DolphinTrack

Visual, Inertial, and Radio Frequency Tracking of Dolphins
Final Capstone Report

Advaith Balaji, Sydney Belt, Thor Helgeson, Terry Tao
Team **tag! you're it!**

May 1, 2025

I. Executive Summary

Fig 1: A Bottle-nose Dolphin wearing the biologging tag "MTag" developed by Dr. Alex Shorter's group. The tag is attached to the animal using suction cups, and can measure their motion [1].

Tracking the motion of marine animals is crucial for understanding their biomechanics and behavior. This proves especially difficult, since aquatic environments introduce significant occlusion and signal attenuation. Our project aims to improve upon the bottle-nose dolphin tracking system developed by Dr. Alex Shorter's lab at the University of Michigan. We do this by improving the visual detection system and introducing a custom radio frequency localization system that allows for tagged dolphin association under uncertainty. This report outlines the motivation and design context behind our work, our design process, and the final design embodiment, which accommodates various technical and stakeholder-driven factors. Our complete system, **DolphinTrack**, is described in detail.

Before proposing solutions to our problem, we first conducted a literature review exploring previous work in video tracking, localization with inertial measurements, and sensor-based positioning systems. In particular, we found various state-of-the-art video segmentation models that could improve upon our sponsor's current visual tracking method [2]. This review of past work, in addition to ongoing conversations with Dr. Alex Shorter and his team, who are the primary stakeholders of our project given their ultimate responsibility of deploying our system in the field for improved understanding of dolphin motion, led us to explore two directions in this design space: visual tracking and radio-frequency localization. We arrived at our final vision models and hardware configurations after utilizing mind-mapping and brain-writing strategies to fully explore our problem context.

The first major video tracking improvement we saw was with SAMURAI, which produces a continuous dolphin track without bounding box dropouts in comparison to our sponsor's existing implementation, which produced sparse dolphin tracks. An example video from our initial SAMURAI testing can be viewed **here**. Since SAMURAI does not support multi-object tracking, to enable visual tracking in scenarios with multiple dolphins, we explored various tracking algorithms, including ByteTrack and SAM2. The largest performance criteria for these models is the frequency of identity switches, wherein the tracked ID of a dolphin changes over time. Identity switches are especially common when tracking multiple dolphins with significant occlusion, close proximity swimming, and agent uncertainty over long periods of free swimming. This association gap is addressed by our radio frequency localization system, which uses an anchor and tag system to localize dolphins upon surfacing. This setup provides robust, high-precision tracking near the surface of the water that enables re-identification of dolphins after an identity switch. This enables a better understanding of the agents' long-term trajectories, which ultimately enables more complete analysis of group swimming behaviors over long periods—our sponsor's goal. The complete system is tested against our critical user requirements and specifications through an experimental verification procedure. We outline the process by which the full system, which we have shown to be operational with grounded human tracking and real-time camera detections, radio frequency re-association, and particle filter position plotting, can be implemented on-site in a dolphin habitat.

CONTENTS

Abstract

Bio-inspired roboticists, marine biologists, and engineers want to understand dolphins' group dynamics and their motion underwater. Existing work has tracked individual dolphins, but not multiple dolphins. We propose **DolphinTrack**, a novel dolphin tracking system using visual tracking and a radio-frequency positioning system to produce precise position estimates for multiple dolphins simultaneously. The camera tracker uses ByteTrack to produce a time-series track of the dolphin in image space, which is converted to the world frame using a homography transform. The radio-frequency positioning system consists of a set of ultra-wideband anchors, which communicate with wearable tags to measure ranging information. This ranging information can be used to produce a pose estimate in world frame using multilateration. These two pose estimates are fused along with the inertial measurements using a particle filter to produce an accurate pose estimates for each dolphin the system is tracking. Most importantly, the radio-frequency positioning system can position the dolphin every time it surfaces to provide the camera tracker with the correct dolphin ID. This ensures that researchers can reliably track multiple individual dolphins simultaneously. Through our tests, we found that the visual tracking system achieved accurate tracking of the majority of dolphin positions, with dropout and identity switch issues being addressed through association from radio-frequency localization. This system could be used to better understand the swimming patterns of dolphins and other marine mammals, which could inspire planning algorithms for swarms of marine robots or improve conservation practices. Code for the full DolphinTrack system can be found at https://github.com/sydneybelt/DolphinTrack.

II. Project Introduction

Fig 2: A dolphin lap trial conducted by Dr. Alex Shorter's group at DolphinQuest Oahu, Hawaii [1]. In each trial, the dolphin swims one lap in a loop. Biologging tags aboard the dolphin collect inertial data, while a camera tracks the dolphin from above. Dr. Shorter uses the collected data to study their dynamics and motion profile.

Engineers, researchers, and scientists can all draw a lot of inspiration from biological systems. Marine animals, in particular, exhibit unique behaviors—by understanding how animals such as dolphins swim in the water, interact socially with their group, and move as a pod with remarkable coordination, we can gain insights that may lead to advancements in areas like swarm robotics, bio-inspired algorithms, and conservation technology. These creatures offer models of efficiency, adaptability, and robustness, which can be applied to a wide range of engineering challenges, from improving robotic designs to enhancing environmental monitoring and conservation efforts. To effectively study biological systems, it is necessary to construct a high-quality tracking system that is capable of accurately capturing the movements and behaviors of the subjects being studied. The case of dolphins offers a unique challenge as they are a species known for their complex social interactions and coordinated swimming patterns. Data collected from them would allow researchers to gain a deeper understanding of their motion and behavior, allowing us to design previously intractable controllers for bio-inspired systems.

The sponsor of our project is Dr. K. Alex Shorter, an associate professor of mechanical engineering at the University of Michigan whose research focuses on the mechanics of marine mammals. This project is part of his efforts to understand how bottlenose dolphins swim, particularly when in the presence of other dolphins. Achieving higher-precision tracking of multiple dolphins will enable characterization of the coordination between dolphins when they swim in a pod. Dr. Shorter believes that understanding how dolphins coordinate could help inspire strategies which could be used to coordinate groups of marine robots efficiently.

As a result, this project aims to improve the precision of the existing dolphin tracking system and enable it to track multiple dolphins continuously and simultaneously. The current tracking method produces a planar pose estimate for the dolphin $\hat{x} \in \mathbb{R}^2$ by fusing a dead-reckoning pose estimate and a camera-based pose estimate using a particle filter [3]. The dead-reckoning pose estimate is obtained using data from a tag attached to the dolphin. The tag is equipped with a TDK ICM-20948, which incorporates a magnetometer that measures the heading of the dolphin [4]. The tag is also equipped with an impeller, which spins as the dolphin swims and measures its speed. The speed and heading of the dolphin define its velocity at each timestep, which is accumulated to estimate the position of the dolphin. The camera pose estimate is obtained from a video captured

from a camera looking at the dolphin enclosure from above. A Faster-RCNN detector is used to detect dolphins in the video frames [2]. The image coordinates of the center of the bounding box returned by the detector is converted to world coordinates, which is then used to update the particle filter.

This method has some limitations, mainly because the task of *continuously* tracking dolphins poses is challenging. With dolphins being evolved to blend into their environment and taking frequent dips deep below the water surface, tracking them visually becomes a problem. Due to this, the visual tracking system often produces sparse track with several detection dropouts. This is especially true for our sponsor's current method, a frame-by-frame detection algorithm, which only takes into consideration the current frame, and not the previous locations. In a scenario where multiple dolphins are swimming in close proximity, once a dolphin is re-detected by the camera system, there is no way to determine whether the dolphin detected was the one that was previously detected.

While a modern object tracking method which produces a continuous trajectory would enable multi-dolphin tracking, the problem of associating camera tracks with the correct dolphin to produce a more reliable pose estimate is nontrivial. Therefore, our sponsor desires a solution that will **enable the researchers to autonomously match an arbitrary camera track to a specific dolphin in the environment**. This will allow them to reliably maintain a visual track for a single dolphin and produce state estimates with less uncertainty, thereby improving the quality of the kinematic data the researchers are drawing inferences upon.

To address these issues, we created a combined radio-frequency and visual tracking system called **DolphinTrack** capable of producing continuous position estimates while maintaining identification of individual dolphins across different frames. This system integrates camera-based tracking with radio-frequency positioning data to provide accurate and reliable localization of multiple dolphines, resulting in a 2D trajectory of each dolphin's movements. These 2D trajectories will allow future researchers to study how dolphins swim in the presence of other dolphins, including how they coordinate their motions and how their movement strategies change when swimming as a group. One application of this research could be to inspire methods for coordinating swarms of marine robots, drawing inspiration from the natural behaviors observed in dolphin groups. Another application of this technology could be improved understanding of dolphin swimming and locomotion strategies, which could assist with dolphin conservation efforts by informing policies on habitat protection and species preservation. A final possible application is to extend this tracking approach to other types of marine mammals, such as whales, manatees, or seals. The ability to precisely track marine mammal motions could benefit marine biologists and bio-inspired roboticists alike, providing new insights into the dynamics of aquatic ecosystems and opening doors to novel robotic designs based on nature's strategies. A detailed description of the various components of our system and how they work together is included in the Design Description section. In order to arrive at our final design, we conducted stakeholder and community analyses, and went through several stages of design iterations, which we outline in the following subsections.

