error (RPE) [138]. ATE is a metric for global consistency, defined as the Root Mean Square Error (RMSE) between the absolute distances of the estimated and ground truth trajectories. RPE complements ATE by focusing on local accuracy and is primarily utilized for evaluating odometry systems. It provides a concise measure of the error between consecutive poses in the estimated trajectory compared to the ground truth trajectory. RPE helps to assess the drift or error accumulation in pose estimation over time, specifically at smaller time intervals.

In addition to these metrics, the RMSE for each trajectory coordinate axis is analyzed, denoted as $E_a$ where $a$ represents $x$, $y$, or $z$. This calculation was carried out to assess the accuracy of the estimation along specific spatial directions. The $\mathrm{RMSE}$ error for each

coordinate axis $a \in \{x, y, z\}$ was calculated using the following formula:

$$E_a = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a_{i,\text{est}} - a_{i,\text{gt}})^2}, \quad a \in \{x, y, z\}$$

In this equation, $N$ represents the total number of poses, while the subscripts $(i, \text{est})$ and $(i, \text{gt})$ correspond to the $i$-th estimated and ground truth poses, respectively. This analysis gave us valuable insights into potential accuracy variations dependent on the spatial direction. By evaluating the error along each axis separately, one could identify any discrepancies in performance for different spatial orientations.

To further assess the suitability of the proposed algorithms for real-time applications, their implementation times were evaluated. For PGO, the average and standard deviation (std) of the time taken for each optimization process was calculated. For the ESKF, the average and std of the time taken to complete one prediction plus update state cycle was recorded.

### 4.2.3 Results

The experiments were performed on a Ubuntu 20.04 laptop with an Intel(R) Core(TM) i9-10885H CPU @ 2.40GHz with 16 cores and 32 Gb of RAM. All codes (PGO and ESKF) are implemented in Python, utilizing the relative interface of GTSAM v4.0 for creating and optimizing the factor graphs. The drone's position and orientation results are obtained from the factor graph based on the final Maximum A Posteriori (MAP) estimate for each node, where nodes are consistently generated at 10 Hz, twice the frequency of TOA measurements. On the other hand, the results for the Error State Kalman Filter (ESKF) approach are acquired at each update stage, coinciding with the reception of TOA measurements. To comprehensively evaluate the performance of the localization algorithm, error metrics are computed by establishing a comparison between the ground truth Vicon pose and the estimated pose that is temporally closest.

The detailed results for the graph-based and ESKF methods are exhaustively documented in Appendix B,

The detailed results for the graph-based and ESKF methods are exhaustively documented

in Appendix B, Table B.1, Table B.2, and Table B.3, corresponding to different datasets from Vicon Room 1 and Vicon Room 2. Specifically, Table B.1 presents results for datasets V101 and V102 from Vicon Room 1, Table B.2 shows results for dataset V103 from Vicon Room 1 and dataset V201 from Vicon Room 2, and Table B.3 contains results for datasets V202 and V203 from Vicon Room 2. These tables provide comprehensive information encompassing ATE, RPE, and translation RMSE for each distinct motion direction in various 5G simulation scenarios for both methods. Additionally, the $\text{ATE}_{\text{avg.}}$ column presents the average ATE for each method across each scenario with 2, 3, 4, and 5 base stations.

Subsequently, the analysis delves into a detailed examination of the obtained results, uncovering the underlying factors contributing to the observed performance trends.

- **Base Stations and Bandwidth Influence:** The results consistently show that the accuracy of TOA-based localization improves when the number of base stations and the available bandwidth (communication scenarios) increase. This is because more base stations provide more reference points for TOA measurements, and higher bandwidth allows for higher resolution TOA measurements. For example, using dataset V101 and a fixed communication scenario (3GPP-38.901-Indoor-LOS), the error decreases from $2.58\ m$ to $0.25\ m$ under graph-based optimization when the number of base stations increases from 2 to 5. However, adding more than four base stations sometimes does not lead to significant or further improvement. This may indicate a lower bound to the error reduction achieved by adding more antennas. Nevertheless, such redundancy may be helpful in those environments where NLOS conditions are more frequent. The effect of bandwidth on the accuracy is also evident, as higher bandwidth leads to higher resolution TOA measurement. In particular, the third scenario (mmMAGIC-Indoor-LOS), which has the highest bandwidth (400 MHz), also offers the highest accuracy for the same antennas.

- **Reduced Accuracy in the Vertical Dimension (z-Axis)** Upon analyzing the Root Mean Square Error (RMSE) along each axis ($E_x$, $E_y$, $E_z$), a prominent observation is the relatively larger errors in the $E_z$ component, indicative of the vertical direction

estimation challenge. This is attributed to the limited offsets provided by base stations in the vertical direction compared to the $x$ and $y$ directions.

- **Limited Impact of TOA Measurements on Attitude Estimation:** In assessing attitude estimation, a lack of discernible patterns is apparent, suggesting that the inclusion of TOA measurements does not notably enhance the accuracy of attitude estimation. This is inherently expected as TOA measurements primarily offer information about position rather than attitude.

- **Superior Performance of Graph-Based Optimization Compared to ESKF:** Notably, the graph-based optimization consistently outperforms ESKF across all datasets, scenarios, and base station configurations, demonstrating the graph-based approach's superior efficacy in enhancing indoor localization accuracy. The main reason behind this advantage lies in the filtering method's practice of marginalizing all older information by limiting itself to the most recent states. In contrast, the graph-based method utilizes the entire historical information, employing all past measurements up to the current one and optimizing the entire trajectory. Notably, graph-based methods tend to perform better when dealing with sparse, low-frequency measurements, such as the current implementation utilizing TOA measurements with a frequency of 5 Hz, which is relatively low. This increased susceptibility of the ESKF to sparse measurements can contribute to more significant performance disparities between the two methods.

- **Robustness Across Different Trajectories and Datasets:** The robustness of the proposed approaches was assessed across datasets with varying levels of complexity, encompassing sequences from two different Vicon rooms: Vicon Room 1 (V101, V102, V103) and Vicon Room 2 (V201, V202, V203). These sequences were categorized into three levels of difficulty: easy (V101, V201), medium (V102, V202), and difficult (V103, V203).

  Despite the diversity in trajectory complexity and environmental conditions, the accuracy of the proposed approaches did not exhibit a clear pattern across the different sequences. Surprisingly, in some instances, presented methods yielded better results

for the more difficult datasets, challenging conventional expectations. This variability underscores the importance of robustness and adaptability in UAV pose estimation algorithms, especially when operating in diverse and dynamic environments.

The findings highlight the effectiveness of the proposed methods in handling the complexities inherent in UAV flight scenarios. The lack of a consistent pattern in accuracy underscores the versatility of the approaches, demonstrating their ability to adapt to various challenges encountered during indoor flight missions. This adaptability is crucial for real-world applications, where UAVs must navigate through unpredictable environments with confidence and precision.

Figures 4.4 and 4.5 provide a comprehensive view of localization performance across all tested configurations. These visualizations clearly demonstrate two key findings: (1) the consistent improvement in accuracy with increasing number of base stations, and (2) the superior performance of higher frequency bands, with 78 GHz delivering the best results. Notably, the graph-based approach achieves sub-meter accuracy with just four base stations at 28 GHz and 78 GHz frequencies, while ESKF requires five base stations to achieve similar performance.

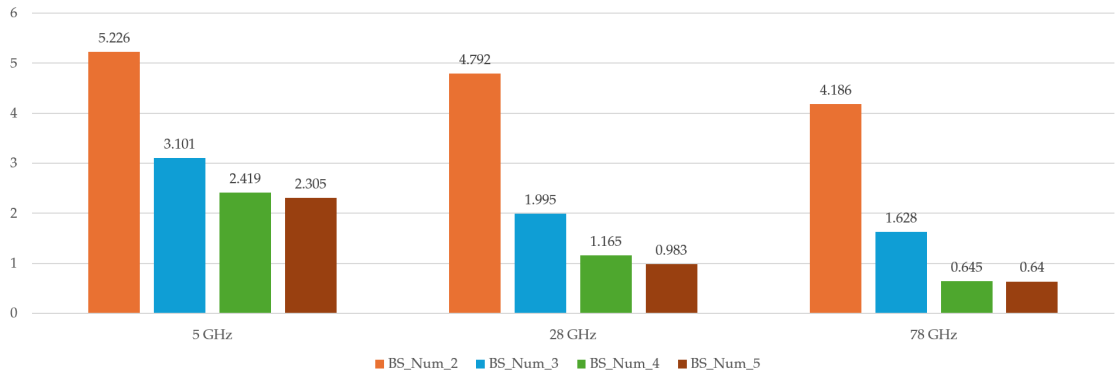



Figure 4.4: ESKF-Based Localization performance showing average ATE (meters) across all datasets. Results demonstrate decreasing error with increasing number of base stations and higher frequency bands, with 78 GHz providing the best performance.

Table 4.3 presents the execution times for both PGO and ESKF, showcasing the best results achieved using five base stations. As evident from the table, both approaches are

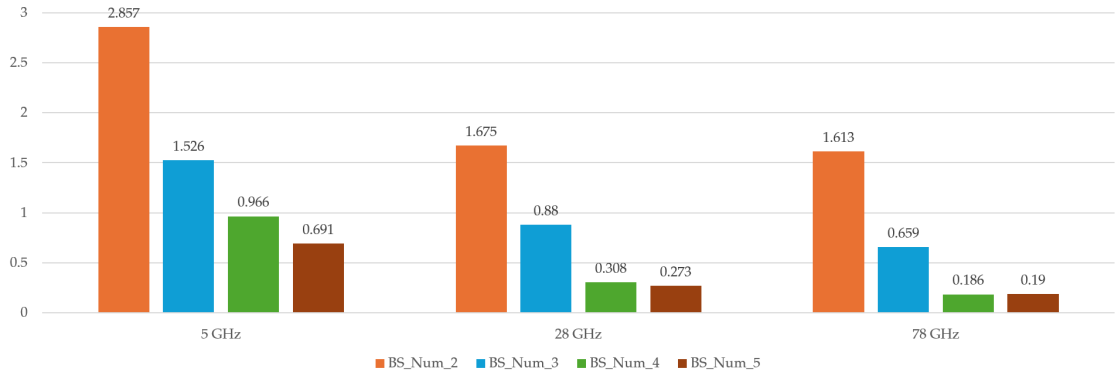


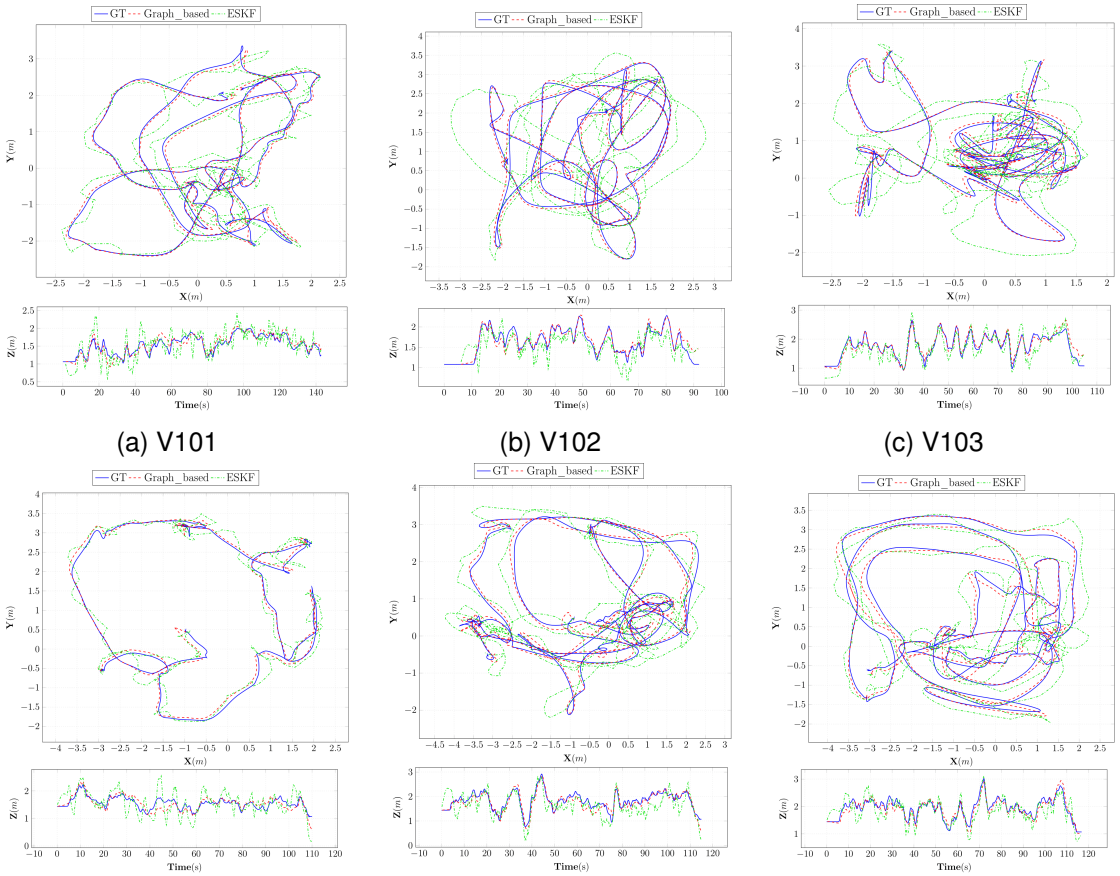

Figure 4.5: Graph-Based Localization performance showing average ATE (meters) across all datasets. The graph-based approach consistently outperforms ESKF (Figure 4.4), particularly at higher frequencies where errors are reduced by up to 70%.

Table 4.3: Computation time comparison for PGO and ESKF in Euroc MAV dataset localization. The table showcases the best results for each dataset, including the average time and standard deviation for each PGO optimization and ESKF (both prediction and update steps). All times are reported in milliseconds.

| | Dataset | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **V101** | **V102** | **V103** | **V201** | **V202** | **V203** |
| Graph-based | 3.08±2.38 | 1.85±1.34 | 2.36±1.69 | 2.89±2.44 | 2.62±2.24 | 2.44±1.72 |
| ESKF-based | 0.39±0.20 | 0.35±0.16 | 0.37±0.14 | 0.39±0.15 | 0.43±0.19 | 0.35±0.15 |

suitable for real-time implementation, with ESKF outperforming PGO, as expected. This observation reinforces ESKF's superior efficiency for real-time applications. For instance, in the V101 dataset, PGO required an average of 3.08 milliseconds per optimization, while ESKF consumed only 0.39 milliseconds for a prediction-update cycle. Since the codes were implemented in Python (with PGO partially implemented in Python), further improvement in real-time performance is expected by converting the codes entirely to C++. Overall, while the ESKF estimation in general runs faster than the PGO, these results also show that both estimation approaches can be implemented for real-time UAV applications, given a suitable on-board processing unit.

To visually illustrate the distinction between the two algorithms, the analysis focused on a specific scenario for each dataset: mmMAGIC-Indoor-LOS configuration with five base stations. In Fig. 4.6, visual representations of the trajectories in the xy-plane (upper




(a) V101 (b) V102 (c) V103

(d) V201 (e) V202 (f) V203

Figure 4.6: Trajectories in the xy-plane (upper low) and z-coordinate over time (lower row) for different datasets. Blue represents ground-truth trajectories, red represents graph-based estimation, and green represents ESKF estimation.

row) and the z-coordinate over time (lower row) for different datasets are presented. Each subfigure corresponds to a specific dataset (V101, V102, etc.), and within each subfigure, three trajectories are displayed: **Blue**: Ground-truth trajectory, **Red**: Graph-based estimation, **Green**: ESKF estimation.

## 4.3 Analysis of Base Station Positions Impact on Localization Accuracy

### 4.3.1 Geometric Dilution of Precision (GDOP)

The impact of base station positioning on localization accuracy, often quantified through Geometric Dilution of Precision (GDOP), has been extensively studied in the literature. GDOP is a well-established metric that evaluates how the geometric arrangement of base stations (or satellites) influences the accuracy of position estimation. The concept of GDOP was originally developed to describe the performance of terrestrial navigation systems and was later adapted to evaluate the geometric effect of satellite configurations on GPS accuracy. When the angular positions of satellites or base stations are close together, the GDOP value increases, resulting in poor positioning performance. Conversely, when the receiver is surrounded by well-distributed satellites or base stations, GDOP is minimized, leading to higher accuracy [139].

As highlighted in [139], GDOP is particularly critical in short-range systems, where ranging errors can significantly impact accuracy. The authors derive analytical expressions for GDOP in various scenarios, showing that GDOP is minimized when the mobile device is surrounded by well-distributed base stations, while poor geometric configurations, such as clustered base stations or when the mobile device is near a base station, result in higher GDOP and degraded positioning performance. Similarly, studies in GPS systems, such as [140], have demonstrated that GDOP plays a critical role in determining positioning accuracy, with low GDOP values corresponding to better performance. These findings underscore the importance of optimizing base station placement to achieve low GDOP, especially in applications requiring high precision, such as drone localization.

Standard GDOP measures the overall geometric quality of the base station configuration, considering all dimensions (x, y, z). It is calculated as the square root of the trace of the inverse of the geometry matrix $G^T G$, where $G$ is the matrix of unit vectors from the receiver to each base station.

The geometry matrix $G$ is constructed as follows:

$$G = \begin{bmatrix} \frac{x_1 - x_r}{d_1} & \frac{y_1 - y_r}{d_1} & \frac{z_1 - z_r}{d_1} & 1 \\ \frac{x_2 - x_r}{d_2} & \frac{y_2 - y_r}{d_2} & \frac{z_2 - z_r}{d_2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_n - x_r}{d_n} & \frac{y_n - y_r}{d_n} & \frac{z_n - z_r}{d_n} & 1 \end{bmatrix},$$

where $(x_i, y_i, z_i)$ are the coordinates of the $i$-th base station, $(x_r, y_r, z_r)$ are the receiver coordinates, and $d_i$ is the distance between the receiver and the $i$-th base station.

The standard GDOP is then computed as:

$$\text{GDOP} = \sqrt{\text{trace}((G^T G)^{-1})}.$$

A lower GDOP value indicates a better geometric configuration, leading to higher positional accuracy. GDOP includes contributions from all dimensions (x, y, z, and time).

Given the extensive research on GDOP and its implications for localization systems, this topic is considered well-understood. In this work, we showcase the effect of base station positioning through five representative configurations (Tetrahedral, Asymmetric, Diamond, Z-Shape, and Clustered) to illustrate the established principles. For a deeper exploration of GDOP and its optimization, readers are referred to [139] and [140], among others.

### 4.3.2 Base Station Configurations and Experimental Setup

To evaluate the impact of base station placement on localization accuracy, experiments were conducted using five different base station configurations across six EuRoC MAV dataset sequences. Five distinct base station configurations (Fig.4.7) with exact coordinates detailed in Table 4.4 across six EuRoC MAV sequences were evaluated to systematically analyze geometric impacts on localization accuracy:

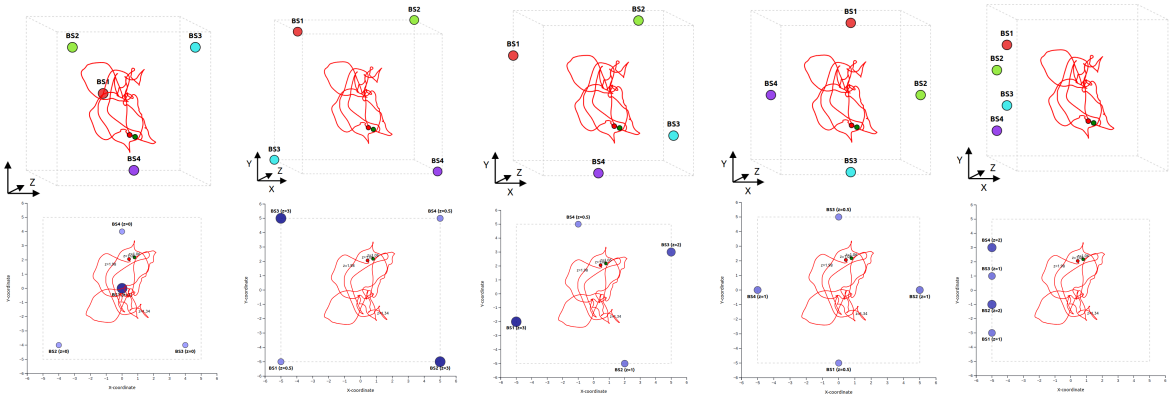
**Tetrahedral**: 3D distributed (ceiling center + ground triangle)

**Z-Shape**: Diagonally alternating heights

**Asymmetric**: Realistic irregular placement

**Diamond**: Distributed evenly at two different heights

**Clustered**: Single-wall co-planar arrangement



(a) Tetrahedral (b) Z-shaped (c) Asymmetric (d) Diamond (e) Clustered

Figure 4.7: Five configurations of base stations (BS1-BS4) with trajectory data from the EuRoC MAV V101 dataset. Top row: 3D representations; Bottom row: corresponding 2D X-Y projections with z-coordinates. Configurations are ordered by performance (best to worst): Tetrahedral (ATE: 0.111m), Z-shaped (0.132m), Asymmetric (0.141m), Corner (0.161m), and Clustered (0.182m).

Table 4.4: Exact 3D coordinates (in meters) of base stations for all evaluated configurations.

| Base Station | Tetrahedral | Diamond | Z-Shape | Asymmetric | Clustered |
|---|---|---|---|---|---|
| BS1 | (0, 0, 3) | (0, -5, 0.5) | (-5, -5, 0.5) | (-5, -2, 3) | (-5, -3, 1) |
| BS2 | (-4, -4, 0) | (5, 0, 1) | (5, -5, 3) | (2, -5, 1) | (-5, -1, 2) |
| BS3 | (4, -4, 0) | (0, 5, 0.5) | (-5, 5, 3) | (5, 3, 2) | (-5, 1, 1) |
| BS4 | (0, 4, 0) | (-5, 0, 1) | (5, 5, 0.5) | (-1, 5, 0.5) | (-5, 3, 2) |

### 4.3.3 Experimental Results and Analysis

Table 4.5 presents the ATE for each configuration across all sequences.

To better understand the relative performance of different configurations, Table 4.6 presents the percentage increase in ATE relative to the best-performing configuration (Tetrahedral).

The bar chart in Figure 4.8 presents the average GDOP values for each base station configuration across all EuRoC MAV dataset sequences, providing direct empirical evidence

Table 4.5: Average ATE (Absolute Trajectory Error) for different base station configurations across datasets.

| Configuration | v101 | v102 | v103 | v201 | v202 | v203 | Average |
|---|---|---|---|---|---|---|---|
| Tetrahedral | 0.0709 | 0.0721 | 0.0849 | 0.1228 | 0.1587 | 0.1501 | 0.1099 |
| Z-Shape | 0.1036 | 0.0960 | 0.0939 | 0.1660 | 0.1802 | 0.1623 | 0.1337 |
| Asymmetric | 0.1155 | 0.1003 | 0.1035 | 0.1723 | 0.1840 | 0.1665 | 0.1403 |
| Diamond | 0.1263 | 0.1045 | 0.1128 | 0.3412 | 0.1819 | 0.1763 | 0.1739 |
| Clustered | 0.1211 | 0.1494 | 0.1431 | 0.2410 | 0.2028 | 0.2244 | 0.1803 |

Table 4.6: Percentage increase in ATE relative to the Tetrahedral configuration (%)

| Configuration | v101 | v102 | v103 | v201 | v202 | v203 | Average |
|---|---|---|---|---|---|---|---|
| Tetrahedral | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Z-Shape | 46.1 | 33.1 | 10.6 | 35.2 | 13.5 | 8.1 | 21.7 |
| Asymmetric | 62.9 | 39.1 | 21.9 | 40.3 | 15.9 | 10.9 | 27.7 |
| Diamond | 78.1 | 44.9 | 32.9 | 177.9 | 14.6 | 17.5 | 58.2 |
| Clustered | 70.8 | 107.2 | 68.6 | 96.3 | 27.8 | 49.5 | 64.1 |

of how geometric arrangement affects positioning quality. These GDOP measurements strongly correlate with the ATE results in Tables 4.5 and 4.6, confirming the theoretical relationship between geometric configuration and localization accuracy.

The Tetrahedral configuration consistently shows the lowest GDOP values (averaging approximately 2.58 across all sequences), which translates directly to its superior localization performance with the lowest ATE (10.99 cm). This configuration's optimal 3D distribution—with one base station at ceiling height and three at ground level in a triangular arrangement—creates a geometry matrix with excellent conditioning, minimizing the trace of $(G^T G)^{-1}$ in all spatial dimensions.

