



XRLoc: Accurate UWB Localization to Realize XR Deployments

Aditya Arun*

University of California San Diego
California, United States
aarun@ucsd.edu

Sureel Shah

University of California San Diego
California, United States
sbs001@ucsd.edu

Shunsuke Saruwatari*

Osaka University
Osaka, Japan
saru@ist.osaka-u.ac.jp

Dinesh Bharadia

University of California San Diego
California, United States
dineshb@eng.ucsd.edu

Abstract

Understanding the location of ultra-wideband (UWB) tag-attached objects and people in the real world is vital to enabling a smooth cyber-physical transition. However, most UWB localization systems today require multiple anchors in the environment, which can be very cumbersome to set up. In this work, we develop XRLoc, providing an accuracy of a few centimeters in many real-world scenarios. This paper will delineate the key ideas that allow us to overcome the fundamental restrictions that plague a single anchor point from localization of a device to within an error of a few centimeters. We deploy a VR chess game using everyday objects as a demo and find that our system achieves 2.4 cm median accuracy and 5.3 cm 90th percentile accuracy in dynamic scenarios, performing at least 8× better than state-of-art localization systems. Additionally, we implement a MAC protocol to furnish these locations for over 10 tags at update rates of 100 Hz, with a localization latency of ~1 ms. We have additionally open-sourced our system's codebase at <https://github.com/ucsdwcsng/xrloc.git>

CCS Concepts

• **Networks** → **Cyber-physical networks**; **Location based services**; • **Human-centered computing** → *Ubiquitous computing*.

Keywords

Indoor Localization; UWB; Single Anchor; TDoA PDoA Combination; XR;

ACM Reference Format:

Aditya Arun, Shunsuke Saruwatari, Sureel Shah, and Dinesh Bharadia. 2023. XRLoc: Accurate UWB Localization to Realize XR Deployments. In *The 21st ACM Conference on Embedded Networked Sensor Systems (SenSys '23)*, November 12–17, 2023, Istanbul, Turkiye. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3625687.3625810>

1 Introduction

Extended Reality (XR), broadly encompassing virtual, augmented, and mixed reality technologies, can potentially revolutionize fields

*Both authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

SenSys '23, November 12–17, 2023, Istanbul, Turkiye

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-84007-0414-7/23/11.

<https://doi.org/10.1145/3625687.3625810>

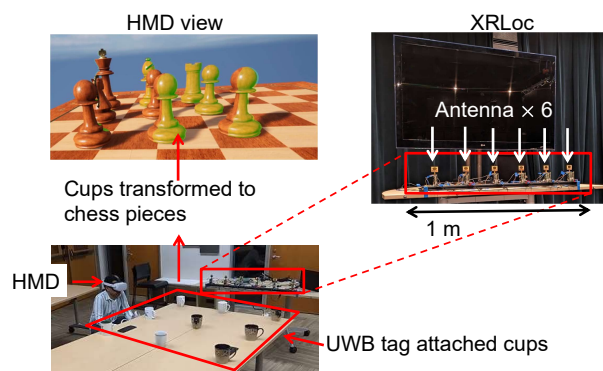


Figure 1: XRLoc enables users to play a life-size chess game with everyday objects. XRLoc localizes mugs retrofitted with off-the-shelf UWB tags from a single vantage point with a few cm of location accuracy, which are then translated to chess pieces in the virtual world.

such as education, healthcare, and gaming [5, 91, 105]. The primary ethos for XR is to provide immersive, interactive, and realistic experiences for users. For example, everyday spaces and objects can be transformed into video game assets (like tennis racquets, swords, or chess pieces) for interactive gaming applications in virtual reality.¹ Or imagine beamforming speakers streaming dedicated and spatial audio to various users to enhance mixed reality applications [42]. Or consider enabling early warning of potential hazards in an environment at a construction site or factory floor via augmented reality [41]. To enable these applications, we find a common thread — any XR system should localize and track objects or users in an environment. Specifically, this object-tracking system needs to satisfy three key requirements to realize XR applications:

R1. Ease of anchor deployment: Any asset localization system must have low deployment efforts, which can potentially be embedded within common electronics like TVs or soundbars. This single module should be smaller than 1 m.²

R2. Accurate and reliable: Assets must be localized to an accuracy within a few centimeters in room-scale scenarios (~5 m). We place a stringent requirement of a few centimeters of accuracy to provide a glitch-free user experience. Providing immersive XR experiences consequently means small user or object tracking errors are more obvious and severely impede the adoption of XR [103]. Specifically,

¹ In this demo, we transform the mugs and a desk in our lab into a life-size chess board (link:<https://youtu.be/WMu4kbsyWE>)

² Most consumer electronics like TVs or soundbars are around 1 m in length.

	Visual	Acoustic	Radar	RFID	Single anchor	XRLoc
R1: Ease of anchor deployment	✓	✓	✓	×	✓	✓
R2: Accuracy and reliability	×	✓	×	✓	×	✓
R3: Multi-asset and low latency	✓	×	✓	✓	✓	✓

Table 1: Existing technologies do not satisfy the 3 key requirements for an XR localization/tracking system.

the localization system must be reliable during movement, under occlusions, and consistently track assets within an accuracy of a few cm.

R3. Multi-asset low latency localization: Finally, an XR system needs to localize multiple objects in an environment in real-time. In dynamic scenarios, this can mean we must localize tens of objects at a rate at which humans visually perceive their surroundings (60–80 Hz) [22], and delays in updates of object locations in a dynamic scenario can break away from an immersive experience.

However, none of the existing asset localization systems meet these three key requirements to deliver XR applications in everyday scenarios (see Table 1). Camera and visual sensors are susceptible to poor lighting and visual occlusions, consequently failing to provide reliable localization (**R2**). Additionally, deploying a camera-based system can be privacy invasive [75] in home and public settings. Acoustic systems [53] provide accurate localization but do not simultaneously localize multiple assets with low latency (**R3**). Radar systems [44, 65, 108] can provide low-latency object tracking from a single module but fail to track occluded objects or those that have small radar cross-sections (RCS). Some RFID systems have succeeded in realizing low latency [55, 82, 111]. Their asymmetric architecture (cost-effective tags and expensive readers) better suits large-scale deployments in retail and industrial sectors. However, long-range RFID systems (~5 m) are expensive and bulky to integrate into consumer electronics, precluding wide-scale deployments (**R1**).

Alternatively, many single RF module localization solutions [16, 26, 28, 32, 33, 46, 51, 58, 59, 85, 100, 101, 112] leveraging WiFi/BLE or ultra-wideband (UWB) are easy to deploy because of transceivers which can be inexpensively deployed in consumer electronics. However, they fail to provide the necessary cm-level accuracy. None of the existing systems simultaneously satisfy all three stringent requirements to enable XR applications, and prior art will be more carefully considered in Sec. 8.

To address the need for XR-compliant localization, we develop XRLoc, which consists of two parts — a localization tag, attachable to objects of interest, and a single localization module to furnish few-cm level locations from a single vantage point. The localization module is less than 1 m and can be easily incorporated within everyday electronics such as televisions or soundbars (satisfying **R1**). It leverages the tag’s single UWB transmission for a few cm accurate localization. An accompanying MAC protocol also supports the localization of multiple tags at an update rate of 100 Hz (satisfying **R3**). An example deployment of XRLoc is showcased in Fig. 1, where beverage cups are attached with off-the-shelf UWB tags. XRLoc is leveraged to transform an office space into a life-sized chess board, with these cups taking the place of chess pieces and localized with cm-level accuracy. A video demo of this case study is also included as well¹. However, to simultaneously meet all the aforementioned requirements, we need to solve four key challenges:

1. Geometric dilution of precision: In most UWB localization systems, three or more UWB anchors need to be placed in diverse locations in a room to localize the UWB tag, increasing deployment efforts and breaking away from **R1**. Alternatively, we can place these UWB anchors within a single localization module constrained to a 1 m space. However, reducing the spatial diversity can worsen the localization accuracy by 10×. This accuracy degradation is called ‘geometric dilution of precision’ [87] (GDOP). A potential strategy to overcome GDOP is to borrow techniques from RFID-systems [55, 82, 111] that achieve real-time cm-scale accuracy from a single RFID reader. However, we observe RFID systems provide 15× better measurement accuracy compared to UWB systems [111], owing to an RFID system’s mono-static architecture (shared clock for transmitter and receiver). Hence, underlying the effect of GDOP is a system’s reduced resilience to measurement noise. This precludes XRLoc from directly borrowing techniques from RFID-based systems.

To reduce our measurement noise, we could increase transmit power to improve signal quality, increase transmission length for better averaging, or choose better hardware with lower noise floors. However, these solutions come at the cost of increased battery consumption at the tag, increased localization latency, or expensive tag design, respectively. Alternatively, XRLoc makes a key observation when looking at the phase difference of the received UWB signal measurements (PDoA) between a pair of anchors — PDoA measurement quality can be improved proportionally to the distance between the pair of anchors. This simple observation forms the cornerstone of XRLoc’s design and allows us to satisfy the first requirement **R1**.

2. Ambiguous location predictions: However, this improved PDoA measurement quality comes at a detrimental cost — increasing the anchor spacing creates multiple ambiguous location predictions as phase measurements wrap around at 2π . The changes in these ambiguities mirror the changes in the true location of the tag, and they do not affect tracking systems [14, 99], which leverage phases to provide cm-level tracking accuracy for handwriting recognition. However, incorrectly choosing an ambiguous absolute location can degrade the accuracy by several tens of centimeters and may create glitches within the XR system.

To predict accurate locations despite phase wrap-around, XRLoc leverages a second observation — unlike phase measurements, time of arrival measurements do not suffer from ambiguity. Specifically, the time difference of arrival (TDoA) between a pair of anchors, although inaccurate in furnishing cm-level localization, can help to detect and filter out ambiguities. By cleverly fusing these time-difference and phase-difference measurements, XRLoc can provide cm-level accurate locations from a single UWB transmission and satisfy the second requirement **R2**.

3. Measurement-bias aware localization: However, as we push the envelope on cm-accurate location predictions, we find that hardware biases can corrupt our location estimates and degrade

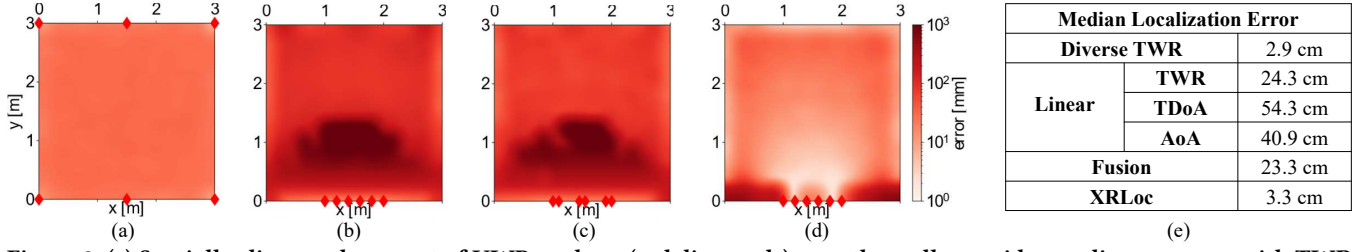


Figure 2: (a) Spatially-diverse placement of UWB anchors (red diamonds) near the walls provides median accuracy with TWR of 2.9 cm (b) when receivers are constrained near the bottom wall, median accuracy degrades by 8× when using TWR (c) fusion of TDoA, TWR, and AoA does not help in these scenarios either, providing median accuracy of 23.3 cm. (d) XRLoc solves the challenges associated with dilution of precision, achieving median accuracy of 3.3 cm (e) Summary of errors when leveraging various UWB measurements and XRLoc.

our location accuracy by over 2×. Specifically, through empirical measurements, and as observed in previous studies [20], UWB modules [21] suffer from a distance-sensitive measurement bias. We model, estimate, and calibrate for these biases via a three-point calibration procedure. We fuse the time and phase measurements with a corrected PDoA and TDoA measurement model by leveraging a particle filter to provide cm-accurate and low-latency location estimates, satisfying **R2**.