A. Related Work

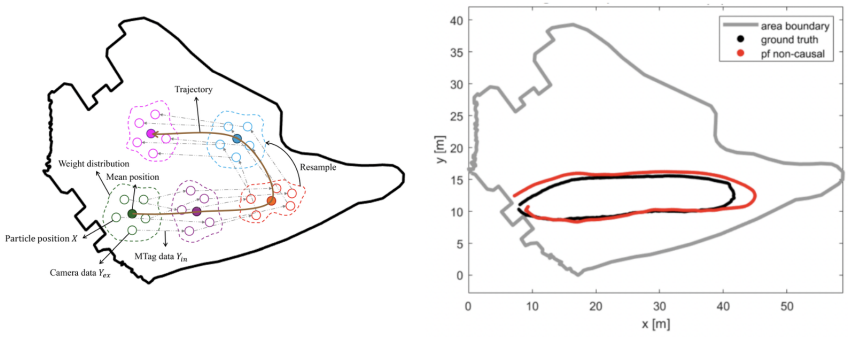


Fig 3: Particle Filter State Estimation for dolphins proposed in Xia et al. [3]

There are various subproblems in marine agent tracking that have already been explored in literature, including work in object tracking, localization, underwater vision, marine dynamics and more. Our project seeks to support the dolphin localization method explored in [3] (and visualized in Figure 3), where the authors employed particle filter localization to combine visual and inertial sensor measurements (from a camera and an inertial measurement unit) to track the planar trajectory of a single dolphin. This method is also reflected in some work in robotics employing visual-inertial localization for mobile robot SLAM (simultaneous localization and

mapping). That work mainly focuses on predicting a highly accurate robot state given observations from a monocular RGB camera, and an IMU onboard the system [5] [6], which is identical to our problem scenario. Previous work has also incorporated different sensor measurements such as acoustics and pressure—which has been explored in various works that explore sensor fusion for localization [7] [8] [9]. Our work attempts to incorporate radio-frequency positioning into the localization pipeline.

One of the key parts of our problem involves producing accurate dolphin position measurement using images from an RGB camera. A relatively simple approach applied in both [10] and [3] uses a Faster R-CNN model to propose bounding boxes to the dolphin in each frame which is used downstream as a sensor measurement in the localization system. There are also methods that only rely on camera data, employing computer vision models (such as YOLO and R-CNNs) to detect and track fish movement across frames [11] [12] [13]. Along with this, it will also be vital to account for the refraction of light due to water to correctly estimate the dolphin's position under the water [14].

Our environment setup introduces some issues that haven't been taken into consideration yet. These issues include dealing with time-series visual tracking, handling occlusions between dolphins, and generalizing across lighting conditions. Thus, we are tasked with creating a tracking system that accounts for these discrepancies.

B. Design Context

The design of our dolphin tracking system is shaped by a variety of social, environmental, educational and economic factors that influence its development and use. We aim for our final system to be deployed on a device that is worn by a bottle-nose dolphin to log data that will support studying its motion profile. While the device is primarily of use to our sponsor's research group, we still take into consideration stakeholders who benefit from such a system to track dolphins precisely as well as study their dynamics. Table I below outlines our key stakeholders and their form of influence.

Stakeholder	**Type**	**Description**
Dr. Alex Shorter (and Co.)	Primary	As the Principal Investigator, Dr. Shorter motivates all research problems and provides funding to realize our final product. His lab students conducting dolphin biomechanics research will directly engage with our solution in the field. Thus, our design is heavily influenced by their needs—we take into consideration the state of their current project and their skillset while constructing the final system.
Research Organizations	Secondary	The funding agencies that support Dr. Shorter's research, such as the Office of Naval Research, provide grants to the lab, and thus have an indirect economic influence on our work. Their goal is to push biomechanics research forward, so our system needs to provide meaningful utility that helps researchers gain new insights into dolphin motion and behavior.
Dolphin Quest Oahu	Secondary	They take on socio-economic influence over our project as they provide the controlled environment to test our solution, while the employees directly interact with the tags and the dolphins. Any proposed solution must be easily integrated into their workspace and comply with their rules and abilties.
Marine & Engineering Researchers	Tertiary	Marine biologists and bio-inspired roboticists stand to benefit in both an educational and societal context, as our system seeks to reveal new physical insights into dolphin locomotion that could inform both biological research and robotic design

Table I: The most important stakeholders involved with our project. Our design and implementation are influenced by them in educational, economic and social contexts. Keeping these stakeholders in mind will ensure our product is beneficial to all

C. Requirements, Specifications & Initial Development

After using our system, the data generated will be used by Dr. Shorter's lab as well as marine biologists studying dolphin gait and behavior. Thus, our generated data will need to satisfy the standards needed for rigorous engineering and science research. The main metric they value is the accuracy of the position measurements, as a lower error means more nuanced gait and behavior analysis. Additionally, if this system needs to be scaled, the cost per unit of trackable area needs to be as low as possible. Therefore, we define several requirements and

specification for the camera tracking system and RF positioning system. Our most important ones are outlined in Table II, where we define some secondary requirements for each subsystem in the relevant Design Testing subsection.

Requirement	Specification	Reasoning
Minimize Camera Detection Dropout	Maximum drop out time is < 5 s	Ensuring the system tracks dolphins without large interruptions to maintain dense position data data. This requirement is more relaxed as the point of the RF System is to help recover from dropouts.
Minimize Dolphin ID Switches	Maximum of 1 identity switch per 10 s of video	Ensuring the system tracks unique dolphins is important to the integrity of the dolphin motion data.
Minimize False Positive Tracks	No more than 5% of tracks should incorrectly identify dolphins	High false positive rates would lead to incorrect identifications, causing incorrect tracking, compromising the integrity of the data. This requirement is also more relaxed as we have the RF system to help recover.
Maximize Position Accuracy	$\|\|c_{cam} - c_{gt}\|\|_2^2$ must be < 15 in pixel space	Maintaining position accuracy is essential for tracking the exact location of dolphins, especially for behavior analysis.
Maximize RF ranging accuracy	Distance measurements will be < 0.1 m from ground truth	Using an ultra-wideband ranging framework has proven to produce distance measurements with < 0.1 m error [15]
Maximize RF positioning accuracy	Position measurements will be < 1 m from ground truth	The allowed variance in the dolphin position estimate was set by the sponsor according to his needs.
Low Cost	Cost per hectare of tracking coverage $< \$1000$	The system must be able to scaled to larger or smaller environments easily. We aim to keep a low cost per anchor, and allow for larger coverage.

Table II: Key requirements and specifications established for our dolphin tracking system. These ensure our system meets standards required for collecting high quality motion data to be used by the researchers.

These key requirements and specifications influenced our design process. We engaged in an iterative design process, exploring multiple potential solutions and checking their performance against our requirements. We continued to engage with our sponsors throughout the process to ensure our work was consistent with their expectations, and took on their feedback regularly to improve upon our design. Our ideation and prototyping journey is described below, starting with our concept generation process, and our alpha design.

1) Ideation

The project's broad scope and applications were explored using mind mapping, a technique that visually organizes the various subproblems into the relevant subcategories: Computer Vision, Hardware, and Kinematics. This helped identify some subproblems, including video tracking, sensor integration, and modeling dolphin motion. To generate solutions, brain writing was used while allowing each team member to ideate on different project components based on their expertise, resulting in a diverse range of ideas. These included time-series detections to improve localization with continuous video data, and end-to-end transformer models to integrate camera and sensor data for seamless tracking, signal processing to derive more information from the measured data, and near-field localization using ultra-wideband technology to improve positional accuracy. These approaches allowed us to address the complex problem of multi-dolphin tracking from both a computational and hardware perspective, laying the foundation for the proposed solutions. By weighing several ideas according to our requirements and specifications, we arrived at an alpha design for our system.

2) Alpha Design

The first iteration of DolphinTrack integrated three primary components: video segmentations, dead-reckoning, and near-field localization. For video segmentation, we use the SAMURAI video segmentation model to process

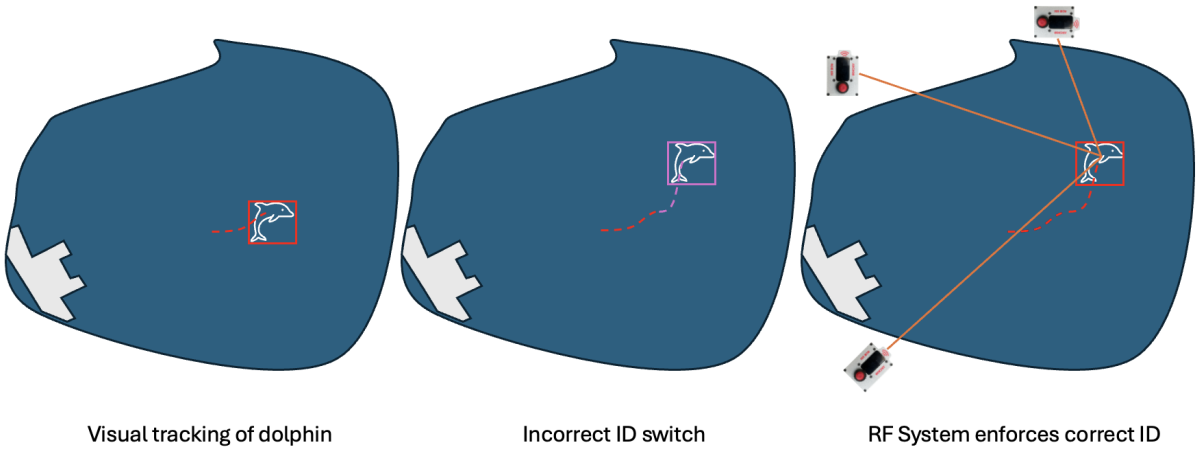


Fig 4: An example of the camera and RF tracking system at work. Box and track colors denote ID. The camera tracking system initially receives continuous detections of the tracked dolphin. Due to the dolphin disappearing from view and returning (or due to a model error), the same dolphin is incorrectly assigned a new ID. To correct this, the tracking system receives an RF positioning ping that re-assigns the correct "red" ID to the dolphin located at the position ping, producing a consistent track.

video streams, where users can prompt the system to track specific dolphins [16]. The dead-reckoning system estimates dolphin position by accumulating velocity from the dolphin's speed and heading, with enhanced accuracy achieved by fusing the compass and gyroscope data from the dolphin's MTag. The near-field localization component calculates dolphin positions based on ranges between custom anchors and the MTag, with each anchor transmitting time-of-flight data. The range information is then uploaded to a LoRA-WAN server, which processes it to obtain position estimates. These estimates are combined and fused in the particle filter to produce a final state estimate.

This alpha design had some key shortcomings. First, it did not address the problem of associating the right dolphin with the camera tracklets during instances of dropout, leading to a potential misidentification. Additionally, the system could not track multiple dolphins simultaneously, which is critical for studying group behaviors. The reliance on the camera for position estimation also introduced limitations, as it could not provide precise estimates independent of visual data, leaving it vulnerable to environmental challenges such as occlusion or poor lighting. After discussions with our sponsor, we arrived at the final design of a combined visual and radio-frequency based positioning system that can easily be used for tracklet association to ensure the researchers can reliable track a single unique dolphin during post-processing. This design is described in detail in the following Design Description section.