The Z-Shape and Asymmetric configurations show moderately higher GDOP values (around 3.25 and 3.56 respectively), corresponding to their intermediate ATE results (13.37 cm and 14.03 cm). These configurations still provide reasonable 3D distribution but lack the optimal surrounding geometry of the Tetrahedral arrangement.

Most striking is the clear division between configurations with good 3D distribution (Tetrahedral, Z-Shape, Asymmetric) and those with poorer spatial diversity (Diamond and Clustered).

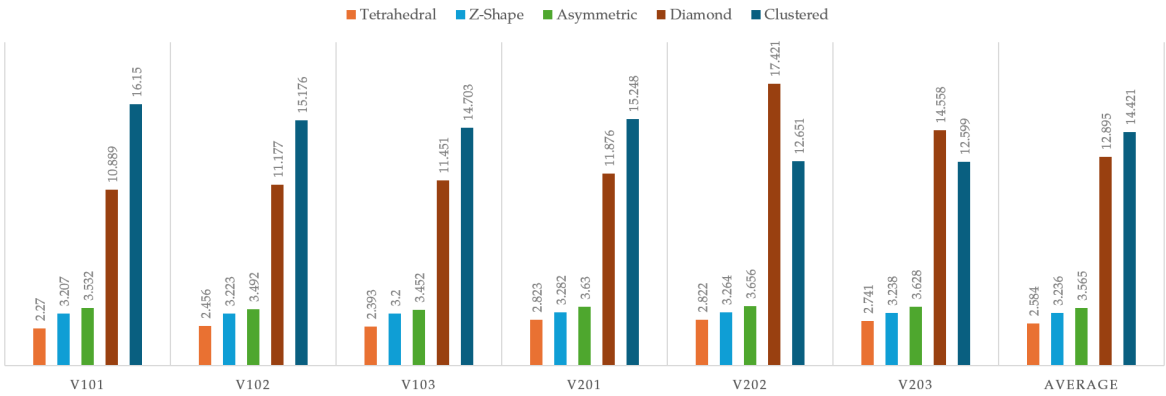



Figure 4.8: Comparison of average Geometric Dilution of Precision (GDOP) values across different base station configurations for all EuRoC MAV dataset sequences. Lower GDOP values indicate better geometric configuration for localization.

The Diamond configuration shows significantly higher GDOP values (averaging about 12.89), particularly in sequence V202 where it peaks at 17.42. This poor geometric conditioning directly corresponds to its degraded localization performance, especially in the challenging V201 sequence where its ATE was 177.9% higher than the Tetrahedral configuration.

The Clustered configuration consistently exhibits the highest GDOP values across all sequences (averaging approximately 14.42), reflecting its fundamentally limited geometric diversity with all base stations positioned along a single wall. This poor geometric conditioning directly translates to the highest average ATE (18.03 cm).

The bar chart visualizes an important practical insight: GDOP varies not just with base station configuration but also with the position of the mobile device within the environment. As the device moves through different regions of the test volume during each sequence, it experiences varying geometric conditions. Configurations with more uniform 3D distribution maintain lower GDOP values throughout the trajectory, while geometrically limited configurations experience significant GDOP spikes in certain regions.

## 4.4 Limitations

The approach used in this study currently has several limitations and hence leaves potential for future work. The 3D position is not fully constrained with only two antennas, making convergence difficult without other measurements. Nevertheless, the UAV's rotation errors primarily result from IMU noise, as the 5G TOA measurements only provide distances to the antennas. The yaw estimation has drift issues because it lacks global measurement to correct it. Integrating other sensors can improve the localization accuracy by observing the rotation around $z$, *e.g.,* employing a magnetometer. Notably, a camera can be incorporated to add other constraints on the 6DoF relative motion based on visual features and loop closures.

Furthermore, the error in the $z$ axis is larger than along $x$ and $y$ axes because of limited offset or variation in the positions of the base stations in the height direction. Integration of the barometer's absolute height measurements is foreseen as a potential solution to address these issues. Additionally, the localization accuracy depends heavily on the quality of the TOA measurements, which can be negatively affected by NLOS conditions. In such cases, correctly setting the measurement uncertainty for each TOA range factor, using Mahalanobis distance to discard outliers, or applying a robust kernel (such as *e.g.,* , a Huber kernel) to the cost function, may be beneficial to alleviate the problem.

Finally, the proposed method assumes that the positions of the base stations are known with high confidence and fixed in the exact location, which may not be the case in real-world scenarios where the stations may be moving or their positions may be completely unknown. However, especially the PGO approach can be adapted to also include an estimation of the unknown position of the BSs, if the TOA measurements can be unambiguously assigned to the sending base station, respectively.

## 4.5 Conclusion

This chapter investigated two distinct methodologies for integrating 5G TOA data with Inertial Measurement Unit (IMU) readings to enhance the pose estimation of a UAV in indoor

environments. Specifically, the research developed and evaluated an Error State Kalman Filter (ESKF) and a Pose Graph Optimization (PGO) framework. Both methods were applied to the EuRoC MAV dataset, which was augmented with simulated 5G TOA data using realistic channel models from the QuaDRiGa simulator.

The experimental results demonstrate that both approaches significantly improve localization accuracy compared to using IMU data alone. Across all configurations, the PGO method consistently outperformed the ESKF. For instance, in dataset V101 using the 28 GHz configuration with 200 MHz bandwidth, the PGO achieved remarkable accuracy improvements as the number of base stations increased, with the error decreasing from 2.58m with two base stations to 0.26m with five base stations, while the ESKF improved from 5.12m to 0.91m. The highest frequency configuration at 78 GHz with 400 MHz bandwidth demonstrated the best performance across all datasets, with PGO achieving errors as low as 0.13m (V101) and ESKF reaching 0.34m in the same conditions.

In terms of computational performance, both methods proved suitable for real-time implementation on standard hardware (tested on an Intel i9-10885H CPU). While the ESKF showed increased susceptibility to sparse measurements due to its reliance on recent state estimates, it demonstrated exceptional efficiency with average processing times of 0.35-0.43ms per prediction-update cycle. The PGO, though computationally more demanding with 1.85-3.08ms per optimization cycle, still operated fast enough for real-time applications while providing better accuracy. These timing results, achieved with Python implementation, suggest even better performance potential with C++ optimization.

The analysis revealed that increasing the number of base stations significantly improved accuracy up to four base stations. While adding a fifth base station sometimes provided marginal improvements, in rare cases it showed slightly reduced performance - though these degradations were minimal. For instance, in dataset V103 with 78 GHz configuration, the PGO's error changed from 0.14m with four base stations to 0.16m with five base stations. This pattern suggests that four base stations might represent an optimal configuration for most deployment scenarios, balancing infrastructure costs and performance gains.

The Root Mean Square Error (RMSE) analysis along individual axes showed a consistent

pattern across all configurations: the vertical (z-axis) errors were significantly higher than horizontal errors. For example, in the 5 GHz configuration (100 MHz bandwidth), while horizontal accuracies (x and y axes) were typically around 0.2-0.4m, the vertical errors reached 0.6-1.0m for both methods. This increased vertical error is attributed to the limited height variation in base station placements, suggesting that strategic vertical distribution of base stations could be crucial for applications requiring precise height estimation.

Despite the promising results, several limitations were identified. The methods struggled with accurate estimation when only two base stations were available, and the z-axis accuracy was consistently lower due to limited vertical constraints. Moreover, the attitude estimation, particularly the yaw angle, suffered from drift issues due to the lack of global rotational measurements. These challenges highlight areas for future work, such as integrating additional sensors like magnetometers or cameras to provide rotational constraints and improve vertical accuracy.

In conclusion, the integration of 5G TOA measurements with IMU data presents a viable path toward enhancing UAV localization in indoor environments. The PGO method, in particular, offers superior accuracy by leveraging full trajectory optimization, achieving sub-meter accuracy in optimal conditions with five base stations and high bandwidth configurations (400 MHz). The findings pave the way for further advancements in sensor fusion techniques, suggesting that future research should focus on addressing the identified limitations and exploring the integration of additional sensing modalities to achieve robust, real-time MAV localization.

# Chapter 5

# Visual SLAM Using 5G: Integrating 5G TOA, IMU, and RGBD Images Using ORB-SLAM3

This chapter explores the integration of Time of Arrival (TOA) measurements into ORB-SLAM3, presenting a comprehensive approach that enhances both global and local localization capabilities for indoor drone navigation. The methodology extends ORB-SLAM3's optimization pipeline to jointly process TOA data from 5G base stations alongside visual and inertial measurements, while simultaneously estimating system biases. This integration not only enables globally referenced trajectories but also resolves fundamental limitations of monocular visual SLAM, such as scale ambiguity and drift. The approach is extensively evaluated across multiple SLAM configurations, including RGB-D, RGB-D-Inertial, Monocular, and Stereo modes, using both the Aerolab and EuRoC MAV datasets. A significant contribution is the investigation of scenarios with unknown base station positions and relaxed connectivity conditions, demonstrating effective operation with intermittent or sequential connections to base stations at different time intervals (e.g., BS1: 10-40s, BS2: 50-70s, BS3: 80-100s). The experimental results reveal several key insights. For minimal sensor configurations like Monocular SLAM, 78 GHz TOA integration provides substantial improvements in local

accuracy (up to 11.4%) while maintaining excellent scale estimation (average error of 0.156%). In sequential base station operation with unknown locations, the system achieves an average improvement of 4.40% in local accuracy with reliable scale estimation (1.30% average error) despite intermittent coverage. A particularly valuable finding is TOA's effectiveness as an alternative to loop closure. When loop closure is disabled, TOA integration from three sequential base stations improves accuracy by 18.6% compared to the no-loop-closure baseline, with continuous TOA measurements achieving even greater enhancements (29.6%). This capability is crucial for real-world scenarios where loop closure may be unavailable due to linear trajectories, exploration of new areas, or resource constraints. Comparative analysis with state-of-the-art approaches, including UWB-VO implementations [4], demonstrates the robustness of the proposed method. Despite using a lower TOA measurement frequency (5 Hz versus 30 Hz in UWB-VO), our approach achieves comparable performance (0.202m mean RMSE) while operating under more relaxed and realistic constraints. This chapter provides comprehensive experimental results across two frequency bands (28 GHz and 78 GHz), revealing that 78 GHz TOA consistently outperforms 28 GHz in both local and global accuracy. The integration enables global positioning across all SLAM modes—a capability absent in baseline configurations—with global accuracy ranging from 0.126m to 0.262m for 78 GHz implementations. Through detailed evaluations and analyses, this research demonstrates that 5G TOA measurements can significantly enhance SLAM performance in complex indoor environments, particularly in challenging scenarios where traditional visual SLAM approaches face fundamental limitations.

## 5.1 Methodology

This section outlines the technical framework and processes to integrate TOA data into ORB-SLAM3 for globally consistent localization. The discussion encompasses the distinction between local and global frames, TOA factor formulation, and the system modifications that enable robust pose estimation across diverse datasets.

### 5.1.1 Local Frame vs. Global Frame in SLAM

In SLAM, the concepts of local and global reference frames are fundamental for accurately representing and interpreting positional data within a mapped environment. A **local frame** is usually established at the beginning of the mission (e.g., at the first keyframe or the sensor's starting point). Once set, all subsequent poses, map points, and sensor measurements are calculated and expressed relative to this local coordinate system. While effective for maintaining local consistency, the local frame lacks alignment with a fixed real-world reference, leading to cumulative drift over time. Additionally, it can limit the system's ability to integrate or compare data from multiple sessions or external sources, which often require a global reference frame for alignment.

In contrast, a **global frame** is a fixed, absolute reference system tied to known landmarks or infrastructure, such as 5G base stations in our setup. This global frame provides a consistent anchor point independent of the SLAM system's starting pose. By aligning the local observations to the global frame, the SLAM system can mitigate drift, maintain long-term positional accuracy, and ensure interoperability with other systems or datasets operating in the same global context.

**Figure 5.1** illustrates the distinction between the local and global frames and highlights the transformation process that bridges them. The local map, represented in red, depicts the environment relative to the initial sensor position. However, this map lacks global context, making it unsuitable for applications requiring absolute positioning. Conversely, the global map, depicted in teal, integrates TOA measurements from fixed base stations, which serve as global reference points. The transformation $Tr$, shown in the figure, aligns the local frame to the global coordinate system, linking the two maps.

This alignment process is essential for achieving a globally consistent SLAM system. The transformation $Tr$ is estimated by leveraging TOA measurements, which provide spatial constraints based on the distances between the robot and the base stations. These constraints refine the local-to-global alignment and ensure that the trajectory and map points are accurately represented within the global frame. The dashed arrow in Figure 5.1 illustrates the

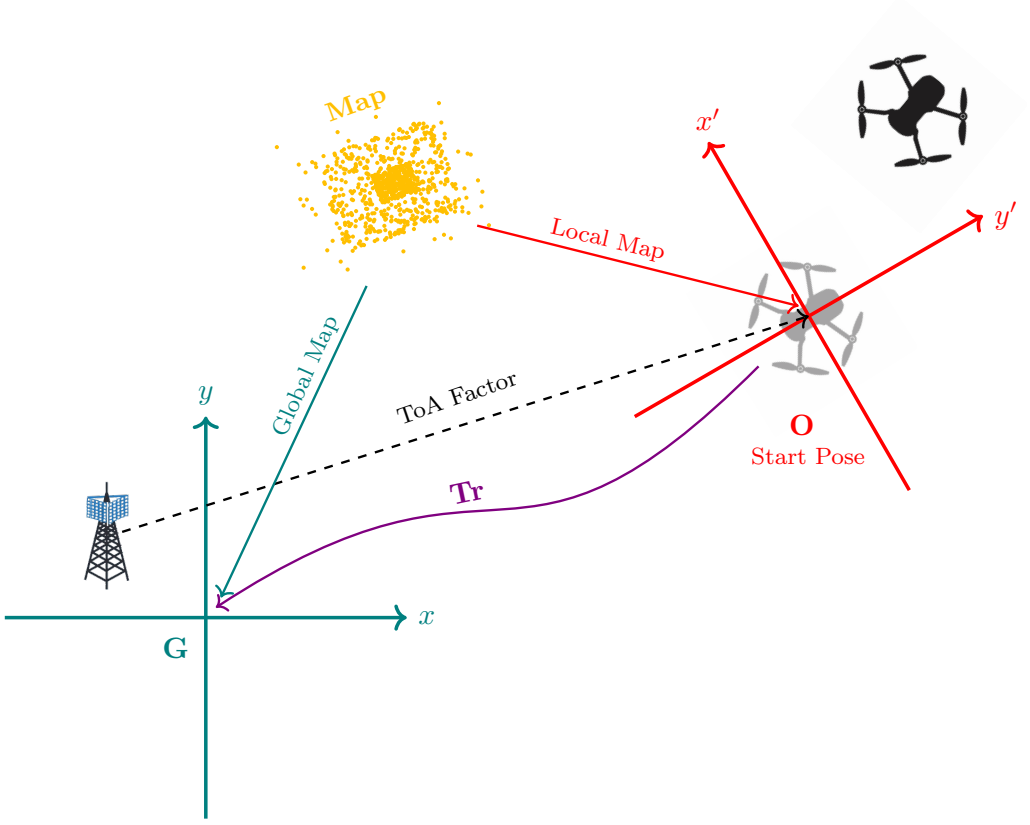



Figure 5.1: Diagram illustrating the relationship between the local and global frames in SLAM. The transformation $\mathbf{Tr}$ aligns the local frame $(x', y')$ with the global coordinate frame $(x, y)$. The single map, shown in both frames, is represented in red for the local frame and teal for the global frame. The initial pose of the drone is faded to indicate its starting position, while the current pose is shown in the local frame. A base station in the global frame provides TOA measurements, aiding in the alignment and refinement of global positioning. The dashed arrow represents the influence of TOA measurements, linking the local and global frames via $\mathbf{Tr}$.

influence of TOA measurements in bridging the local and global scales.

The integration of a global reference frame offers several advantages:

- **Drift Mitigation**: Periodic alignment with the global frame reduces cumulative drift inherent in local-only SLAM systems.
- **Multi-Agent Collaboration**: By operating within the same global frame, multiple robots or drones can share maps and coordinate actions effectively.
- **Robustness in Challenging Environments**: In scenarios with limited visual or inertial cues (e.g., featureless or dynamic environments), the global reference frame, established through methods such as 5G base stations or GPS, provides consistent positional information, ensuring robustness and accuracy.

This distinction between local and global frames underpins the methodology used in this work, particularly in the integration of TOA data within ORB-SLAM3.

### 5.1.2 TOA Factor Formulation

TOA factor integrates distance measurements from base stations into the SLAM framework through a formulated error model. The factor connects multiple optimization vertices: the camera pose, local-to-global transformation, TOA measurement bias, and, in monocular cases, a scale factor. Figure 5.2 illustrates the structure of the TOA factor, highlighting how each component interacts within the optimization graph. The double-bordered base station nodes indicate that these parameters are fixed and their positions are known.

**Mathematical Formulation:** Let $d_{\text{meas}}$ represent the measured distance affected by noise $\boldsymbol{\eta}^{\text{Dist}}$ and bias $\tau$:

$$d_{\text{meas}} = d_{\text{true}} + \boldsymbol{\eta}^{\text{Dist}} + \tau \tag{5.1}$$

The factor involves the following key components:

- Camera pose in the SLAM frame (local frame): $\mathbf{T}_{oc}$

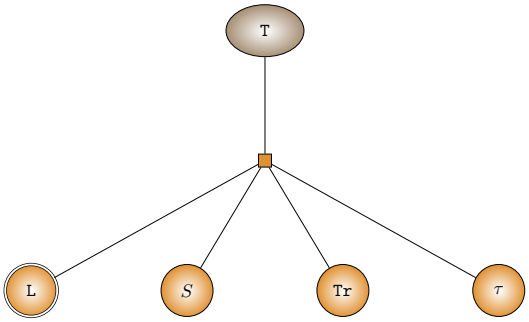



Figure 5.2: Structure of the TOA factor, illustrating the key components: camera pose node ($\mathrm{T}$), scale factor ($s$), local-to-global transformation node ($\mathrm{Tr}$), base station position node ($\mathrm{L}$), and bias node ($\tau$). Double-bordered nodes indicate fixed parameters during optimization.

- Local-to-global transformation: $\mathbf{T}_{go}$
- Base station position in global frame: $L_{\mathsf{G}}$
- Scale factor (for monocular systems): $s$
- Bias of TOA distance measurement: $\tau$

The transformation chain to calculate the keyframe pose in the global frame is given by:

$$\mathbf{T}_{gc} = \mathbf{T}_{go} \cdot \mathbf{T}_{oc} \tag{5.2}$$

The calculated distance between the camera and the base station is:

$$d_{\mathsf{calculated}} = ||s \cdot t_{\mathsf{gc}} - L_{\mathsf{G}}||_2 \tag{5.3}$$

where $t_{\mathsf{gc}}$ is the translation component of $\mathbf{T}_{gc}$, representing the camera's position in the global frame and $s$ is the scaling parameter applied to account for scale differences, crucial in monocular SLAM systems where scale ambiguity exists. By introducing $s$ into the error computation, the optimization can estimate the true metric scale of the environment:

- If $s$ is known (e.g., in stereo or RGB-D systems), it can be fixed at $s = 1$.
- In monocular systems, $s$ becomes an additional variable to estimate during optimization.

The TOA factor error is then computed as:

$$e = d_{\text{calculated}} - (d_{\text{meas}} - \tau) \tag{5.4}$$

The optimization framework minimizes a combined cost function for TOA measurements. For a single base station, the objective is to minimize the squared error between the measured and estimated TOAs:

$$\min_{\mathbf{T}_{oc}, \mathbf{T}_{go}, \tau, s} \sum_{i} (e_i)^2 \tag{5.5}$$

where $i$ indexes the individual TOA measurement, and $e_i$ is the error between the measured and estimated TOA for the i-th measurement. For multiple base stations, the objective function is extended to minimize the sum of squared errors across all base stations.

$$\min_{\mathbf{T}_{oc}, \mathbf{T}_{go}, \tau_j, s} \sum_{i} \sum_{j} (e_{i,j})^2 \tag{5.6}$$

where the summation spans all TOA measurements and their corresponding optimization variables.

- $j$ indexes the base stations
- $\tau_j$ is the bias for the $j$-th base station
- $e_{i,j}$ is the error between the measured and estimated TOA for the $i$-th measurement at the $j$-th base station

Note that for pure local SLAM with unknown base station positions, we can modify the system architecture by removing the fixed transformation node entirely or maintaining it as fixed while changing the base station nodes from fixed to non-fixed status. This configuration eliminates the requirement for prior knowledge of base station locations, though it consequently removes global positioning capabilities. By making base station positions estimable parameters within the optimization process, the system can still leverage TOA measurements for scale estimation and trajectory refinement without requiring pre-surveyed infrastructure.

**Uncertainty Propagation:** The information matrices (inverse covariance matrices) of both the scale parameter and the local-to-global transformation are suitably updated through the optimization process. These matrices are computed using the Hessian approximation derived from the Jacobians of the error terms with respect to the corresponding vertices. The updated uncertainties serve as prior information in subsequent optimization iterations, enabling a more robust convergence by:

- Providing appropriate weighting for new measurements based on accumulated certainty
- Preventing aggressive updates based on noisy or conflicting measurements to well-established estimates
- Allowing faster adaptation when uncertainty is high

### 5.1.3 System Components and Integration

**System Overview:** The proposed system integrates TOA measurements into the ORB-SLAM3 framework to achieve globally consistent localization and mapping. The main components of the system include tracking, local mapping, loop closing & TOA-based global map refinement. These components interact in a multi-threaded architecture to ensure real-time performance.

Figure 5.3 illustrates the system architecture, highlighting how TOA measurements are incorporated into the SLAM pipeline. Several existing components have been modified, including the Local Bundle Adjustment module, which now incorporates TOA measurements for improved optimization accuracy. The system diagram uses color coding to distinguish between modifications to the original ORB-SLAM3 framework: new components are highlighted in green, while modified components are shown in yellow.

**System Integration:** The TOA measurements are integrated at multiple levels within the ORB-SLAM3 framework to maximize their utility while maintaining real-time performance, i.e., Tracking, Local Mapping, Loop Closing & TOA-based global map refinement. The following sections highlight key changes and additions to the system.

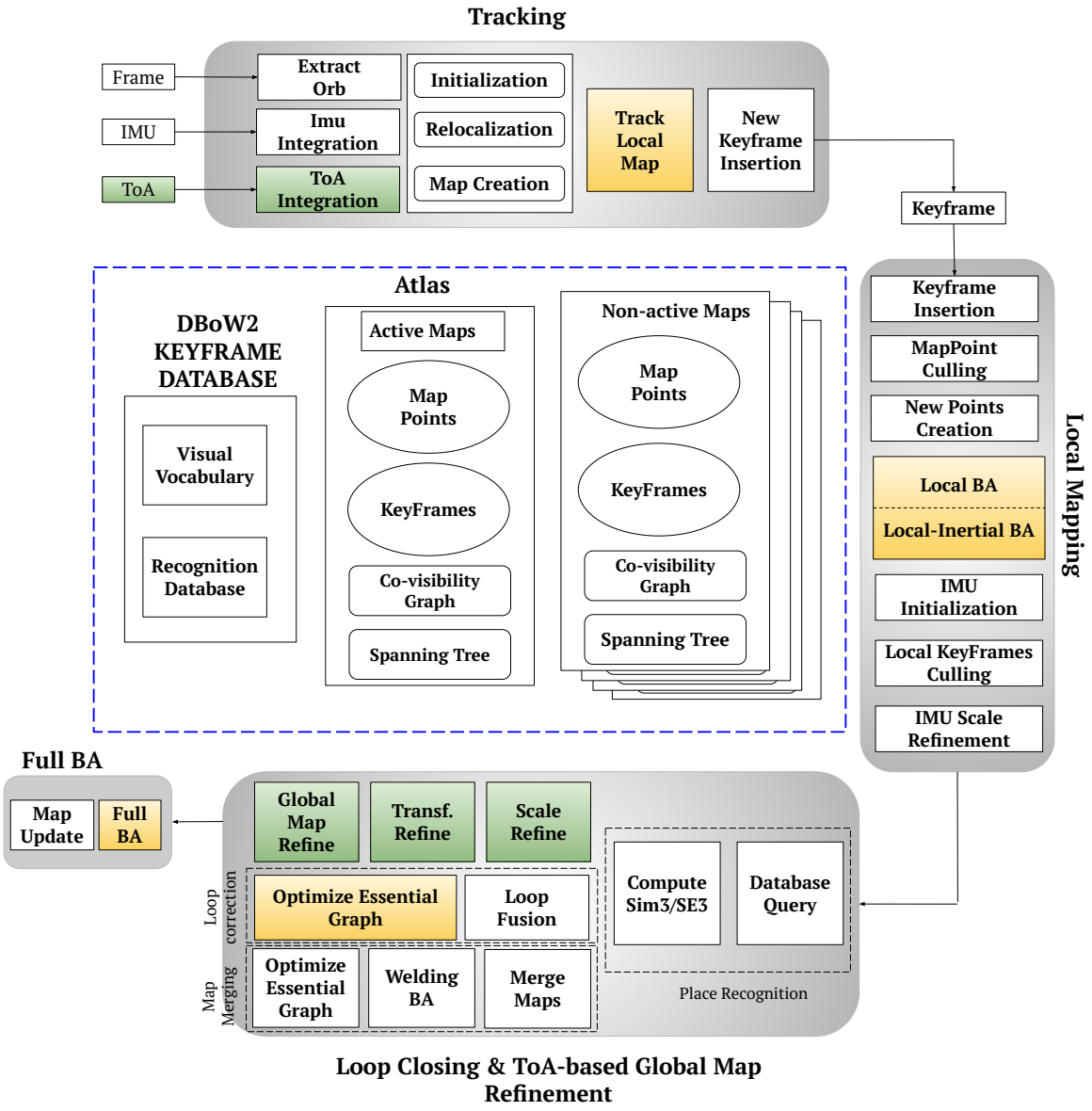



Figure 5.3: Diagram of ORB-SLAM3 with TOA integration, illustrating the pipeline across Tracking, Local Mapping, and Loop Closing threads. New components are shown in green, including TOA-based global map refinement, while modified components are in yellow, such as Local Bundle Adjustment.