4. High update rate multi-tag operation: In addition to providing low-latency localization, XRLoc must furnish locations for multiple objects in the environment. Often, the UWB transmissions for localization from multiple tags in an environment can cause packet collisions at XRLoc’s module. The collision causes localization failure 25% of the time. We leverage a low-power wireless side channel to alleviate packet collisions to design a power-efficient medium access control (MAC) protocol. Specifically, XRLoc deploys a LoRa-based MAC to support consistent localization for tens of tags at over 80 Hz, satisfying **R3**.

XRLoc brings together these key techniques to build a 1 m sized module, consisting of 6 Decawave DW1000 [69] UWB modules for localization, along with a Semtech LoRa SX1272 [81] to furnish a side-channel for the MAC protocol. Additionally, we prototype a simple UWB + LoRa Tag using the Decawave EVB1000 and a LoRa SX1272. Through extensive evaluations, we find that XRLoc satisfies all the three stringent requirements with

- (1) Static localization error with median and 90th percentile accuracy of 1.5 cm and 5.5 cm, an improvement of 9.5× and 5.2× from state-of-art systems [113].
- (2) Dynamic localization error with median and 90th percentile accuracy of 2.4 cm and 5.3 cm, an improvement of 11× and 8× from state-of-art systems [113].
- (3) Localization failure rate of 0.5% when using the MAC protocol as compared to a failure rate of 25% without a MAC protocol, a 50× improvement, for 10 tags operating simultaneously at 100 Hz location update rate.
- (4) Location compute latency of 1 ms, allowing for real-time localization (60 Hz) of 16 tags.

2 Why is this problem hard?

We have established the need for localizing users and objects within a few centimeters of a single vantage point. In this section, we will find that restricting our sensing to within a space of 1 m reduces our geometric diversity leading to localization errors of many 10’s of

centimeters. This phenomenon is commonly referred to as geometric dilution of precision. We will explore the use of three common UWB measurements — two-way-ranging (TWR), time-difference-of-arrival (TDoA), and angle-of-arrival (AoA) — and find systems that rely on these measurements fail to furnish the required accuracy. Additionally, we’ll explore fusing and jointly optimizing for these measurements to improve localization accuracy. However, even this measurement fusion is insufficient. To test this hypothesis, we build a simple simulation environment described below.

Simulation environment: We perform extensive simulation in a 3 × 3 environment, a standard room size, to find the best case localization accuracy. We use 6 UWB transceivers, placed either diversely in the environment (red diamonds in Fig. 2 (a)) or in a limited space near the bottom wall (see Fig. 2 (b, c, d)). Next, we divide this space into a 1 mm grid and place tags in each position to measure the location accuracy. The pixels of the ‘heatmaps’ represent these tag locations, and the pixel color intensity quantifies the median localization accuracy across 100 simulated trials.

Simulating TWR: Many UWB radios measure the time of flight (ToF) of the signal between the transmitter and receiver up to a resolution of 15.6 ps [23]. The ToF is measured via multiple packet exchanges, taking at least 0.3 ms [18]. And clock drifts at the receiver during this TWR event can lead to a ToF measurement deviation of 150 ps for a 0.5 ppm clock crystal. Hence, we characterize our simulated TWR measurements with a zero-mean Gaussian with a standard deviation of 150 ps.

Simulating TDoA: Instead of an absolute time of flight measurement, we can measure the difference in the time of arrivals across a pair of synchronized receivers. However, TDoA measurements depend on the receivers’ clock synchronization accuracy. Our measurements, independently verified by Decawave [70], show clock-sync errors in best-case wired synchronization can cause a TDoA measurement deviation of 140 ps. Hence, our simulated TDoAs are Gaussian distributed with a standard deviation of 140 ps.

Simulating AoA: Some UWB systems [38, 113] alternatively measure the angle of arrival of a signal between a pair of receivers placed half-wavelength apart (see close pairs of red diamonds in Fig. 2(c)). We can measure AoA with noise deviation of 1.5°, as independently verified in [38, 113]. Consequently, we simulate our AoA measurements as zero-mean Gaussian with 1.5° standard deviation.

2.1 Quantifying localization errors

TWR, or distance measurements between a tag and multiple receivers placed diversely in an environment, can be used to trilaterate

the tag's position to achieve a few cm-level accuracy. From Fig. 2(e), we find that the median localization error is 2.9 cm. Additionally, this error is consistent (with a variation of a few centimeters) across the space (see heatmap in Fig. 2(a)). However, when we place all the receivers within a 1 m linear form factor to satisfy **R1** (ease of deployability), we find that the accuracy degrades by over 8× as compared to the diverse antenna placement. Additionally, we observe a non-uniform performance with errors as large as 1 m. To meet **R1**, we have made our localization system too erroneous to be usable.

The fundamental reason for the performance degradation is the reduced geometric diversity when the antennas are closer. With the antennas placed around the environment, trilateration is more resilient to errors in distance measurements. We similarly quantify the localization errors by leveraging TDoA or AoA measurements and summarize the results in Fig. 2(e). We find under low geometric diversity, the median localization errors are 54.4 cm and 40.9 cm for TDoA and AoA measurements, respectively.

2.2 Fusing all measurements

Similar to many robotics applications [4], we can use TWR, AoA, and TDoA measurements to provide higher accuracy. This fusion is done by jointly optimizing the error function from TWR, AoA, and TDoA [52] measurements. Specifically, in Fig. 2(c), we measure 6 TWR measurements from each receiver (red diamonds), 3 AoA measurements from each closely-spaced pair of UWB receivers, and 3 TDoA measurements between one antenna from each of these paired groups. The measurement-fusion efforts provide median localization of 23.3 cm. However, it still fails to meet our criteria of a few-cm error in localization.

None of the existing states of art systems can surmount the challenge of localizing from a single vantage point and deliver the stringent requirements set forth by our application use case. In XR-Loc, we develop the algorithm (Sec. 3) and prototype a system (Sec. 4 and 5) to achieve this small-form-factor, high accuracy (median accuracy of 3.3 cm as seen from Fig. 2(d) and Fig. 2(e)), and multi-asset localization system, for use within VR systems and immersive audio applications. In the following section, we will delineate the key ideas which allow XR-Loc to circumvent the challenges posed by geometric dilution of precision.

3 Circumventing low-spatial diversity

First, we will explore improving our phase measurements to improve location accuracy by increasing the antenna separation (Sec. 3.1). However, this comes with the unintended side-effect of introducing ambiguities to our location prediction. So, we consider the use of time difference of arrival (TDoA) measurements to combat these ambiguities (Sec. 3.2). Finally, we explore fusing these measurements in an accurate and low-latency fashion by leveraging a particle filter (Sec. 3.3). By exploring the key ideas here, XR-Loc will fulfill **R2** and furnish few-cm level localization.

3.1 Improving localization resolution

As discussed in Sec. 2, we reduce our resiliency to noise when we try to localize tags from a single vantage point. Lacking spatial diversity adds vulnerability to the optimization creating large outlier measurements and preventing few-cm scale localization. However, when we have two closely (less than half-wavelength) separated

antennas, we can find the phase difference ($\Delta\phi$) between this pair as

$$\Delta\phi = \frac{2\pi d}{\lambda} \sin(\theta)$$

where θ is the incoming angle of arrival w.r.t. to the normal of this pair of antennas, $d = \frac{\lambda}{2}$ is the distance between them, and λ is the wavelength at 3.5 GHz UWB center frequency.³ However, the typical UWB phase has a resolution of around 8 bits,⁴ which provides a phase resolution of 1.4°, and consequently a localization resolution of 2.1 cm at a distance of 3 m from the localization module. However, increasing the inter-antenna separation, d , linearly increases the measured phase difference. We can leverage this to improve our localization resolution to the ~1 mm limit when the antenna separation is 1 m.

Prior works [14, 99] have leveraged this fact to increase accuracy for handwriting tracking purposes. But, widening this separation comes at the cost of introducing more phase ambiguities. This is apparent when we return to the AoA equation and observe that our phase-difference measurements, $\Delta\phi$, wrap over 2π for a larger separation than half-wavelength separation for angles between $-90^\circ < \theta < 90^\circ$. This is not an issue for tracking purposes, where the changes in location of these ambiguities mirror the true changes in the location and continue to provide a similar trajectory estimate. However, for XR-Loc, we find predicting and tracking these incorrect locations can degrade the localization accuracy by an order of magnitude to several tens of centimeters.

3.2 Ruling out ambiguities

To overcome ambiguities, a simple solution is adding more antennas between the two we have placed so far. These additional antennas will help eliminate phase ambiguities by reducing the consecutive antenna distance while employing a 1-m antenna array aperture. Fig. 3 (a, b) depicts these ambiguities that exist in such a system by showing the likely positions of the tag. Considering the simulation environment from Sec. 2, we deploy two arrays with spacing 33.3 cm and 25 cm for the same antenna aperture of 1 m. Next, we deploy a tag at the center of the space and predict its potential locations (pixel color intensities) in both scenarios. We observe that keeping the same aperture of 1 m, we have similar measurement errors (peak widths) in both cases, consistent with our previous findings, but reducing separation creates fewer ambiguities. Deploying 23 antennas within this 1 m, each spaced half-wavelength apart, will remove all our ambiguities at the cost of increased hardware complexity.

Alternatively, we observe TDoA measurements are free from ambiguities and can potentially be leveraged to disambiguate the predictions from PDoA. Similarly to the previous PDoA images, in Fig. 3 (c, d), we only show the tag's location likelihoods when relying on TDoA measurements. The TDoA peak, although very erroneous (larger peak widths), is unambiguous. Additionally, increasing the number of antennas reduces this error/peak width. To recap, by reducing the antenna separation (or increasing the number of antennas), we increase the separations between the ambiguities coming from PDoA measurements and tighten our peak

³We develop this intuition assuming far field, but later in Sec. 3.3 we consider the exact phase difference measurement

⁴Although DW1000 have a 16-bit real and imaginary part in each CIR sample, it has a to be correct by a 7-bit RCPHASE measurement, reducing its overall phase resolution.

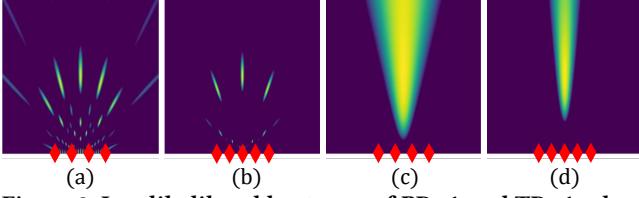


Figure 3: Log-likelihood heat map of PDoA and TDoA when changing the number of antennas N .

widths coming from TDoA. Consequently, at the correct antenna spacing, our ambiguous peaks will be wide enough to be rejected by our TDoA measurements. We find this sweet spot when we use 6 antennas, 4× fewer antennas than would have otherwise been required.

3.3 Jointly optimizing for TDoA and PDoA measurements

We can now extend the key intuitions to leverage TDoA and PDoA to develop a localization algorithm to meet our few-cm accuracy requirement. As further explained in Sec. 4, via careful engineering and hardware design choices, we measure PDoA with a standard deviation $\sigma_\theta = 5^\circ$ and TDoA with a standard deviation of $\sigma_t = 150$ ps. This *measurements* can be modeled as a zero-mean Gaussian:

$$\text{TDoA between Rx } i \text{ and } j : t_{i,j} \sim \mathcal{N}(0, \sigma_t)$$

$$\text{PDoA between Rx } i \text{ and } j : \theta_{i,j} \sim \mathcal{N}(0, \sigma_\theta).$$

Additionally, given a candidate tag location, $\vec{p}$, and receiver locations $\vec{x}_i, \forall i \in [1, 2, \dots, N]$ we can also compute the *expected* PDoA and TDoA as

$$\begin{aligned} \text{TDoA} : \hat{t}_{ij} &= \frac{|\vec{p} - \vec{x}_i|}{c} - \frac{|\vec{p} - \vec{x}_j|}{c} \\ \text{PDoA} : \hat{\theta}_{ij} &= \text{mod} \left(2\pi \left(\frac{|\vec{p} - \vec{x}_i|}{\lambda} - \frac{|\vec{p} - \vec{x}_j|}{\lambda} \right), 2\pi \right) \end{aligned} \quad (1)$$

where $\vec{p}$ is the location of the tag and $\vec{x}_i/\vec{x}_j$ are the locations of the 6 UWB antennas placed within a linear 1 m array. c and λ are the speed of light and UWB wavelength, respectively. Note here we forgo the far-field assumption made in Sec. 3.1.