III. Design Description

We propose our system **DolphinTrack**, an improved dolphin tracking system that leverages an RF positioning system and time-series visual tracking models to produce a more accurate and reliable track for a given unique dolphin. The system consists of two key components: the RF positioning system, which utilizes ultra-wideband anchors to provide precise ranging information, and the visual tracking component, which uses an object tracking model to track the dolphin's motion based on camera data. By integrating these two components, DolphinTrack is able to produce continuous and reliable position estimates that are more robust and resilient to occlusions and multiple dolphins compared to traditional single-source tracking methods. The camera system will be responsible for tracking a dolphin's position continuously. In the case of a dropout or identity switch, the RF positioning system will help enforce correct identification for each dolphin according to its position. This framework is visualized in Figure 4. The following sections include a detailed design description of each subcomponent.

A. Visual Tracking System

The visual tracking system explores various model architectures for tracking dolphins amidst occlusions and dropout. It also includes various pipeline components to allow for faster implementation, iteration, and evaluation for users.

1) Model Architectures

YOLO: You Only Look Once [17] is a convolutional neural network model for object detection. It can be applied to individual video frames to identify dolphin instances. It was chosen as part of our pipeline for its efficiency, ease of implementation and re-training, and its use as a detection backbone for other Ultralytics-supported trackers (including ByteTrack and BoT-SORT) [18].

ByteTrack: ByteTrack [19] leverages all detections in frames to make associations for multi-object tracking, allowing it to pick up on potential occlusions or low confidence detections. It was chosen as part of our pipeline for this information maintenance capability which is especially useful in our environment, and for its inference speed and computational efficiency.

BoT-SORT: BoT-SORT [20] incorporates object motion and visual features, camera re-orientation, Kalman filtering for more accurate object detections. It was chosen as part of our pipeline to serve as a baseline of comparison to ByteTrack with similar principles of tracking overlaid on YOLO detections, but it shows relatively minimal performance improvement in static environments that don't warrant the additional computational load for the test videos we have used this far. This model can also motivate future work on camera positioning and dynamic environments.

SAM2: Segment Anything Model 2 [21] uses initial frame prompting to generate initial segmentation masks, which are then propagated through videos using stored frame memory. It was chosen as part of our pipeline for its zero-shot capabilities, providing accurate segmentations from single frame prompts for which video propagation can be improved with fine-tuning. The continuous mask mitigates identity switches and false detections. Its largest bottleneck is the RAM requirements for memory storage, but this can be addressed with reasonable compute (High-RAM A100 GPU, available on Google Colab) and batching video segments.

2) Training Pipeline

The training pipeline consists of dataset labeling tools for fine-tuning custom models. For the architectures included in our method, this can be done with YOLO, SAM2, and COCO backbones. This condensed re-training for detection components allows for faster iteration and broader applications of newly retrained models and curated datasets.

3) Inference and Evaluation Pipeline

The inference evaluation pipeline consists of an end to end system that takes in a raw video and allows users to generate ground truth and tracked annotations based on a chosen architecture and trained model. The ground truth centroid script runs a user interface for plotting ground truth coordinates at the desired frame rate (to reduce manual load for longer videos) and saves centroids in a format that can directly be compared to centroids generated by the model evaluation scripts included for each architecture. The final script generates metrics we used in the validation stage of our computer vision algorithms and saves additional information that can be used for further analysis and improvements.

4) Utility Scripts

To add more usability and efficiency to the process, we have included additional scripts that support batching videos for running on memory-constrained hardware, and running a particle filter to plot world-frame coordinates. This is helpful to run powerful models like SAM2 on longer video, reducing the constraint from memory to time which is less relevant for offline processing, and extrapolate trajectories under conditions of dropout in the vision system.

B. Radio-Frequency Positioning System

The radio-frequency positioning system consists of tags and anchors (see Appendix A for drawings and schematics). The tags are intended to be integrated with the MTags on the dolphins, while the anchors are placed at known locations in the environment. The anchors use two-way ranging to measure their distance to each tag, which they then transmit over LoRa to a central computer. This computer tracks the most recent range measurement between each tag and anchor pair and implements a multilateration algorithm to compute the position of each tag.

1) Two-way ranging

Two-way ranging is a technique that uses the propagation speed of radio waves to calculate the distance between two transceivers. In this system, our transceiver is the Qorvo DWM1000 ultra-wideband module [22]. The round-trip time for sending an initiation message and receiving a response is used to calculate the propagation time

of the signal. If only one set of messages is used, clock offset in the transceiver modules can cause significant error in the propagation time estimate. For this reason, we use double-sided two-way ranging, which involves measuring two round-trips. This can be achieved with a minimum of three messages, with the tag initiating the exchange and the anchor completing the ranging estimation.

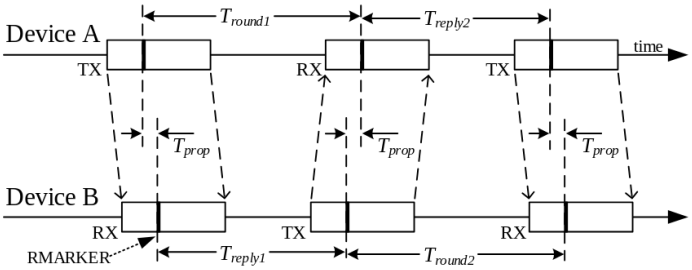


Fig 5: Diagram depicting the messages and time measurements involved in double-sided two-way ranging. Taken from [22]

We implemented the double-sided two-way ranging algorithm detailed in the DW1000 user manual [22]. The process involves an initialization message from the tag, a response message from the anchor, and then a final message from the tag (see Figure 5). The round trip time from the transmission of the initialization message to the reception of the response message is denoted T_{round1}. The round trip time from the transmission of the response message to the reception of the final message is denoted T_{round2}. The time from the anchor receiving the initialization message to sending the response message is denoted T_{reply1}. The time from the tag receiving the response message to transmitting the final message is denoted T_{reply2}. From these times, we can compute the propagation time of the radio signal.

$$\hat{T}_{prop} = \frac{T_{round1}T_{round2} - T_{reply1}T_{reply2}}{T_{round1} + T_{round2} + T_{reply1} + T_{reply2}} \tag{1}$$

To compute the range $r_i \in \mathbb{R}$ between the tag and anchor i, we simply multiply $\hat{T}_{prop}$ by the speed of light in air.

$$\begin{aligned} r_i &= \hat{T}_{prop} \cdot c_{air} \\ c_{air} &= 299,702,547\frac{m}{s} \end{aligned} \tag{2}$$

2) Positioning via multilateration

The positioning portion of the radio-frequency system operates on the principle of multilateration. Given range measurements between a tag and at least three anchors with known positions, we can calculate the position of the tag. The algorithm we adopted allows for computing position with an arbitrary number of ranges [23]. Given anchor positions $(x_1, y_1, z_1), (x_2, y_2, z_2)...(x_n, y_n, z_n) \in \mathbb{R}^3$ and ranges $r_1, r_2...r_n \in \mathbb{R}$ we can set up a system of equations for the tag position $(x, y, z) \in \mathbb{R}^3$.

$$\begin{bmatrix} 1 & -2x_1 & -2y_1 & -2z_1 \\ 1 & -2x_2 & -2y_2 & -2z_2 \\ 1 & -2x_3 & -2y_3 & -2z_3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -2x_n & -2y_n & -2z_n \end{bmatrix} \begin{bmatrix} x^2+y^2+z^2 \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_1^2 - x_1^2 - y_1^2 - z_1^2 \\ r_2^2 - x_2^2 - y_2^2 - z_2^2 \\ r_3^2 - x_3^2 - y_3^2 - z_3^2 \\ \vdots \\ r_n^2 - x_n^2 - y_n^2 - z_n^2 \end{bmatrix} \tag{3}$$

This system can be written in the form $Ax = b$, which we solve using the Moore-Penrose pseudo-inverse.

$$\hat{x} = (A^T A)^{-1} b \tag{4}$$

From $\hat{x}$, we only use $\hat{x}_1$ and $\hat{x}_2$, which correspond to the x-y planar pose. Since the MTag already computes depth with sufficient accuracy, the z value of the dolphin pose does not need to be tracked by this system.

3) Data collection from anchors via LoRa

Since our ranging is completed by the anchors, to compute position the ranges need to be sent to a central computer. We decided to use the LoRa protocol to achieve this, since it provides reliable transmission over long distances [24]. The anchors each have a Seeed Studio Wio-E5 LoRa transceiver module [25]. A Seeed Studio Wio-E5 mini is attached via USB connection to the central computer. The anchors are set up to transmit over LoRa after each ranging exchange, and the LoRa module attached to the positioning computer is set up to receive all incoming messages. The message sent over LoRa includes the anchor ID, the tag ID, and the distance between them. Each of these pieces of information is repeated five times to allow for naive error correction. In our testing, a delay of about 100 to 150 ms was necessary between LoRa transmissions to avoid interference. On the actual dolphin tracking system, to avoid this delay, we would advise replacing the LoRa transmission with a fourth ultra-wideband transmission back to the tag from the anchor in each ranging exchange containing the range information. The tag can then compute its position and store it on the MTag or store the raw range data for later post-processing. This is because LoRa transmissions are much slower than ultra-wideband transmissions. While the maximum LoRa transmission rate is about 50 kbps, we ran our DW1000 ultra-wideband transceivers at 850 kbps, and they can run at up to 6.8 Mbps [22].

C. Full System Integration

1) Motivation

The primary purpose of combining a visual and radio frequency tracking system is to handle associations across tracked segments once dolphins surface. As shown in Figure 6, identity switches can occur when there are instances of dropout or close proximity swimming, both of which are present with our current vision model architectures.