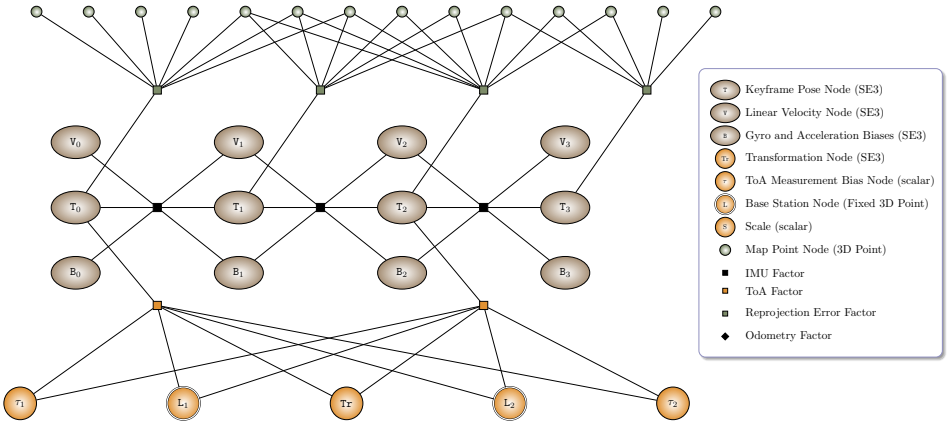
**Tracking Thread** TOA factors are incorporated into frame-to-frame pose optimization, providing additional constraints during tracking. While this integration offers modest improvements, it can enhance robustness during rapid motion or feature-poor sequences. Figure 5.4c illustrates the optimization structure in the Tracking thread, showing the interconnected nodes for the coming frame pose, map points, the local-to-global transformation, TOA biases, and base station positions (two are shown in the example). Two main factor types are integrated into the optimization process: TOA distance factors, as discussed earlier, and reprojection error factors. Reprojection error factors assess the difference between the observed location of a map point in an image and its projected position based on the estimated camera pose. By minimizing this error, the camera pose's alignment with the map is optimized.

**Local Mapping Thread** Beyond the Tracking thread, TOA data is integrated into the Local Mapping thread to refine keyframe poses, map points, and the local-to-global transformation. This is achieved through two optimization components:

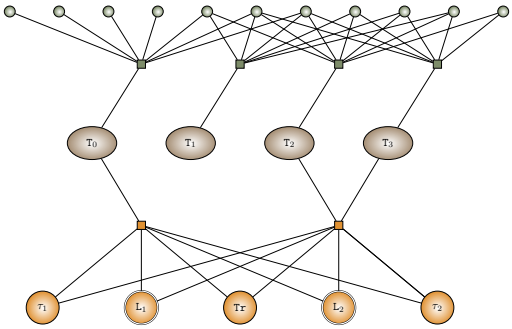
- For systems without IMU data, the Local Bundle Adjustment incorporates visual features alongside the TOA measurements to optimize the poses of keyframes within the local optimization window, as well as the associated map points and the local-to-global transformation. This is illustrated in Figure 5.4b, where the optimization graph includes nodes for keyframe poses, map points, and the transformation between the local and global coordinate frames.
- Local-Inertial Bundle Adjustment: For systems equipped with an IMU, the Local-Inertial Bundle Adjustment jointly considers visual, inertial, and TOA measurements in the optimization process. This combined optimization leads to more accurate pose and map estimation, especially in the presence of significant IMU noise. The structure of this optimization graph is shown in Figure 5.4a, where the additional IMU-related factors are incorporated alongside the visual and TOA-based factors.

By leveraging TOA measurements in these Local Mapping optimizations, the ORB-SLAM3 system can further refine the estimates of keyframe poses, map points, and the local-to-

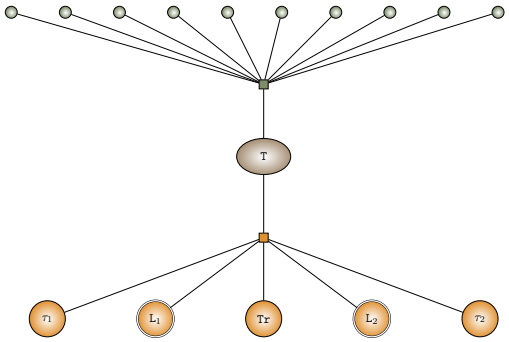
global transformation, leading to improved overall accuracy and robustness. The inclusion of TOA data is particularly beneficial in scenarios where the visual information alone may be insufficient, such as in the presence of significant IMU noise or in feature-poor environments.




(a) Local Inertial BA

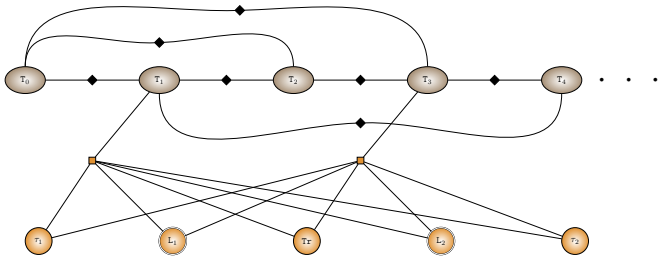


(b) Local Bundle Adjustment

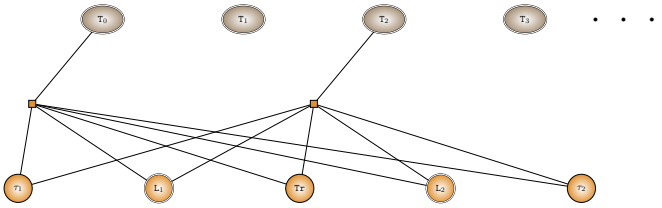


(c) Tracking Pose Optimization

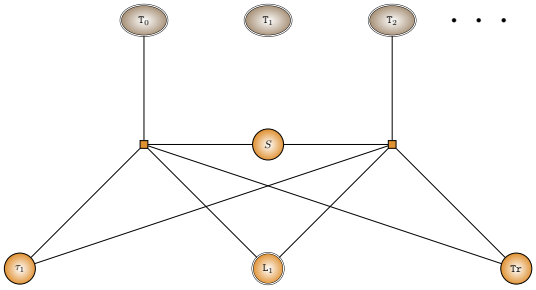
Figure 5.4: Optimization graph structures (Part 1): (a) Local Inertial BA integrates visual, inertial, and ToA measurements to refine keyframe poses and map points while addressing IMU noise; (b) Local Bundle Adjustment refines keyframes, map points, ToA biases, and local-to-global transformations within a local window; (c) Tracking Pose Optimization enables real-time camera pose estimation using reprojection errors and ToA constraints. Double-bordered nodes represent fixed vertices during optimization.



(a) Global Map Refinement



(b) Transformation Refinement



(c) Scale Refinement

Figure 5.5: Optimization graph structures (Part 2): (d) Global Map Refinement ensures global consistency using odometry, co-visibility, loop closure, and ToA edges to improve global map consistency and keyframe accuracy; (e) Transformation Refinement periodically optimizes the local-to-global transformation with fixed keyframe poses, enhancing robustness against IMU noise and drift; (f) Scale Refinement optimizes the scale factor for monocular SLAM, ensuring consistent keyframe poses and map points. Double-bordered nodes represent fixed vertices during optimization.

**Loop Closing & TOA-based Global Map Refinement** One key component in Loop closing is the optimization of the Essential Graph which is responsible for Loop correction in which TOA edges are also added. The Essential Graph is a subset of the full map, containing all the keyframes and four types of edges:

- Odometry edges: These edges connect consecutive keyframes, representing the relative pose change between them.
- Covisibility edges: These edges connect keyframes that share a significant number of map points, representing the visual constraints between them.
- Loop closure edges: These edges are added when a loop closure is detected, providing additional constraints to correct drift and maintain global consistency.
- TOA-related edges

By incorporating TOA measurements into the Essential Graph optimization, the accuracy of keyframe poses is improved by leveraging the additional spatial information from the TOA data and refining the local-to-global transformation.

**New TOA-based Components** In addition to these modifications, new components are added to the system, which are explained in the following section.

**Global Map Refinement:** A periodic global optimization process is performed to maintain the global consistency of the map using TOA. The optimization process is triggered by multiple conditions:

- Excessive TOA distance errors beyond measurement covariance thresholds
- Significant accumulated motion since last optimization
- Time-based triggers ensuring regular refinement
- Keyframe count thresholds

Figure 5.5a illustrates the optimization graph structure. It comprises keyframe pose nodes, transformation nodes, and bias nodes, interconnected by three types of edge types, Odometry edges (straight lines), Covisibility edges (angled lines connecting from above or below), and TOA-related edges.

This integrated approach combines odometry, visual, and TOA constraints to improve the accuracy of the estimated keyframe poses. After the keyframe poses are updated, the map points are also updated accordingly to maintain the consistency of the global map.

**Transformation Refinement for Inertial Systems:** In systems equipped with IMU, the initial estimation of the local-to-global transformation can be particularly challenging. When the initial guess for this transformation is far from the true value, the local inertial optimization in the Local Mapping thread can struggle to effectively leverage the TOA measurements to converge to the correct transformation. To mitigate this issue, a dedicated optimization process is included that focuses solely on refining the local-to-global transformation while keeping the keyframe poses fixed. This approach offers several advantages:

- Periodic Optimization of the local-to-global Transformation: The transformation refinement optimization is performed periodically within the Loop Closing & TOA-based Global Map Refinement thread, ensuring that the local-to-global transformation is regularly updated and maintained.
- Fixed Keyframe Poses: By keeping the keyframe poses fixed during this optimization, the system can ensure a stable convergence of the transformation estimates, without the additional complexity of jointly optimizing the poses and transformation.
- Enhanced Robustness against IMU Noise and Drift: The dedicated transformation refinement optimization, with fixed keyframe poses, is more robust against the effects of IMU noise and drift. This is particularly important in scenarios where the IMU measurements contain significant errors, as the optimization can focus solely on refining the transformation without being overly influenced by noisy inertial data.

The optimization graph for this transformation refinement process maintains a straightforward structure, incorporating only TOA and odometer edges. Figure 5.5b illustrates this graph structure.

**Scale Refinement for Monocular Systems:** Monocular visual SLAM systems inherently suffer from scale ambiguity. To address this, a dedicated optimization component for global scale factor estimation and update is incorporated into the system.

This scale refinement is performed in addition to the local-to-global transformation while keeping keyframe poses fixed. By simplifying the graph to include only TOA edges, this approach enables global optimization, leveraging information from the entire keyframe map for a more accurate scale estimate. Figure 5.5c illustrates the structure of the scale refinement optimization graph. Key elements include: Keyframe pose nodes, scale factor node (s), TOA biases, and the Transformation node.

Optimizing the scale factor while keeping keyframe poses fixed efficiently estimates the correct global scale without introducing additional uncertainties or instabilities. The estimated scale factor is then propagated to update all keyframe poses and map point positions, ensuring a consistent and accurate global map representation.

The dedicated scale refinement optimization plays a crucial role in addressing the scale ambiguity inherent to monocular SLAM. While one might consider using the local mapping thread for this purpose, such an approach would be suboptimal for two key reasons. First, the optimization needs to handle significantly higher uncertainty levels, which can lead to instability when processed in local mapping. Second, updating both transformation and scale parameters through local mapping is inefficient, as it operates on a limited map section rather than leveraging the global map information available.

From an implementation standpoint, the existing three threads within ORB-SLAM3 are utilized without creating a new one. Specifically, the TOA-related global optimizations are integrated into the Loop Closing thread, which proved to be a natural and effective choice.

The integration of TOA measurements into the Loop Closing thread was selected for several reasons:

1. Efficient Asynchronous Execution without Impacting Real-Time Tracking: By incorporating the TOA-based optimizations into the asynchronous Loop Closing thread, the system ensures that computationally intensive global optimizations are performed without affecting the real-time tracking performance in the Tracking thread. This approach efficiently utilizes system resources, preventing other threads from being burdened and maintaining overall system performance.

2. Natural Integration with Global Map Maintenance and Separation of Concerns: The Loop Closing thread is responsible for global map consistency tasks like loop closure detection and correction. Integrating the TOA-based optimizations into this thread allows us to seamlessly combine global map refinement processes, leveraging existing infrastructure and workflows. This also maintains a clear separation between local and global optimization tasks—the Tracking thread focuses on local pose optimization, and the Local Mapping thread handles local keyframe and map point refinement—ensuring that each thread operates effectively without interference.

### 5.1.4 System Operation with Unknown Base Station Positions

While our primary implementation assumes known base station locations to enable global positioning, we also investigate scenarios where these positions are unknown. In this configuration, we modify our optimization framework by changing the base station nodes from fixed to non-fixed status while maintaining a fixed transformation node (TR). This adjustment effectively eliminates the requirement for prior knowledge of base station locations.

The modified optimization graph structure still incorporates ToA measurements, but instead of using them to align the local frame with a globally defined coordinate system, they now serve primarily to constrain the relative geometry of the trajectory and enhance local consistency. While this configuration precludes absolute global positioning capabilities, it remains valuable for improving SLAM performance, particularly in minimal sensor setups where scale ambiguity and drift represent significant challenges.

## 5.2 Experiment

This section presents a comprehensive evaluation of our proposed methodology. We designed experiments to rigorously assess the performance of our SLAM system across diverse configurations, with particular emphasis on the novel integration of simulated 5G
toa measurements with traditional visual and inertial data sources. Our evaluation employs both the widely-recognized EuRoC MAV dataset as a public benchmark and our custom-

collected Aerolab dataset featuring RGBD-inertial data.

### 5.2.1 Experimental Setup and Data Collection

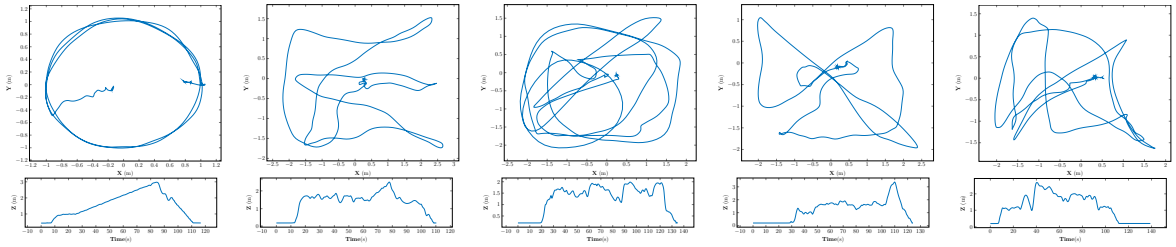
**Aerolab Dataset:** The primary experiments were conducted in a controlled indoor environment, the Aerolab, specifically designed to support drone flights and data collection under realistic conditions for SLAM evaluation. The core equipment included an Intel RealSense D435i camera for RGB-D (color and depth) and IMU data acquisition, and the OptiTrack motion capture system with a 12-camera configuration that provided high-precision ground truth data. This comprehensive setup enabled the acquisition of synchronized multimodal data essential for rigorous SLAM evaluation.

The Intel RealSense D435i camera, mounted on the drone, captured RGB-D data at 30 Hz and IMU data at 200 Hz, with automatic synchronization between the visual and inertial data streams. To support diverse testing scenarios, we collected five unique datasets by flying the drone through varied trajectories that mimic realistic navigation patterns:

- **Dataset 0**: 116 seconds
- **Dataset 1**: 110 seconds
- **Dataset 2**: 135 seconds
- **Dataset 3**: 124 seconds
- **Dataset 4**: 139 seconds

The trajectories for each dataset are visualized in Figure 5.6, illustrating both the x-y trajectories (top row) and the time vs. z plots (bottom row). The drone operated within a netted 5x5x5 meter flight area, ensuring safety while providing sufficient space for complex maneuvers.

**EuRoC MAV Dataset:** In addition to our custom Aerolab dataset, we employed the EuRoC MAV dataset as used in the previous chapter. For our experiments, we focused on:



(a) Dataset0 (b) Dataset1 (c) Dataset2 (d) Dataset3 (e) Dataset4

Figure 5.6: Trajectories for the five Aerolab datasets. Each subfigure contains the x-y trajectory (top) and the time vs. z trajectory (bottom) for a specific dataset.

- **Vicon Room sequences (V101-V103, V201-V202)**: These sequences feature a drone flying in a room equipped with a Vicon motion capture system, providing high-precision ground truth for trajectory evaluation. We augmented these sequences with realistic 5G TOA simulations following the methodology explained in the previous chapter.
- **Machine Hall sequences (MH01-MH05)**: Recorded in a large industrial environment, these sequences present more challenging conditions with varied lighting, longer trajectories, and more complex motion patterns.

The EuRoC dataset allowed us to evaluate our system on standardized benchmarks and compare our results directly with state-of-the-art methods in the literature, particularly for the comparison with UWB-based approaches.

### 5.2.2 Calibration and Data Synchronization

For the Aerolab dataset, comprehensive calibration of the camera and IMU was conducted using the `Kalibr` toolbox and `allan_variance_ros`. The calibration procedure involved collecting data with checkerboard patterns and performing specific drone motions, resulting in intrinsic parameters (focal length, distortion coefficients) and the transformation matrix between the camera and IMU.

Data synchronization was handled differently for each dataset:

- **Aerolab**: The RealSense D435i camera provided synchronized RGB-D and IMU data, which was time-aligned with the OptiTrack ground truth.

- **EuRoC MAV**: This dataset already provides time-synchronized stereo images and IMU data with corresponding ground truth poses.

### 5.2.3 Augmenting Datasets with Simulated 5G TOA Measurements

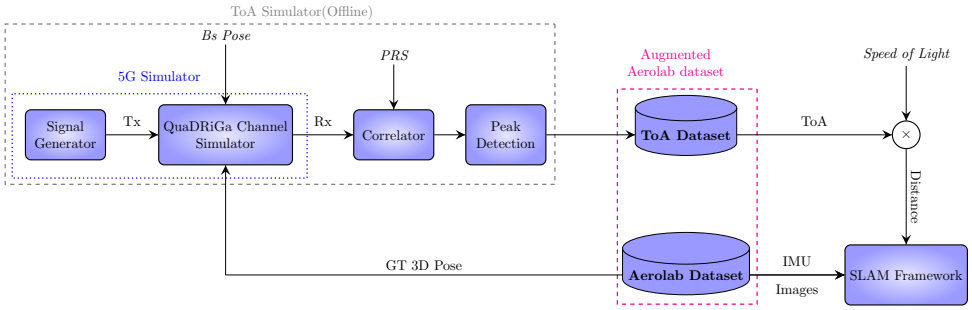



Figure 5.7: Architecture of the dataset generation pipeline, showing the integration of simulated 5G TOA measurements with experimental data. The pipeline consists of a 5G signal simulator using QuaDRiGa channel models, signal processing components, and the final integration with sensor data from both datasets. The combined data serves as input to the SLAM framework, with TOA measurements scaled by the speed of light to obtain distance measurements.

To overcome infrastructure limitations, we developed a high-fidelity simulation pipeline to generate realistic 5G TOA measurements. Figure 5.7 illustrates our end-to-end process that augments both the Aerolab and EuRoC datasets with simulated TOA data. Following the same methodology established in the previous chapter, we used ground truth trajectories from each dataset as inputs to our 5G signal simulator. For signal generation, we maintained the same two 5G FR2 millimeter wave configurations:

- **28 GHz band**: Utilizing 200 MHz bandwidth with 120 kHz subcarrier spacing
- **78 GHz band**: Utilizing 400 MHz bandwidth with 240 kHz subcarrier spacing

Both configurations employed identical transmission parameters (0 dBm transmit power, 10 dB SNR) with PRS generated at 10 Hz. The QuaDRiGa channel simulator modeled realistic wireless propagation effects, and the resulting TOA measurements were synchronized with the visual and inertial data from each dataset. For base station deployment, we implemented multiple configurations:

- **Aerolab Configuration**: Four base stations strategically positioned at: BS1: $(2.5, -2.5, 4.5)$, BS2: $(2.5, 2.5, 4.0)$, BS3: $(-2.5, 2.5, 5.0)$, and BS4: $(-6.5, -2.5, 2.0)$ within the OptiTrack global coordinate frame.
- **EuRoC Vicon Room Configuration**: For these sequences, we utilized the same base station configuration as detailed in the previous chapter.
- **EuRoC Machine Hall Configuration**: To enable direct comparison with recent literature, we implemented:
  - Single base station positioned at [10, 10, 10] matching Li et al. [4] but operating at 5 Hz instead of 30 Hz for practicality
  - Sequential two-base station scenario with BS1 active during 20-55s and BS2 active during 60-100s (Figure 5.10)

### 5.2.4 Evaluation Framework and Scenarios

Our evaluation framework encompasses multiple complementary scenarios designed to comprehensively assess the benefits of 5G ToA integration with traditional SLAM across varying conditions and constraints:

- **Global SLAM with Known Base Stations**: We evaluated the integration of 5G ToA measurements (at both 28 GHz and 78 GHz) with different SLAM configurations (RGB-D, RGB-D Inertial, Monocular, and Stereo) using known base station positions. This scenario examines ToA's ability to enable absolute positioning in global coordinates, a capability traditionally absent in vision-only SLAM.
- **SLAM with Unknown Base Station Locations**: We relaxed operational constraints by estimating both the trajectory and base station positions simultaneously - a more realistic scenario for practical deployments. This included:
  - Reduced number of active base stations
  - Sequential line-of-sight connections where only one base station maintains LOS during specific time intervals (BS1: 10-40s, BS2: 50-70s, BS3: 80-100s)

  - Limited base station coverage spanning only portions of the trajectory

- **ToA as Loop Closure Alternative**: We investigated ToA's potential as a substitute for traditional loop closure, which frequently becomes unavailable in realistic scenarios (linear trajectories, previously unvisited areas, or environments with appearance changes). By manually disabling loop closure in sequences where it contributes significantly to drift correction, we evaluated how ToA integration compensates for this limitation.

- **Comparative Analysis with State-of-the-Art**: Using the EuRoC Machine Hall dataset, we directly compared our approach with UWB-VO and other established methods. While replicating the experimental setup from recent literature, we adjusted parameters to reflect more realistic operating conditions (5 Hz measurement frequency versus 30 Hz), enabling fair yet practical evaluation.

For each scenario, we employed the Absolute Trajectory Error (ATE) metric with two distinct evaluation approaches: (1) Local Evaluation with optimal alignment to assess trajectory consistency, and (2) Global Evaluation without alignment to measure absolute positioning accuracy. For monocular configurations, we additionally reported scale error percentages to quantify the system's ability to recover correct scale information. To ensure robust statistical evaluation, each experiment was conducted five times, and the average is reported as the final value.

## 5.3 Results

### 5.3.1 Global SLAM Performance with 5G TOA Integration

This section presents the experimental results evaluating global SLAM performance with 5G TOA integration at two frequencies: 28 GHz and 78 GHz. We evaluate the system on two distinct datasets:

1. The Aerolab dataset: Performance metrics for RGB-D, RGB-D Inertial, and Monocular SLAM configurations (detailed in Appendix B.2.1, Table B.4).