The location ($\vec{p}$) which gives the closest expected measurements to the actual measurements is the likely tag location,

$$\min_{\vec{p}} \begin{bmatrix} \vec{e}_t^T & \vec{e}_\theta^T \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \vec{e}_t^T & \vec{e}_\theta^T \end{bmatrix} \quad (2)$$

where $\vec{e}_t$ and $\vec{e}_\theta$ measure the error between our predictions and the actual measurements, and

$$\Sigma = \text{diag}(\sigma_t^2, \dots, \sigma_t^2, \sigma_\theta^2, \dots, \sigma_\theta^2)$$

is a diagonal covariance matrix containing the TDoA and PDoA measurements standard deviations. Note here that since each receiver on XRLoc's localization module is independently measuring the TDoA and PDoA, we have a diagonal covariance matrix.

The simplest way to find this best tag location is to perform a grid search over our space to find the minimum point for Eq. 2. Aiming for cm-level localization, we choose a grid size of 1×1 mm. But this exhaustive search can be time-consuming (around 61.2 s / location on a 12-core CPU), precluding real-time localization in dynamic situations. Alternatively, we can leverage gradient descent-based optimization techniques [27] to arrive at the most likely tag position. However, these techniques fail when we do not have a

good initial estimate of the location, which is the case when looking to localize a tag in a large environment [29].

To surmount this challenge, we provide the final insight — selectively searching over the large space instead can reduce the computation complexity for localization. The brute force approach unnecessarily searches over each grid point for every packet. We can instead sample our environment more sparsely and slowly converge to our ideal location over a few packets. This is, in fact, the key idea behind particle filters [2], which are commonly used in state estimation scenarios with highly non-convex error functions and poor initialization.

Armed with this insight, for the first packet we receive, we uniformly distribute a set of particles (500 particles/m²) in our environment and compute the likelihood of these positions. When we receive consecutive packets, we can re-sample the set of particles with the highest probability and continue converging to our true locations. However, despite the fewer likelihood computations required, particle filters commonly furnish non-real-time estimates (with a latency of 7.2 ms on a 12-core CPU). To combat this problem, XRLoc adaptively re-samples and reduces the number of particles based on the current confidence of the estimate. As we do not know the tag's location, many particles are initially required to sample the search space uniformly. However, our particles converge close to the true location over time, improving our confidence in the location estimate. We can reduce the number of particles needed as we no longer need to explore the space uniformly. Empirically, this adaptive particle filter implementation converges within five measurements and provides a location estimate with a 1.2 ms latency on a 12-core CPU.

4 Challenges with prototyping XRLoc

Additional considerations arise when employing the ideas from Sec. 3 while prototyping XRLoc using off-the-shelf components. First, we need to acquire low-noise phase measurements. In Sec. 4.1, selecting the right clock is imperative to ensure a low phase noise. Second, due to hardware imperfections, we find that the expected PDoA measurements (Eq. 1) do not match the real-world measurements. To account for the offsets, we devise a calibration scheme and re-consider the formulation of the expected PDoA measurements in Sec. 4.2. Finally, we explore the effects of multipath reflection on the TDoA measurements in Sec. 4.3.

4.1 Acquiring accurate time and phase

Before prototyping XRLoc, we conducted extensive simulations to investigate the minimum phase and time acquisition accuracy needed to achieve few-centimeter positioning accuracy, assuming 6 antennas were equally spaced in a 1-meter region. In a 3×3 environment, we implemented the algorithm presented in Sec. 3.3 at varying phase and time acquisition noise levels. Our simulation results are presented in Fig. 4(a), where the horizontal axis represents the standard deviation of the phase error, and the vertical axis represents the 50 percentile of the localization error. Each line shows the standard deviation of the time error.

From this simulation, we make two key observations. First, we see that time errors between 3–250 ps provide similar localization accuracy, and these lines are grouped in the plot. However, exceeding 300 ps in time error significantly increases localization error, as TDoA fails to segregate ambiguity made by PDoA. Second,

these simulations clarify that few-cm level accurate localization requires high phase accuracy. Specifically, the red vertical line marks a threshold of 5° of standard deviation in phase measurement needed to achieve few-cm accurate locations.

The synchronization clock is the main factor affecting this phase noise in our system. The phase of the UWB signal is measured by first down-converting the received signal with the carrier signal. It is measured relative to this carrier signal by the baseband processing unit [23]. And when we consider measuring the PDoA, we look at the difference in the phase of any two receivers. In this situation, if both receivers share the same carrier clock, then the PDoA they measure will be induced purely from the relative distance traveled by the signals to each receiver (see Eq. 1). A simple way to achieve this is to connect the two receive antennas to the same UWB module [14]. However, we observed the overhead of extracting the complete CIR when implementing these systems is large (~ 1.2 ms), precluding low-latency localization. Specifically, we have the API overhead to measure the data and the data extraction overhead over USB, requiring $599 \mu\text{s}$ and $612 \mu\text{s}$, respectively.

Alternatively, we prototype our system using independent UWB modules [69] for each receiver, eliminating the need to export CIR measurements. This reduces the data acquisition latency by $\sim 4\times$ to $\sim 340 \mu\text{s}$. However, we cannot synchronize the carrier clocks on these independent modules, but instead, synchronize a lower 38.4 MHz clock leading to phase measurement errors. Via measurements with different clocks, we find that the phase noise in this input clock can largely influence the noise in the PDoA measurements. Specifically, from the oscillator's data sheet [19], we can obtain the phase noise of the oscillator, $N_\phi(f_{\text{offset}})$ where f_{offset} is the frequency offset from the center frequency of the oscillator. Using the $N_\phi(f_{\text{offset}})$, the standard deviation of clock jitter, σ_{jitter} , can be expressed as follows.

$$\sigma_{\text{jitter}} = \frac{\sqrt{2}}{2\pi f_{\text{osc}}} \sqrt{\Delta f N_\phi(f_{\text{offset}})} \quad (3)$$

where, Δf is the bandwidth of the measurement and f_{osc} is the oscillator frequency. We measure the standard deviation of the phase error (σ_ϕ) and time stamping error (σ_t) as:

$$\sigma_\phi = \frac{c}{\lambda} \frac{f_{\text{osc}}}{2\pi f_s} \sigma_{\text{jitter}}; \quad \sigma_t = \frac{f_{\text{osc}}}{f_t} \sigma_{\text{jitter}}$$

where, f_s is the sampling frequency, f_t is the frequency of the clock used for to measure time-of-arrival and c is the speed of light. We can choose an appropriate clock to meet our phase and time measurement thresholds by modeling this noise behavior. Many off-the-shelf [1, 19] clocks satisfy these requirements at reasonable price points and employ [19] in prototyping XRLoc. For example, according to the datasheets provided by Crystek [19] and Abracon [1], their respective phase noise values at 100 kHz offset are -160 dBc/Hz and -150 dBc/Hz, while their respective phase noise values at 100 Hz offset are -115 dBc/Hz and -109 dBc/Hz.

4.2 Combating hardware biases

In Eq. 1, we provided an expression for the expected PDoA measurement if we know the underlying tag and receiver locations. In reality, however, we see a large deviation when we compare the expected PDoA measurements with true PDoA measurements. To verify this, we perform an experiment varying the distance of a tag from XRLoc's localization module. In Fig. 4(b), the green 'RAW'

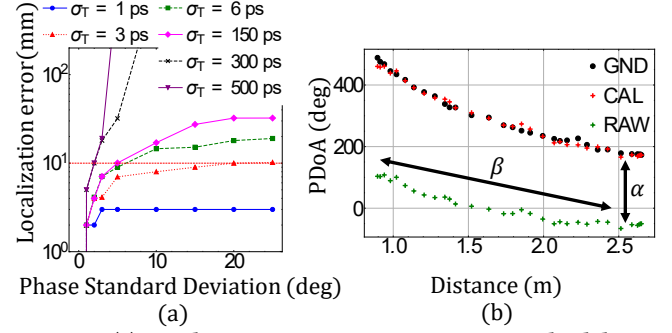


Figure 4: (a) Localization error vs. PDoA error standard deviation, with TDoA error standard deviations as each line in the legend. For few-cm level localization, the threshold, per the red line, is $\sigma_t = 5^\circ$ and $\sigma_t = 150$ ps. (b) Phase measurements (green) deviate from ideal (black) measurements. Performing appropriate calibration fixes these deviations (red).

measurements are shifted from black ground truth 'GND' measurements. Visually, we observe three deviations — a constant additive bias (α) which contributes to a downward shift, a multiplicative bias (β) w.r.t. distance affecting the slope of the line, and an exponential bias (γ) w.r.t. distance affecting the curvature (non-visualized in the figure). We assume these biases result from the ADC saturation when the distances are too close and propose a 3-point calibration to compute these hardware-specific calibrations below. Subsequently, we modify our *expected* PDoA measurements from Eq. 1 as

$$\hat{\phi}_{i,j} = \text{mod} \left(\left\{ \frac{2\pi d_i}{\lambda} - \alpha_i - \beta_i d_i^{\gamma_i} \right\} - \left\{ \frac{2\pi d_j}{\lambda} - \alpha_j - \beta_j d_j^{\gamma_j} \right\}, 2\pi \right)$$

where, $\alpha_i, \beta_i, \gamma_i$ are the calibration parameters and $d_i = |\vec{p} - \vec{x}_i|$ is the distance between the tag and UWB receiver. We replace Eq. 1 with this updated expected PDoA equation for the particle filter described in Sec. 3.3.

To estimate these calibration parameters, we perform a three-point calibration. First, we model the phase (Φ) measured at each UWB module according to these biases as

$$\tilde{\Phi}_i = \Phi_i + \alpha_i + \beta_i (d_i)^{\gamma_i}, \quad i \in [1, N],$$

where $\tilde{\Phi}$ is the calibrated phase. Next, we measure the received phase (Φ) at each UWB receiver for three *known* locations within our space. Finally, we use regression to find the expected calibration parameters, which minimize the deviation between the measured and expected phases according to the above equation.

4.3 Handling multipath reflections

However, in common indoor settings, reflections of the RF signal can potentially lead to ambiguities in TDoA measurement [79]. Despite our best efforts to acquire bias-corrected PDoA measurements, the presence of multipath can prevent us from ruling out ambiguous location predictions. However, UWB signals sample at the rate of 1 GHz, implying a time resolution of 1 ns. This fine-time resolution implies we are only corrupted by reflected paths whose additional travel distance is within 30 cm. In indoor environments, finding such close-by reflected paths is unlikely, and we find that our direct path and reflected signals are separable in the time domain. With this in mind, we measure the time of arrival and phase of the signals

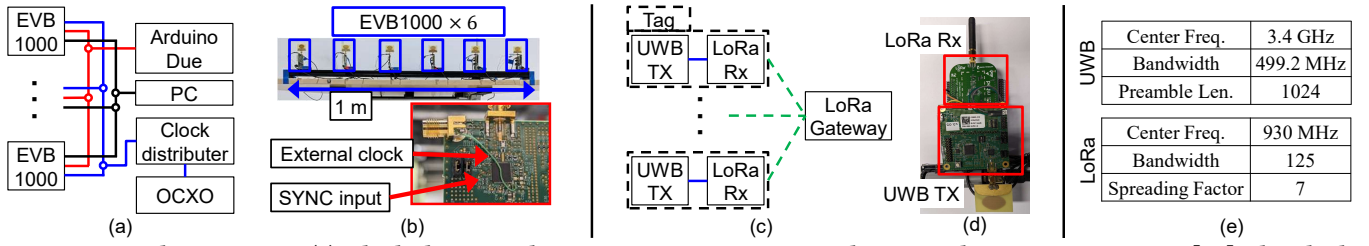


Figure 5: Implementation: (a) Block diagram showcasing interconnections between the 6 UWB receivers [69], the clock synchronization scheme (blue), “SYNC” implementation (red), and data back-haul via USB (black). (b) real-world implementation of block diagram; *inset*: external modification to UWB receiver. (c) block diagram for Tag showcasing the UWB and LoRa radios, the interrupt line (blue) to schedule UWB transmission and LoRa clock-sync broadcasts (dotted green) (d) real-world implementation of Tag. (e) UWB/LoRa radio parameters.

at the hardware reported first peak index, FPI [72], at the 6 UWB receivers in XRLoc’s localization module.