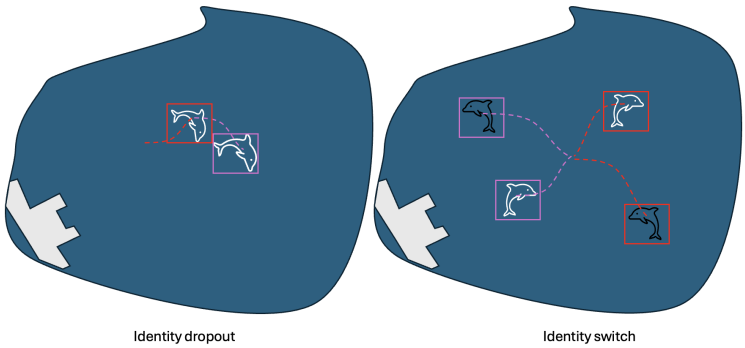


Fig 6: Cases of Identity Mismatch

When the same dolphin assumes a new ID, we need to revert it back to the original to maintain a complete understanding of its trajectory over time. This consistent ID is invoked when the RF tag daughter board transmits pings to generate position information.

2) Live Person Tracking Demonstration

To demonstrate the efficacy of this system in a controlled environment, given constraints on traveling for on-site testing, we set up our combined visual tracking and particle filter localization in the Robotics Atrium. The setup is depicted in Figure 7, with four anchors positioned around the space, an established world coordinate frame for transformations, and a camera positioned one floor above ground level.

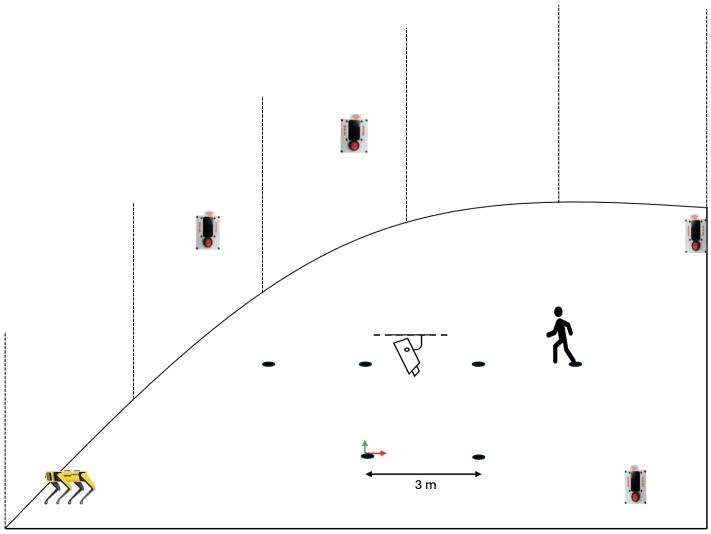


Fig 7: Full System Setup in Robotics Atrium

All people in the field of view of the camera are tracked in real time, running an object tracking algorithm on an NVidia Jetson Xavier. We display overlayed bounding boxes and ID numbers. One person holds an tag, which sends pings from anchors to determine its position in the world frame. This position is then compared to the camera detections in the world frame (calculated through a homography transformation), and the visual bounding box and ID remain at red color and 0 respectively for the tagged person. We then send these coordinates over a socket to display the particle filter for world frame motion.

3) Implementation in the Field

While the person tracking demonstration shows how we can make associations with our two system components, especially when it comes to rectifying identity switches, it fails to replicate the conditions of the actual environment. Validation of this system will require the vision system working on dolphins and the RF system transmitting just above the surface of the water to demonstrate full applicability.

IV. Design Testing

A. Visual Tracking System

To effectively test the design for the computer vision component of our system, we had to choose which model architectures we would be comparing. The general features of each of our proposed architectures are summarized in Table III. It is important to note that while all model architectures are capable of object detection, only time-series based models can track objects by assigning unique IDs that are carried across frames. It is also noteworthy that while all model architectures are capable of single dolphin detection, not all of them can handle multiple detections at once. SAMURAI, which has produces continuous tracks for single dolphins by applying a Kalman filter to segmentations generated by SAM2, is not able to handle multiple dolphins. The last major differentiator is whether these models require pre-training. While all of them could be improved by fine-tuning the underlying detection algorithm on a custom dataset, and all are supported by Roboflow annotations as outlined in our final design, SAM2 and SAMURAI by extension demonstrate zero-shot performance with just a single point prompt on the initial frame.

Model	Detection	ID Tracking	Single Dolphin	Multi Dolphin	Pre-Training
Faster R-CNN	Yes	No	Yes	Yes	Yes
YOLO	Yes	No	Yes	Yes	Yes
ByteTrack	Yes	Yes	Yes	Yes	Yes
BoT-SORT	Yes	Yes	Yes	Yes	Yes
SAMURAI	Yes	No	Yes	No	No
SAM2	Yes	Yes	Yes	Yes	No

Table III: Comparison of different models across various capabilities

Given these tradeoffs among expected performance and computational load for these models, we decided to evaluate on YOLO, ByteTrack, BoT-SORT, and SAM2. Faster R-CNN was slow to train and was producing

sparse detections, and SAMURAI, while consistently accurate, has a lack of support for multiple dolphins at once which would severely limit the application of our design.

The experiment design holds the following constant: evaluation videos, training data, and compute resources. This is critical to get an unbiased understanding of performance and efficiency of each model. The test sequences included over 3 minutes of data across 2 videos, totaling 11,524 frames.

To properly analyze these models in the context of our requirements and specifications, we developed a custom evaluation pipeline as part of the DolphinTrack System. There is a UI to manually label ground truth coordinates across frames for each dolphin and save the results to a JSON file for parsing. Inference results are also set up to output tracked coordinates to a JSON file. These files are then compared to generate results for the metrics outlined in IV, whose performance is summarized in V.

Specification	Experiment Metric	Explanation
Maximum drop out time is < 5 s	Dropout	Average dropout across all dolphins (in seconds).
Maximum of 1 identity switch per 10 s	ID Switches	Average number of identity switches per dolphin.
$\|\|c_{cam} - c_{gt}\|\|_2^2$ must be < 15 in pixel space	Position Accuracy	Average accuracy of tracks in pixels (Euclidean distance between ground truth and tracked position).
No more than 5% of tracks should incorrectly identify dolphins	Incorrect tracks	Percentage of tracks that were not dolphins.
At least 50% of dolphin positions must be tracked	Positions Tracked	Percentage of ground truth dolphin positions that were tracked.
Inference time must be less than 3 times the length of the input video	Inference Time	Time it took to run the model on an A100 GPU (meeting our requirement of modest compute as it is available on Google Colab).

Table IV: Specifications evaluated through computer vision component testing.

To illustrate whether each model meets our requirements, V highlights metrics that meet requirements in green and those that are far from meeting as red. It is important to note that YOLO can't be evaluated on identity switches as object detections are not assigned an identity. There is a clear performance dropoff as the video segment gets longer, which is critically important for a system intended to monitor swimming behavior for hours at a time. SAM2, without further refinement, is not feasible for long-term application as it is unable to re-identify dolphins once it has lost a track due to the prompted nature of the model. ByteTrack and BoT-SORT provide promising results that can be further improved with fine-tuning of confidence and IoU thresholds for frame-to-frame associations, with ByteTrack offering more efficiency with some slight reductions in performance. Overall, ByteTrack will be used as the primary vision component of our system.

Trial	Model	Dropout	ID Switches	Position Error	Incorrect Tracks	Positions Tracked	Inference Time
Video 1 (3 sec)	YOLO	1.29 sec	N/A	7.95 px	0%	28.41%	6 sec
	ByteTrack	0.77 sec	2.75	17.45 px	0%	51.66%	6 sec
	BoT-SORT	0.68 sec	2	17.50 px	0%	50.18%	13 sec
	SAM2	1.27 sec	0	11.08 px	1.62%	89.67%	30 sec
Video 2 (3 min)	YOLO	8.65 sec	N/A	11.57 px	2.70%	32.32%	6 min
	ByteTrack	8.34 sec	22.25	12.16 px	5.16%	47.47%	6 min
	BoT-SORT	8.33 sec	20.75	12.17 px	5.12%	47.81%	10 min
	SAM2	179 sec	0	7.68 px	0%	3.59%	20 min

Table V: Computer Vision Model Performance

This experiment influenced several changes in our design and provided further justification for the combination of the computer vision and RF localization components. The difficulty of re-identifying prompts for SAM2 and SAMURAI motivated the development of a labeling pipeline that would allow for re-prompting in addition to ground truth identification. This will help to address the shortcomings in dropout, along with reductions in the overlap threshold necessary to maintain tracks across frames. Dropout under severe occlusion that can't be overcome by vision fine-tuning can be addressed with our particle filter localization system, which can fill in detection gaps. The prevalence of identity switches is a primary motivator for our RF localization system to handle associations. This will allow us to re-associate visually detected tracklets for tagged dolphins upon

surfacing. Incorrect tracks is a large concern of our sponsor, who is keen on avoiding misleading detections that could disrupt movement behavior analysis. This can be mitigated by increasing the confidence threshold necessary to identify a detection and maintaining consistently positioned detections across frames to prevent object switching. Inference time is largely dependent on compute resources and is not a primary concern as there are no real-time detections being made. Efficiency is an important consideration, however, when allocating resources for analysis.

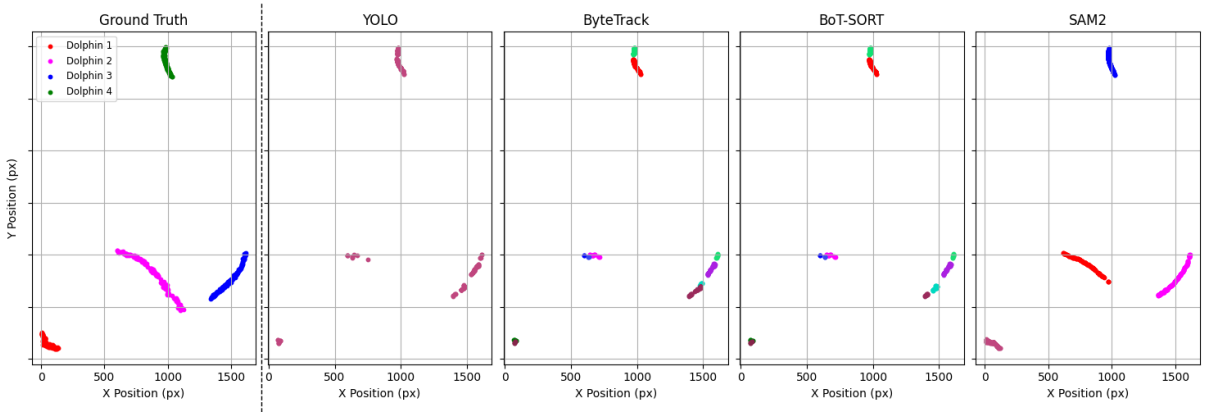


Fig 8: Ground Truth and Detection Positions from CV Models for Video 1

To visualize the results of these models, we include a coordinate plotting capability from the centroid files we created during evaluation. As shown in Figure 8, positional plots are consistent with performance metrics for the first (shorter) video. YOLO generates accurate tracks without identities, ByteTrack and BoT-SORT generate accurate tracks with identity switches, and SAM2 generates continuous tracks without identity switches.