2. The EuRoC MAV dataset: Stereo and Monocular SLAM performance across five sequences (V101-V103, V201-V202) (detailed in Appendix B.2.2, Table B.5). For each configuration, we report:

- **Local error (meters)**: Trajectory accuracy after post-alignment.
- **Global error (meters)**: Absolute positioning accuracy without post-alignment.
- **For monocular configurations:**
  - **Local error**: After scale correction.
  - **Unscaled local error**: Before scale correction.
  - **Scale error (%)**: Deviation from true scale.
- **Local improvement percentage** compared to baseline.

Results represent averages across five runs for statistical reliability. To ensure statistical reliability, each experiment was conducted five times, and the values presented represent the average ATE across these runs. The tables include baseline measurements (without TOA integration) and results with TOA integration at both frequency bands, along with mode averages for each SLAM configuration.

Before delving into the analysis, it is crucial to understand the difference between local and global evaluations and the implications of including baseline global results. In the **local evaluation**, the transformation (rotation and translation) between the estimated trajectory and the ground truth is assumed to be unknown and is calculated offline after the SLAM process. This alignment is typically performed using methods like Horn's absolute orientation algorithm, which finds the best-fit transformation that minimizes the trajectory error. Essentially, the local evaluation focuses on the consistency of the trajectory shape, disregarding the absolute positioning and orientation in the global frame.

In contrast, the **global evaluation** does not perform any post-processing alignment. The SLAM algorithm must estimate both the trajectory and the transformation directly, operating entirely in the global coordinate frame. This approach is more challenging because it requires

the SLAM system to maintain accurate global positioning and orientation throughout the operation without relying on offline corrections. This capability is demonstrated in the 3D plot shown in Fig. 5.8, which showcases one experiment (Dataset 0). The RGB-D trajectory enhanced with 78 GHz TOA measurements in the global frame closely matches the ground truth, demonstrating how TOA integration enables consistent global positioning.

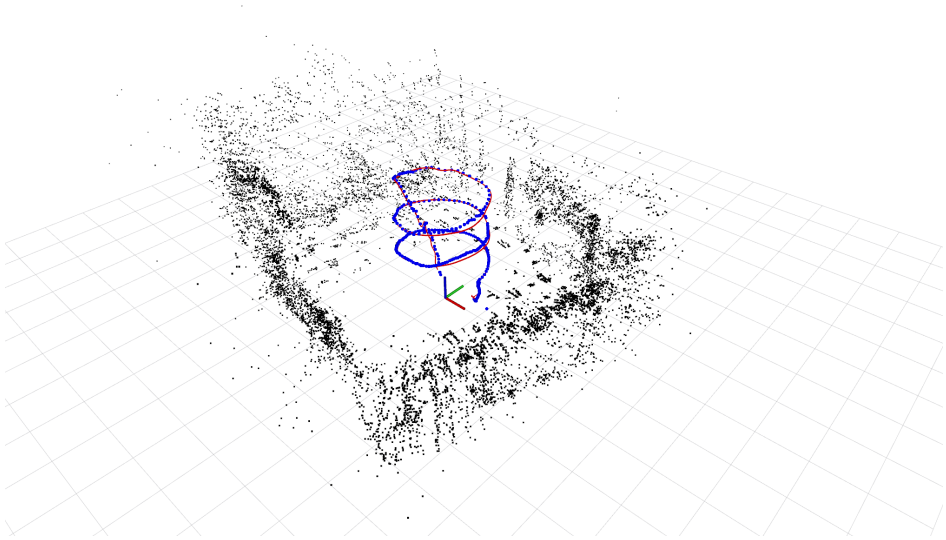


Figure 5.8: RGB-D trajectory and map point estimates in the global frame for Dataset0 with 78 GHz TOA integration. The SLAM results show keyframe pose estimates (blue lines), mapped features (black dots), and the ground truth trajectory (red line). Coordinate axes are color-coded: red for X, green for Y, and blue for Z. The trajectory estimate, enhanced with 78 GHz TOA data, aligns closely with the ground truth in the global frame, demonstrating the system's ability to maintain accurate global positioning without post-processing alignment.

**Analysis of SLAM Modes with TOA Integration:** The integration of Time-of-Arrival (TOA) measurements into various SLAM modes reveals several key insights across both the Aerolab and EuRoC datasets. For RGB-D SLAM, the baseline already achieves good local accuracy (0.115m average), while TOA integration shows inconsistent effects—78 GHz TOA yields a modest improvement of 1.3% in local accuracy, whereas 28 GHz TOA causes a 5.2% degradation. Dataset0 exhibits the most significant improvement (14.1%) with 78 GHz TOA, contrasting with Dataset4's 12.3% degradation using the same configuration.

RGB-D Inertial SLAM follows a similar pattern with more pronounced variations: 78 GHz TOA shows an average local improvement of 5.97%, while 28 GHz TOA offers minimal improvement (0.42%). Dataset2 presents anomalous degradation with both TOA frequencies (-6.9% and -69.4%, excluded from averages), indicating potential trajectory-specific challenges.

In Monocular SLAM, where scale estimation is inherently problematic, 78 GHz TOA integration provides a substantial 11.4% improvement in local accuracy while maintaining excellent scale estimation (0.156% average error). Meanwhile, 28 GHz TOA integration results in a 7.8% degradation despite reasonable scale estimation (0.300% error). On the EuRoC dataset, both Stereo (8.2% improvement with 78 GHz) and Monocular modes (3.1% improvement with 78 GHz) demonstrate consistent benefits from 78 GHz TOA integration.

The most significant advantage of TOA integration lies in enabling global positioning across all SLAM modes—a capability absent in baseline configurations. The 78 GHz TOA consistently outperforms 28 GHz TOA in global accuracy: RGB-D (0.133m vs. 0.198m), RGB-D Inertial (0.262m vs. 0.280m), and Monocular (0.126m vs. 0.267m) on the Aerolab dataset. This performance disparity likely stems from the higher precision of 78 GHz measurements.

Environmental factors and trajectory characteristics significantly influence TOA integration performance, as evidenced by the inconsistent results across datasets. Dataset2 particularly demonstrates how specific conditions can challenge TOA integration across multiple SLAM modes. Despite these variations, the addition of global positioning capabilities represents a substantial enhancement for applications requiring absolute positioning, even when local accuracy improvements remain marginal or inconsistent.

### 5.3.2 SLAM Performance with Unknown Base Station Configurations

To evaluate SLAM performance with unknown base station locations, we modified our system architecture by changing the base station nodes from fixed to non-fixed status while maintaining a fixed transformation node. This configuration eliminates the requirement for prior knowledge of base station locations, though it does preclude global positioning capabilities. We further relaxed operational conditions by:

- Reducing the number of active base stations
- Implementing sequential line-of-sight (LOS) connections where only one base station maintains LOS during specific, non-overlapping time intervals
- Limiting base station coverage to partial trajectory segments rather than the complete path

We evaluated these scenarios across both minimal and rich sensor configurations under 78 GHz 5G network. For minimal sensor setups, we focused on monocular configurations using both the Aerolab and EuRoC MAV datasets, as TOA integration typically shows the most pronounced effects with limited sensor data. For rich sensor configurations (RGB-D and stereo), we investigated TOA's potential as an alternative to loop closure. Loop closure may be unavailable in many practical scenarios, such as:

- Linear trajectories without path intersections
- Exploration of previously unvisited areas
- Time-critical operations that prevent revisiting locations
- Environments with significant appearance changes
- Resource-constrained systems where loop closure detection is computationally prohibitive

While ideally we would test on extended trajectories without natural loop closures, our available datasets had limited duration. Instead, we simulated loop closure unavailability by manually disabling it in sequences where it contributes to drift correction (Aerolab Dataset0, V102,V103, V202, and V203), then evaluated how TOA integration could compensate for this limitation.

For comparison with recent literature, we also replicated the experimental setup from [4] using the EuRoC Machine Hall dataset. While their work used simulated TOA measurements from one base station at 30 Hz ($\sigma = 5$ cm Gaussian noise, base station at [10,10,10]), we

modified the scenario to operate at a more practical 5 Hz measurement frequency. This allows direct comparison while reflecting more realistic operational constraints.

**Sequential Unknown Base Station Operation (Monocular):**

Table 5.1: MONO SLAM with Sequential Unknown Base Stations

| Sequence | Configuration | Local (m) | Local (Unscaled) (m) | Scale | Local Imp. |
|---|---|---|---|---|---|
| V101 | Baseline | 0.0950 | N/A | N/A | – |
| | Sequential BS | 0.0944 | 0.1222 | 3.81% | 0.63% |
| V102 | Baseline | 0.1004 | N/A | N/A | – |
| | Sequential BS | 0.1001 | 0.1020 | 0.71% | 0.30% |
| V103 | Baseline | 0.1160 | N/A | N/A | – |
| | Sequential BS | 0.1068 | 0.1144 | 1.63% | 7.93% |
| V201 | Baseline | 0.0860 | N/A | N/A | – |
| | Sequential BS | 0.0848 | 0.0924 | 1.33% | 1.39% |
| V202 | Baseline | 0.1760 | N/A | N/A | – |
| | Sequential BS | 0.1748 | 0.1803 | 0.38% | 0.68% |
| Dataset1 | Baseline | 0.0550 | N/A | N/A | – |
| | Sequential BS | 0.0534 | 0.0709 | 0.90% | 2.90% |
| Dataset3 | Baseline | 0.0740 | N/A | N/A | – |
| | Sequential BS | 0.0612 | 0.0714 | 0.31% | 17.29% |
| **Average** | Baseline | **0.1003** | N/A | N/A | – |
| | Sequential BS | **0.0965** | **0.1077** | **1.30%** | **4.40%** |

Sequential BS: Three unknown base stations operating in different time intervals (BS1: 10-40s, BS2: 50-70s, BS3: 80-100s).
Local: Post-scale alignment error; Local (Unscaled): Pre-scale alignment error.
Scale: Absolute percentage error in scale estimation.
Improvement calculated relative to baseline performance.
Note: V203, Dataset0, Dataset2, and Dataset4 lose tracking before 100 seconds, while the sequential BS operates up to 100 seconds.

Table 5.1 presents results for scenarios where three base stations operate in different time intervals (BS1: 10-40s, BS2: 50-70s, BS3: 80-100s) with unknown locations at frequency 78 GHz as illustrated in Figure 5.9. For the monocular configuration, results demonstrate consistent improvement across datasets despite the challenging sequential operation scenario. We observed an average improvement of 4.40% in local accuracy across all sequences, with particularly strong improvements in Dataset3 (17.29%) and V103 (7.93%). The system maintained reliable scale estimation with an average error of 1.30% across all datasets. These improvements are especially noteworthy considering the lack of continuous base station coverage, suggesting that even intermittent TOA measurements can significantly enhance SLAM performance in minimal sensor configurations.

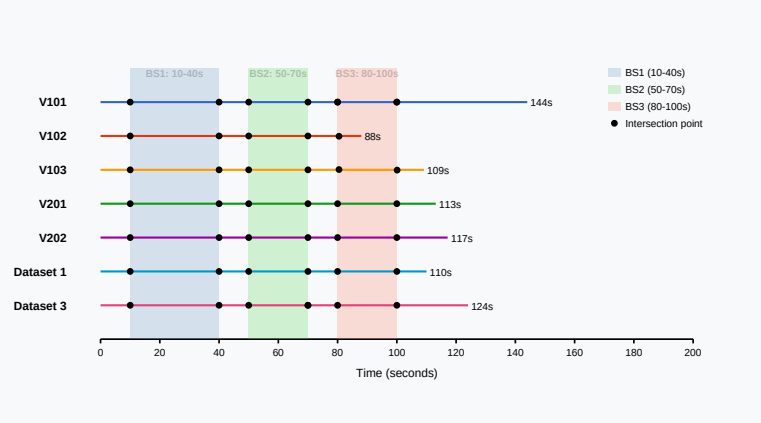



Figure 5.9: Bastion Station Activity Over Trajectories: This plot illustrates the active periods of three bastion stations (BS1: 10–40s, BS2: 50–70s, BS3: 80–100s) over the trajectories of various datasets (V101, V102, V103, V201, V202, Dataset 1, and Dataset 3). Each horizontal bar represents the total duration of a dataset, with colored segments highlighting the intervals during which each bastion station is active

**TOA Integration as an Alternative to Loop Closure:** In typical SLAM implementations, loop closure serves as a critical mechanism for drift correction by recognizing previously visited locations. However, this capability faces limitations in many real-world scenarios. Our investigation focuses on whether TOA measurements can provide a viable alternative when traditional loop closure is unavailable or unreliable. This study is conducted using RGBD for the Aerolab dataset and Stereo for the Euroc Mav dataset.

Table 5.2 presents a comprehensive comparison between baseline SLAM with loop closure enabled, the same system with loop closure disabled, and two TOA integration configurations using unknown base station locations: sequential operation of three base stations and continuous operation of three base stations. The results reveal a compelling case for TOA integration as a loop closure alternative.

When loop closure is disabled, we observe significant degradation in localization accuracy across most sequences, with the average ATE increasing from 0.294m to 0.395m (a 34.35% degradation). TOA integration effectively mitigates this performance loss, with three sequential base stations improving accuracy by 18.6% compared to the no-loop-closure baseline. The improvement becomes even more pronounced with continuous TOA measurements from

Table 5.2: Effect of TOA on Local Accuracy (SE3 Post-Alignment)

| Dataset | Baseline (LC) | Baseline (No LC) | TOA (Seq) | TOA (Cont) | Improvement (Seq / Cont) (%) |
|---|---|---|---|---|---|
| Dataset0 | 0.064 | 0.153 | 0.138 | 0.129 | 9.8% / 15.7% |
| V102 | 0.099 | 0.272 | 0.205 | 0.142 | 24.6% / 47.8% |
| V103 | 0.161 | 0.507 | 0.299 | 0.257 | 41.0% / 49.3% |
| V202 | 0.190 | 0.213 | 0.230 | 0.208 | -8.0% / 2.3% |
| V203 | 0.960 | 0.832 | 0.621 | 0.557 | 25.4% / 33.1% |
| **Average** | **0.294** | **0.395** | **0.299** | **0.259** | **18.6% / 29.6%** |

Baseline (LC): Baseline results with loop closure enabled.

Baseline (No LC): Baseline results with loop closure manually deactivated.

TOA (Seq): Results with 3 sequential base stations (freqseq).

TOA (Cont): Results with 3 continuous base stations (freq78).

Improvement (%) is calculated relative to the baseline without loop closure.

V203 shows higher error due to tracking losses and relocalization, included to showcase system behavior under challenging conditions.

three base stations, achieving a 29.6% enhancement and reducing the average error from 0.395m to 0.259m.

The improvements are particularly striking in challenging sequences where loop closure is typically critical for maintaining baseline performance. For example, in sequence V104, continuous TOA measurements deliver a remarkable 49.3% improvement in accuracy over the no-loop-closure baseline, while sequence V203 shows a 33.1% improvement. These significant gains demonstrate that TOA integration can effectively replace loop closure in preserving trajectory consistency, even without prior knowledge of base station locations. Interestingly, in sequence V203, the baseline without loop closure achieves a slightly better performance (0.832 meters) compared to the baseline with loop closure (0.960 meters). This counterintuitive result may be attributed to suboptimal pose graph optimizations, particularly in this challenging sequence, which involves tracking losses and relocalization events. A similar pattern is observed in dataset V202, where the baseline with loop closure (0.190 meters) marginally outperforms the baseline without loop closure (0.213 meters). In this case, using Time of Arrival (TOA) as a substitute for loop closure provides only minimal improvements or even degradation.

**[:** Comparison with UWB-VO and Other State-of-the-Art]Comparison with UWB-VO and Other State-of-the-Art Methods

Table 5.3 presents a detailed comparison, introducing both Local (scaled) and Unscaled

Table 5.3: Comparison of SLAM Accuracy and Scale Consistency on EuRoC MAV Machine Hall Sequences

| Dataset | UWB-VO | VINS-mono | OKVINS | ROVIO | Ours (1BS) | | Ours (2 Seq BS) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Local | Unscaled | Local | Unscaled |
| MH01 | 0.120 | 0.270 | 0.160 | 0.210 | **0.102** *(1.26%)* | 0.116 | 0.103 *(1.84%)* | 0.129 |
| MH02 | 0.130 | 0.150 | 0.220 | 0.250 | **0.119** *(0.26%)* | 0.120 | 0.120 *(0.66%)* | 0.125 |
| MH03 | 0.280 | **0.130** | 0.240 | 0.250 | 0.208 *(1.20%)* | 0.215 | 0.208 *(0.31%)* | 0.212 |
| MH04 | 0.360 | **0.250** | 0.340 | 0.490 | 0.289 *(2.53%)* | 0.349 | 0.286 *(1.35%)* | 0.317 |
| MH05 | 0.300 | 0.350 | 0.470 | 0.520 | 0.300 *(3.93%)* | 0.460 | **0.294** *(1.82%)* | 0.375 |
| **Mean** | 0.238 | 0.230 | 0.286 | 0.344 | 0.204 *(1.84%)* | 0.252 | **0.202** *(1.20%)* | 0.232 |

1BS: Single base station at [10,10,10] with 5Hz measurements
2BS: Two sequential base stations active during 20-55s and 60-100s respectively
Local: Results with scale correction applied
Unscaled: Results without scale correction
Bold values indicate best RMSE performance for each sequence
Values in italics show scale errors for our methods
UWB-VO reported scaling errors < 1% but did not provide detailed results

RMSE values. The Local results reflect the accuracy with scale correction applied (Sim(3) alignment), while the Unscaled results show raw SE(3) alignment-based errors, making direct scale comparisons possible.

The UWB-VO approach, proposed by Li et al., is a tightly coupled SLAM framework that integrates monocular visual odometry with UWB Time-of-Arrival (TOA) constraints to address scale ambiguity and reduce drift. This method employs joint optimization, where UWB distance residuals are fused with visual reprojection errors during both tracking and local mapping. The system assumes the presence of a single static UWB anchor and was evaluated using the Machine Hall dataset from the EuRoC MAV benchmark. Notably, their experiments rely on simulated UWB TOA data, generated with a Gaussian distribution at a standard deviation of 5 cm and a frequency of 30Hz.

Unlike UWB-VO, which relies on a single simulated UWB anchor at 30 Hz, our approach evaluates both single and sequential base station setups (Fig 5.10) at a lower, more practical 5 Hz measurement rate. The sequential two-base-station approach improves performance in challenging sequences (MH04, MH05).

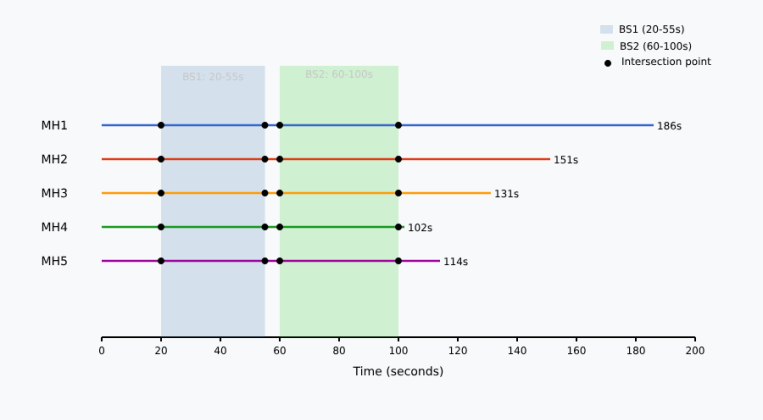



Figure 5.10: Bastion Station Activity Over Trajectories (Euroc MAV Machine Hall): This plot illustrates the active periods of two bastion stations (BS1: 20-55s, BS2: 60-100s) over the trajectories of Machine Hall (MH01-MH05). Each horizontal bar represents the total duration of a dataset, with colored segments highlighting the intervals during which each bastion station is active

Our implementation, like UWB-VO, is based on ROS (Robot Operating System), ensuring

consistency in the software framework. However, while our results were obtained through direct experiments on our system, the results for UWB-VO, VINS-mono, OKVINS, and ROVIO are taken from their originally reported implementations as published in their respective works.

Despite using a lower TOA measurement frequency, our method achieves a comparable 0.202m mean RMSE. Even in Unscaled form, the RMSE remains competitive, demonstrating strong inherent scale estimation without external correction. In contrast, UWB-VO reports <1% scale error but does not provide per-sequence details, making direct comparisons difficult.

These results highlight that our sequential base station method and lower measurement rate still yield robust performance, outperforming several established methods while operating under more relaxed and realistic constraints.

## 5.4 Analysis of Base Station Configurations for SLAM Performance

The previous Chapter examined the impact of base station configurations on localization accuracy. This section extends this investigation to assess how various base station geometries influence SLAM performance across multiple datasets when fusing TOA measurements with visual data.

### 5.4.1 Experimental Setup

The experiments were conducted using the same five base station configurations introduced in the previous section (refer to Figure 4.7 for visual representations and Table 4.4 for their exact coordinates):

- **Tetrahedral**: A 3D distributed arrangement with one base station at the ceiling center and three base stations forming a triangle on the ground.
- **Z-Shape**: Base stations arranged in diagonally alternating heights.

- **Asymmetric**: A realistic irregular placement of base stations.
- **Diamond**: Base stations evenly distributed at two different heights.
- **Clustered**: A co-planar arrangement with all base stations mounted on a single wall.

These configurations were evaluated across both EuRoC MAV and Aerolab Dataset for Stereo and RGB-D respectively.

### 5.4.2 Results on EuRoC MAV Dataset

Table 5.4 presents the average ATE values for both local and global accuracy across all configurations and EuRoC MAV dataset sequences.

Table 5.4: ATE for different base station configurations across EuRoC MAV datasets (meters). Each cell contains local accuracy / global accuracy values.

| Configuration | V101 | V102 | V103 | V201 | V202 | Average |
|---|---|---|---|---|---|---|
| Tetrahedral | 0.078/0.083 | 0.124/0.135 | 0.140/0.158 | 0.124/0.158 | 0.184/0.189 | 0.130/0.145 |
| Z-Shape | 0.079/0.082 | 0.108/0.129 | 0.177/0.203 | 0.087/0.135 | 0.188/0.199 | 0.128/0.150 |
| Diamond | 0.078/0.083 | 0.186/0.286 | 0.148/0.162 | 0.084/0.118 | 0.177/0.188 | 0.135/0.167 |
| Asymmetric | 0.079/0.083 | 0.105/0.118 | 0.186/0.239 | 0.131/0.207 | 0.183/0.203 | 0.137/0.170 |
| Clustered | 0.084/0.095 | 0.137/0.155 | 0.223/0.628 | 0.096/0.245 | 0.189/0.204 | 0.146/0.265 |

The results demonstrate the impact of the base station geometry on both local and global SLAM accuracy, with the Z-Shape configuration showing the best average local accuracy (0.128m) and the Tetrahedral configuration achieving the best average global accuracy (0.145m). The Clustered configuration showed the poorest performance in both metrics, with a particularly significant degradation in global accuracy (0.265m, 82.8% worse than the Tetrahedral configuration).

### 5.4.3 Results on AeroLab Dataset

The same evaluation of different base station configurations was conducted using the AeroLab dataset. Table 5.5 presents the average ATE values.

The AeroLab results show almost the same pattern as the EuRoC dataset. The Z-Shape configuration achieved the best average local accuracy (0.113m), closely followed by the

Table 5.5: ATE for different base station configurations with 5G TOA integrated RGB-D across AeroLab datasets (meters). Each cell contains local accuracy / global accuracy values.

| Configuration | Dataset 0 | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Average |
|---|---|---|---|---|---|---|
| Tetrahedral | 0.063/0.092 | 0.110/0.115 | 0.143/0.150 | 0.092/0.095 | 0.165/0.166 | 0.115/0.124 |
| Z-Shape | 0.062/0.095 | 0.107/0.125 | 0.146/0.155 | 0.095/0.105 | 0.157/0.165 | 0.113/0.129 |
| Diamond | 0.062/0.097 | 0.110/0.138 | 0.143/0.152 | 0.097/0.117 | 0.165/0.172 | 0.115/0.135 |
| Asymmetric | 0.068/0.112 | 0.109/0.142 | 0.155/0.160 | 0.099/0.103 | 0.172/0.176 | 0.121/0.138 |
| Clustered | 0.069/0.131 | 0.113/0.144 | 0.148/0.159 | 0.091/0.096 | 0.176/0.184 | 0.119/0.143 |

Tetrahedral and Diamond configurations (both 0.115m). For global accuracy, the Tetrahedral configuration maintained its advantage (0.124m), with the Z-Shape configuration a close second (0.129m). The Clustered configuration again showed the poorest performance overall, particularly in global accuracy (0.143m).