5 Enabling multi-tag operation

Through the ideas presented in Sec. 3 and 4, XRLoc fulfills the first two requirements for a localization system to be compatible with XR applications — ease of deployment (**R1**) and accuracy (**R2**). However, when we extend the current system to localize multiple tags in an environment, packet collisions amongst various tags can detrimentally affect our localization rates, resulting in a packet drop of 25%. Alternative to allowing tags to transmit arbitrarily, we can schedule individual tags at specific time intervals and leverage time-division multiple access (TDMA) to prevent collisions.

We seek to enable a total localization rate of 1000 Hz at XRLoc’s receiver means localizing a 1000 tags at a rate of 1 Hz or 10 tags at 100 Hz. Specifically, we explore leveraging low-power wireless technologies [36, 78] as a side channel for MAC protocol operation. A MAC controller needs to perform three tasks — onboarding new tags, providing time synchronization, and applying corrections to tags that deviate from their time slots. Existing systems [9, 57, 93] leverage UWB signals for providing this MAC control. However, we observe when a large number of tags need to be onboarded or corrections to the tag’s time slots need to be made, frequent collisions between UWB signals for localization and MAC control can exacerbate the problem we seek to solve. Furthermore, UWB is known to have high power consumption (e.g., about 416 mW for DW1000) during reception due to its use of wide bandwidth and despreading processing [10].

Alternatively, we propose using an additional side-channel leveraging low-power wireless technologies [36, 78] to simplify the MAC control and allow for independent tag management and localization functions. Using LoRa (e.g., about 20 mW for SX1280) or BLE (e.g., about 16 mW for nRF52832) as a side channel can also reduce power consumption of the tag. Hence, we prototype our system with LoRa as a side channel to furnish reliable and low-power MAC control for multiple tags. LoRa and UWB are at 900 MHz and 3.5 GHz, allowing them to co-exist with minimal interference.

The MAC protocol consists of two components — a LoRa MAC controller (gateway), which is deployed along with the localization module we have built so far, and a LoRa Receiver (LoRa RX) connected to the UWB tag. The gateway performs the three core functions of the MAC protocol.

Discovery and Onboarding: New tags introduced to a system transmit beacon packets to announce their presence. Subsequently,

the gateway invites these new tags to join the network by assigning a specific transmit time slot to transmit the UWB localization packets. The number and duration of a transmit slot are determined by the maximum number of tags and their localization rate. Currently, we support 1000 slots with a 1 ms slot width. Fig. 5(c) illustrates a block diagram of operation.

Global Time sync: Each tag must have a consistent notion of time slots, which requires a global time synchronization within the accuracy of at least half the slot width. Previous works [73] have showcased microsecond-level accuracy in clock synchronization, and we leverage these works to provide time synchronization. The gateway broadcasts synchronization packets every 100 s. This is the time it takes for a 5 ppm clock to drift by 500 μ s, half the slot width for each tag. The LoRa RX (co-located with the UWB tag) receives these sync packets and corrects for its clock drift.

Correcting erroneous tags: Finally, as a precautionary measure, we develop a correction mechanism to re-slot colliding tags. There may be a time-sync failure at tags, resulting in transmission at an incorrect time slot, leading to consistent collisions among groups of tags. By tracking the tags which suffer consistent collisions, the gateway broadcasts a correction packet over LoRa to re-slot the erroneous tag.

6 Implementation

We have seen XRLoc consists of three core components — the localization module, the LoRa MAC handler, and the UWB+LoRa tag. This section will take a closer look at prototyping these components.

Localization module: XRLoc’s primary contribution is a single-vantage point localization module using off-the-shelf components with a size of 1 m. This small size allows the localization module to be deployed within common electronics like TV’s our soundbars. Fig.5 shows the implemented prototype. The prototype is built with 6 UWB receivers EVB1000 [69], with Fig. 5(e) detailing the configuration parameters. We synchronize the UWB modules to a common clock (OCXO [19]) via a clock distributor module [90] as shown by the ‘blue’ path in Fig/ 5(a). Additional to the clock modification discussed in Sec. 4.1, we expose the EVB1000’s ‘SYNC’ pin to reset the time on the UWB modules to reduce bias in TDoA measurements. This sync is handled by an Arduino Due and is indicated in the ‘red’ path, with additional details provided in [70]. When each EVB1000 receives a single “blink” signal for localization from the UWB Tag, the receiver reports the first-peak-index (FPI) of the direct path in the channel impulse response’s peak, the signal

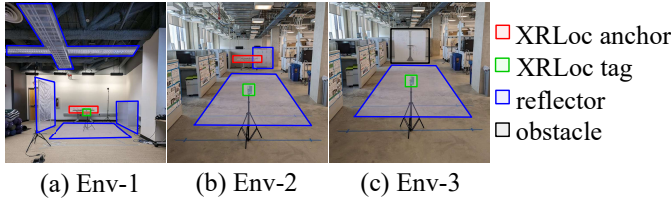


Figure 6: Evaluated in three spaces referred to as (a) Env-1: office-like, (b) Env-2: large-scale, and (c) Env-3: Non-line-of-sight condition. The tag, XRLoc’s module, and relevant regions are marked.

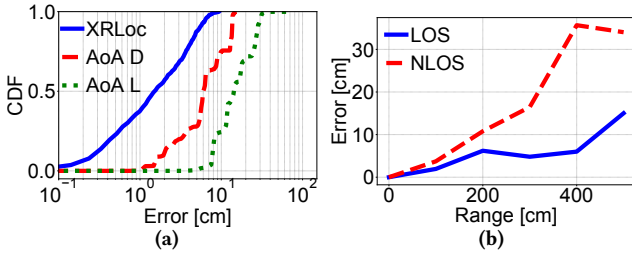


Figure 7: XRLoc’s Localization performance. (a) Static localization error, (b) localization accuracy vs. range with LOS (Env-2) and NLOS (Env-3) conditions.

phase at this point, time of arrival (RXTIME), and a carrier phase correction (RCPHASE) via the data path (shown in black).

LoRa MAC gateway: The LoRa gateway is the central controller to initialize, discover, and onboard all the tags in the environment. It is prototyped with a LoRa SX1272 [81] transmitter. This handler maintains the MAC state machine and performs all the functions described in Sec. 5.

Tags: We prototype the tag (shown in Fig. 5(d)) using the EVB1000 [69] and program it with the parameters in Fig. 5(e). The tag transmits ‘blink’ packets at 60 Hz, with each transmitted frame having 14 bytes of payload, including packet number and MAC address, to facilitate and test the MAC protocol. Operating in parallel, we have the LoRa SX1272 receiving time-sync packets from the Gateway module maintaining the UWB transmit slots and providing medium access control. An interrupt pin is raised by LoRa RX (shown in blue in Fig. 5(d)) to initiate a UWB ‘blink’ transmission at the accurate time slot.

7 Evaluation

XRLoc takes strides in achieving a few cm-scale localization in static and dynamic conditions. We rigorously test the system over eight different moving datasets and at multiple static points in various environments, including line-of-sight (LOS at Env-1 and 2) and non-line-of-sight (NLOS at Env-3) conditions as shown in Fig. 6. To make the NLOS condition in Env-3, a wooden board 2.5 cm thick was placed 30 cm forward from the XRLoc anchor. Additionally, we re-implement state-of-art AoA-based UWB localization system ULoc [113] based on their open-source documentation. We place 3 anchors in a diverse scenario, as a triangle in this space, and a constrained linear scenario, in a 1 m straight line. We test ULoc with the same static and dynamic positions.

7.1 Static localization accuracy

One of the key use cases targeted in XRLoc is to provide accurate locations of real-world objects and place them in the virtual realm. These objects of interest could be tagged with inexpensive and long-lasting UWB tags, which will relay their location to the VR system. To simulate this use case, we place multiple tags in the environment with the simple goal of recreating a life-size chess game. In this static scenario, from Fig. 7(a), we observe a median and 90th percentile error of 1.5 cm and 5.5 cm, respectively. We additionally observe XRLoc provides a 9.5× and 4.0× improvement at the median over using ULoc in a linear (AoA-L) and diverse (AoA-D) placement scenario which have (median, 90th%) of (14.6 cm, 28.7 cm) and (6.1 cm, 13.7 cm), respectively. The evaluation of different ranges shows median errors of 6.8 cm and 15.2 cm at 4m and 5 m in the LOS condition, respectively, and 35.3 cm and 34.0 cm in the NLOS condition as shown in Fig. 7(b).

7.2 Moving localization accuracy

Continuing with the motivation of playing a life-size chess game, we characterize XRLoc’s localization accuracy in dynamic scenarios. Fig. 8(a) and 8(b) showcase two characteristic movement patterns we tested. We tested 8 movements, as shown in the demo video¹, and achieved median and 90th errors of 2.4 cm and 5.3 cm, respectively, as shown in Fig. 8(c). We observe an 11× and 3.2× improvement at median over using ULoc in a linear (AoA-L) and diverse (AoA-D) placement scenario, which have (median, 90th%) of (26.0, 43.3 cm) and (7.5 cm, 17.4 cm), respectively.

In Fig. 8(d), we show the time-series error of localization for the ‘Fig. 8(b)’ movement scenario (Fig. 8(c)). We note that opting to use a particle filter over a brute force approach provides a localization latency of 1 ms, compared to exhaustive grid search’s latency of 61.2 s on a 12 Core CPU as explained in Sec. 3.3. However, because the particle filter performs a sparse sampling over the entire space, XRLoc may initialize the tag’s location incorrectly. This is visible in the inset shown in Fig. 8(d). But, throughout 5 received packets, we can see the location converges to the true location, and XRLoc subsequently provides accurate location predictions.

7.3 MAC protocol efficacy

In the previous sections, we have shown XRLoc can achieve a few-cm level localization from a single localization module, meeting the first two requirements (R1 and R2). To allow multiple tags to be localized with this accuracy, XRLoc leverages a LoRa side-channel to develop a power-efficient MAC protocol as described in Sec. 5. To evaluate its efficacy, we set up 10 tags to transmit at 100 Hz for a half-hour period. Fig. 9(a) showcases the packet success ratio, and we find over 99.5% of the packets are received by XRLoc’s localization module. Alternatively, when we do not have a MAC protocol, we have an average success rate of 76%, ranging between 56% – 87%. Specifically, considering the best and worst tag, we plot the packet arrival rate in Fig. 9(b) over the 30 min period and observe there are large periods when packets from Tag 09 are not received, likely due to collision from either Tag 02 or any of the other tags in the environment. Alternatively, we see a consistent packet arrival rate using a MAC protocol. Clearly, a MAC protocol is necessary to achieve multi-tag tracking and localization at high rates and fulfill R3.