B. Radio-Frequency Positioning System

We performed several experiments to verify the capabilities of the radio-frequency positioning system. We tested the ranging accuracy, the affect of water on ranging, and the positioning accuracy.

1) Ranging accuracy

To test the ranging accuracy of the radio-frequency system, we placed an anchor and tag at three fixed distances and recorded the mean and standard deviation of the estimated position. The anchor and tag devices used in the experiment were kept constant, and both were kept in a fixed position and orientation. We recorded the range for about ten seconds for each tag.

Actual Distance (m)	Mean (m)	Standard Deviation (m)
1.00	0.74	0.02
5.00	4.72	0.02
10.00	9.82	0.03

Table VI: Test of ranging accuracy at three distances. The range measurements exhibit acceptable error, but the delays could be tuned further to achieve higher accuracy.

The error for the ranging system was between one and three decimeters for all the distances we tested. While we did not meet our specification of less than one decimeter of error, this level of error ultimately ended up being satisfactory for making assocations between tag and camera position estimates. If the antenna delays of the DW1000 modules were tuned more precisely, this error could potentially be reduced. However, for our purposes, a few decimeters of error was acceptable. The standard deviations of the ranges were low, which leads to a reliable range estimate over time. Overall, the ranging capabilities of the system were satisfactory.

2) Underwater performance

Since the goal of this project is to localize dolphins, we needed to assess how radio-frequency positioning would be affected by water. While the system is only intended to provide position updates when dolphins surface, we wanted to characterize how the depth of the tag underwater would affect the rate of the ranging transactions.

We set up the tag in a waterproof bag and fastened it to a bamboo steamer with a glass dish attached for ballast (Figure 9). We placed this setup into a bathtub and adjusted the depth of the tag by adding more water to the

bathtub. We transmitted at a fixed rate of about 60 Hz and held the anchor at a fixed distance from the tag of about one meter. We recorded the number of ranging transactions completed by the anchor per second using a laptop computer connected over USB.

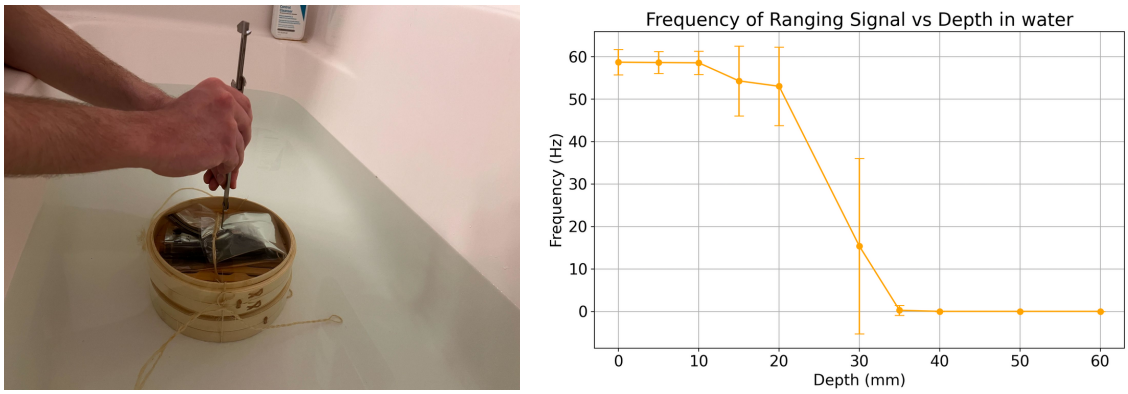


Fig 9: Our experimental setup and results for our underwater ranging test. Ranging cuts out completely below about three centimeters underwater.

Figure 9 shows the results of our underwater ranging test. As the depth of the tag underwater increased, the mean ranging frequency decreased and the standard deviation of the frequency increased. Below three centimeters underwater, the ranging cuts out completely. This is a much more drastic reduction in ranging capability than we initially expected, but we believe that RF localization still shows promise for this problem domain. These results verify that the radio-frequency positioning system will only be useful when the dolphins surface.

3) Positioning accuracy

To test the positioning accuracy of the radio-frequency system, we set up an experiment to compare a known trajectory with the estimated trajectory. We set up our tags in the Robotics Atrium and recorded the position estimates from our tag as we walked through a proscribed trajectory. We measured the points of the known trajectory in the coordinate frame of the radio-frequency system for comparison to the tag trajectory. The tag was set to transmit to the anchors at a fixed rate. We attempted to walk with the tag at a roughly constant speed and hold it at an approximately fixed height and orientation.

The results of the positioning experiment are displayed in Figure 10. The estimated tag position tracks the ground truth position fairly accurately. The trajectory error had a mean of 36.8 cm and a standard deviation of 34.5 cm. When the tag becomes co-linear with the two leftmost anchors, the ranging accuracy deviates from the ground truth. This shows the importance of anchor placement for accurate positioning. In order to resolve this issue, all anchors should be placed so that they fully surround the region where the tag will operate.

V. Discussion

Throughout the design and implementation of **DolphinTrack**, several areas became apparent where further work could improve the system's performance. The design decisions made were influenced by the need to balance accuracy and computational efficiency, but there are clear opportunities for improvement in both the tracking algorithms and hardware integration.

We believe a major roadblock was a lack of access to the actual environment our proposed solution will be used in. Since the testing environments at Brookfield Zoo and Dolphin Quest Oahu were not accessible during development, we had to make several assumptions about the environmental conditions. This lack of immersion in the actual workspace required us to ignoring key factors, like varying water and weather conditions and dolphin surfacing behavior, which are critical for the robustness of the tracking algorithms. For example, if we had access to the environment, we would have spent time exploring unique data collection methods, and collecting high quality camera data across weather conditions to help our vision models generalize better. We also ended up finding that to achieve high accuracy with the RF-system, it would require a calibration pipeline in the real environment, which we would explore and perfect if we had access. Some of our main design flaws are identified below.

Visual tracking improvements: The new models showed improvements in tracking accuracy, reducing dropout times and identity switches compared to earlier approaches, but still struggled over longer video segments, indicating room for further refinement. However, it is important to note that this is not due to model limitations,

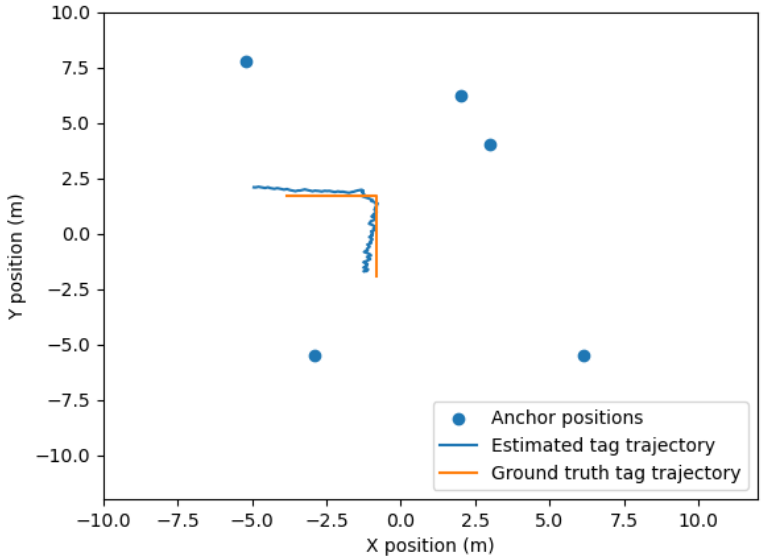


Fig 10: A chart showing the results of our positioning test. The ground-truth and estimated trajectories are shown, along with anchor positions. When the tag is co-linear with two of the anchors at the edge of the anchor constellation, the positioning becomes less accurate.

but rather due to lack of time and resources. It is completely feasible and possible to track multiple dolphins, continuously, uniquely, and all at once. SAMURAI, in particular, demonstrated exceptional capabilities in maintaining a continuous, accurate track [16]. However, it does not yet support multi-object tracking. The best use of time and resources would be to implement a multi-object version of SAMURAI, and run it using a high performance GPU - this would provide the highest accuracy and continuity. An implementation of this would run n instances of SAMURAI for n dolphins in the environment, where each model would be responsible for tracking exactly one dolphin. Since the number of dolphins are constant, this would demonstrate the greatest performance if given enough compute.

Poor underwater performance of RF Positioning: The RF system was always meant to only work when the dolphin surfaces. However, from literature, we expected that it may work at least 50 cm below the water surface. Our tests revealed that it only worked 3 cm underwater, significantly failing expectations. For RF positioning slightly further underwater, we would redesign to incorporate signals that would work slightly better underwater. Given more time and resources, we would focus on experimenting with lower frequency signals, or magnetic field based localization [26].

Lack of testing in the real environment: One of the design's main flaws is not in the design itself, but rather in the testing. While both subsystems have been benchmarked (to a certain degree) on data in the water, the complete system remains untested underwater or outdoors. We know the baseline capabilities of the RF system from the underwater ranging test, but we still do not know the effects that continuous deep underwater motion will have on the positioning system. We also do not know the effects of multiple tracked dolphins on the RF system and the camera's ability to track in any viewpoint.

VI. Reflection

In the course of this project, we have gained a deeper understanding of both the technical and societal impacts of our work. Initially, our focus was on achieving high precision in dolphin tracking. However, as we progressed, we started to see broader applications of this framework extending beyond the immediate research goals.