### 5.4.4 Cross-Dataset Analysis

Comparing the results across both datasets reveals several consistent patterns and some interesting differences, as shown in Table 5.6.

Table 5.6: Combined average RMSE across both datasets (meters)

| Configuration | EuRoC Local/Global | AeroLab Local/Global | Combined Local/Global |
|---|---|---|---|
| Tetrahedral | 0.130/0.145 | 0.115/0.124 | 0.123/0.135 |
| Z-Shape | 0.128/0.150 | 0.113/0.129 | 0.121/0.140 |
| Diamond | 0.135/0.167 | 0.115/0.135 | 0.125/0.151 |
| Asymmetric | 0.137/0.170 | 0.121/0.138 | 0.129/0.154 |
| Clustered | 0.146/0.265 | 0.119/0.143 | 0.133/0.204 |

The combined results highlight several key insights:

- The Z-Shape configuration shows the best overall local accuracy (0.121m), with the Tetrahedral configuration a close second (0.123m).
- The Tetrahedral configuration delivers the best overall global accuracy (0.135m), outperforming other configurations by at least 3.7%.
- The Clustered configuration performs consistently worse across both datasets, with significantly degraded global accuracy (0.204m, 51.1% worse than Tetrahedral).

- The performance gap between configurations is more pronounced in the EuRoC dataset than in the AeroLab dataset, suggesting that environment complexity and motion dynamics play important roles in determining the impact of base station geometry.

### 5.4.5 Global-to-Local Accuracy Ratio

The ratio between global and local accuracy provides insight into how well each configuration maintains global consistency. Table 5.7 presents these ratios for both datasets.

Table 5.7: Global-to-local accuracy ratio for different base station configurations

| Configuration | EuRoC Ratio | AeroLab Ratio | Combined Ratio |
|---|---|---|---|
| Tetrahedral | 1.12 | 1.08 | 1.10 |
| Z-Shape | 1.17 | 1.14 | 1.16 |
| Diamond | 1.24 | 1.17 | 1.21 |
| Asymmetric | 1.24 | 1.14 | 1.19 |
| Clustered | 1.82 | 1.20 | 1.53 |

The Tetrahedral configuration shows the lowest ratios across both datasets (1.10 combined), indicating that its local trajectory accuracy translates well to global positioning. The Clustered configuration has the highest ratio (1.53 combined), demonstrating significant degradation when moving from local to global accuracy assessment. This degradation is particularly severe in the EuRoC dataset (ratio of 1.82), highlighting the vulnerability of the Clustered configuration to challenging conditions.

### 5.4.6 Relationship Between GDOP and SLAM Performance

Referencing the GDOP analysis from the previous section (Figure 4.8), a strong correlation between GDOP values and SLAM performance across both datasets is observed. The Tetrahedral configuration's superior global accuracy directly corresponds to its lowest average GDOP (2.58) in the EuRoC dataset.

Similarly, as shown in Figure 5.11, the AeroLab dataset exhibits the same pattern, with the Tetrahedral configuration having the lowest average GDOP (2.06), followed by Z-Shape (3.21), Asymmetric (3.66), Diamond (11.36), and Clustered (13.63).

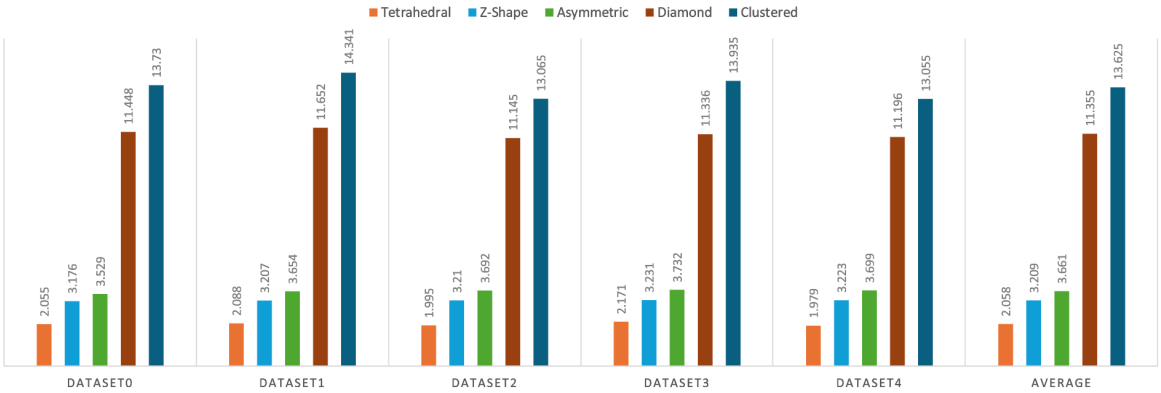



Figure 5.11: Comparison of average Geometric Dilution of Precision (GDOP) values across different base station configurations for all AeroLab dataset sequences. Lower GDOP values indicate better geometric configuration for localization.

This relationship confirms that geometric principles applied to localization accuracy extend to SLAM performance, particularly for global accuracy. However, the fusion with visual data helps mitigate some of the geometric limitations, especially for local trajectory accuracy, resulting in smaller percentage differences between configurations compared to pure localization.

### 5.4.7 Analysis of Unkown Base Station position for Scale Ambiguity Resolution

To evaluate the impact of base station positioning on scale estimation and 5G-enhanced monocular SLAM, experiments were conducted using the MH03 sequence from the EuRoC MAV dataset. The primary objective was to assess whether the placement of a single base station with unknown position could effectively resolve scale ambiguity in monocular SLAM and to analyze how the base station's position influences estimation accuracy.

The experiment involved testing over 150 different base station configurations systematically. For each configuration, the base station was positioned at random coordinates:

- X-coordinate: uniformly sampled at random from the interval $[-5, 15]$ meters
- Y-coordinate: uniformly sampled at random from the interval $[-5, 10]$ meters

- Z-coordinate: fixed at 10 meters for all configurations

This approach allowed sampling a broad area around the trajectory, simulating different possible mounting locations for a base station within an indoor environment. By keeping the Z-coordinate fixed, realistic ceiling-mounted or overhead positioning scenarios that would be practical in real-world applications were simulated.

The raw data contained 150 complete records of experimental results. Before conducting the analysis, a robust outlier removal procedure was applied:

1. For each metric (RMSE, RMSE_GT, and Scale), we calculated the interquartile range (IQR):
    - First quartile ($Q_1$): representing the 25th percentile
    - Third quartile ($Q_3$): representing the 75th percentile
    - IQR = $Q_3 - Q_1$
2. We established boundary thresholds for outlier identification:
    - Lower bound = $Q_1 - 1.5 \times$ IQR
    - Upper bound = $Q_3 + 1.5 \times$ IQR
3. Data points falling outside these boundaries were considered outliers and excluded from the statistical analysis to prevent skewing of the results.

This procedure removed approximately 15% of data points for Local ATE, 16% for Local ATE (unscaled), and 6% for Scale values, ensuring that our analysis focused on the representative behavior of the system.

The results are shown in Figure 5.12 and Figure 5.13.

The analysis of the processed data revealed several key findings:

**Scale Ambiguity Resolution:** The results demonstrate that a single strategically placed base station can effectively resolve scale ambiguity in monocular SLAM. The scale factor distribution after outlier removal had:

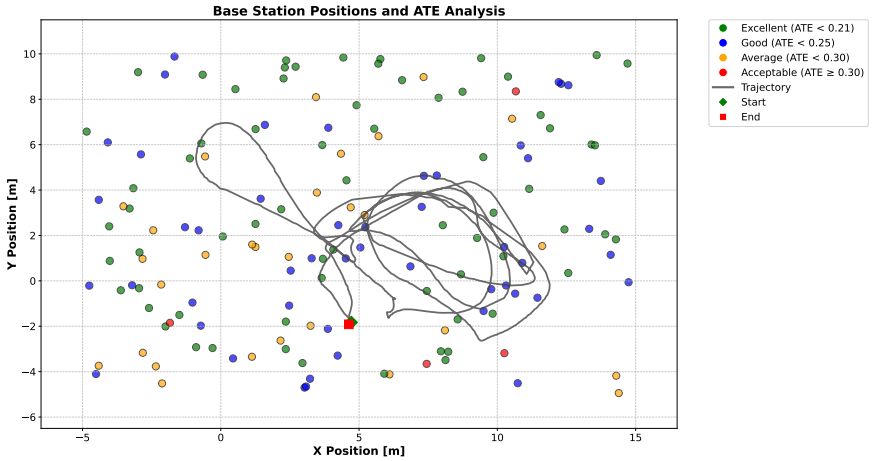



Figure 5.12: Distribution of base station positions and their corresponding ATE performance. The gray line represents the trajectory with a green diamond marking the start point and a red square marking the end point. Base stations are color-coded by performance: green (Excellent, $ATE < 0.21m$), blue (Good, $ATE < 0.25m$), orange (Average, $ATE < 0.30m$), and red (Acceptable, $ATE \geq 0.30m$). The distribution shows that 45.3% achieved excellent performance, 32.7% good performance, 19.3% average performance, and only 2.7% acceptable performance.

- Mean value: $1.0044 \pm 0.0261$
- Median value: $1.0002$
- Range: $0.9189$ to $1.1904$

The proximity of the mean and median values to 1.0 indicates that the scale estimation was highly accurate across most base station configurations. This suggests that the system can recover the true scale of the environment regardless of the specific base station position, within the tested range.

**Error Analysis:** The Local ATE and Local ATE unscaled distributions showed consistent performance too. After outlier removal:

- Mean RMSE: $0.2174 \pm 0.0129$ meters
- Mean RMSE_GT: $0.2124 \pm 0.0120$ meters

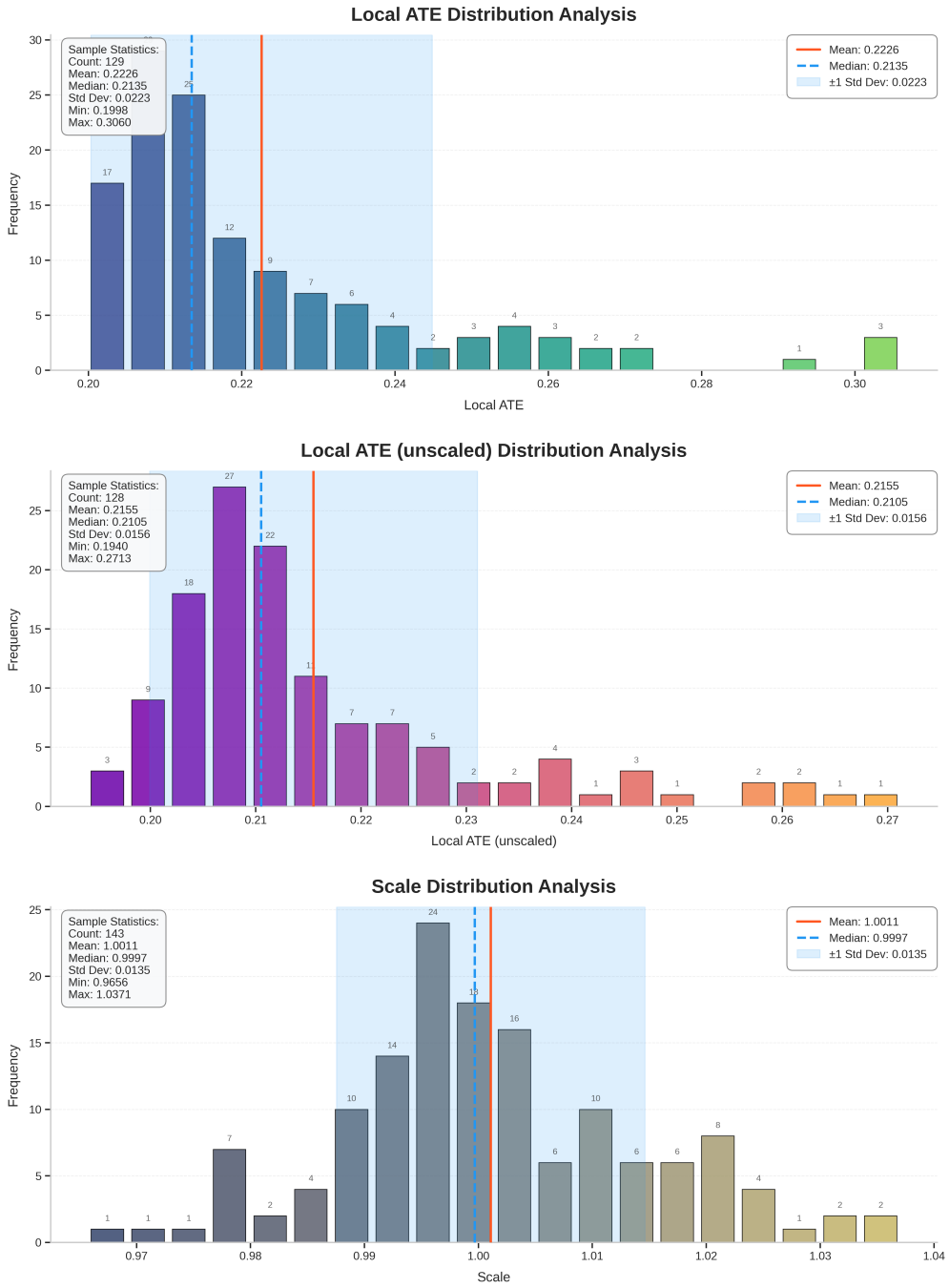



Figure 5.13: Bar charts showing the frequency distributions of Local Accuracy, Local accuracy (unscaled), and Scale after outlier removal. The dashed red line represents the mean value.

The consistent sub-25cm error across most configurations demonstrates that the system provides reliable trajectory estimation regardless of the specific base station placement. This robustness to base station positioning is a significant advantage for practical deployments.

**Effect of Base Station Positioning:** Remarkably, analysis revealed that the system performed well across almost all tested base station positions. While some configurations yielded marginally better results than others, the overall performance remained consistent throughout the sampled space.

## 5.5 System Limitations and Future Enhancements

Despite the advances made in implementing TOA-enhanced SLAM with relaxed connectivity and unknown base station positions, several important limitations remain that affect the system's real-world applicability. While we have successfully demonstrated operation with unknown base station positions and sequential connectivity patterns, there remains an inherent trade-off between this flexibility and global positioning capabilities. Our implementation with non-fixed base station nodes eliminates the requirement for infrastructure surveying but consequently removes global localization capabilities. This represents a fundamental constraint rather than a framework limitation—global SLAM inherently requires some fixed and known reference points to establish absolute positioning.

A significant limitation of our current implementation involves TOA measurement fidelity in complex environments. The simulation framework assumes ideal Line-of-Sight (LoS) conditions, while real-world Non-Line-of-Sight (NLoS) scenarios, caused by obstacles and reflections, can significantly degrade TOA measurement accuracy. NLoS conditions introduce positive biases and increased variance in TOA measurements, potentially compromising the improvements demonstrated in our controlled evaluations.

This challenge is compounded by our use of static information matrices for both TOA and visual-inertial measurements, which fail to capture the dynamic nature of measurement uncertainties in real environments. An adaptive uncertainty model that could identify and appropriately weigh measurements based on environmental conditions would enhance

robustness but requires sophisticated signal quality estimation capabilities not present in our current implementation. Moreover, while simulation using the QuaDRiGa framework provides a controlled testing environment, it may not fully capture the complexities of real-world signal propagation, hardware noise, and environmental dynamics. The simulated measurements at 28 GHz and 78 GHz frequencies reflect theoretical capabilities rather than measurements from physical hardware, which would introduce additional complexities and error sources.

These limitations highlight the need for future validation with experimental TOA measurements from actual 5G infrastructure and the development of more sophisticated uncertainty models that can adapt to changing environmental conditions. Despite these constraints, the demonstrated improvements in both unknown base station scenarios and sequential connectivity patterns represent significant advances toward practical deployment in real-world settings with imperfect infrastructure knowledge and connectivity.

## 5.6 Conclusion

This chapter presented the integration of 5G TOA measurements into the ORB-SLAM3 visual-inertial SLAM framework, successfully enhancing both global and local localization capabilities across various operational scenarios. By incorporating TOA measurements as additional constraints, our system addresses fundamental challenges in visual SLAM while demonstrating effectiveness under relaxed conditions previously considered impractical. Our comprehensive evaluation across the AeroLab and EuRoC MAV datasets revealed several key findings. The most significant improvements were observed in minimal sensor configurations, particularly Monocular SLAM, where 78 GHz TOA integration provided an 11.4% improvement in local accuracy while maintaining excellent scale estimation (0.156% average error). Even more compelling was the system's ability to function effectively with unknown base station positions and sequential connectivity patterns, achieving an average improvement of 4.40% in local accuracy with reliable scale estimation (1.30% average error) despite intermittent coverage. A particularly valuable contribution is the demonstration of TOA measurements as an effective alternative to loop closure. In scenarios where loop

closure was disabled, TOA integration from three sequential base stations improved accuracy by 18.6% compared to the no-loop-closure baseline, with continuous TOA measurements achieving even greater enhancements (29.6%). This capability addresses critical limitations in real-world applications involving linear trajectories, exploration of new areas, or resource-constrained platforms. Our comparative analysis with state-of-the-art approaches like UWB-VO confirmed the robustness of our method. Despite using a lower TOA measurement frequency (5 Hz versus 30 Hz in UWB-VO), our approach achieved comparable performance (0.202m mean RMSE) while operating under more relaxed and realistic constraints, including unknown base station positions and sequential connectivity. Across frequency bands, 78 GHz TOA consistently outperformed 28 GHz in both local and global accuracy. For RGB-D SLAM, 78 GHz TOA yielded a modest 1.3% improvement in local accuracy (compared to 5.2% degradation with 28 GHz) while enabling global positioning with 0.133m accuracy (versus 0.198m for 28 GHz). These patterns held across most configurations, though with varying degrees of advantage. Despite these advances, important limitations remain. The current implementation assumes ideal Line-of-Sight conditions and uses static information matrices that don't fully capture the dynamic nature of measurement uncertainties in complex environments. Additionally, while simulation provides a controlled testing environment, it may not completely represent the complexities of real-world signal propagation and hardware performance. In conclusion, this research demonstrates the significant potential of integrating 5G TOA measurements with visual-inertial SLAM frameworks, particularly with our novel contributions in relaxed connectivity requirements and unknown base station positioning. These advances substantially increase the system's applicability to real-world deployment scenarios where infrastructure knowledge may be limited or connectivity intermittent, while maintaining robust performance that challenges or exceeds state-of-the-art alternatives.

# Chapter 6

# Conclusion

This thesis has explored the integration of 5G-based data with advanced sensor fusion techniques to enhance Unmammed Aerial Vehicles (UAV) localization and SLAM in indoor environments. By addressing the limitations of traditional localization methods, the research demonstrated how 5G Time of Arrival (TOA) measurements can be effectively fused with inertial and visual data to achieve precise and robust localization.

## 6.1 Key Findings and Results

The experimental results highlight several key advancements in both pose estimation and SLAM integration. The PGO-based approach, which consistently outperformed ESKF, achieved a superior accuracy of 13 cm with five base stations compared to ESKF's 34 cm. Even with minimal infrastructure, both methods maintained sub-meter accuracy with just three base stations, demonstrating their robustness in resource-limited scenarios. Furthermore, real-time implementation was successfully validated, with ESKF offering exceptional computational efficiency.

A significant breakthrough was achieved in implementing SLAM with unknown base station positions and relaxed connectivity conditions. The system demonstrated reliable performance with sequential base station connections, where only one base station maintained line-of-sight during specific, non-overlapping time intervals (BS1: 10-40s, BS2: 50-70s, BS3: 80-100s).

These relaxed conditions showed consistent improvement across datasets, with an average improvement of 4.40% in local accuracy for monocular configurations while maintaining reliable scale estimation (average error of 1.30%).

Perhaps most notably, the research established TOA measurements as an effective alternative to loop closure in rich sensor configurations. When loop closure was disabled (simulating scenarios like linear trajectories or exploration of new areas), TOA integration from three sequential base stations improved accuracy by 18.6% compared to the no-loop-closure baseline, with continuous TOA measurements achieving even greater enhancements (29.6%). Comparative analysis with state-of-the-art approaches, including UWB-VO implementations, confirmed the robustness of our method. Despite using a lower TOA measurement frequency (5 Hz versus 30 Hz in UWB-VO), our approach achieved comparable performance (0.202m mean RMSE) while operating under more relaxed and realistic constraints.

The research further confirmed the advantages of higher frequency bands, with 78 GHz configurations demonstrating superior performance in most cases. Despite the increased complexity, real-time performance was maintained across all configurations, validating the framework's applicability in practical scenarios. These findings emphasize the potential of combining 5G TOA measurements with traditional visual-inertial data to enhance localization and mapping capabilities in challenging indoor environments, even with limited infrastructure knowledge and intermittent connectivity.

## 6.2 Research Impact and Significance

The research presented in this thesis has made several significant contributions to the field of UAV navigation and localization:

### Technical Advancement

- Development of a novel global SLAM framework through innovative deep integration of 5G TOA measurements with visual-inertial data, enabling consistent localization across extended trajectories

- Implementation of diverse state-of-the-art techniques for TOA integration, including an efficient ESKF framework and a robust PGO approach, demonstrating flexibility across different computational requirements
- Achievement of centimeter-level accuracy in challenging indoor environments without requiring extensive infrastructure through careful fusion of TOA constraints
- Demonstration of reliable operation with sequential base station connections with unknown positions, where only one base station maintains line-of-sight during specific non-overlapping time intervals
- Development of TOA integration as an effective alternative to loop closure, improving accuracy by up to 29.6% in scenarios where traditional loop closure is unavailable

### Practical Implications

- Comprehensive evaluation of 5G TOA performance across diverse scenarios, including varying base station densities, network configurations, connectivity patterns, and sensor setups, establishing clear understanding of capabilities and limitations
- Relaxation of traditional operational constraints, particularly through sequential base station connectivity and unknown base station positioning, significantly increasing the system's applicability to real-world deployment scenarios
- Provision of viable solutions for environments where loop closure may be unavailable due to linear trajectories, exploration of new areas, or resource constraints
- The systematic evaluation across diverse setups provides clear guidelines for practical implementation choices based on specific application requirements
- The demonstrated real-time performance and robustness across different configurations establish a foundation for practical deployment in real-world applications
- The ability to maintain accurate localization with minimal infrastructure (three base stations) makes the system economically viable for widespread adoption

- The framework's effectiveness with both 28 GHz and 78 GHz frequencies provides flexibility in implementation based on available infrastructure

### Foundation for Future Research

- The established performance benchmarks provide reference points for future research in 5G-aided navigation systems
- The successful integration of 5G TOA with various sensor modalities opens new possibilities for hybrid localization approaches
- The demonstrated improvements in challenging scenarios suggest potential applications in other complex indoor environments
- The systematic characterization of system behavior under different conditions lays a solid foundation for future research directions and improvements

### Broader Implications

Beyond UAV localization, this research contributes to broader fields such as robotics, telecommunications, and smart city applications. The integration of 5G TOA data with visual-inertial sensors provides a roadmap for advancing autonomous systems, enabling more robust navigation in GPS-denied environments. The findings also underscore the potential of 5G technology as a transformative tool for diverse applications, including indoor logistics, disaster response, and industrial automation.

## 6.3 Limitations

While the research demonstrates significant advances in 5G-enabled drone localization and SLAM, several fundamental limitations are worth consideration. The global framework requires fixed and known base station positions: this requirement stems from the fundamental properties of TOA measurements rather than our specific implementation. When all the base station positions are unknown, global SLAM becomes impossible as there are no fixed

landmarks to establish absolute positioning in the global frame. however, the system could be extended to handle scenarios where some base station positions are uncertain or dynamic while maintaining global positioning capabilities through a subset of known fixed stations that serve as anchors.

The system's practical implementation faces additional challenges in TOA measurement modeling. Current simulations assume perfect line-of-sight conditions between the drone and base stations, neglecting real-world signal obstruction common in complex indoor environments. These conditions could degrade TOA measurement accuracy, especially in non-line-of-sight scenarios, and the use of static uncertainty models fails to account for the dynamic nature of real-world signal propagation. Additionally, the framework assumes consistent signal reception and stable network conditions, which may not hold in real-world scenarios due to interference, temporary base station failures, or coverage gaps.

System performance presents further technical challenges. The pose graph optimization approach, while achieving high accuracy, introduces significant computational overhead that may challenge real-time operation on resource-constrained platforms. Additionally, geometric limitations in base station placement, particularly in height diversity, affect vertical accuracy.