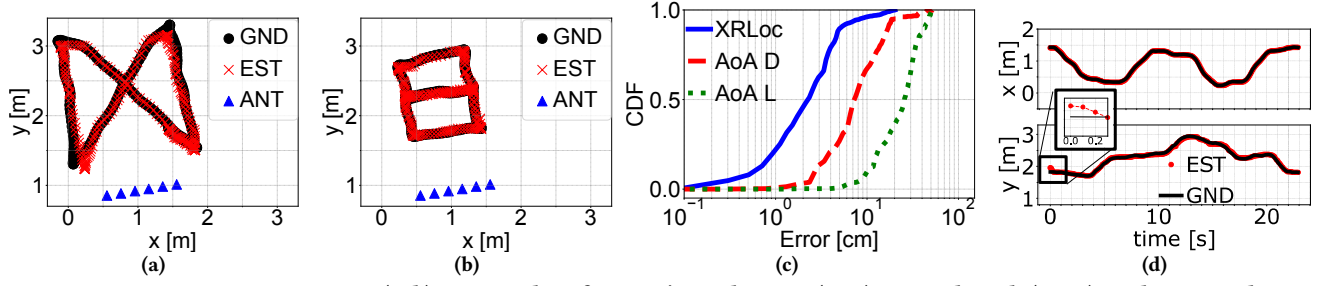


Figure 8: Dynamic testing scenario: (a, b) Scatter plot of XRLoc’s predictions (EST), ground truth (GND), and antenna locations (ANT). More examples can be found in demo video¹. (c) CDF of errors compared to AoA-based localization when three anchors [113] are placed diversely around the room (AoA-D) and constrained to a 1 m single line (AoA-L). (d) time-series errors of movement in (b), with the inset showcasing particle filter convergence within five packets.

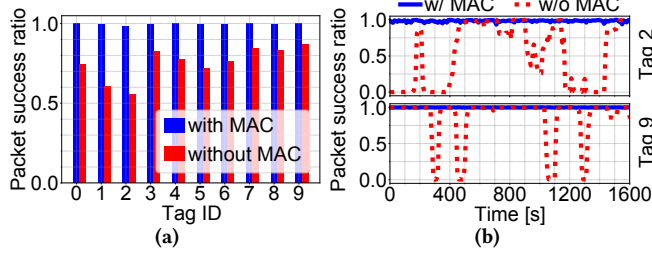


Figure 9: MAC protocol performance: (a) Packet success ratio across ten tags with (blue) and without (red) mac protocol. (b) packet success ratio over time for Tag 2 (best performing without MAC) and Tag 9 (worst performing without MAC). In all cases, MAC protocol provides a success rate of over 99.5%.

7.4 Justifying design choices

The evaluations from the previous sections prove XRLoc’s ability to fulfill the stringent requirements set for Sec. 1. In the following section, we will answer key questions about the design choices made when developing XRLoc.

Why TDoA and PDoA are both needed?

As we have discussed, a system relying purely on time-based measurements will not meet the stringent requirements of few-cm localization accuracy. We further evaluate this on our datasets in Fig. 10(a). We see a median localization accuracy of 49.1 cm, deviating over an order of magnitude from our few-cm level accuracy requirement. This re-iterates the challenge of achieving single-vantage point localization. We also claimed in Sec. 3.2 TDoA measurements play an important role in ruling out ambiguous initialization caused by PDoA-only localization. To confirm this, we see in the same figure when PDoA is solely used for localization, and we have a median accuracy of 74.8 cm. Clearly, ambiguities from phase wrap-around can be detrimental to XRLoc’s performance, emphasizing TDoA’s role. Through this micro-benchmark, it is apparent TDoA and PDoA work hand-in-hand to provide few-cm location accuracy.

How does the aperture effect the localization?

In Sec. 3.1, we discussed the importance of the antenna aperture in bringing resilience to phase measurement error. Consequently, a wider distance between the first and last antenna helps to improve localization accuracy. To ensure easy integration within everyday consumer electronics (like TVs or soundbars), we restrict XRLoc’s size to less than 1 m wide. However, how important is antenna

aperture to our localization performance? For this, we reduce the maximum antenna aperture to 80, 60, and 40 cm and report the results in Fig. 10(b). Clearly, a reduction in the aperture size affects the localization accuracy, with median localization accuracy reducing to 9.6, 19.3, and 35.0 cm, respectively. In fact, we see a steep drop-off in accuracy when we have an aperture of 40 cm. Furthermore, we see that a minimum aperture of 1 m is required to achieve the required localization accuracy. Under space constraints, smaller apertures may be used at the cost of lower accuracy.

How many antennas are needed?: Clearly, a minimum aperture of 1 m is needed. However, within this aperture, how many antennas are needed to meet the localization requirements? This is an important question to consider to make XRLoc cost-effective. In the previous localization accuracy analysis, we consider an array with 6 antennas. In Fig. 10(c), we reduce the number of antennas placed within the 1 m aperture. For 6, 5, and 4 antennas, we see the median location accuracy of 4.7, 6.9, and 28.7 cm, respectively. As few as 4 antennas are enough to meet the required few-cm localization accuracy at the median. Although, we observe a sharp reduction in localization accuracy in the 90th percentile. More antennas provide a better averaging effect and reduce erroneous TDoA and PDoA measurements, hence improving the localization performance at higher percentiles. From these experiments, we empirically observe choosing at least 6 antennas meets the required few-cm level accuracy required for XR applications.

Are there better antenna spacing we can choose?: So far, we have considered placing our antennas in a uniform linear array (ULA), separated by 20 cm. However, many works [94, 99] showcase antenna patterns that are more optimal than a ULA. To investigate the improvements from these co-prime antenna arrays, we leverage our simulator from Sec. 2 to carry out extensive simulations and showcase the results in Fig. 10(d). We see slight degradation of error when using co-prime arrays. However, co-prime arrays can be levered to reduce the number of antennas required by XRLoc to achieve similar location accuracy.

Why do we need fine-grained bias compensation?: We evaluate the system-level measurements. In XRLoc, we choose the appropriate clock sources to achieve the required accuracy in both TDoA and PDoA measurements (Sec. 4.1) and additionally calibrate for TDoA and PDoA hardware biases via a 3-point calibration scheme (Sec. 4.2). In Fig. 11(a), we showcase the importance of this bias calibration, observing median localization accuracy degrade by 1.8× to a median accuracy of 2.4 cm without applying appropriate bias

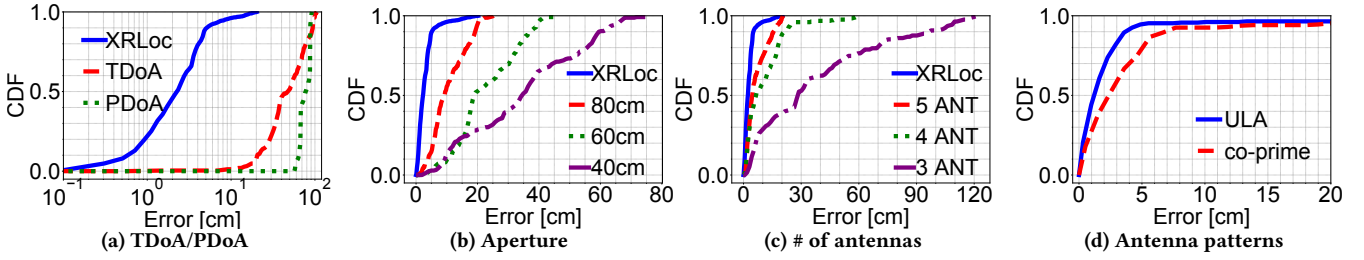


Figure 10: Microbenchmarks: (a) Using TDoA or PDoA only as opposed to a fusion (XRLoc). (b) Reducing aperture from 1 m (XRLoc). (c) Reducing the number of antennas while keeping the aperture at 1 m. (d) Leveraging co-prime antenna array as opposed to uniform linear array (ULA).

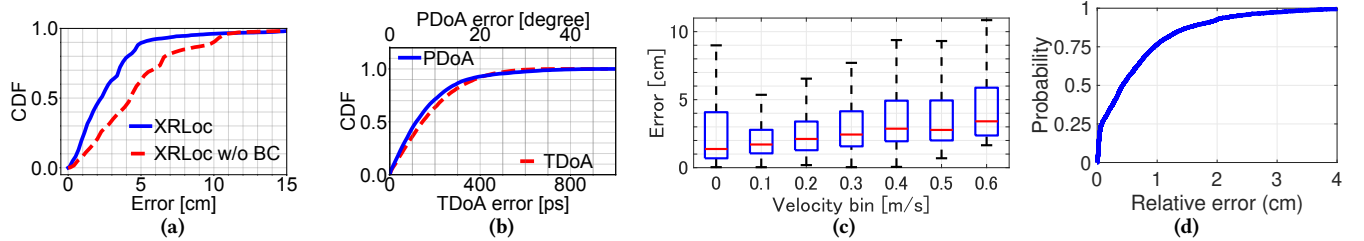


Figure 11: (a) Localization error with and without bias calibration; (b) measured PDoA and TDoA errors; (c) Localization error over different tag velocities showcasing XRLoc's resilience to movement; (d) relative localization error indicating smoothness of XRLoc's trajectory predictions.

calibration. In Fig. 11(b), we also observe an average TDoA error of 180.7 ps and PDoA error of 8.2°.

How does XRLoc's tracking performance over varying speeds?

It is imperative XRLoc provides accurate location and tracking predictions as the tag moves in the environment. The collected datasets involved tag movements at speeds of up to 68.1 cm/s, with the median error performance degrading from 2.4 cm to 3.3 cm, as shown in Fig. 11(c). Our data further indicates that errors in the relative change in position ("jumps") at the median and 90th percentile are 0.3 cm and 1.7 cm, respectively, which ensures a smooth tracking experience as shown in Fig. 11(d).

8 Related Works

Providing indoor location information for people and various inanimate objects is a well-studied problem. This section will broadly cover various techniques leveraged to address this problem. We will find that none of the existing techniques meets the stringent requirements we set up earlier in Sec. 1. Recall that we seek to provide easy-to-deploy (R1), few-cm accurate localization (R2) in dynamic scenarios for multiple people or objects of interest (R3). A few key technologies which can be considered are:

Visual sensing: Under this broad umbrella, we have many distinct technologies. Existing VR systems utilize external IR-based sensors [12], specialized cameras [97] or LED-based tags [83] to furnish accurate ground truth locations. There are also works that deploy a single Lidar [37] for person tracking or utilize headset-mounted cameras [63]. However, these systems are sensitive to visual occlusions, hindering a user experience. Recent works [54, 76, 114] which leverage machine learning to track objects despite occlusions. Alternatively, other studies [35, 50, 104, 106, 109] seek to deploy multiple cameras, let tag equips with a camera, or utilize special light sources to be robust to occlusions. However, no studies have simultaneously solved all the problems of ease of anchor deployment (R1),

accuracy (R2), and the risk of security and privacy [98]. Moving away from deploying privacy-invasive cameras, other works [62] seek to use the cameras on-board VR setups fused with occlusion-resilient radio-frequency (RF) signals like ultra-wideband. These systems have a low deployment cost but do not achieve a few-cm level accuracy.

Acoustic sensing: Alternative to these systems, various acoustic localization systems [13, 24, 49, 60, 66, 102] take advantage of the lower speed of sound (340 m/s) for fine-grained localization and meet the required localization accuracy. However, acoustic sensing has a few fundamental drawbacks [48]. First, acoustic systems [53] are difficult to provide both multi-asset and low latency localization simultaneously because of narrow bandwidth, deviating from R3. Second, they hinder music and audio playback, precluding immersive XR applications. Third, acoustic signals that employ ultrasound (> 20 kHz) for sensing have considerable audio leakage in the audible frequency range, affecting user experience.