A. Product Influence

There are several areas our solution could have impact; we believe this solution could cause a cascading improvement to engineering and environmental research. Our project has no direct link to public safety and welfare, but does, however, have a close link to environmental safety and protection. Our **DolphinTrack** gives the user the ability to precisely follow the position of a moving marine agent while maintaining tracking reliably. The consolidated system itself is low maintenance, minimally invasive to the animal, and also scales easily (by adding more anchors to larger areas). As such a system is developed further (whether its by the

sponsor or by an interested party), researchers' ability to collect high quality data to study animals will greatly increase. This could greatly benefit the environmental research community. Additionally, having stronger data to study dolphins could accelerate the development of bio-inspired machines and algorithms. This has significant potential for industries such as underwater exploration, marine research, and conservation, impacting the global research community and marketplace. An important negative impact this product could have is the generation of e-waste. Any users must take care to keep the anchors and tags well maintained to prevent excessive construction of new PCBs. Any improvements made to the boards must also be safe for the dolphins and those working with the dolphin.

B. Social Dynamics Within Our Project

The synergy between our different expertise is exemplified in the interesting way we approach the localization problem. In this case, our diverse *technical* backgrounds produced an interesting, novel system for dolphin tracking. Between all the team members, we had experience in computer vision, 3D geometry and algebra, systems programming, PCB Design, wireless communication and more. The combination of all of this background realized a product that used all of this knowledge collectively. We proposed a system that integrated computer vision, geometry, and radio communication to solve our positioning problem. While there were no particular cultural dissimilarities we had to navigate, we did have to navigate through each members different works styles. Some of us tend to focus on rapid development and prototyping, while some of us focus on rigorous testing of ideas and effective communication. So we often used our weekly meetings to bridge gaps and sync up.

Our main design was mostly influenced by our sponsor. Dr. Shorter has the best understanding of the dolphins' behaviors and their environment, so he served as our most reliable advisor regarding any design choices. This dynamic could have over-constrained our prototyping process, but we preferred to spend less time ideating and more time focusing on building a product that would provide utility to the researchers. The sponsor's invaluable recommendations on constructing the radio-frequency positioning system and communicating the needs of the visual-tracking system gave us a roadmap towards solving the main problem of project.

C. Inclusion and Equity

Inclusion of all team members' was very important to us, as we all cared about learning new things through this course. We took all ideas as potential solutions, and had discussions about their pros and cons. For some situations, we consulted a decision matrix. This was one way we maintained inclusivity of technical skills throughout the semester. To ensure team members were heard, we practiced active listening and took note of all action items that were discussed, and assigned them to the team member with the most relevant skills. While there were definitely cultural differences between all of us, they did not affect our project, as we believe we all focused on producing a high quality project by the end of the semester.

D. Ethical Considerations

There were some ethical considerations we addressed, particularly regarding the well-being of dolphins in captivity. Our primary goal was to make data collection as unobtrusive as possible, prioritizing methods that minimize behavioral disruption and avoid physical contact. To that end, we deliberately avoided sonar-based approaches, since dolphins rely on echolocation, and instead selected RF-based methods using ultra-wideband (3.1–10.6 GHz) and LoRa (915 MHz), which do not interfere with the animals' natural communication or sensing. Furthermore, all localization hardware was integrated into the existing MTag housing, ensuring no significant increase in the device's mass or footprint. This allowed us to maintain the established suction-cup mounting system without affecting the dolphins' comfort or movement.

VII. Future Revision Recommendations

A. Radio-Frequency Positioning

Based on system testing and deployment constraints, we identified several specific areas for improvement in future iterations of the RF localization system:

- Use a fourth DW1000 message (final message from anchor to tag) to transmit the calculated range back to the tag. Since ultra-wideband supports a much higher data rate than LoRa, this would allow the MTag to log its position data to the SD card at a higher frequency than the LoRa server can receive.
- Replace the STM32F0 microcontroller with a faster alternative. The DW1000 SPI clock is currently set to 18MHz, but this can be increased to at least 20MHz. Several developers (see posts by AndyA on the

Decawave forums) have reported success using 30MHz. This change would reduce communication latency with the DW1000 and improve overall throughput.

- Add a glass cover to the anchor enclosure to prevent damage to the IPS screen during handling or deployment.
- Develop an automatic anchor position calibration tool. The calibration process should be GUI-based, where the user inputs rough anchor positions and the software refines them using actual inter-anchor TWR measurements. Anchors would enter calibration mode via a downlink LoRa message, then perform TWR with the next-highest anchor ID using a polling sequence and timeout.
- Port the codebase to ESP32 by modifying port.h to use the ESP HAL and removing all STM32 HAL references. An alternative is to switch to the thotro/arduino-dw1000 library, although we haven't validated it for this use case.

B. Recommended Anchor Placement at Dolphin Quest Oahu

Our recommended anchor placement aims to cover the entire lap-swimming environment with five anchors placed around the lagoon. Ideally, the anchors should be placed outside the swimming area and a few meters above the water level. Here, we also propose placing anchors 1 and 2 on the roof, and anchors 3, 4, and 5 mounted on tripods 11 12. Placing the anchors above the water level will increase their functional range as there will be less attenuation.

Fig 11: Dolphin Quest Oahu top view with recommended anchor placements

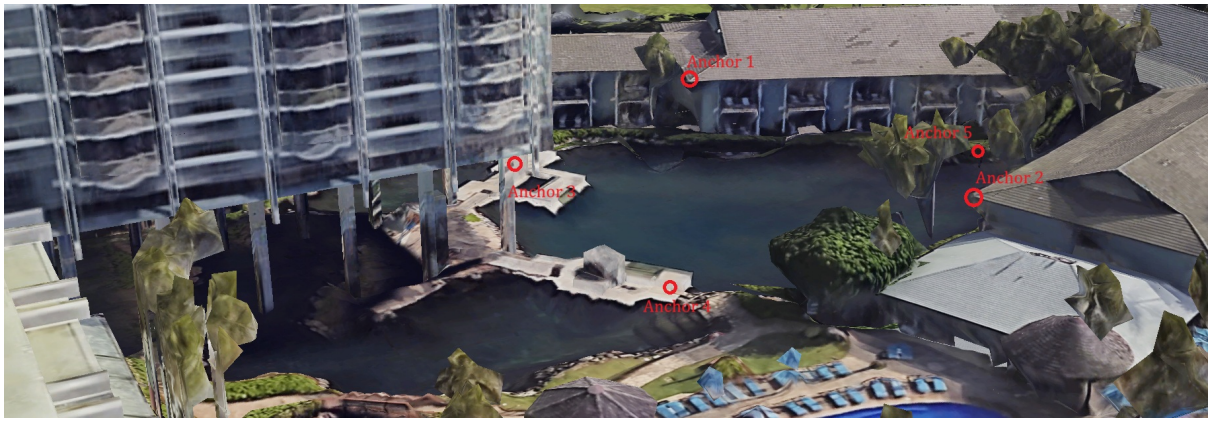


Fig 12: Dolphin Quest Oahu side view with recommended anchor placements

C. Computer Vision

With greater access to on-site data for training and validation, there are several areas for improvement in the computer vision system.

1) Data Acquisition and Training

For supervised learning models, data is critical for improving model performance and generalization. This involves varied, representative, and comprehensive samples of the environment, which limited our ability to dictate our sample set. A strong recommendation would be to explore training models under varying weather conditions and incorporating data augmentation as a pre-processing step to artificially change lighting conditions and potential occlusion sources. To evaluate performance in different environments, we would also recommend annotating more data both for dolphins in Dolphin Quest Oahu and Brookfield Zoo Chicago. We were severely limited in model training and fine-tuning by the amount of data we were able to annotate. Having refined models in different conditions, especially with efficient data collection methods, would allow for greater evaluation capabilities. From a purely visual perspective, it would seem like the distinction between dolphins and water is clearer at Brookfield. It would also be interesting to analyze the extent to which models could generalize toward other environments by applying different combinations of data to the lagoon and to the aquarium.

2) Model Architectures

There are always new model architectures being released that handle real-time object tracking, addressing various shortcomings of others that might be relevant to our problem. We would recommend keeping up to date with these, exploring especially those that are promoted by Roboflow to streamline the fine-tuning process. From the architectures explored in our project, we would recommend extending SAMURAI to work with multiple dolphins given it's superior performance for single dolphin tracking. The underlying principle of applying a Kalman Filter to segmentation masks could reasonably be extended to multiple instances. Another direction we would recommend exploring is unsupervised learning, which would eliminate the need for labeled data, allowing for faster iteration and broader generalization. The amount of passive data being collected through free swimming recordings could be leveraged to discern patterns from the motion itself. An approach could be extending the idea of optical flow, which wouldn't be directly applicable to us because of noise from wave and other object motion but would provide a solid foundation for finding motion across frames in an otherwise relatively static environment.

3) Efficient Tuning

To continue on the idea of efficiency, which is critical for model development, tuning hyperparameters can and should be optimized to further push the limits of our existing models. As mentioned in response to the verification testing, some values to start with would be the confidence threshold (which determines how confident the model has to be in the detection to classify something as a dolphin), the IOU threshold (which determines how much overlap there needs to be between detected and ground truth bounding boxes to classify something as a dolphin), and frame-by-frame overlap requirements (which determine whether to continue a certain dolphin identity throughout multiple frames of dolphin tracks. These, along with core machine learning model hyperparameters like the learning rate and number of epochs, can be automatically traversed by sampling within a proposed range of values.

4) Automated Re-Prompting

Given the promising nature of SAMURAI and SAM2, both of which rely on an initial bounding box indication and would be greatly improved by re-prompting throughout a video. Automating the process of re-prompting would make for significantly more impactful models, especially when running on long videos taken from free swimming sessions that aren't closely monitored on the tracking side. One way to do this would be to determine whether there has been a significant period of time with no detections, at which point re-prompting could be done with a single frame classification model like YOLO, or alerted to a user who could leverage our custom UI for ground truth position labeling.