The validation methodology imposes further constraints. The reliance on simulated 5G measurements, while providing controlled testing conditions, may not fully capture real-world signal propagation effects and hardware imperfections. Furthermore, the evaluation, primarily conducted in controlled environments, does not address the system's performance in more dynamic and unpredictable scenarios, such as environments with temporary signal obstructions, base station outages, or changing network topologies.

## 6.4 Future Work

Building on the limitations identified in this research, several promising directions for future work are outlined below. These directions aim to enhance measurement integration, refine algorithms, and bridge the gap between simulation and real-world implementation.

### Enhanced Measurement Integration

Future advancements in measurement integration can address key challenges in localization accuracy and robustness. Combining diverse data sources has the potential to mitigate current limitations:

**Hybrid TOA-AOA Fusion:** Developing algorithms that integrate Time of Arrival (TOA) and Angle of Arrival (AOA) measurements will enhance both position and orientation estimation. Angular data, especially from 5G beamforming capabilities and massive MIMO arrays, can resolve heading ambiguities and improve overall precision. Additionally, leveraging multipath reflections—currently treated as interference—may provide valuable localization information.

**Multi-Modal Sensor Fusion:** Expanding the sensor suite to include barometric pressure sensors, magnetometers, and other modalities can significantly enhance localization. Absolute height references from barometric data can address vertical accuracy challenges, particularly in scenarios without visual data. Magnetic heading measurements can reduce yaw drift over long durations. Dynamic fusion strategies that adapt sensor contributions based on environmental conditions and data quality will further improve robustness.

### Advanced Algorithmic Development

To tackle inherent limitations in TOA-based localization and SLAM, further algorithmic innovations are required:

**Robust Measurement Models:** Introducing dynamic uncertainty models that adapt to signal quality metrics and environmental conditions will enhance measurement reliability. Additionally, non-line-of-sight (NLoS) detection and mitigation techniques, possibly leveraging machine learning, can ensure accurate localization even in complex environments.

**Joint Localization and Base Station Estimation:** Extending the global SLAM framework to estimate both UAV and base station positions simultaneously will enable operation in partially unknown infrastructure settings. This joint estimation approach will allow the system to adapt to dynamic or mobile base station configurations, increasing its applicability in diverse scenarios.

**Handling Dynamic Networks:** Algorithms that accommodate changing network topologies, such as mobile base stations, signal blockages, or failures, will further enhance system robustness. Developing methods to dynamically manage these scenarios ensures reliable performance in unpredictable environments.

**Computational Optimization:** Improving the efficiency of pose graph operations, particularly through sparsity exploitation and incremental updates, can reduce computational overhead and make real-time implementation feasible even on resource-constrained platforms.

### Real-World Validation and Deployment

Closing the gap between simulation and deployment is a critical next step.

**Validation with Real 5G Hardware:** Testing the system with actual 5G infrastructure will verify the accuracy and reliability of simulation-based findings. Practical challenges, such as hardware-induced signal imperfections and environmental effects, can be addressed during this phase.

**Environmental Testing:** Conducting evaluations across diverse indoor settings—ranging from open areas to complex, obstacle-rich environments—will provide insights into system performance under realistic conditions. This includes studying the impact of multipath effects, NLoS scenarios, and varying network topologies.

**Dynamic Deployment Scenarios:** Exploring use cases involving mobile base stations, temporary signal obstructions, and network reconfiguration will test the system's adaptability and highlight areas for further improvement.

## 6.5 Final Remarks

In conclusion, this thesis demonstrated the potential of integrating 5G TOA measurements with traditional localization methods to overcome the challenges of indoor UAV navigation. The proposed framework provides a solid foundation for future advancements in robust and scalable autonomous navigation systems, paving the way for innovative applications in

robotics, smart cities, and beyond

# Appendix A

# Detailed Measurement Statistics

## A.1 Time of Arrival Measurements for EuRoC MAV Dataset

Table A.1: Error statistics of the estimated TOA distance to BSs for different scenarios and given in meters.

| Dataset | 5G Sim. Scenario | Statistic | TOA #1 | TOA #2 | TOA #3 | TOA #4 | TOA #5 |
|---|---|---|---|---|---|---|---|
| V101 | QuaDRiGa-Industrial-LOS | Mean | 0.129 | -0.045 | 0.006 | -0.081 | -0.023 |
| | | Std. | 0.568 | 0.81 | 0.763 | 0.872 | 0.718 |
| | 3GPP-38.901-Indoor-LOS | Mean | -0.024 | -0.021 | -0.059 | 0.041 | -0.06 |
| | | Std. | 0.344 | 0.368 | 0.352 | 0.394 | 0.369 |
| | mmMAGIC-Indoor-LOS | Mean | 0.002 | 0.01 | -0.008 | 0.003 | -0.01 |
| | | Std. | 0.185 | 0.171 | 0.173 | 0.159 | 0.176 |
| V102 | QuaDRiGa-Industrial-LOS | Mean | 0.16 | -0.033 | -0.135 | -0.129 | -0.156 |
| | | Std. | 0.645 | 0.874 | 0.722 | 0.739 | 0.677 |
| | 3GPP-38.901-Indoor-LOS | Mean | -0.104 | 0.104 | 0.106 | -0.122 | -0.052 |
| | | Std. | 0.358 | 0.39 | 0.404 | 0.367 | 0.322 |
| | mmMAGIC-Indoor-LOS | Mean | 0.037 | -0.011 | -0.018 | 0.016 | -0.046 |
| | | Std. | 0.174 | 0.153 | 0.154 | 0.16 | 0.193 |
| V103 | QuaDRiGa-Industrial-LOS | Mean | 0.043 | -0.065 | 1.232 | -0.066 | -0.387 |
| | | Std. | 0.775 | 0.784 | 1.628 | 0.772 | 1.275 |
| | 3GPP-38.901-Indoor-LOS | Mean | -0.042 | 0.053 | 0.008 | -0.033 | -0.022 |
| | | Std. | 0.353 | 0.382 | 0.387 | 0.36 | 0.369 |
| | mmMAGIC-Indoor-LOS | Mean | 0.001 | -0.011 | 0.012 | 0.0 | -0.013 |
| | | Std. | 0.176 | 0.166 | 0.17 | 0.18 | 0.183 |
| V201 | QuaDRiGa-Industrial-LOS | Mean | 0.059 | 0.108 | -0.182 | -0.154 | -0.27 |
| | | Std. | 0.751 | 0.897 | 0.592 | 0.986 | 0.79 |
| | 3GPP-38.901-Indoor-LOS | Mean | 0.025 | -0.07 | -0.045 | 0.054 | 0.12 |
| | | Std. | 0.364 | 0.379 | 0.392 | 0.302 | 0.367 |
| | mmMAGIC-Indoor-LOS | Mean | -0.012 | 0.026 | 0.015 | -0.019 | 0.015 |
| | | Std. | 0.164 | 0.177 | 0.163 | 0.18 | 0.192 |
| V202 | QuaDRiGa-Industrial-LOS | Mean | 0.027 | 0.141 | 0.072 | 0.082 | -0.204 |
| | | Std. | 0.716 | 0.674 | 0.908 | 0.933 | 0.631 |
| | 3GPP-38.901-Indoor-LOS | Mean | 0.052 | -0.053 | -0.039 | 0.012 | 0.043 |
| | | Std. | 0.391 | 0.348 | 0.427 | 0.359 | 0.351 |
| | mmMAGIC-Indoor-LOS | Mean | -0.007 | 0.007 | -0.018 | 0.013 | 0.011 |
| | | Std. | 0.178 | 0.168 | 0.17 | 0.19 | 0.192 |
| V203 | QuaDRiGa-Industrial-LOS | Mean | 0.067 | -0.017 | 0.273 | -0.13 | -0.045 |
| | | Std. | 0.754 | 0.73 | 1.343 | 0.724 | 0.684 |
| | 3GPP-38.901-Indoor-LOS | Mean | 0.009 | -0.022 | -0.046 | 0.043 | 0.02 |
| | | Std. | 0.376 | 0.351 | 0.317 | 0.37 | 0.358 |
| | mmMAGIC-Indoor-LOS | Mean | 0.018 | 0.017 | 0.002 | 0.004 | -0.009 |
| | | Std. | 0.17 | 0.181 | 0.174 | 0.178 | 0.174 |

## A.2 Time of Arrival Measurements for AeroLab Dataset

Table A.2: TOA Mean and Standard Deviation for 28 GHz and 78 GHz Frequencies Across Five Datasets and Four Base Stations

| Dataset | Frequency (GHz) | Base Station | Std (cm) | Mean (cm) |
|---|---|---|---|---|
| Dataset0 | 28 | 1 | 32.92 | 4.20 |
| | | 2 | 36.05 | -1.21 |
| | | 3 | 31.99 | 4.19 |
| | | 4 | 41.32 | 8.28 |
| | 78 | 1 | 19.58 | -2.71 |
| | | 2 | 19.44 | 1.34 |
| | | 3 | 19.10 | 0.31 |
| | | 4 | 14.25 | -1.72 |
| Dataset1 | 28 | 1 | 36.31 | -6.88 |
| | | 2 | 29.60 | -3.56 |
| | | 3 | 36.60 | 1.88 |
| | | 4 | 32.22 | -2.68 |
| | 78 | 1 | 15.95 | -2.83 |
| | | 2 | 17.01 | -2.29 |
| | | 3 | 16.08 | 0.56 |
| | | 4 | 15.64 | -0.86 |
| Dataset2 | 28 | 1 | 38.44 | -8.24 |
| | | 2 | 31.94 | -4.93 |
| | | 3 | 33.25 | 7.97 |
| | | 4 | 30.49 | -0.50 |
| | 78 | 1 | 15.83 | 0.87 |
| | | 2 | 18.92 | -4.97 |
| | | 3 | 16.38 | -0.80 |
| | | 4 | 16.63 | -0.59 |
| Dataset3 | 28 | 1 | 40.14 | -14.56 |
| | | 2 | 32.19 | -6.68 |
| | | 3 | 31.17 | 4.49 |
| | | 4 | 27.64 | -4.97 |
| | 78 | 1 | 15.03 | 0.87 |
| | | 2 | 17.61 | -8.71 |
| | | 3 | 14.69 | -1.49 |
| | | 4 | 15.94 | -1.06 |
| Dataset4 | 28 | 1 | 38.23 | -19.41 |
| | | 2 | 29.54 | -2.18 |
| | | 3 | 31.23 | -3.09 |
| | | 4 | 40.62 | -8.31 |
| | 78 | 1 | 14.40 | 3.12 |
| | | 2 | 16.80 | -4.67 |
| | | 3 | 15.56 | 0.57 |
| | | 4 | 18.76 | -1.08 |

# Appendix B

# Detailed Experimental Results

## B.1 Localization by Fusing 5G TOA and IMU using ESKF and PGO

Table B.1: Evaluation of the Error State Kalman Filter (ESKF) vs. Graph-Based Optimization Localization under Different Scenarios and Configurations for Vicon Room 1 (Datasets V101 and V102). **Best values for each dataset are shown in bold.**

| Dataset | 5G sim. scenario | Approach | BS Num | ATE(m) | $E_X$(m) | $E_Y$(m) | $E_Z$(m) | $RPE_T$ (m) | $RPE_R$ (deg) | $ATE_{avg.}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| V101 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 3.7290 | 0.7326 | 1.1775 | 3.4615 | 0.0174 | 0.3979 | 1.438 |
| | | | 3 | 1.2566 | 0.2978 | 0.1910 | 1.2058 | 0.0113 | 0.3984 | |
| | | | 4 | 1.5274 | 0.2230 | 0.1588 | 1.5026 | 0.0091 | 0.4175 | |
| | | | 5 | 0.6791 | 0.2431 | 0.1553 | 0.6147 | 0.0082 | 0.4142 | |
| | | ESKF-based | 2 | 6.6928 | 3.2562 | 4.3922 | 3.8600 | 0.5830 | 0.0286 | 3.17 |
| | | | 3 | 2.9209 | 0.7723 | 0.6821 | 2.7331 | 0.2571 | 0.0211 | |
| | | | 4 | 3.3593 | 0.6545 | 0.5430 | 3.2498 | 0.2672 | 0.0244 | |
| | | | 5 | 2.8782 | 0.5868 | 0.4830 | 2.7760 | 0.2644 | **0.0201** | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 2.5805 | 1.0242 | 1.6213 | 1.7266 | 0.0182 | 0.3979 | 0.867 |
| | | | 3 | 1.1189 | 0.1935 | 0.1225 | 1.0952 | 0.0083 | 0.3979 | |
| | | | 4 | 0.3787 | 0.1757 | 0.1147 | 0.3152 | 0.0070 | 0.4114 | |
| | | | 5 | 0.2583 | 0.1717 | 0.0970 | 0.1668 | 0.0066 | 0.4487 | |
| | | ESKF-based | 2 | 5.1227 | 2.5365 | 2.8306 | 3.4345 | 0.3837 | 0.0333 | 1.934 |
| | | | 3 | 2.2327 | 0.5324 | 0.3538 | 2.1392 | 0.2465 | 0.0209 | |
| | | | 4 | 1.4097 | 0.4463 | 0.2755 | 1.3084 | 0.1865 | 0.0356 | |
| | | | 5 | 0.9072 | 0.4194 | 0.2373 | 0.7687 | 0.1763 | 0.0266 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 1.9861 | 0.7814 | 1.0087 | 1.5220 | 0.0104 | 0.3978 | **0.701** |
| | | | 3 | 1.2447 | 0.1228 | 0.0522 | 1.2375 | 0.0065 | 0.4378 | |
| | | | 4 | 0.1432 | 0.0576 | 0.0468 | 0.1225 | 0.0055 | 0.4468 | |
| | | | 5 | **0.1312** | **0.0516** | **0.0422** | **0.1131** | **0.0052** | 0.4571 | |
| | | ESKF-based | 2 | 5.0005 | 2.2481 | 3.3849 | 2.9143 | 0.4272 | 0.0255 | 1.504 |
| | | | 3 | 1.7167 | 0.2226 | 0.1824 | 1.6924 | 0.2070 | 0.0337 | |
| | | | 4 | 0.4643 | 0.1745 | 0.1481 | 0.4040 | 0.1596 | 0.0472 | |
| | | | 5 | 0.3400 | 0.1602 | 0.1305 | 0.2699 | 0.1436 | 0.0475 | |
| V102 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 1.5352 | 0.8783 | 0.4125 | 1.1897 | 0.0110 | 0.4361 | 0.744 |
| | | | 3 | 0.8399 | 0.2819 | 0.1810 | 0.7702 | 0.0108 | 0.4373 | |
| | | | 4 | 0.7252 | 0.2178 | 0.1785 | 0.6683 | 0.0102 | 0.4365 | |
| | | | 5 | 0.6216 | 0.2006 | 0.1540 | 0.5678 | 0.0094 | 0.4373 | |
| | | ESKF-based | 2 | 4.6070 | 3.4935 | 0.9800 | 2.8389 | 0.4711 | 0.0394 | 2.154 |
| | | | 3 | 2.1624 | 0.9664 | 0.7710 | 1.7742 | 0.3553 | 0.0672 | |
| | | | 4 | 2.0928 | 0.8675 | 0.7558 | 1.7482 | 0.3709 | 0.1039 | |
| | | | 5 | 1.9070 | 0.7949 | 0.7851 | 1.5455 | 0.3664 | 0.1051 | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 0.7864 | 0.4475 | 0.1228 | 0.6349 | 0.0094 | 0.4361 | 0.523 |
| | | | 3 | 1.3843 | 0.1719 | 0.1388 | 1.3666 | 0.0123 | 0.4363 | |
| | | | 4 | 0.2257 | 0.1620 | 0.0983 | 0.1226 | 0.0083 | 0.4381 | |
| | | | 5 | 0.2171 | 0.1588 | 0.0946 | 0.1139 | 0.0082 | 0.4449 | |
| | | ESKF-based | 2 | 5.0163 | 4.3755 | 1.2167 | 2.1302 | 0.5266 | **0.0345** | 1.625 |
| | | | 3 | 1.2971 | 0.6606 | 0.4610 | 1.0166 | 0.3490 | 0.0819 | |
| | | | 4 | 0.9720 | 0.4812 | 0.5191 | 0.6661 | 0.3181 | 0.0796 | |
| | | | 5 | 0.8386 | 0.4801 | 0.4566 | 0.5142 | 0.2920 | 0.0794 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 0.3714 | 0.1816 | 0.0661 | 0.3172 | 0.0084 | 0.4362 | **0.166** |
| | | | 3 | 0.1981 | 0.0640 | 0.0451 | 0.1820 | 0.0074 | 0.4418 | |
| | | | 4 | 0.1297 | **0.0527** | 0.0441 | 0.1100 | 0.0074 | 0.4653 | |
| | | | 5 | **0.1286** | 0.0603 | **0.0365** | **0.1076** | **0.0072** | 0.4549 | |
| | | ESKF-based | 2 | 4.3145 | 3.3932 | 0.5532 | 2.6068 | 0.5891 | 0.0454 | 1.409 |
| | | | 3 | 1.7085 | 0.3224 | 0.3215 | 1.6467 | 0.3436 | 0.0482 | |
| | | | 4 | 0.5381 | 0.2799 | 0.3225 | 0.3275 | 0.2720 | 0.1526 | |
| | | | 5 | 0.4848 | 0.2804 | 0.2686 | 0.2902 | 0.2544 | 0.1653 | |

Table B.2: Evaluation of the Error State Kalman Filter (ESKF) vs. Graph-Based Optimization Localization under Different Scenarios and Configurations for Vicon Room 1 (Dataset V103) and Vicon Room 2 (Dataset V201). **Best values for each dataset are shown in bold.**

| Dataset | 5G sim. scenario | Approach | BS Num | ATE(m) | $\mathbf{E_X}$(m) | $\mathbf{E_Y}$(m) | $\mathbf{E_Z}$(m) | $\mathbf{RPE_T}$ (m) | $\mathbf{RPE_R}$ (deg) | $\mathbf{ATE_{avg.}}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| V103 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 1.7425 | 0.8096 | 0.4779 | 1.4671 | 0.0185 | 0.7368 | 1.247 |
| | | | 3 | 2.1882 | 0.4163 | 0.2007 | 2.1388 | 0.0141 | 0.7372 | |
| | | | 4 | 0.9904 | 0.3355 | 0.1479 | 0.9200 | 0.0156 | 0.7381 | |
| | | | 5 | 1.3146 | 0.6974 | 0.3194 | 1.0676 | 0.0173 | 0.7443 | |
| | | ESKF-based | 2 | 4.7003 | 3.7447 | 1.2424 | 2.5547 | 0.4716 | 0.0373 | 3.237 |
| | | | 3 | 4.4011 | 0.8701 | 0.9770 | 4.2022 | 0.3801 | 0.0397 | |
| | | | 4 | 3.2805 | 0.6970 | 0.8300 | 3.0963 | 0.3068 | 0.0657 | |
| | | | 5 | 3.8017 | 1.1043 | 0.9159 | 3.5206 | 0.3981 | 0.0690 | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 0.7494 | 0.4310 | 0.3216 | 0.5219 | 0.0133 | 0.7363 | **0.355** |
| | | | 3 | 0.4769 | 0.2144 | 0.0727 | 0.4198 | 0.0117 | 0.7412 | |
| | | | 4 | 0.2759 | 0.2287 | 0.0685 | 0.1383 | 0.0113 | 0.7446 | |
| | | | 5 | 0.2733 | 0.2242 | 0.0538 | 0.1467 | 0.0113 | 0.7492 | |
| | | ESKF-based | 2 | 4.8341 | 3.9437 | 1.2121 | 2.5193 | 0.3909 | **0.0355** | 2.028 |
| | | | 3 | 2.4700 | 0.8539 | 0.6351 | 2.2290 | 0.2908 | 0.0964 | |
| | | | 4 | 1.6018 | 0.8075 | 0.5747 | 1.2584 | 0.2673 | 0.1084 | |
| | | | 5 | 1.2320 | 0.8089 | 0.4765 | 0.7978 | 0.2530 | 0.0887 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 1.2840 | 0.2598 | 0.1496 | 1.2485 | 0.0112 | 0.7362 | 0.558 |
| | | | 3 | 1.2075 | 0.0970 | 0.0603 | 1.2021 | 0.0120 | 0.7453 | |
| | | | 4 | **0.1365** | **0.0738** | 0.0510 | **0.1029** | 0.0105 | 0.7720 | |
| | | | 5 | 0.1614 | 0.0754 | **0.0398** | 0.1370 | **0.0102** | 0.7636 | |
| | | ESKF-based | 2 | 3.8474 | 2.9339 | 1.3210 | 2.1094 | 0.4834 | 0.0596 | 1.388 |
| | | | 3 | 1.4482 | 0.6386 | 0.3389 | 1.2548 | 0.2525 | 0.1338 | |
| | | | 4 | 0.7638 | 0.5816 | 0.3572 | 0.3426 | 0.2104 | 0.1402 | |
| | | | 5 | 0.8789 | 0.6446 | 0.3093 | 0.5111 | 0.2092 | 0.1180 | |
| V201 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 5.2234 | 3.7356 | 2.7495 | 2.4020 | 0.0399 | 1.2588 | 2.319 |
| | | | 3 | 2.6128 | 0.5229 | 0.5995 | 2.4887 | 0.0246 | 1.2620 | |
| | | | 4 | 2.0649 | 0.4159 | 0.3468 | 1.9926 | 0.0204 | 1.2601 | |
| | | | 5 | 1.6947 | 0.4047 | 0.2766 | 1.6223 | 0.0188 | 1.2640 | |
| | | ESKF-based | 2 | 5.1291 | 3.1488 | 2.4413 | 3.2299 | 0.4138 | **0.0454** | 2.939 |
| | | | 3 | 3.9739 | 0.7953 | 0.9333 | 3.7800 | 0.3330 | 0.0483 | |
| | | | 4 | 2.6749 | 0.5321 | 0.6420 | 2.5416 | 0.2257 | 0.0476 | |
| | | | 5 | 2.9173 | 0.6072 | 0.4912 | 2.8108 | 0.2127 | 0.0800 | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 2.3833 | 1.5583 | 1.4689 | 1.0461 | 0.0239 | 1.2580 | **0.875** |
| | | | 3 | 1.2636 | 0.2385 | 0.2003 | 1.2246 | 0.0153 | 1.2581 | |
| | | | 4 | 0.3906 | 0.1807 | 0.1700 | 0.3017 | 0.0133 | 1.2771 | |
| | | | 5 | 0.3375 | 0.1792 | 0.1300 | 0.2548 | 0.0130 | 1.2656 | |
| | | ESKF-based | 2 | 4.1948 | 2.1934 | 1.5755 | 3.2099 | 0.4360 | 0.0623 | 1.68 |
| | | | 3 | 2.1001 | 0.5156 | 0.3570 | 2.0043 | 0.2055 | 0.0501 | |
| | | | 4 | 1.2208 | 0.3549 | 0.3474 | 1.1152 | 0.1684 | 0.0495 | |
| | | | 5 | 0.8838 | 0.3682 | 0.2477 | 0.7644 | 0.1535 | 0.0528 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 4.1182 | 2.6694 | 2.0691 | 2.3564 | 0.0355 | 1.2578 | 1.142 |
| | | | 3 | 1.1002 | 1.1678 | 0.0602 | 1.0856 | 0.0158 | 1.2626 | |
| | | | 4 | 0.2598 | **0.0783** | 0.0549 | 0.2416 | 0.0128 | 1.2730 | |
| | | | 5 | **0.2307** | 0.0797 | **0.0475** | **0.2113** | **0.0123** | 1.2633 | |
| | | ESKF-based | 2 | 3.9157 | 2.1876 | 2.2021 | 2.3870 | 0.3262 | 0.0907 | 1.371 |
| | | | 3 | 1.8596 | 0.2765 | 0.2222 | 1.8254 | 0.1780 | 0.0509 | |
| | | | 4 | 0.6011 | 0.2386 | 0.1851 | 0.5197 | 0.1489 | 0.0554 | |
| | | | 5 | 0.4789 | 0.2505 | 0.1974 | 0.3573 | 0.1259 | 0.0577 | |