Radar-based sensing: Mm-wave radars near the 60 GHz and 77 GHz bands have gained recent interest. Many works [44, 108] have looked at furnishing human pose with these radars from a single radar. Recent work [65] has shown that the human body can act as a strong blockage at these frequencies. These blockages can hinder tracking multiple people and objects in an environment and affect user experience. Additionally, tracking and identifying smaller assets in an environment can be challenging as radar reflections depend on an object's radar cross-sectional area (RCS). Alternatively, many works [86] propose placing retro-reflective tags on objects with small RCS to guarantee their detection; however, these systems suffer from poor localization accuracy.

		tag				anchor				
		cost	energy per loc	one tag $\max(f_{loc})$	multiple tags $\max(\sum f_{loc,i})$	RF cost	size per anchor	# of anchors	RF power consumption	easy to deploy
UWB	XRLoc	\$19	37 uJ	5556 Hz	5556 Hz	\$42	100 cm	1	2268 mW	Yes
	SALMA [32]	\$12	161 uJ	847 Hz	156 Hz	\$12	7 cm	1	378 mW	Yes
	ULoc [113]	\$12	30 uJ	5556 Hz	1023 Hz	\$244	17 cm	4	12096 mW	No
	SnapLoc [31]	\$12	118 uJ	1664 Hz	307 Hz	\$48	7 cm	4	1512 mW	No
	Pozyx [68]	\$12	644 uJ	212 Hz	39 Hz	\$48	7 cm	4	1512 mW	No
RFID	TurboTrack [55]	\$0.2	0 uJ	134 Hz	134 Hz	\$10000s	187 cm	1	75000 mW	Yes
	Tagoram [111]	\$0.2	0 uJ	40 Hz	40 Hz	\$1000s	27 cm	4	46000 mW	No

Table 2: Cost, size, and power consumption of related technologies. See Appendix A for details on each value.

RF-based sensing: The robustness of sub-6 GHz RF-signals to occlusions [84] and low privacy risk makes it a promising technology to consider. The common mode of operation is for multiple RF radios to jointly localize an active RF transmitter or a passive RF reflector (tags). Many works have looked at leveraging WiFi [45, 58, 64, 95, 107], LoRa [39], or BLE [8] to achieve robust user localization. However, these systems fail to provide the required localization accuracy due to bandwidth limitations.

RFID has a strong asymmetry in the reader-tag relationship, and the transmitter and receiver share the same clock, which allows for highly accurate phase acquisition. According to [55, 82, 111], RFID systems do not have carrier and sampling frequency offset and enjoy a phase measurement accuracy of 0.085° [111], $15\times$ better than the UWB, which provides an accuracy of 1.4° . Using the highly accurate phase, [40, 55, 56, 99, 111] has succeeded with tracking or localization at the few cm levels. However, due to the asymmetric nature, RFID readers whose range is several meters are not suitable for embedding into consumer electronics (**R1**) because of their power-hungriness and expensiveness (ex. Impinj Speedway R420 costs \$1666). The main target of RFID is industrial or retail store settings where thousands of tags must be deployed inexpensively, and readers' one-time cost is justifiable. For instance, [82] looks at item ordering in manufacturing lines, retail stores, or libraries. [55, 111] examine industrial robotics or baggage handling tasks.

Unlike RFID, Ultra-wideband provides a more symmetric architecture where localization modules can cost \$10–100. Consequently, we have seen their increased adoption in smartphones and smart tags. It provides over 500 MHz of bandwidth and a time resolution of 1 ns, providing localization accuracy to a few tens of centimeters. Many current UWB-localization schemes leverage the accurate time-resolution for Two-Way Ranging (TWR) [3, 11, 25, 30, 43, 47, 67, 115] and localize objects via trilateration. However, these multiple-packet exchanges increase localization latency and prevent real-time tracking of multiple objects of interest (**R3**). Many works instead leverage the TDoA or PDoA of the UWB signal to multiple time-synchronized anchors [14, 15, 31, 92, 96], or AoA measurements [23, 38, 113] at multiple anchors to furnish locations using a single packet. Some works [110] employ alternative transmission schemes to TWR to reduce the packet overhead. However, these systems only meet the necessary localization accuracy when the UWB anchors are placed in diverse locations, increasing deployment efforts and deviating from **R1**.

As discussed in Sec. 2, few-cm accurate localization is challenging due to geometric dilution of precision. To circumvent this problem, three common techniques are leveraged. First, by leveraging reflected paths in the environment, many systems [16, 32, 46, 51, 59, 85, 112] create additional “virtual” radios in the environment. These “virtual” radios provide the spatial diversity to localize an object of interest. However, multipath is often unreliable [6], leading to localization failure and poor user experience. Second, many works [26, 74, 100, 101] look at fusing TWR, TDOA, and AoA information to provide single anchor localization solutions. However, some systems cannot furnish the few-cm accurate localization requirement or rely on TWR measurements, increasing the system's latency. Finally, some works develop switched beam antennas [28, 33], which selectively sense signals approaching the anchor from different directions. However, these systems lack the required angular resolution to provide localization accuracy of a few cm.

9 Discussion

Table 2 compares XRLoc and other localization systems with cost, size, and power consumption. The calculation of each value is based on the values in the data sheets, and the conditions are aligned as much as possible. Note that each value is for reference purposes only because each value varies depending on the target performance and implementation details. For UWB, we compared with four different typed systems for reference: SALMA [32] as a single receiver system, ULoc [113] assumed as a 3D AoA system, SnapLoc [31] as a concurrent ranging system, and Pozyx [68] as a TWR-based system. For RFID, two extreme examples were used as a comparison, as the technology is fundamentally different from UWB: Tagoram [111], which is based on COTS RFID readers, and TurboTrack [55], which uses software-defined radio equipment.

UWB tag comparison: XRLoc incurs a slight increase in cost due to the inclusion of side-channel functionality. The side-channel function also increases energy consumption, but the increase in power consumption is limited, as the side channel can maintain continuous synchronization with one reception every 100 seconds. As for the maximum localization frequency in the case of a single tag ($\max(f_{loc})$), XRLoc achieves the same frequency as the state-of-the-art ULoc, since localization is possible using only a blink frame. Furthermore, when multiple tags exist, XRLoc achieves about 5.4 times more frequent localization than ULoc because it employs TDMA using the side channel. If TDMA is not used, UWB will communicate in pure ALOHA [89], reducing throughput to about 18.4 % in the presence of many terminals.

UWB anchor comparison: XRLoc is comparable to other localization systems in Table 2 except SALMA [32]. There is a clear trade-off between SALMA and XRLoc. XRLoc achieves about 5.3 times better localization accuracy and about 35.7 times better localization frequency in the presence of multiple tags than SALMA. In exchange for this performance improvement, XRLoc increases anchor cost by a factor of 3.5, anchor size by a factor of 14.3, and power consumption by a factor of 6.0 compared to SALMA. The future challenge for XRLoc is how to reduce the size of the anchor while maintaining the current localization accuracy and frequency.

Tag size: Thanks to semiconductor technology, it is already possible and realistic to combine multiple wireless communication functions on a single chip. The Semtech SX1280 [80], for example, combines BLE and LoRa functionality in a chip in 4 mm × 4 mm size. Apple AirTag [7] also embeds UWB, BLE, and NFC functionality in a 32 mm diameter module. For reference, the typical size of the RFID is approximately 16 mm × 43 mm × 0.2 mm.

UWB vs. RFID: A key technological difference between RFID and UWB is the tag-reader asymmetry. RFID tags are simple, easily deployable, and inexpensive, whereas readers are commonly power-hungry and expensive (~\$1900). Whereas UWB is more symmetric, the reader and tag are equivalent in power, cost (~\$12), and complexity. They hence serve different applications.

RFID is applied to industrial or retail store settings [55, 82, 111] where thousands of tags must be deployed inexpensively, and readers' one-time cost is justifiable. On the flip side, employing RFID readers in consumer electronics like televisions or smart speakers can often double or triple the cost of these consumer electronics.

Alternatively, UWB is a localization solution for consumer electronics. Here, less-than-\$5 UWB chips can be incorporated with smartphones, televisions, or smart speakers to support the localization of asset tags with a 2–3 years battery life. UWB anchor plugins compatible with Cisco routers to support large-scale UWB localization in large spaces [77] are also being rolled out. Consequently, we see an increasing need for UWB systems in AR/VR devices [61, 88] and sports IoT applications [30]. To satisfy these needs, we develop an XRLoc prototype anchor (costing ~\$42) and tag (costing ~\$19).

10 Conclusion and Future Work

XRLoc overcomes fundamental challenges arising from geometric dilution of precision to deliver cm-level accurate localization by developing an easy-to-deploy and low-latency localization module. Through this development, we are one step closer to achieving immersive XR experiences. However, a few possibilities of future work can be explored to build upon XRLoc.

Extensions to 3D: XRLoc focuses on localizing people and assets on a 2D floor plane, which is required in various XR applications. However, these ideas can be extended to the 3D domain by incorporating a vertical array of antennas in conjunction with the current horizontal linear array, such as an L-shaped or T-shaped antenna array [113].

Improving power efficiency of XRLoc's localization module: Various works [10] have noted the 10× higher power consumption of UWB reception than transmission. Keeping this in mind, we designed a system that requires only a single transmission from the tag for localization to ensure long battery life. However, using 6 receivers on XRLoc's wall-powered localization module is power

inefficient. To rectify this, antenna switching schemes [34] can be employed, or multiple antennas can be combined to connect to a single receiver [14] to reduce the number of receivers. However, unlike XRLoc's system, these alternatives will not be FiRa compliant [17].

Miniaturized tag design: We prototype our tag from off-the-shelf EVB1000 [69] and LoRa [81] evaluation boards. Future work can look towards miniaturizing these tag designs. Since these radios we employ are centered at 3.4 GHz and 930 MHz, it allows us to place these modules in close proximity with limited RF interference.

Acknowledgement

We are grateful to the anonymous reviewers and our shepherd for their insightful feedback. We thank Neil Smith at UCSD for his help with the VR Demo, Kazuhiro Kizaki at Osaka University, and the members of WCSNG Lab. A part of this research is supported by JST PRESTO #JPMJPR2032, and JSPS KAKENHI #23H00470.

A Basis for calculating the values in Table 2

The costs of tag and anchor were calculated based on the price of a 1,000-unit lot of key components only. We totaled, for UWB, the cost of the DW1000 IC chip, the SX1261 IC chip, the STM32F105RC CPU, and the chip antenna. For RFID, Table 2 accumulated the prices of the circularly polarized patch antenna, USRP X310, and USRP N210. All systems require a computational function to control each anchor, collect information from the tags, and calculate their positions. In addition, systems with multiple anchors require cabling costs to connect the anchors to the calculation function. Since the calculation function and cabling costs depend highly on the situation and application, Table 2 does not reflect the costs.

For *power consumption*, Table 2 totaled the values listed in the data sheets for the same major components used to calculate the costs. However, as with costs, we did not add the power required for the computational function. The *energy per loc* is the energy required for one localization. The frame size and transmit/receive switching overhead are calculated consistent with the values used in the beacon and TWR descriptions listed in [71] and [21].

One tag max(f_{loc}) is the maximum frequency of localization when there is only one tag. Since the processing time depends on the hardware, the localization frequency was calculated purely using the protocol overhead. As with *energy per loc*, the values are consistent with those listed in [71] and [21] for comparison. *Multiple tag max($\sum f_{loc}$)* is the maximum of the total number of localization frequencies for all tags when collisions are considered in the case of a large number of tags. When centrally controlled, as in XRLoc's TDMA and RFID, the performance is the same as one tag max(f_{loc}), but collisions occur when using pure ALOHA [89]. Since the theoretical maximum throughput of pure ALOHA is 18.4 % of the total, we multiplied one tag max(f_{loc}) by 0.184.

The *size per anchor* is the length of the individual anchor. For those using UWB, we calculated it based on the size of the EVB1000 [69] if the system can be implemented with EVB1000. For those using RFID, we calculated it from the information provided in the paper. As for *# of anchors*, 1 was used for a single anchor system, and 4 for a multiple anchors system. As for *easy-to-deploy*, we chose "No" for a multiple anchor system and "Yes" for a single anchor system. However, among the systems marked "Yes", those with smaller *size per anchor* are easier to deploy.