VIII. Conclusion

Tracking biological systems in marine environments presents several key localization challenges, including visual occlusion and signal attenuation. To fully understand group swimming behavior of dolphins, we must maintain long-term associated tracks which can be achieved with a combination of visual detection and radio frequency positioning. This report presents an overview of our visual-inertial tracking problem, as well as a detailed description of the final iteration of our system's design. We have outlined the key stakeholders that

influence our project and have broadly explored the solution space in different contexts. This has allowed us to prove our design to address our sponsor's (and to a lesser extent, our stakeholders') needs and expectations.

Our final system embodiment, DolphinTrack, is a fused vision and radio frequency tracking system that handles position updates and re-association for long-term dolphin tracking in group swimming environments. The vision tracking system makes conservative detections to ensure identification is limited to dolphins, but still achieves accurate positioning for half of ground truth positions. Its major shortcoming is the susceptibility to dropout and identity switching, which is inevitable in marine environments regardless of model improvements and hyperparameter tuning. To address this, our radio frequency system generates position signals that are evaluated in comparison to frame detections to re-associate visually tracked dolphins with a tagged identification. This system is demonstrated to be applicable in these scenarios as it achieves 30 cm position accuracy, which is within the margin of error for unique camera detections and will therefore produce an overlapping estimation within a certain threshold. Ranging is achieved up to 3 cm depth, meaning re-association occurs when dolphins surface. This is a reasonable assumption given the length of free swimming videos that would otherwise be unmonitored. We hand off this design to Dr. Shorter's lab to implement in the field and gain a more thorough understanding of dolphin dynamics.

IX. Bill Of Materials

Table VII: Anchor PCB BOM

Name	Quantity	Manufacturer	Manufacturer Part Number	Supplier Unit Price	Supplier Subtotal
1uF	10	Murata	GCM188R71E105KA64J	0.088	0.88
12pF	2	KEMET	C0603C120J5GACTU	0.007	0.014
10uF	1	Murata	GRM188R60J106ME47D	0.065	0.065
52746-1871	1	Molex	52746-1871	2.36032	2.36032
2171790001	1	Molex	2171790001	0.75	0.75
CONN HEADER 2POS	2	JST Sales America Inc.	S2B-PH-SM4-TB	0.44	0.88
22R	3	TE Connectivity Neohm	CPF0603B22RE1	0.12	0.36
10k	3	Vishay Dale	CRCW060310K0FKEC	0.019	0.057
1k	1	Panasonic	ERJ-3EKF1001V	0.145	0.145
100k	1	Yageo	RC0603FR-13100KL	0.00899	0.08985
200k	1	Yageo	RC0603JR-07200KL	0.003	0.003
5k1	2	Yageo	RC0603FR-075K1L	0.009	0.09
SWITCH SLIDE SPDT	1	E-Switch	500SSP1S1M6QEA	3.52	3.52
IC REG LINEAR 3.3V	1	Texas Instruments	TLV70033DSER	0.28	0.28
STM32F072CBT6TR	1	STMicroelectronics	STM32F072CBT6TR	4.54	4.54
DWM1000	1	Qorvo	DWM1000	23.62	23.62
RF TXRX SMD LORA	1	Seeed Technology	317990687	6.83	6.83
1-CELL 1-A LINEAR	1	Texas Instruments	BQ25185DLHR	1.62	1.62
CRYSTAL 8.0000MHZ	1	Abracon LLC	ABC2-8.000MHZ-4-T	2.09	2.09
300R	1	Panasonic	ERJ-3GEYJ301V	0.1	0.1
RF ANT SOLDER SMD	1	TE Connectivity Linx	ANT-916-CHP-T	2.91	2.91
13k	1	Panasonic	ERA-3AEB133V	0.1	0.1

Table VIII: Tag Daughterboard PCB BOM

Name	Quantity	Manufacturer	Manufacturer Part Number	Supplier Unit Price	Supplier Subtotal
1uF	6	Murata	GCM188R71E105KA64J	0.088	0.88
2171790001	1	Molex	2171790001	0.75	0.75
22R	2	TE Connectivity Neohm	CPF0603B22RE1	0.12	0.24
10k	1	Vishay Dale	CRCW060310K0FKEC	0.019	0.019
5k1	2	Yageo	RC0603FR-075K1L	0.009	0.09
TLV70033DSER	1	Texas Instruments	TLV70033DSER	0.263	0.263
STM32F072CBT6TR	1	STMicroelectronics	STM32F072CBT6TR	4.54	4.54
DWM1000	1	Qorvo	DWM1000	23.62	23.62

Table IX: Miscellaneous BOM

Description	Quantity	Supplier Unit Price	Supplier Subtotal
Adafruit			
EYESPI Cable - 18-pin 50mm long Flex PCB (FPC) A-B Type	10	0.75	7.5
1.47 320x172 Rectangle Color IPS TFT Display - ST7789	5	17.5	87.5
Mcmaster			
90 Degree Countersink M2 x 0.40 mm Thread 20 mm Long	11	1	11
90 Degree Countersink Angle M2 x 0.40 mm Thread 6 mm Long	1	8.75	8.75
Black-Oxide Steel Hex Nut M2 x 0.40 mm Thread	1	7.35	7.35
Amazon			
20 Sets Mini Micro Jst 2.0 Ph 2-Pin Connector Plug Male with 150mm Cable & Female	1	8.59	8.59
KBT 3.7V 1200mAh Li-Polymer Battery	1	42.99	42.99
JST PH 2.0 mm Pitch 2-Pin	1	5.99	5.99
Digikey			
SWITCH PB SPST-NO 0.125A 125V	5	4.73	23.65
PCBWay			
Anchor PCB Blank 4 layers HASL 1.6mm 5 Pieces	1	77.71	77.71
Tag Daughterboard PCB Blank 4 layers HASL 1.6mm 5 Pieces	1	50.06	50.06

Table X: Combined BOM

Item	Quantity	Unit Subtotal	Total
Anchor	5	51.30	256.52
Tag	3	30.40	91.21
Anchor Case 5 Pieces	1	203.32	203.32
PCB Blanks	1	127.77	127.77
Combined Total			**678.82**

X. Team Member Bios

Advaith Balaji is from Ann Arbor and is currently pursuing a degree in Robotics with a minor in Computer Science at UMich. His passion for robotics stems from early involvement in robotics through the FIRST program in high school. In addition, the small close-knit UM Robotics undergrad community gave him an opportunity to be immersed in Robotics & AI from day one, which led him to explore projects in robot navigation, computer vision, and manipulation. He has a strong foundation in programming, and enjoys exploring mathematics for Robotics.

Advaith Balaji's research lies at the intersection of robot manipulation, semantic reasoning, and object geometry, focusing on integrating language and vision for intelligent robotic grasping. He has worked on language-guided object search, 3D object localization with SDFs, grape localization, and task-oriented grasp recovery, with publications in ICRA and the Michigan AI Symposium. In the future, he plans to pursue a PhD in robotics, advancing robot perception and decision-making in unstructured environments. Outside of robotics, he enjoys cooking and experimenting in the kitchen, as well as mountain climbing.

Sydney Belt is a student at the University of Michigan College of Engineering, pursuing a degree in Robotics. She is from Arlington, Virginia and went to a local technology-focused high school that sparked her passion for robotics and deep learning. This was also when she started teaching classes that aimed to expand access to computer science education. She has extended this passion as a member of the teaching team for ROB 330: SLAM & Navigation and ROB 599: Deep Learning for Robot Perception.

Sydney's current research involves data attribution for imitation learning as part of the UMich Mapping and Motion Lab. She is also the current President and former Computer Vision Lead of the Autonomous Robotic Vehicle Team on campus. After graduation, she plans to complete her master's degree in robotics and build upon her industry experience in machine learning for embedded systems at Tesla and Apple. In her free time, she enjoys hiking, weightlifting, and playing soccer.

Thor Helgeson is a robotics student at the University of Michigan College of Engineering, pursuing a degree in Robotics with a minor in Art History. He is originally from Ann Arbor, Michigan. He first became interested in robotics after competing on his high school's FIRST Robotics Competition team. His research so far has focused on human factors in space. Specifically, he has investigated how viewpoint planning parameters affect manual space station inspection performance. He also helped develop the curriculum for the University of Michigan Robotics course ROB 204: Introduction to Human-Robot Systems. After graduation, he plans to complete a Master of Science in Robotics at the University of Michigan.

Terry Tao is a robotics student at the University of Michigan College of Engineering, pursuing a degree in Robotics. Terry was originally from Long Island, and decided to be a robotics major because he did robotics in High School. He decided neither mechanical engineering nor electrical engineering fit him completely, he wanted something more interdisciplinary, which is why he joined the MRacing Formula Student team his freshman year. Now Terry is the Autonomous Director for MRacing and is responsible for making sure the driverless car competes at MIS this summer.

REFERENCES

[1] "New activity trackers for dolphin conservation – mechanical engineering."

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[3] M. Xia, J. Zhang, N. Wang, G. Antoniak, N. West, D. Zhang, and K. A. Shorter, "Cornering in the water: An investigation of dolphin swimming performance," *arXiv preprint arXiv:2411.17688*, 2024.

[4] T. Corporation, "World's lowest power 9-axis mems motiontracking™ device," *ICM-20948 datasheet*, 2024.

[5] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, p. 1004–1020, Aug. 2018.

[6] G. Huang, "Visual-inertial navigation: A concise review," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 9572–9582, 2019.

[7] J. Zhu, H. Li, and T. Zhang, "Camera, lidar, and imu based multi-sensor fusion slam: A survey," *Tsinghua Science and Technology*, vol. 29, p. 415–429, Apr. 2024.

[8] Y. Wang, C. Xie, Y. Liu, J. Zhu, and J. Qin, "A multi-sensor fusion underwater localization method based on unscented kalman filter on manifolds," *Sensors*, vol. 24, no. 19, p. 6299, 2024.

[9] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[10] D. Zhang, J. Gabaldon, L. Lauderdale, M. Johnson-Roberson, L. J. Miller, K. Barton, and K. A. Shorter, "Localization and tracking of uncontrollable underwater agents: Particle filter based fusion of on-body imus and stationary cameras," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 6575–6581, IEEE, 2019.