Table B.3: Evaluation of the Error State Kalman Filter (ESKF) vs. Graph-Based Optimization Localization under Different Scenarios and Configurations for Vicon Room 2 (Datasets V202 and V203). **Best values for each dataset are shown in bold.**

| Dataset | 5G sim. scenario | Approach | BS Num | ATE(m) | $E_X$(m) | $E_Y$(m) | $E_Z$(m) | $RPE_T$ (m) | $RPE_R$ (deg) | $ATE_{avg.}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| V202 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 2.6935 | 1.5963 | 1.2222 | 1.7925 | 0.0367 | 2.7873 | 1.453 |
| | | | 3 | 2.4520 | 0.6049 | 0.4874 | 2.3256 | 0.0321 | 2.7877 | |
| | | | 4 | 1.4682 | 0.3855 | 0.2990 | 1.3847 | 0.0300 | 2.7900 | |
| | | | 5 | 0.6496 | 0.3859 | 0.2206 | 0.4737 | 0.0283 | 2.7902 | |
| | | ESKF-based | 2 | 5.6998 | 3.5542 | 3.1184 | 3.1830 | 0.6514 | 0.1137 | 3.038 |
| | | | 3 | 3.4033 | 1.2179 | 0.9019 | 3.0472 | 0.4557 | 0.0987 | |
| | | | 4 | 3.3419 | 0.7543 | 0.5956 | 3.2007 | 0.3051 | 0.0986 | |
| | | | 5 | 2.7431 | 0.8620 | 0.5499 | 2.5454 | 0.3003 | 0.1049 | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 1.7066 | 0.7871 | 0.7965 | 1.2878 | 0.0348 | 2.7873 | 0.738 |
| | | | 3 | 1.0422 | 0.2209 | 0.1824 | 1.0020 | 0.0294 | 2.7881 | |
| | | | 4 | 0.5193 | 0.1608 | 0.1479 | 0.4712 | 0.0277 | 2.7915 | |
| | | | 5 | 0.4228 | 0.1740 | 0.1319 | 0.3621 | 0.0274 | 2.7903 | |
| | | ESKF-based | 2 | 3.9675 | 2.3513 | 1.9427 | 2.5373 | 0.4971 | **0.0962** | 1.634 |
| | | | 3 | 2.1767 | 0.5870 | 0.4274 | 2.0520 | 0.3171 | 0.1224 | |
| | | | 4 | 1.0396 | 0.4805 | 0.3843 | 0.8380 | 0.2640 | 0.1300 | |
| | | | 5 | 0.9861 | 0.5676 | 0.3499 | 0.7264 | 0.2564 | 0.1245 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 1.4226 | 0.7650 | 0.6926 | 0.9792 | 0.0317 | 2.7873 | **0.559** |
| | | | 3 | 0.9065 | 0.1162 | 0.1054 | 0.8929 | 0.0282 | 2.7887 | |
| | | | 4 | 0.2417 | **0.1000** | 0.0961 | 0.1980 | 0.02682 | 2.7960 | |
| | | | 5 | **0.2244** | 0.1012 | **0.0909** | **0.1785** | **0.02680** | 2.8014 | |
| | | ESKF-based | 2 | 4.5127 | 3.1000 | 2.6701 | 1.9040 | 0.4324 | 0.1106 | 1.462 |
| | | | 3 | 1.4249 | 0.4480 | 0.3224 | 1.3136 | 0.2629 | 0.2437 | |
| | | | 4 | 0.7655 | 0.4097 | 0.2682 | 0.5884 | 0.2236 | 0.2483 | |
| | | | 5 | 0.6050 | 0.2987 | 0.2638 | 0.4551 | 0.2163 | 0.1951 | |
| V203 | QuaDRiGa-Industrial-LOS | Graph-based | 2 | 3.4606 | 2.3413 | 2.2257 | 1.2410 | 0.0442 | 3.2686 | 1.258 |
| | | | 3 | 0.9769 | 0.3589 | 0.3384 | 0.8432 | 0.0274 | 3.2689 | |
| | | | 4 | 1.2247 | 0.2967 | 0.3169 | 1.1452 | 0.0295 | 3.2717 | |
| | | | 5 | 0.6272 | 0.3255 | 0.2812 | 0.4565 | 0.0269 | 3.2758 | |
| | | ESKF-based | 2 | 5.5247 | 2.9753 | 2.8260 | 3.6992 | 0.6252 | **0.1131** | 2.758 |
| | | | 3 | 3.0756 | 0.9551 | 0.9707 | 2.7577 | 0.3711 | 0.1402 | |
| | | | 4 | 3.0525 | 0.8027 | 0.6954 | 2.8618 | 0.3478 | 0.1223 | |
| | | | 5 | 2.1353 | 0.9533 | 0.6151 | 1.8090 | 0.3513 | 0.1257 | |
| | 3GPP-38.901-Indoor-LOS | Graph-based | 2 | 3.0104 | 1.8298 | 1.8757 | 1.4819 | 0.0342 | 3.2686 | 0.962 |
| | | | 3 | 0.9943 | 0.1643 | 0.1753 | 0.9648 | 0.0265 | 3.2686 | |
| | | | 4 | 0.4270 | 0.1577 | 0.1328 | 0.3740 | 0.0258 | 3.2754 | |
| | | | 5 | 0.3771 | 0.1526 | 0.1100 | 0.3268 | 0.0258 | 3.2775 | |
| | | ESKF-based | 2 | 4.0704 | 2.2813 | 1.9862 | 2.7238 | 0.4975 | 0.1227 | 1.568 |
| | | | 3 | 1.7312 | 0.5890 | 0.4748 | 1.5571 | 0.2855 | 0.1418 | |
| | | | 4 | 1.0466 | 0.4893 | 0.4362 | 0.8159 | 0.2687 | 0.1405 | |
| | | | 5 | 0.9896 | 0.4989 | 0.4095 | 0.7501 | 0.2498 | 0.1412 | |
| | mmMAGIC-Indoor-LOS | Graph-based | 2 | 1.5031 | 0.5633 | 0.6366 | 1.2397 | 0.0290 | 3.2686 | **0.475** |
| | | | 3 | 0.4342 | 0.1250 | 0.0858 | 0.4068 | 0.0263 | 3.2697 | |
| | | | 4 | 0.2509 | **0.1202** | 0.0855 | 0.2030 | 0.0257 | 3.2752 | |
| | | | 5 | **0.1878** | 0.1224 | **0.0782** | **0.1191** | **0.0254** | 3.2771 | |
| | | ESKF-based | 2 | 4.2958 | 2.2371 | 2.5190 | 2.6654 | 0.5333 | 0.1149 | 1.392 |
| | | | 3 | 1.6151 | 0.2677 | 0.3013 | 1.5640 | 0.2796 | 0.1407 | |
| | | | 4 | 0.6045 | 0.2814 | 0.2341 | 0.4811 | 0.2218 | 0.1482 | |
| | | | 5 | 0.4459 | 0.2554 | 0.2195 | 0.2923 | 0.1992 | 0.1569 | |

## B.2 5G-Enhanced Visual-inertial SLAM

### B.2.1 AeroLab Dataset Results

Table B.4: Average Errors and Improvements for Different SLAM Modes on Aerolab Dataset.
*Note: For the monocular mode with TOA integration, results for Datasets 2 and 4 are calculated up to 75 s and 65 s respectively, as the system lost tracking beyond these durations.*

(a) RGB-D SLAM (SE3 Post-Alignment)

| Dataset | Configuration | Local (m) | Global (m) | Local Imp. |
|---|---|---|---|---|
| Dataset0 | Baseline | 0.064 | N/A‡ | – |
| | 28 GHz TOA | 0.066 | 0.237 | -3.1% |
| | 78 GHz TOA | 0.055 | 0.101 | 14.1% |
| Dataset1 | Baseline | 0.116 | N/A | – |
| | 28 GHz TOA | 0.117 | 0.165 | -0.9% |
| | 78 GHz TOA | 0.117 | 0.133 | -0.9% |
| Dataset2 | Baseline | 0.155 | N/A | – |
| | 28 GHz TOA | 0.160 | 0.197 | -3.2% |
| | 78 GHz TOA | 0.148 | 0.158 | 4.5% |
| Dataset3 | Baseline | 0.095 | N/A | – |
| | 28 GHz TOA | 0.102 | 0.149 | -7.4% |
| | 78 GHz TOA | 0.094 | 0.104 | 1.1% |
| Dataset4 | Baseline | 0.146 | N/A | – |
| | 28 GHz TOA | 0.162 | 0.242 | -11.0% |
| | 78 GHz TOA | 0.164 | 0.171 | -12.3% |
| **Mode Average** | Baseline | **0.115** | N/A | – |
| | 28 GHz TOA | **0.121** | **0.198** | **-5.2%** |
| | 78 GHz TOA | **0.116** | **0.133** | **1.3%** |

(b) RGB-D Inertial SLAM (SE3 Post-Alignment)

| Dataset | Configuration | Local (m) | Global (m) | Local Imp. |
|---|---|---|---|---|
| Dataset0 | Baseline | 0.089 | N/A | – |
| | 28 GHz TOA | 0.092 | 0.471 | -3.4% |
| | 78 GHz TOA | 0.075 | 0.276 | 15.7% |
| Dataset1 | Baseline | 0.119 | N/A | – |
| | 28 GHz TOA | 0.118 | 0.206 | 0.8% |
| | 78 GHz TOA | 0.118 | 0.315 | 0.8% |
| Dataset2 | Baseline | 0.216 | N/A | – |
| | 28 GHz TOA | 0.231† | 0.267 | -6.9% † |
| | 78 GHz TOA | 0.366† | 0.426 | -69.4% † |
| Dataset3 | Baseline | 0.065 | N/A | – |
| | 28 GHz TOA | 0.064 | 0.134 | 1.5% |
| | 78 GHz TOA | 0.062 | 0.125 | 4.6% |
| Dataset4 | Baseline | 0.108 | N/A | – |
| | 28 GHz TOA | 0.105 | 0.321 | 2.8% |
| | 78 GHz TOA | 0.105 | 0.168 | 2.8% |
| **Mode Average** | Baseline | **0.095** | N/A | – |
| | 28 GHz TOA | **0.94** | **0.280** | **0.42%** |
| | 78 GHz TOA | **0.90** | **0.262** | **5.97%** |

(c) Monocular SLAM (Both SE3 and Sim3 Post-Alignment for TOA)

| Dataset | Configuration | Local (m) | Global (m) | Local (Unscaled) | Scale | Local Imp. |
|---|---|---|---|---|---|---|
| Dataset0 | Baseline | Failed | N/A | N/A | N/A | – |
| | 28 GHz TOA | Failed | Failed | – | – | – |
| | 78 GHz TOA | Failed | Failed | – | – | – |
| Dataset1 | Baseline | 0.055 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.055 | 0.380 | 0.059 | 0.93% | 0.6% |
| | 78 GHz TOA | 0.057 | 0.121 | 0.058 | 0.37% | -3.3% |
| Dataset2 | Baseline | 0.058 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.084 | 0.186 | 0.089 | 1.58% | -44.0%† |
| | 78 GHz TOA | 0.074 | 0.152 | 0.075 | 0.62% | -26.6%† |
| Dataset3 | Baseline | 0.074 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.077 | 0.162 | 0.084 | 1.97% | -3.8% |
| | 78 GHz TOA | 0.056 | 0.127 | 0.058 | 0.76% | 23.8% |
| Dataset4 | Baseline | 0.062 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.075 | 0.260 | 0.084 | 2.44% | -20.2% |
| | 78 GHz TOA | 0.054 | 0.131 | 0.061 | 1.90% | 13.8% |
| **Mode Average** | Baseline | **0.064** | N/A | N/A | N/A | – |
| | 28 GHz TOA | **0.069** | **0.267** | **0.076** | **0.30%** | **-7.8%** |
| | 78 GHz TOA | **0.056** | **0.126** | **0.059** | **0.15%** | **11.4%** |

† Dataset2's RGB-D Inertial and monocular with TOA results were excluded from average calculations due to anomalous degradation.

‡ N/A denotes baseline systems without global positioning and scaling capability.

Monocular-inertial configurations failed across all datasets due to initialization issues, preventing successful tracking.

For RGB-D and RGB-D Inertial modes, accuracy is evaluated using SE3 post-alignment. For Monocular mode, baseline uses Sim3 post-alignment, while TOA configurations provide both SE3 and Sim3 post-alignment results.

### B.2.2 EuRoC MAV Dataset Results

Table B.5: Average Errors and Improvements for Different SLAM Modes on EuRoC MAV Dataset.
*Note: Results for V203 are excluded due to tracking failures.*

(a) Stereo SLAM (SE3 Post-Alignment for TOA)

| Sequence | Configuration | Local (m) | Global (m) | Local Imp. |
|---|---|---|---|---|
| V101 | Baseline | 0.084 | N/A | – |
| | 28 GHz TOA | 0.084 | 0.232 | 0.0% |
| | 78 GHz TOA | 0.081 | 0.183 | 3.6% |
| V102 | Baseline | 0.099 | N/A | – |
| | 28 GHz TOA | 0.113 | 0.585 | -14.1% |
| | 78 GHz TOA | 0.106 | 0.139 | -7.2% |
| V103 | Baseline | 0.161 | N/A | – |
| | 28 GHz TOA | 0.182 | 0.359 | -13.0% |
| | 78 GHz TOA | 0.126 | 0.190 | 21.7% |
| V201 | Baseline | 0.113 | N/A | – |
| | 28 GHz TOA | 0.094 | 0.335 | 17.2% |
| | 78 GHz TOA | 0.086 | 0.177 | 23.8% |
| V202 | Baseline | 0.190 | N/A | – |
| | 28 GHz TOA | 0.196 | 0.274 | -3.0% |
| | 78 GHz TOA | 0.192 | 0.265 | -1.0% |
| **Average*** | Baseline | **0.129** | N/A | – |
| | 28 GHz TOA | **0.134** | **0.357** | **-2.6%** |
| | 78 GHz TOA | **0.118** | **0.191** | **8.2%** |

(b) Monocular SLAM (Both SE3 and Sim3 Post-Alignment for TOA)

| Sequence | Configuration | Local (m) | Global (m) | Local (Unscaled) (m) | Scale | Local Imp. |
|---|---|---|---|---|---|---|
| V101 | Baseline | 0.095 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.094 | 0.247 | 0.099 | 1.55% | 1.1% |
| | 78 GHz TOA | 0.094 | 0.149 | 0.094 | 0.05% | 1.1% |
| V102 | Baseline | 0.100 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.100 | 0.438 | 0.102 | 0.40% | 0.0% |
| | 78 GHz TOA | 0.098 | 0.128 | 0.098 | 0.59% | 2.0% |
| V103 | Baseline | 0.116 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.103 | 0.224 | 0.111 | 2.24% | 11.2% |
| | 78 GHz TOA | 0.102 | 0.415 | 0.102 | 0.01% | 12.1% |
| V201 | Baseline | 0.086 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.085 | 0.669 | 0.125 | 3.75% | 1.2% |
| | 78 GHz TOA | 0.085 | 0.247 | 0.085 | 0.17% | 1.2% |
| V202 | Baseline | 0.176 | N/A | N/A | N/A | – |
| | 28 GHz TOA | 0.178 | 0.579 | 0.179 | 0.74% | -1.1% |
| | 78 GHz TOA | 0.178 | 0.248 | 0.181 | 1.36% | -1.1% |
| **Average*** | Baseline | **0.115** | N/A | N/A | N/A | – |
| | 28 GHz TOA | **0.112** | **0.431** | **0.123** | **1.74%** | **2.5%** |
| | 78 GHz TOA | **0.111** | **0.237** | **0.112** | **0.20%** | **3.1%** |

# References


[1] Beiya Yang and Erfu Yang. “A survey on radio frequency based precise localisation technology for UAV in GPS-denied environment”. In: *Journal of Intelligent & Robotic Systems* 103 (2021), pp. 1–30.

[2] Faheem Zafari, Athanasios Gkelias, and Kin K Leung. “A survey of indoor localization systems and technologies”. In: *IEEE Communications Surveys & Tutorials* 21.3 (2019), pp. 2568–2599.

[3] Ping Zhang et al. “Cooperative localization in 5G networks: A survey”. In: *Ict Express* 3.1 (2017), pp. 27–32.

[4] Kun Li et al. “UWB-VO: Ultra-Wideband Anchor Assisted Visual Odometry”. In: *2023 IEEE International Conference on Unmanned Systems (ICUS)*. IEEE. 2023, pp. 943–950.

[5] Yong Tian, Zhenan Tang, and Yan Yu. “Third-Order Channel Propagation Model-Based Indoor Adaptive Localization Algorithm for Wireless Sensor Networks”. In: *IEEE Antennas and Wireless Propagation Letters* 12 (2013). Conference Name: IEEE Antennas and Wireless Propagation Letters, pp. 1578–1581. ISSN: 1548-5757. DOI: 10.1109/LAWP.2013.2293578.

[6] Alejandro Martinez-Sala et al. “An accurate radio channel model for wireless sensor networks simulation”. In: *Journal of communications and networks* 7.4 (2005). Publisher: KICS, pp. 401–407.

[7] Byeong-ho Lee et al. “Genetic Algorithm for Path Loss Model Selection in Signal Strength Based Indoor Localization”. In: *IEEE Sensors Journal* (2021). Conference Name: IEEE Sensors Journal, pp. 1–1. ISSN: 1558-1748. DOI: 10.1109/JSEN.2021.3110971.

[8] Tsung-Han Lin et al. “A microscopic examination of an RSSI-signature-Based indoor localization system”. In: (2008), pp. 2–6.

[9] Hoang M. Le, Jean-Pierre Rossi, and Dirk Slock. “A Geometric Interpretation of Trilateration for RSS-based Localization”. In: *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1797–1801.

[10] Abdelmoumen Norrdine. “An algebraic solution to the multilateration problem”. In: *Proceedings of the 15th international conference on indoor positioning and indoor navigation, Sydney, Australia*. Vol. 1315. 2012.

[11] Apidet Booranawong et al. “RSSI-Based Indoor Localization Using Multi-Lateration With Zone Selection and Virtual Position-Based Compensation Methods”. In: *IEEE Access* 9 (2021). Conference Name: IEEE Access, pp. 46223–46239. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3068295.

[12] Thomas Janssen, Rafael Berkvens, and Maarten Weyn. “Comparing machine learning algorithms for RSS-based localization in LPWAN”. In: *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. 2019, pp. 726–735.

[13] Kuo Yang et al. “RSS-Based Indoor Localization Using Min-Max Algorithm With Area Partition Strategy”. In: *IEEE Access* 9 (2021). Conference Name: IEEE Access, pp. 125561–125568. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3111650.

[14] Suppachai Monta, Sathaporn Promwong, and Vongkeo Kingsakda. “Evaluation of ultra wideband indoor localization with trilateration and min-max techniques”. In: *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2016, pp. 1–4.

[15] Shaoguo Xie, Yanjun Hu, and Yi Wang. “An improved E-Min-Max localization algorithm in wireless sensor networks”. In: *2014 IEEE International Conference on Consumer Electronics-China*. IEEE, 2014, pp. 1–4.

[16] Fernando Seco et al. “A survey of mathematical methods for indoor localization”. In: *2009 IEEE International Symposium on Intelligent Signal Processing*. Aug. 2009, pp. 9–14. DOI: 10.1109/WISP.2009.5286582.

[17] Nasir Saeed et al. “A State-of-the-Art Survey on Multidimensional Scaling-Based Localization Techniques”. In: *IEEE Communications Surveys Tutorials* 21.4 (2019). Conference Name: IEEE Communications Surveys Tutorials, pp. 3565–3583. ISSN: 1553-877X. DOI: 10.1109/COMST.2019.2921972.

[18] HaiQing Zhai and Yang Zhang. “A recursive weighted least squares optimization algorithm based on RSS in wireless sensor networks”. en. In: *Internet Technology Letters* 4.6 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/itl2.313, e313. ISSN: 2476-1508. DOI: 10.1002/itl2.313. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/itl2.313 (visited on 04/11/2022).

[19] SeYoung Kang, TaeHyun Kim, and WonZoo Chung. “Hybrid RSS/AOA Localization using Approximated Weighted Least Square in Wireless Sensor Networks”. en. In: *Sensors* 20.4 (Jan. 2020). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 1159. ISSN: 1424-8220. DOI: 10.3390/s20041159. URL: https://www.mdpi.com/1424-8220/20/4/1159 (visited on 04/11/2022).

[20] Xinrong Li. “RSS-based location estimation with unknown pathloss model”. In: *IEEE Transactions on Wireless Communications* 5.12 (2006). Publisher: IEEE, pp. 3626–3633.

[21] Yinghao Sun et al. “Robust RSS-Based Source Localization With Unknown Model Parameters in Mixed LOS/NLOS Environments”. In: *IEEE Transactions on Vehicular Technology* 70.4 (Apr. 2021). Conference Name: IEEE Transactions on Vehicular Technology, pp. 3926–3931. ISSN: 1939-9359. DOI: 10.1109/TVT.2021.3064444.

[22] Zengfeng Wang et al. “Cooperative RSS-Based Localization in Wireless Sensor Networks Using Relative Error Estimation and Semidefinite Programming”. In: *IEEE Transactions on Vehicular Technology* 68.1 (Jan. 2019). Conference Name: IEEE Transactions on Vehicular Technology, pp. 483–497. ISSN: 1939-9359. DOI: 10.1109/TVT.2018.2880991.

[23] Jiong Shi, Gang Wang, and Liping Jin. “Least Squared Relative Error Estimator for RSS Based Localization With Unknown Transmit Power”. In: *IEEE Signal Processing*

*Letters* 27 (2020). Conference Name: IEEE Signal Processing Letters, pp. 1165–1169. ISSN: 1558-2361. DOI: 10.1109/LSP.2020.3005298.

[24] Radim Zemek et al. “A Joint Estimation of Target Location and Channel Model Parameters in an IEEE 802.15.4-based Wireless Sensor Network”. In: *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*. ISSN: 2166-9589. Sept. 2007, pp. 1–5. DOI: 10.1109/PIMRC.2007.4394355.

[25] Angelo Coluccia and Fabio Ricciato. “On ML estimation for automatic RSS-based indoor localization”. In: *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*. May 2010, pp. 495–502. DOI: 10.1109/ISWPC.2010.5483724.

[26] Xiaojun Mei et al. “RSS-Based Byzantine Fault-Tolerant Localization Algorithm Under NLOS Environment”. en. In: *IEEE Communications Letters* 25.2 (Feb. 2021), pp. 474–478. ISSN: 1089-7798, 1558-2558, 2373-7891. DOI: 10.1109/LCOMM.2020.3027904. URL: https://ieeexplore.ieee.org/document/9210109/ (visited on 09/22/2021).

[27] Naiti Jiang and Ning Zhang. “Expectation Maximization-Based Target Localization From Range Measurements in Multiplicative Noise Environments”. In: *IEEE Communications Letters* 25.5 (May 2021). Conference Name: IEEE Communications Letters, pp. 1524–1528. ISSN: 1558-2558. DOI: 10.1109/LCOMM.2021.3050455.

[28] Seuk-Yen Phoong and Mohd Tahir Ismail. “A comparison between Bayesian and maximum likelihood estimations in estimating finite mixture model for financial data”. In: *Sains Malaysiana* 44.7 (2015), pp. 1033–1039.

[29] Di Jin et al. “Bayesian cooperative localization using received signal strength with unknown path loss exponent: Message passing approaches”. In: *IEEE Transactions on Signal Processing* 68 (2020). Publisher: IEEE, pp. 1120–1135.

[30] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[31] Alessandro Benini, Adriano Mancini, and Sauro Longhi. “An IMU/UWB/Vision-based Extended Kalman Filter for Mini-UAV Localization in Indoor Environment using 802.15.4a Wireless Sensor Network”. en. In: *Journal of Intelligent & Robotic Systems* 70.1-4

(Apr. 2013), pp. 461–476. ISSN: 0921-0296, 1573-0409. DOI: 10.1007/s10846-012-9742-1. URL: http://link.springer.com/10.1007/s10846-012-9742-1 (visited on 08/30/2021).

[32] Bing-Fei Wu and Cheng-Lung Jen. "Particle-Filter-Based Radio Localization for Mobile Robots in the Environments With Low-Density WLAN APs". In: *IEEE Transactions on Industrial Electronics* 61.12 (Dec. 2014). Conference Name: IEEE Transactions on Industrial Electronics, pp. 6860–6870. ISSN: 1557-9948. DOI: 10.1109/TIE.2014.2327553.

[33] Haohao Yin et al. "UWB-based indoor high precision localization system with robust unscented Kalman filter". In: *2016 IEEE International Conference on Communication Systems (ICCS)*. Dec. 2016, pp. 1–6. DOI: 10.1109/ICCS.2016.7833646.

[34] E.A. Wan and R. Van Der Merwe. "The unscented Kalman filter for nonlinear estimation". en. In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*. Lake Louise, Alta., Canada: IEEE, 2000, pp. 153–158. ISBN: 978-0-7803-5800-3. DOI: 10.1109/ASSPCC.2000.882463. URL: http://ieeexplore.ieee.org/document/882463/ (visited on 04/12/2022).