References

- [1] Abracon LLC. 2022. ASTXR-12-38.400MHZ-514054-T. <https://abracon.com/Oscillators/ASTXR-12-38.400MHZ-514054-T.pdf>. Accessed: 2023-02-14.
- [2] Michiel Aernouts, Noori BnILam, Nico Podevijn, David Plets, Wout Joseph, Rafael Berkvens, and Maarten Weyn. 2020. Combining TDoA and AoA with a particle filter in an outdoor LoRaWAN network. In *IEEE/ION Position, Location and Navigation Symposium (PLANS)*. 1060–1069.
- [3] Abdulrahman Alarifi, AbdulMalik Al-Salman, Mansour Alsaleh, Ahmad Al-nafessah, Suheer Al-Hadhrani, Mai Al-Ammar, and Hend Al-Khalifa. 2016. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors* 16, 5 (2016), 707:1–36.
- [4] Mary Alatise and Gerhard Hancke. 2020. A review on challenges of autonomous mobile robot and sensor fusion methods. *IEEE Access* 8 (2020), 39830–39846.
- [5] Sepehr Alizadehsalehi, Ahmad Hadavi, and Joseph Chuenhui Huang. 2020. From BIM to extended reality in AEC industry. *Automation in Construction* 116 (2020), 103254:1–13.
- [6] Peter Almers, Fredrik Tufvesson, and Andreas Molisch. 2006. Keyhole effect in MIMO wireless channels: Measurements and theory. *IEEE Transactions on Wireless Communications* 5, 12 (2006), 3596–3604.
- [7] Apple. 2021. AirTag. <https://www.apple.com/airtag/>. [Accessed: 2023-03-11].
- [8] Roshan Ayyalasomayajula, Deepak Vasisht, and Dinesh Bharadia. 2018. BLoc: CSI-based accurate localization for BLE tags. In *ACM International Conference on emerging Networking EXperiments and Technologies (CoNEXT)*. 126–138.
- [9] Jan Bauwens, Nicola Macoir, Spilios Giannoulis, Ingrid Moerman, and Eli De Poorter. 2021. UWB-MAC: MAC protocol for UWB localization using ultra-low power anchor nodes. *Ad Hoc Networks* 123 (2021), 102637:1–11.
- [10] Andreas Biri, Neal Jackson, Lothar Thiele, Pat Pannuto, and Prabal Dutta. 2020. SociTrack: Infrastructure-free interaction tracking through mobile sensor networks. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–14.
- [11] Francisco Bonnin-Pascual and Alberto Ortiz. 2019. An UWB-based System for Localization inside Merchant Vessels. In *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. 1559–1562.
- [12] Miguel Borges, Andrew Symington, Brian Coltin, Trey Smith, and Rodrigo Ventura. 2018. HTC Vive: Analysis and accuracy improvement. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2610–2615.
- [13] GaoShuai Cao, Kuang Yuan, Jie Xiong, Panlong Yang, Yubo Yan, Hao Zhou, and Xiang-Yang Li. 2020. EarphoneTrack: Involving earphones into the ecosystem of acoustic motion tracking. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 95–108.
- [14] YiFeng Cao, Ashutosh Dhekne, and Mostafa Ammar. 2021. ITrackU: Tracking a pen-like instrument via UWB-IMU fusion. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 453–466.
- [15] Hui Chen, Tarig Ballal, Nasir Saeed, Mohamed-Slim Alouini, and Tareq Al-Naffouri. 2020. A joint TDoA-PDoA localization approach using particle swarm optimization. *IEEE Wireless Communications Letters* 9, 8 (2020), 1240–1244.
- [16] Zhe Chen, Guorong Zhu, Sulei Wang, Yuedong Xu, Jie Xiong, Jin Zhao, Jun Luo, and Xin Wang. 2019. M³: Multipath assisted Wi-Fi localization with a single access point. *IEEE Transactions on Mobile Computing* 20, 2 (2019), 588–602.
- [17] Dieter Coppens, Eli De Poorter, Adnan Shahid, Sam Lemey, and Chris Marshall. 2022. An overview of ultra-wideband (UWB) standards (IEEE 802.15.4, FiRa, Apple): Interoperability aspects and future research directions. *IEEE Access* 10 (2022), 70219–70241.
- [18] Pablo Corbalán and Gian Pietro Picco. 2020. Ultra-wideband concurrent ranging. *ACM Transactions on Sensor Networks* 16, 4 (2020), 1–41.
- [19] Crystek Corp. 2022. Crystek XTAL OSC VCXO 122.8800MHZ. <https://www.crystek.com/home/crystek/default.aspx>. Accessed: 2023-02-14.
- [20] Andreas De Preter, Glenn Goysens, Jan Anthonis, Jan Swevers, and Goele Pipeleers. 2019. Range bias modeling and autocalibration of an UWB positioning system. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 1–8.
- [21] Decawave. 2020. Datasheet: DW1000 IEEE802.15.4-2011 UWB Transceiver. Version 2.22.
- [22] Michael Deering. 1998. The limits of human vision. In *International Immersive Projection Technology Workshop*. 1–6.
- [23] Igor Dotlic, Andrew Connell, Hang Ma, Jeff Clancy, and Michael McLaughlin. 2017. Angle of arrival estimation using Decawave DW1000 integrated circuits. In *IEEE Workshop on Positioning, Navigation and Communications (WPNC)*. 1–6.
- [24] Alireza Famili, Angelos Stavrou, Haining Wang, and Jung-Min Jerry Park. 2022. Pilot: High-precision indoor localization for autonomous drones. *IEEE Transactions on Vehicular Technology* 72, 5 (2022), 6445–6459.
- [25] Enrique García, Pablo Poudereux, Álvaro Hernández, Jesús Ureña, and David Gualda. 2015. A robust UWB indoor positioning system for highly complex environments. In *IEEE International Conference on Industrial Technology (ICT)*. 3386–3391.
- [26] Feng Ge and Yuan Shen. 2021. Single-anchor ultra-wideband localization system using wrapped PDoA. *IEEE Transactions on Mobile Computing* 21, 12 (2021), 4609–4623.
- [27] Amparo Gil, Javier Segura, and Nico Temme. 2007. *Numerical methods for special functions*. SIAM.
- [28] Gianni Giorgetti, Alessandro Cidronali, Sandeep Gupta, and Gianfranco Manes. 2009. Single-anchor indoor localization using a switched-beam antenna. *IEEE Communications Letters* 13, 1 (2009), 58–60.
- [29] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 249–256.
- [30] Mahanth Gowda, Ashutosh Dhekne, Sheng Shen, Romit Roy Choudhury, Lei Yang, Suresh Golwalkar, and Alexander Essanian. 2017. Bringing IoT to Sports Analytics. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 499–513.
- [31] Bernhard Großwindhager, Michael Stocker, Michael Rath, Carlo Alberto Boano, and Kay Römer. 2019. SnapLoc: An ultra-fast UWB-based indoor localization system for an unlimited number of tags. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 61–72.
- [32] Bernhard Großwindhager, Michael Rath, Josef Kulmer, Mustafa S Bakr, Carlo Alberto Boano, Klaus Witrals, and Kay Römer. 2018. SALMA: UWB-based single-anchor localization system using multipath assistance. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 132–144.
- [33] Mateusz Groth, Krzysztof Nyka, and Lukasz Kulas. 2021. Calibration-free single-anchor indoor localization using an ESPAR antenna. *Sensors* 21, 10 (2021), 3431:1–21.
- [34] Zhihao Gu, Taiwei He, Junwei Yin, Yuedong Xu, and Jun Wu. 2021. TyrLoc: A low-cost multi-technology MIMO localization system with a single RF chain. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 228–240.
- [35] Yundong Guo, Zhenyu Liu, Hao Luo, Huijie Pu, and Jianrong Tan. 2022. Multi-person multi-camera tracking for live stream videos based on improved motion model and matching cascade. *Neurocomputing* 492 (2022), 561–571.
- [36] Naresh Kumar Gupta. 2016. *Inside Bluetooth low energy*. Artech House.
- [37] Mahmudul Hasan, Junichi Hanawa, Riku Goto, Ryota Suzuki, Hisato Fukuda, Yoshinori Kuno, and Yoshinori Kobayashi. 2022. LiDAR-based detection, tracking, and property estimation: A contemporary review. *Neurocomputing* 206 (2022), 393–405.
- [38] Milad Heydariaan, Hossein Dabirian, and Omprakash Gnawali. 2020. AnguLoc: Concurrent angle of arrival estimation for indoor localization with UWB radios. In *IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*. 112–119.
- [39] Bashima Islam, Md Tamzeed Islam, Jasleen Kaur, and Shahriar Nirjon. 2019. Lorain: Making a case for lora in indoor localization. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 423–426.
- [40] Chengkun Jiang, Yuan He, Xiaolong Zheng, and Yunhao Liu. 2018. Orientation-aware RFID tracking with centimeter-level accuracy. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 290–301.
- [41] Tero Kaarlela, Sakari Pieskä, and Tomi Pitkähö. 2020. Digital twin and virtual reality for safety training. In *2020 11th IEEE international conference on cognitive infocommunications (CogInfoCom)*. IEEE, 000115–000120.
- [42] S Keeping and A Woolls-King. 2007. Beyond physics for superior sound. *Maxwell: Periodiek der Electrotechnische Vereeniging*, 11 (1), 2007 (2007).
- [43] Benjamin Kempke, Pat Pannuto, and Prabal Dutta. 2016. Harmonium: Asymmetric, bandstitched UWB for fast, accurate, and robust indoor localization. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 1–12.
- [44] Hao Kong, Xiangyu Xu, Jiadi Yu, Qilin Chen, Chenguang Ma, Yingying Chen, Yi-Chao Chen, and Linghe Kong. 2022. m³Track: mmWave-based multi-user 3D posture tracking. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 491–503.
- [45] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. SpotFi: Decimeter level localization using Wi-Fi. In *ACM Conference on Special Interest Group on Data Communication (SIGCOMM)*. 269–282.
- [46] Manikanta Kotaru and Sachin Katti. 2017. Position tracking for virtual reality using commodity Wi-Fi. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 68–78.
- [47] Anton Ledergerber and Raffaello D'Andrea. 2019. Ultra-wideband angle of arrival estimation based on angle-dependent antenna transfer function. *Sensors* 19, 20 (2019), 4466:1–21.
- [48] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: Practical problems for acoustic sensing. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 381–390.
- [49] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-Track: Pushing the limits of contactless multi-target tracking using acoustic signals. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 150–163.
- [50] Jing Li, Jing Xu, Fangwei Zhong, Xiangyu Kong, Yu Qiao, and Yizhou Wang. 2020. Pose-assisted multi-camera collaboration for active object tracking. In