[35] Johannes Schmitz, Florian Schröder, and Rudolf Mathar. "TDOA fingerprinting for localization in non-line-of-sight and multipath environments". In: *2015 International Symposium on Antennas and Propagation (ISAP)*. Nov. 2015, pp. 1–4.

[36] Jingzhi Tan and Hui Zhao. "UAV Localization with Multipath Fingerprints and Machine Learning in Urban NLOS Scenario". In: *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. Dec. 2020, pp. 1494–1499. DOI: 10.1109/ICCC51575.2020.9345143.

[37] Lei Yu et al. "Fingerprinting localization based on neural networks and ultra-wideband signals". In: *2011 IEEE international symposium on signal processing and information technology (ISSPIT)*. IEEE, 2011, pp. 184–189.

[38] Geun Yeop Ha et al. “LoRa ToA-based localization using fingerprint method”. In: *2019 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2019, pp. 349–353.

[39] Marcelo Nogueira de Sousa and Reiner S. Thomä. “Applying Random Forest and Multipath Fingerprints to Enhance TDOA Localization Systems”. In: *IEEE Antennas and Wireless Propagation Letters* 18.11 (Nov. 2019). Conference Name: IEEE Antennas and Wireless Propagation Letters, pp. 2316–2320. ISSN: 1548-5757. DOI: 10.1109/LAWP.2019.2934466.

[40] Chen Wei et al. “Joint AOA-RSS Fingerprint Based Localization for Cell-Free Massive MIMO Systems”. In: *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. Dec. 2020, pp. 590–595. DOI: 10.1109/ICCC51575.2020.9344979.

[41] Jiajun He and Hing Cheung So. “A Hybrid TDOA-Fingerprinting-Based Localization System for LTE Network”. In: *IEEE Sensors Journal* 20.22 (Nov. 2020). Conference Name: IEEE Sensors Journal, pp. 13653–13665. ISSN: 1558-1748. DOI: 10.1109/JSEN.2020.3004179.

[42] Chenglong Li et al. “CRLB-based Positioning Performance of Indoor Hybrid AoA/RSS/ToF Localization”. en. In: *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Pisa, Italy: IEEE, Sept. 2019, pp. 1–6. ISBN: 978-1-72811-788-1. DOI: 10.1109/IPIN.2019.8911771. URL: https://ieeexplore.ieee.org/document/8911771/ (visited on 04/13/2022).

[43] Jukka Talvitie, Markku Renfors, and Elena Simona Lohan. “Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals”. In: *IEEE transactions on vehicular technology* 64.4 (2015). Publisher: IEEE, pp. 1340–1353.

[44] Jingxue Bi et al. “Fast Radio Map Construction by using Adaptive Path Loss Model Interpolation in Large-Scale Building”. en. In: *Sensors* 19.3 (Jan. 2019). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 712. ISSN: 1424-8220. DOI:

10.3390/s19030712. URL: https://www.mdpi.com/1424-8220/19/3/712 (visited on 03/24/2022).

[45] Wei Sun et al. “Augmentation of Fingerprints for Indoor WiFi Localization Based on Gaussian Process Regression”. In: *IEEE Transactions on Vehicular Technology* 67.11 (Nov. 2018). Conference Name: IEEE Transactions on Vehicular Technology, pp. 10896–10905. ISSN: 1939-9359. DOI: 10.1109/TVT.2018.2870160.

[46] Simon Yiu et al. “Wireless RSSI fingerprinting localization”. en. In: *Signal Processing* 131 (Feb. 2017), pp. 235–244. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2016.07.005. URL: https://www.sciencedirect.com/science/article/pii/S0165168416301566 (visited on 03/24/2022).

[47] Piyush Tiwary, Ankur Pandey, and Sudhir Kumar. “Differential d-Vectors for RSS based Localization in Dynamic IoT Networks”. In: *2021 International Conference on COMmunication Systems NETworkS (COMSNETS)*. ISSN: 2155-2509. Jan. 2021, pp. 82–85. DOI: 10.1109/COMSNETS51098.2021.9352896.

[48] Yimao Sun et al. “An Improved Closed-Form Solution for Differential RSS-based Localization”. In: *2020 IEEE Radar Conference (RadarConf20)*. ISSN: 2375-5318. Sept. 2020, pp. 1–5. DOI: 10.1109/RadarConf2043947.2020.9266537.

[49] Mikkel Baun Kjærgaard. “Indoor location fingerprinting with heterogeneous clients”. en. In: *Pervasive and Mobile Computing* 7.1 (Feb. 2011), pp. 31–43. ISSN: 15741192. DOI: 10.1016/j.pmcj.2010.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S157411921000043X (visited on 01/26/2022).

[50] Xuming Fang et al. “Optimal weighted K-nearest neighbour algorithm for wireless sensor network fingerprint localisation in noisy environment”. In: *Iet Communications* 12.10 (2018). Publisher: IET, pp. 1171–1177.

[51] Boyuan Wang et al. “A Novel Weighted KNN Algorithm Based on RSS Similarity and Position Distance for Wi-Fi Fingerprint Positioning”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 30591–30602. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2973212.

[52] Hui Zhang et al. “Weighted Adaptive KNN Algorithm with Historical Information Fusion for Fingerprint Positioning”. In: *IEEE Wireless Communications Letters* (2022). Conference Name: IEEE Wireless Communications Letters, pp. 1–1. ISSN: 2162-2345. DOI: 10.1109/LWC.2022.3152610.

[53] Shuzhi Liu, Rashmi Sharan Sinha, and Seung-Hoon Hwang. “Clustering-Based Noise Elimination Scheme for Data Pre-Processing for Deep Learning Classifier in Fingerprint Indoor Positioning System”. en. In: *Sensors* 21.13 (Jan. 2021). Number: 13 Publisher: Multidisciplinary Digital Publishing Institute, p. 4349. ISSN: 1424-8220. DOI: 10.3390/s21134349. URL: https://www.mdpi.com/1424-8220/21/13/4349 (visited on 03/28/2022).

[54] Xin Li et al. “Integrated WiFi/PDR/Smartphone Using an Adaptive System Noise Extended Kalman Filter Algorithm for Indoor Localization”. en. In: *ISPRS International Journal of Geo-Information* 5.2 (Feb. 2016). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 8. ISSN: 2220-9964. DOI: 10.3390/ijgi5020008. URL: https://www.mdpi.com/2220-9964/5/2/8 (visited on 03/28/2022).

[55] Ankur Pandey, Ryan Sequeira, and Sudhir Kumar. “SELE: RSS Based Siamese Embedding Location Estimator for a Dynamic IoT Environment”. In: *IEEE Internet of Things Journal* (2021). Conference Name: IEEE Internet of Things Journal, pp. 1–1. ISSN: 2327-4662. DOI: 10.1109/JIOT.2021.3098356.

[56] Yongxing Wang et al. “Target Positioning Algorithm Based on RSS Fingerprints of SVM of Fuzzy Kernel Clustering”. en. In: *Wireless Personal Communications* 119.4 (Aug. 2021), pp. 2893–2911. ISSN: 0929-6212, 1572-834X. DOI: 10.1007/s11277-021-08377-4. URL: https://link.springer.com/10.1007/s11277-021-08377-4 (visited on 03/02/2022).

[57] Amira Chriki, Haifa Touati, and Hichem Snoussi. “SVM-based indoor localization in Wireless Sensor Networks”. In: *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. ISSN: 2376-6506. June 2017, pp. 1144–1149. DOI: 10.1109/IWCMC.2017.7986446.

[58] Kenny Fong Peng Wye et al. “RSS-Based Fingerprinting Localization with Artificial Neural Network”. In: 1755.1 (2021), p. 012033.

[59] Daoud Burghal et al. “A comprehensive survey of machine learning based localization with wireless signals”. In: *arXiv preprint arXiv:2012.11171* (2020).

[60] Rashmi Sharan Sinha and Seung-Hoon Hwang. “Comparison of CNN Applications for RSSI-Based Fingerprint Indoor Localization”. en. In: *Electronics* 8.9 (Sept. 2019). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 989. ISSN: 2079-9292. DOI: 10.3390/electronics8090989. URL: https://www.mdpi.com/2079-9292/8/9/989 (visited on 03/28/2022).

[61] Xudong Song et al. “A Novel Convolutional Neural Network Based Indoor Localization Framework With WiFi Fingerprinting”. In: *IEEE Access* 7 (2019). Conference Name: IEEE Access, pp. 110698–110709. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2933921.

[62] Brahim El Boudani et al. “Implementing deep learning techniques in 5G IoT networks for 3D indoor positioning: DELTA (DeEp Learning-Based Co-operaTive Architecture)”. In: *Sensors* 20.19 (2020). Publisher: Multidisciplinary Digital Publishing Institute, p. 5495.

[63] Anetta Nagy et al. “RSS-based Localization for Directional Antennas”. en. In: *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Vienna, Austria: IEEE, Sept. 2020, pp. 774–781. ISBN: 978-1-72818-956-7. DOI: 10.1109/ETFA46521.2020.9212055. URL: https://ieeexplore.ieee.org/document/9212055/ (visited on 10/25/2021).

[64] Javier Schloemann and R. Michael Buehrer. “On the Value of Collaboration in Location Estimation”. In: *IEEE Transactions on Vehicular Technology* 65.5 (May 2016). Conference Name: IEEE Transactions on Vehicular Technology, pp. 3585–3596. ISSN: 1939-9359. DOI: 10.1109/TVT.2015.2442173.

[65] U. Nazir et al. “Classification of localization algorithms for wireless sensor network: A survey”. en. In: *2012 International Conference on Open Source Systems and Technologies*. Lahore, Pakistan: IEEE, Dec. 2012, pp. 1–5. ISBN: 978-1-4673-3097-8 978-1-4673-3094-7 978-1-4673-3096-1. DOI: 10.1109/ICOSST.2012.6472830. URL: http://ieeexplore.ieee.org/document/6472830/ (visited on 09/22/2021).

[66] Chih-Yu Wen and Yu-Cheng Hsiao. “Decentralized anchor-free localization for wireless ad-hoc sensor networks”. In: *2008 IEEE International Conference on Systems, Man and Cybernetics*. ISSN: 1062-922X. Oct. 2008, pp. 2777–2785. DOI: 10.1109/ICSMC.2008.4811717.

[67] Faheem Zafari, Athanasios Gkelias, and Kin Leung. “A Survey of Indoor Localization Systems and Technologies”. en. In: *arXiv:1709.01015 [cs]* (Jan. 2019). arXiv: 1709.01015. URL: http://arxiv.org/abs/1709.01015 (visited on 09/22/2021).

[68] Jyoti Kumari, Prabhat Kumar, and Sunil Kumar Singh. “Localization in three-dimensional wireless sensor networks: a survey”. en. In: *The Journal of Supercomputing* 75.8 (Aug. 2019), pp. 5040–5083. ISSN: 0920-8542, 1573-0484. DOI: 10.1007/s11227-019-02781-1. URL: http://link.springer.com/10.1007/s11227-019-02781-1 (visited on 02/17/2022).

[69] Safar M. Maghdid. “A Comprehensive Review of Indoor/Outdoor Localization Solutions in IoT era: Research Challenges and Future Perspectives”. In: *TechRxiv. Preprint* (2021). Publisher: TechRxiv.

[70] Andrew J Davison et al. “MonoSLAM: Real-time single camera SLAM”. In: *IEEE transactions on pattern analysis and machine intelligence* 29.6 (2007), pp. 1052–1067.

[71] Georg Klein and David Murray. “Parallel tracking and mapping for small AR workspaces”. In: *2007 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE. 2007, pp. 225–234.

[72] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. “ORB-SLAM: a versatile and accurate monocular SLAM system”. In: *IEEE transactions on robotics* 31.5 (2015), pp. 1147–1163.

[73] Raul Mur-Artal and Juan D Tardós. “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras”. In: *IEEE transactions on robotics* 33.5 (2017), pp. 1255–1262.

[74] Carlos Campos et al. “Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam”. In: *IEEE Transactions on Robotics* 37.6 (2021), pp. 1874–1890.

[75] Ethan Rublee et al. “ORB: An efficient alternative to SIFT or SURF”. In: *2011 International conference on computer vision*. Ieee. 2011, pp. 2564–2571.

[76] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. “DTAM: Dense tracking and mapping in real-time”. In: *2011 international conference on computer vision*. IEEE. 2011, pp. 2320–2327.

[77] Jakob Engel, Thomas Schöps, and Daniel Cremers. “LSD-SLAM: Large-scale direct monocular SLAM”. In: *European conference on computer vision*. Springer. 2014, pp. 834–849.

[78] Jakob Engel, Vladlen Koltun, and Daniel Cremers. “Direct sparse odometry”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.3 (2017), pp. 611–625.

[79] Christian Forster et al. “SVO: Semidirect visual odometry for monocular and multicamera systems”. In: *IEEE Transactions on Robotics* 33.2 (2016), pp. 249–265.

[80] Jon Zubizarreta, Iker Aguinaga, and Jose Maria Martinez Montiel. “Direct sparse mapping”. In: *IEEE Transactions on Robotics* 36.4 (2020), pp. 1363–1370.

[81] Anastasios I. Mourikis and Stergios I. Roumeliotis. “A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (2007), pp. 3565–3572.

[82] Choudhury Paul and Stergios I. Roumeliotis. “MSCKF 2.0: Multi-state constraint Kalman filter with observability-based consistent EKF for robust visual-inertial navigation”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (2017), pp. 2421–2428.

[83] Michael Bloesch et al. “Robust visual inertial odometry using a direct EKF-based approach”. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2015), pp. 298–304.

[84] Stefan Leutenegger et al. “Keyframe-based visual–inertial odometry using nonlinear optimization”. In: *The International Journal of Robotics Research* 34.3 (2015), pp. 314–334.

[85] Tong Qin, Peiliang Li, and Shaojie Shen. “Vins-mono: A robust and versatile monocular visual-inertial state estimator”. In: *IEEE Transactions on Robotics* 34.4 (2018), pp. 1004–1020.

[86] Tong Qin et al. “A general optimization-based framework for global pose estimation with multiple sensors”. In: *arXiv preprint arXiv:1901.03642* (2019).

[87] Christian Gentner et al. “Multipath assisted positioning with simultaneous localization and mapping”. In: *IEEE Transactions on Wireless Communications* 15.9 (2016), pp. 6104–6117. DOI: 10.1109/TWC.2016.2574999.

[88] Erik Leitinger et al. “Factor graph based simultaneous localization and mapping using multipath channel information”. In: *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE. 2017, pp. 652–658.

[89] Erik Leitinger et al. “A belief propagation algorithm for multipath-based SLAM”. In: *IEEE transactions on wireless communications* 18.12 (2019), pp. 5613–5629.

[90] Xinghe Chu et al. “Vehicle localization via cooperative channel mapping”. In: *IEEE Transactions on Vehicular Technology* 70.6 (2021), pp. 5719–5733.

[91] Rico Mendrzik, Henk Wymeersch, and Gerhard Bauch. “Joint localization and mapping through millimeter wave MIMO in 5G systems”. In: *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2018, pp. 1–6.

[92] Guillermo Bielsa et al. “Indoor localization using commercial off-the-shelf 60 GHz access points”. In: *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE. 2018, pp. 2384–2392.

[93] Panagiotis T Karfakis, Micael S Couceiro, and David Portugal. “Nr5g-sam: A slam framework for field robot applications based on 5g new radio”. In: *Sensors* 23.11 (2023), p. 5354.

[94] Ruben Morales Ferre, Gonzalo Seco-Granados, and Elena Simona Lohan. “Positioning Reference Signal design for positioning via 5G”. In: *National Committee for Radiology in Finland* (2019).

[95] José A del Peral-Rosado et al. “Feasibility study of 5G-based localization for assisted driving”. In: *2016 International conference on localization and GNSS (ICL-GNSS)*. IEEE. 2016, pp. 1–6.

[96] Sharief Saleh, Amr S El-Wakeel, and Aboelmagd Noureldin. “5G-Enabled Vehicle Positioning Using EKF With Dynamic Covariance Matrix Tuning”. In: *IEEE Canadian Journal of Electrical and Computer Engineering* 45.3 (2022), pp. 192–198.

[97] Jukka Talvitie et al. “Positioning of high-speed trains using 5G new radio synchronization signals”. In: *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2018, pp. 1–6.

[98] Zhaohan Zhang et al. “AoA-and-Amplitude Fingerprint Based Indoor Intelligent Localization Scheme for 5G Wireless Communications”. In: *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE. 2021, pp. 1–5.

[99] Kimia Shamaei and Zaher M Kassas. “Receiver design and time of arrival estimation for opportunistic localization with 5G signals”. In: *IEEE Transactions on Wireless Communications* 20.7 (2021), pp. 4716–4731.

[100] Satyam Dwivedi et al. “Positioning in 5G networks”. In: *arXiv:2102.03361 [cs]* (Feb. 2021). arXiv: 2102.03361. URL: http://arxiv.org/abs/2102.03361 (visited on 04/06/2022).

[101] Samira Hayat et al. “Edge computing in 5G for drone navigation: What to offload?” In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2571–2578.

[102] José A del Peral-Rosado et al. “Feasibility study of 5G-based localization for assisted driving”. In: *2016 International conference on localization and GNSS (ICL-GNSS)*. IEEE. 2016, pp. 1–6.

[103] Sharief Saleh, Amr S El-Wakeel, and Aboelmagd Noureldin. “5G-Enabled Vehicle Positioning Using EKF With Dynamic Covariance Matrix Tuning Positionnement de véhicules à l’aide de la 5G utilisant un EKF avec réglage dynamique de la matrice de covariance”. In: *IEEE Canadian Journal of Electrical and Computer Engineering* (2022).

[104] Mengguan Pan et al. “A Joint DOA and TOA Estimation Scheme for 5G Signals Under Array Modeling Errors”. In: ().

[105] Stephan Jaeckel et al. “QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials”. In: *IEEE transactions on antennas and propagation* 62.6 (2014), pp. 3242–3256.

[106] Mengguan Pan et al. “A Low-Complexity Joint AOA and TOA Estimation Method for Positioning with 5G Signals”. In: ().

[107] Maximilian Stahlke et al. “Transfer Learning to adapt 6G AI-based Fingerprint Localization across Environments”. In: *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*. IEEE. 2022, pp. 1–5.

[108] Jiayi Meng et al. “A study of network-side 5G user localization using angle-based fingerprints”. In: *2020 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN*. IEEE. 2020, pp. 1–6.

[109] Estifanos Yohannes Menta et al. “On the performance of AoA–based localization in 5G ultra–dense networks”. In: *IEEE Access* 7 (2019), pp. 33870–33880.

[110] Amal Sellami, Leila Nasraoui, and Leila Najjar Atallah. “Multi-stage localization for massive MIMO 5G systems”. In: *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE. 2020, pp. 1–6.

[111] Roman Klus, Jukka Talvitie, and Mikko Valkama. “Neural network fingerprinting and GNSS data fusion for improved localization in 5G”. In: *2021 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE. 2021, pp. 1–6.

[112] Fuxi Wen et al. “A survey on 5G massive MIMO localization”. In: *Digital Signal Processing* 94 (2019), pp. 21–28.

[113] Zhongliang Deng et al. “A TDOA and PDR fusion method for 5G indoor localization based on virtual base stations in unknown areas”. In: *IEEE Access* 8 (2020), pp. 225123–225133.

[114] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC press, 2012.

[115] Thomas G Dietterich. “Ensemble methods in machine learning”. In: *International workshop on multiple classifier systems*. Springer. 2000, pp. 1–15.

[116] Antonio Albanese, Vincenzo Sciancalepore, and Xavier Costa-Pérez. “First Responders Got Wings: UAVs to the Rescue of Localization Operations in Beyond 5G Systems”. In: *IEEE Communications Magazine* 59.11 (Nov. 2021). arXiv: 2109.03180, pp. 28–34. ISSN: 0163-6804, 1558-1896. DOI: 10.1109/MCOM.101.2100273. URL: http://arxiv.org/abs/2109.03180 (visited on 04/04/2022).

[117] Peng Huang et al. “Edge Robotics: Edge-Computing-Accelerated Multi-Robot Simultaneous Localization and Mapping”. In: *IEEE Internet of Things Journal* (2022).

[118] Peng Huang et al. “ColaSLAM: Real-Time Multi-Robot Collaborative Laser SLAM via Edge Computing”. In: *2021 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE. 2021, pp. 242–247.

[119] Arthur Charléty et al. “2D Phase-based RFID localization for on-site landslide monitoring”. In: *Remote Sensing* 14.15 (2022), p. 3577.

[120] Priyanka Sinha and Ismail Guvenc. “Impact of Antenna Pattern on TOA Based 3D UAV Localization Using a Terrestrial Sensor Network”. In: *IEEE Transactions on Vehicular Technology* (2022).

[121] Nozhan Hosseini et al. “UAV command and control, navigation and surveillance: A review of potential 5G and satellite systems”. In: *2019 IEEE Aerospace Conference*. IEEE. 2019, pp. 1–10.

[122] S Jaeckel et al. “QuaDRiGa-quasi deterministic radio channel generator, user manual and documentation”. In: *Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2. 0.0* (2017).

[123] Qiuming Zhu et al. “3GPP TR 38.901 channel model”. In: *the wiley 5G Ref: the essential 5G reference online*. Wiley Press Hoboken, NJ, USA, 2021, pp. 1–35.

[124] Zihao Fu et al. “Channel simulation and validation by QuaDRiGa for suburban microcells under 6 GHz”. In: *2018 12th international symposium on antennas, propagation and EM theory (ISAPE)*. IEEE. 2018, pp. 1–4.

[125] Stephan Jaeckel et al. “Industrial indoor measurements from 2-6 GHz for the 3GPP-NR and QuaDRiGa channel model”. In: *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE. 2019, pp. 1–7.

[126] Peter J Huber. “Robust estimation of a location parameter”. In: *The Annals of Mathematical Statistics* 35.1 (1964), pp. 73–101.

[127] Averil Burton Chatfield. *Fundamentals of high accuracy inertial navigation*. Vol. 174. Aiaa, 1997.

[128] Anastasios I. Mourikis and Stergios I. Roumeliotis. “A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation”. In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. 2007, pp. 3565–3572. DOI: 10.1109/ROBOT.2007.364024.

[129] Surachai Panich. “Indirect kalman filter in mobile robot application”. In: *J. Math. Stat* 6.381384.43 (2010).

[130] Frank Dellaert and Michael Kaess. *Factor Graphs for Robot Perception*. Foundations and Trends in Robotics, Vol. 6, 2017. URL: http://www.cs.cmu.edu/~kaess/pub/Dellaert17fnt.pdf.

[131] Christian Forster et al. *IMU preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation*. 2015.

[132] Michael Kaess et al. “iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering”. In: *2011 IEEE International Conference on Robotics and Automation*. 2011, pp. 3281–3288. DOI: 10.1109/ICRA.2011.5979641.

[133] Frank Dellaert and GTSAM Contributors. *borglab/gtsam*. Version 4.2a8. May 2022. DOI: 10.5281/zenodo.5794541. URL: https://github.com/borglab/gtsam).

[134] Yuichi Tazaki. “A Spanning Tree-Based Multi-Resolution Approach for Pose-Graph Optimization”. In: *IEEE Robotics and Automation Letters* 7.4 (2022), pp. 10033–10040.

[135] Michael Burri et al. “The EuRoC micro aerial vehicle datasets”. In: *The International Journal of Robotics Research* (2016). DOI: 10.1177/0278364915620033. eprint: http://ijr.sagepub.com/content/early/2016/01/21/0278364915620033.full.pdf+html. URL: http://ijr.sagepub.com/content/early/2016/01/21/0278364915620033.abstract.

[136] 3GPP. *Study on channel model for frequencies from 0.5 to 100 GHz*. 2018.

[137] Letícia Carneiro de Souza et al. “A study on propagation models for 60 GHz signals in indoor environments”. In: *Frontiers in Communications and Networks* 2 (2022), p. 757842.

[138] David Prokhorov et al. “Measuring robustness of visual slam”. In: *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE. 2019, pp. 1–6.

[139] Ian Sharp, Kegen Yu, and Y Jay Guo. “GDOP analysis for positioning system design”. In: *IEEE Transactions on Vehicular Technology* 58.7 (2009), pp. 3371–3382.

[140] Chih-Hung Wu, Wei-Han Su, and Ya-Wei Ho. “A Study on GPS GDOP Approximation Using Support-Vector Machines”. In: *IEEE Transactions on Instrumentation and Measurement* 60.1 (2011), pp. 137–145. DOI: 10.1109/TIM.2010.2049228.