- AAAI Conference on Artificial Intelligence (AAAI), Vol. 34. 759–766.
- [51] Ze Li, Zengshan Tian, Zhongchun Wang, and Zhenyuan Zhang. 2020. Multipath-assisted indoor localization using a single receiver. *IEEE Sensors Journal* 21, 1 (2020), 692–705.
 - [52] Congfeng Liu, Jie Yang, and Fengshuai Wang. 2013. Joint TDoA and AoA location algorithm. *Journal of Systems Engineering and Electronics* 24, 2 (2013), 183–188.
 - [53] Manni Liu, Linsong Cheng, Kun Qian, Jiliang Wang, Jin Wang, and Yunhao Liu. 2020. Indoor acoustic localization: A survey. *Human-centric Computing and Information Sciences* 10 (2020), 1–24.
 - [54] Donghai Luo, Daobo Wang, Shengji Xia, and Tingting Bai. 2022. A novel approach for visual tracking based on occlusion recognition. *Highlights in Science, Engineering and Technology* 7 (2022), 124–133.
 - [55] Zhihong Luo, Qiping Zhang, Yunfei Ma, Manish Singh, and Fadel Adib. 2019. 3D backscatter localization for fine-grained robotics. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 765–782.
 - [56] Yunfei Ma, Nicholas Selby, and Fadel Adib. 2017. Minding the billions: Ultra-wideband localization for deployed RFID tags. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 248–260.
 - [57] Nicola Macoir, Matteo Ridolfi, Jen Rossey, Ingrid Moerman, and Eli De Poorter. 2018. MAC protocol for supporting multiple roaming users in multi-cell UWB localization networks. In *IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 588–599.
 - [58] Alex T Mariakakis, Souvik Sen, Jeongkeun Lee, and Kyu-Han Kim. 2014. Sail: Single access point-based indoor localization. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 315–328.
 - [59] Paul Meissner and Klaus Witrals. 2012. Multipath-assisted single-anchor indoor localization in an office environment. In *International Conference on Systems, Signals and Image Processing (IWSSIP)*. 22–25.
 - [60] Massimo Merenda, Luca Catarinucci, Riccardo Colella, Demetrio Iero, Francesco Della Corte, and Riccardo Carotenuto. 2022. RFID-based indoor positioning using edge machine learning. *IEEE Journal of Radio Frequency Identification* 6 (2022), 573–582.
 - [61] SPARK Microsystems. 2022. UWB Puts the Mobility in the Metaverse. <https://www.sparkmicro.com/uwb-puts-the-mobility-in-the-metaverse/>. Accessed: 2023-10-10.
 - [62] John Miller, Elahe Soltanaghai, Raewyn Duvall, Jeff Chen, Vikram Bhat, Nuno Pereira, and Anthony Rowe. 2022. Cappella: Establishing multi-user augmented reality sessions using inertial estimates and peer-to-peer ranging. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 428–440.
 - [63] Riccardo Monica and Jacopo Aleotti. 2022. Evaluation of the Oculus Rift S tracking system in room scale virtual reality. *Virtual Reality* 26, 4 (2022), 1335–1345.
 - [64] Sherif Mostafa, Khaled Harras, and Moustafa Youssef. 2022. A survey of indoor localization systems in multi-floor environments. *TechRxiv* (2022), 1–31.
 - [65] Swagato Mukherjee, Gregory Skidmore, Tarun Chawla, Anmol Bhardwaj, Camillo Gentile, and Jelena Senic. 2022. Scalable modeling of human blockage at millimeter-wave: A comparative analysis of knife-edge diffraction, the uniform theory of diffraction, and physical optics against 60 GHz channel measurements. *IEEE Access* 10 (2022), 133643–133654.
 - [66] Hiroaki Murakami, Masanari Nakamura, Hiromichi Hashizume, and Masanori Sugimoto. 2019. 3-D localization for smartphones using a single speaker. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 1–8.
 - [67] A. Poulou, Z. Emersic, O. Steven Eyobu, and D. Seog Han. 2020. An accurate indoor user position estimator for multiple anchor UWB localization. In *International Conference on Information and Communication Technology Convergence (ICTC)*. 478–482.
 - [68] Pozyx. 2021. Multi Technology RTLS – Indoor & Outdoor. <https://www.pozyx.io/>. [Accessed: 2023-03-11].
 - [69] Qorvo, Inc. 2016. EVB1000: Ultra-Wideband (UWB) Transceiver Evaluation Kit. <https://www.qorvo.com/products/p/EVK1000>. Accessed: 2023-02-14.
 - [70] Qorvo, Inc. 2014. Wired Synchronization of Anchor Nodes in a TDoA Real Time Location System. APS007, version 1.10.
 - [71] Qorvo, Inc. 2015. The Implementation of Two-way Ranging with the DW1000. APS013, version 2.3.
 - [72] Qorvo, Inc. 2016. DW1000 Metrics for Estimation of Non Line Of Sight Operating Conditions. APS006, version 1.10.
 - [73] Ceferino Gabriel Ramirez, Anton Sergeyev, Assya Dyussenova, and Bob Iannucci. 2019. LongShoT: Long-range synchronization of time. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 289–300.
 - [74] Mohammad Rostami and Karthikeyan Sundaresan. 2022. Enabling high accuracy pervasive tracking with ultra low power UWB tags. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 459–472.
 - [75] Kyle Russell. 2013. People are worried Microsoft’s new Xbox will be able to spy on you. *Business Insider* (2013).
 - [76] Gozde Sahin and Laurent Itti. 2023. HOOT: Heavy occlusions in object tracking benchmark. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 4830–4839.
 - [77] Ali Samioglu and Richard Jang. 2021. Cisco and Wipolot Present the First Real-time Location System (RTLS) with an App Hosting Solution Using Ultra Wide-Band (UWB). <https://blogs.cisco.com/networking/cisco-and-wipolot-first-uw-b-based-location-system-with-app-hosting>. Accessed: 2023-10-10.
 - [78] Ramon Sanchez-Iborra and Maria-Dolores Cano. 2016. State of the art in LP-WAN solutions for industrial IoT services. *Sensors* 16, 5 (2016), 708:1–14.
 - [79] Jan Scheuing and Bin Yang. 2006. Disambiguation of TDoA estimates in multipath multi-source environments (DATEMM). In *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (ICASSP)*, Vol. 4. 837–840.
 - [80] Semtech Corp. 2022. LoRa Connect Long Range Low Power LoRa 2.4GHz RF Transceiver With Ranging. <https://www.semtech.com/products/wireless-rf/lora-connect/sx1280>. Accessed: 2023-10-10.
 - [81] Semtech Corp. 2022. LoRa Connect Mbed shield, SX1272, 868 and 915MHz. <https://www.semtech.com/products/wireless-rf/lora-connect/sx1272mb2das>. Accessed: 2023-03-17.
 - [82] Longfei Shangguan and Kyle Jamieson. 2016. The design and implementation of a mobile RFID tag sorting robot. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 31–42.
 - [83] Rahul Anand Sharma, Adwait Dongare, John Miller, Nicholas Wilkerson, Daniel Cohen, Vyas Sekar, Prabal Dutta, and Anthony Rowe. 2020. All that Glitters: Low-power Spoof-resilient Optical Markers for Augmented Reality. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 289–300.
 - [84] Christopher Slezak and Sundeep Rangan. 2022. Measurement-based indoor millimeter wave blockage models. *IEEE Transactions on Wireless Communications* 21, 8 (2022), 6774–6786.
 - [85] Elahe Soltanaghai, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. Multipath triangulation: Decimeter-level WiFi localization and orientation with a single unaided receiver. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 376–378.
 - [86] Elahe Soltanaghai, Akarsh Prabhakara, Artur Balanuta, Matthew Anderson, Jan Rabaey, Swarun Kumar, and Anthony Rowe. 2021. Millimetro: mmWave retro-reflective tags for accurate, long range localization. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 69–82.
 - [87] James Spilker Jr, Penina Axelrad, Bradford Parkinson, and Per Enge. 1996. *Global positioning system: Theory and applications, volume I*. American Institute of Aeronautics and Astronautics.
 - [88] Dean Takahashi. 2023. Emerge Partners with Disney to Bring Multi-sensory Communication to Homes. <https://venturebeat.com/games/xsolla-rolling-out-new-developer-opportunities-partnerships-and-more/>. Accessed: 2023-10-10.
 - [89] Andrew S. Tanenbaum. 2012. *Computer Networks 5th Edition*. Pearson.
 - [90] Texas Instruments, Inc. 2017. LMK04832EVM: LMK04832 evaluation module for ultra-low-noise, 3.2-GHz, 15-output, JESD204B clock jitter cleaner. <https://www.ti.com/tool/LMK04832EVM>. Accessed: 2023-02-14.
 - [91] Bruce Thomas. 2012. A survey of visual, mixed, and augmented reality gaming. *Computers in Entertainment* 10, 1 (2012), 1–33.
 - [92] Janis Tiemann, Fabian Eckermann, and Christian Wietfeld. 2016. Atlas: an open-source tdoa-based ultra-wideband localization system. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 1–6.
 - [93] Janis Tiemann, Yehya Elmasry, Lucas Koring, and Christian Wietfeld. 2019. ATLAS Fast: Fast and simple scheduled TDOA for reliable ultra-wideband localization. In *IEEE International Conference on Robotics and Automation (ICRA)*. 2554–2560.
 - [94] P. Vaidyanathan and Piya Pal. 2010. Sparse sensing with coprime arrays. In *Asilomar Conference on Signals, Systems and Computers*. 1405–1409.
 - [95] Deepak Vasishth, Swarun Kumar, and Dina Katabi. 2016. Decimeter-level localization with a single WiFi access point. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 165–178.
 - [96] Davide Vecchia, Pablo Corbalán, Timofei Istomin, and Gian Pietro Picco. 2019. TALLA: Large-scale TDoA localization with ultra-wideband radios. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 1–8.
 - [97] Vicon Motion Systems Ltd. 2021. Vicon. <https://www.vicon.com/>. Accessed: 2021-04-29.
 - [98] Neil Vigdor. 2019. Somebody’s watching: Hackers breach ring home security cameras. *The New York Times* (2019).
 - [99] Jue Wang, Deepak Vasishth, and Dina Katabi. 2014. RF-IDraw: Virtual touch screen in the air using RF signals. *ACM SIGCOMM Computer Communication Review* 44, 4 (2014), 235–246.
 - [100] Tianyu Wang, Hanying Zhao, and Yuan Shen. 2019. An efficient single-anchor localization method using ultra-wide bandwidth systems. *Applied Sciences* 10, 1 (2019), 57:1–17.
 - [101] Tianyu Wang, Hanying Zhao, and Yuan Shen. 2019. High-accuracy localization using single-anchor ultra-wide bandwidth systems. In *IEEE/CIC International Conference on Communications in China (ICCC)*. 59–63.
 - [102] Weiguo Wang, Jinming Li, Yuan He, and Yunhao Liu. 2020. Symphony: Localizing multiple acoustic sources with a single microphone array. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 82–94.

- [103] Xiangyu Wang and Phillip Dunston. 2007. Design, strategies, and issues towards an augmented reality-based construction training platform. *Journal of Information Technology in Construction* 12, 25 (2007), 363–380.
- [104] Yanxiang Wang, Jiawei Hu, Hong Jia, Wen Hu, Mahbub Hassan, Ashraf Uddin, Brano Kusy, and Moustafa Youssef. 2023. Spectral-Loc: Indoor localization using light spectral information. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (2023), 1–26.
- [105] Nannan Xi, Juan Chen, Filipe Gama, Marc Riar, and Juho Hamari. 2022. The challenges of entering the metaverse: An experiment on the effect of extended reality on workload. *Information Systems Frontiers* (2022), 1–22.
- [106] Pengjin Xie, Lingkun Li, Jiliang Wang, and Yunhao Liu. 2020. LiTag: localization and posture estimation with passive visible light tags. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 123–135.
- [107] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-grained indoor location system. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 71–84.
- [108] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. 2021. mmMesh: Towards 3D real-time dynamic human mesh construction using millimeter-wave. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 269–282.
- [109] Shangyao Yan, Zhimeng Yin, and Guang Tan. 2021. CurveLight: An accurate and practical indoor positioning system. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 152–164.
- [110] Jing Yang, BaiShun Dong, and Jiliang Wang. 2022. VULoc: Accurate UWB localization for countless targets without synchronization. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–25.
- [111] Lei Yang, Yekui Chen, Xiang Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 237–248.
- [112] Xianan Zhang, Lieke Chen, Mingjie Feng, and Tao Jiang. 2022. Toward reliable non-line-of-sight localization using multipath reflections. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–25.
- [113] Minghui Zhao, Tyler Chang, Aditya Arun, Roshan Ayyalasomayajula, Chi Zhang, and Dinesh Bharadia. 2021. ULoc: Low-power, scalable and cm-accurate UWB-tag localization and tracking for indoor applications. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–31.
- [114] Fangwei Zhong, Peng Sun, Wenhan Luo, Tingyun Yan, and Yizhou Wang. 2021. Towards distraction-robust active visual tracking. In *International Conference on Machine Learning (ICML)*. 12782–12792.
- [115] Lukasz Zwirello, Tom Schipper, Marlene Harter, and Thomas Zwick. 2012. UWB localization system for indoor applications: Concept, realization and analysis. *Journal of Electrical and Computer Engineering* 2012 (2012), 1–12.