[11] L. Yang, Y. Liu, H. Yu, X. Fang, L. Song, D. Li, and Y. Chen, "Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: A review," *Archives of Computational Methods in Engineering*, vol. 28, p. 2785–2816, June 2021.

[12] Y. Liu, B. An, S. Chen, and D. Zhao, "Multi-target detection and tracking of shallow marine organisms based on improved yolo v5 and deepsort," *IET Image Processing*, 2024.

[13] Q. Wang, X. Huang, H. Tao, Z. Xia, C. Liu, and X. Ren, "Marine navigation radar multi-target tracking using adaptive innovation sequence-based joint probability data association," in *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 511–516, IEEE, 2024.

[14] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 51–65, 2011.

[15] H. Soganci, S. Gezici, and H. V. Poor, "Accurate positioning in ultra-wideband systems," *IEEE Wireless Communications*, vol. 18, p. 19–27, Apr. 2011.

[16] C.-Y. Yang, H.-W. Huang, W. Chai, Z. Jiang, and J.-N. Hwang, "Samurai: Adapting segment anything model for zero-shot visual tracking with motion-aware memory," Nov. 2024. arXiv:2411.11922.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016.

[18] "Multi-object tracking with ultralytics yolo." https://docs.ultralytics.com/modes/track/.

[19] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," 2022.

[20] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," 2022.

[21] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. K. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. B. Girshick, P. Doll'ar, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," *ArXiv*, vol. abs/2408.00714, 2024.

[22] *DW1000 User Manual*, 2017.

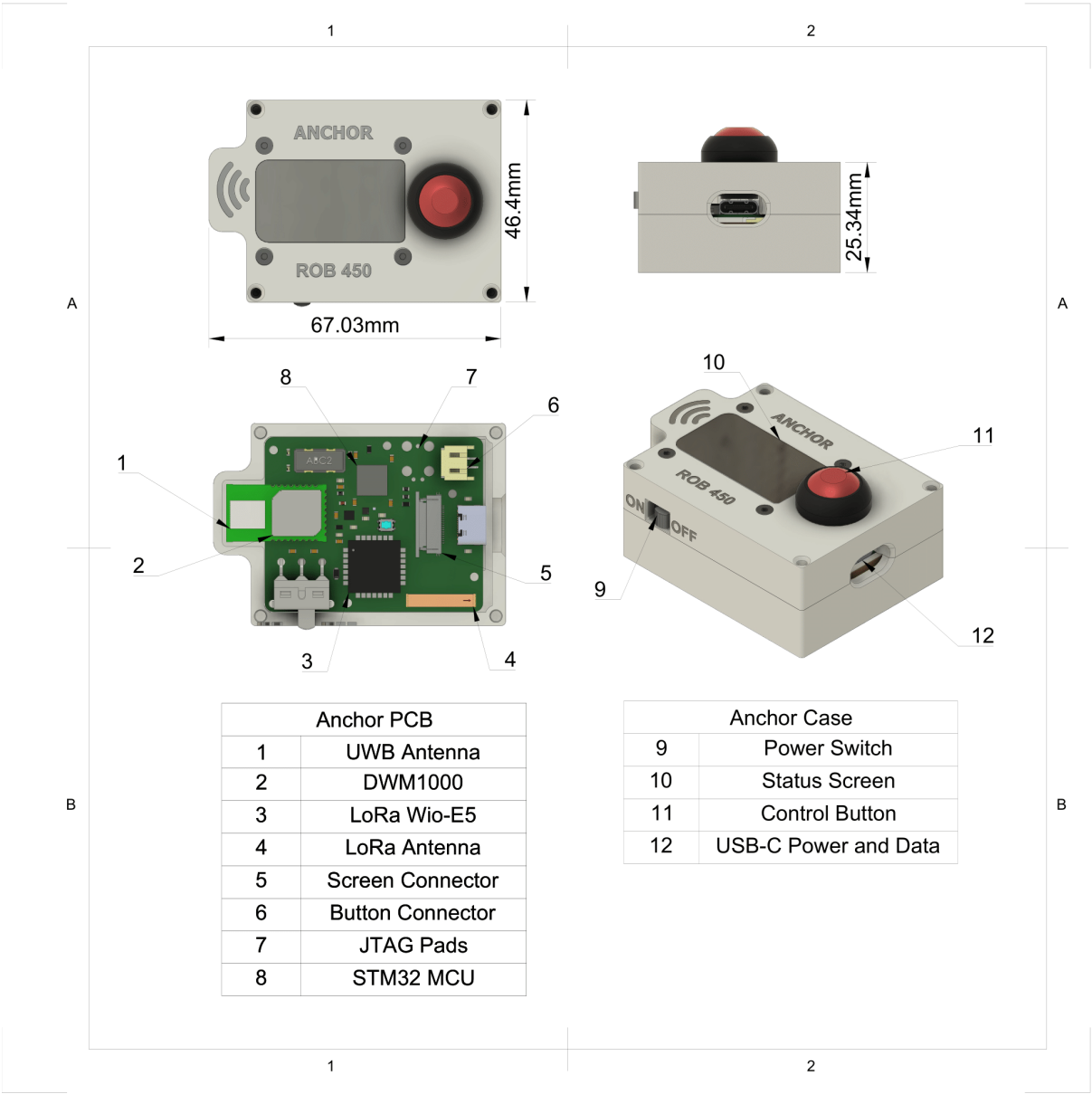
[23] A. Norrdine, "An algebraic solution to the multilateration problem," 04 2015.

[24] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for iot-based applications," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7680–7699, 2022.

[25] *Wio-E5 Datasheet*, 2023.

[26] S. Bian, P. Hevesi, L. Christensen, and P. Lukowicz, "Induced magnetic field-based indoor positioning system for underwater environments," *Sensors*, vol. 21, no. 6, 2021.

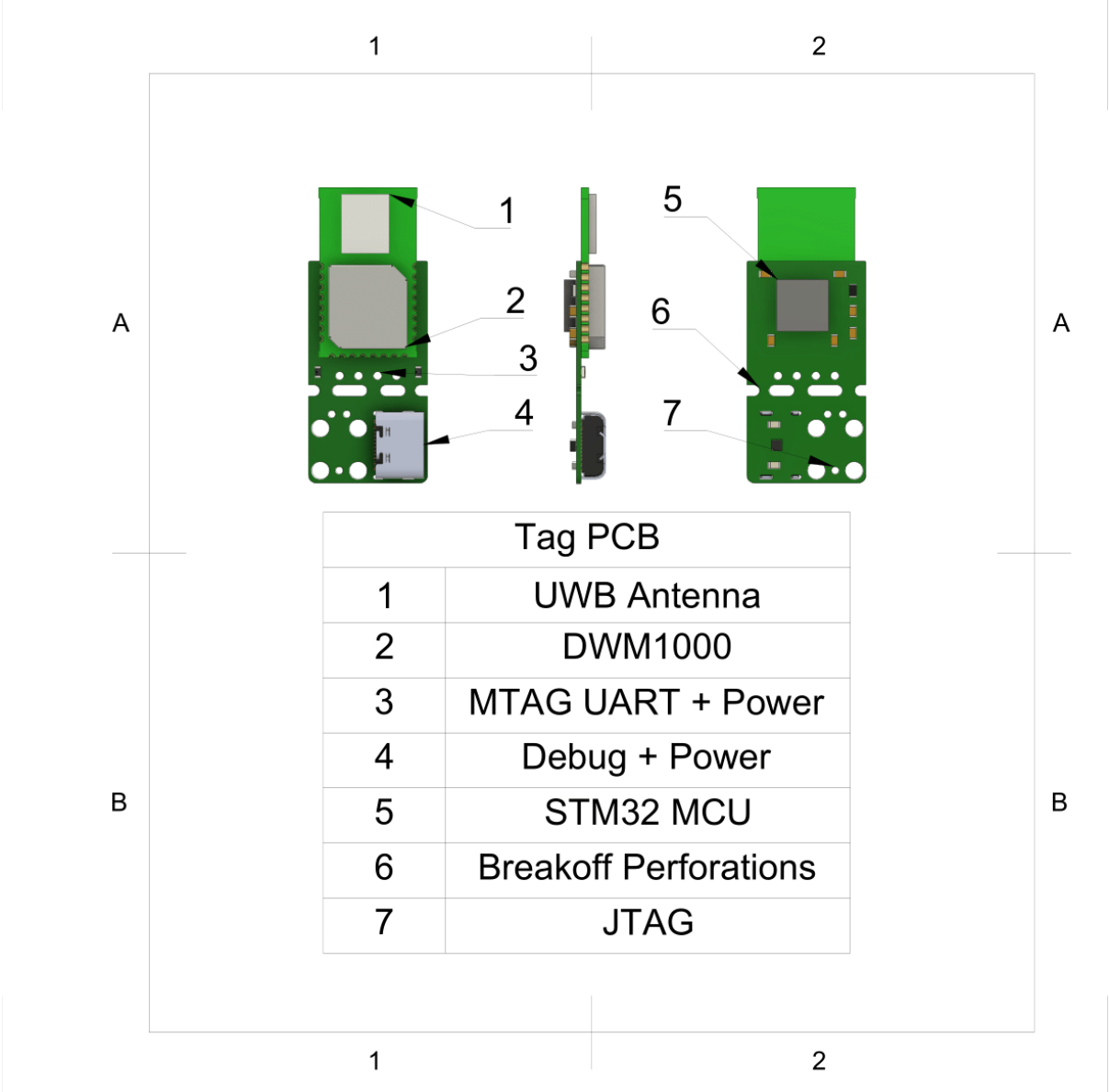
APPENDIX A
ANCHOR & TAG SCHEMATICS



Anchor PCB	
1	UWB Antenna
2	DWM1000
3	LoRa Wio-E5
4	LoRa Antenna
5	Screen Connector
6	Button Connector
7	JTAG Pads
8	STM32 MCU

Anchor Case	
9	Power Switch
10	Status Screen
11	Control Button
12	USB-C Power and Data

Fig 13: Mechanical CAD for the Anchor



Tag PCB	
1	UWB Antenna
2	DWM1000
3	MTAG UART + Power
4	Debug + Power
5	STM32 MCU
6	Breakoff Perforations
7	JTAG

Fig 14: Mechanical CAD for the MTag Daughterboard

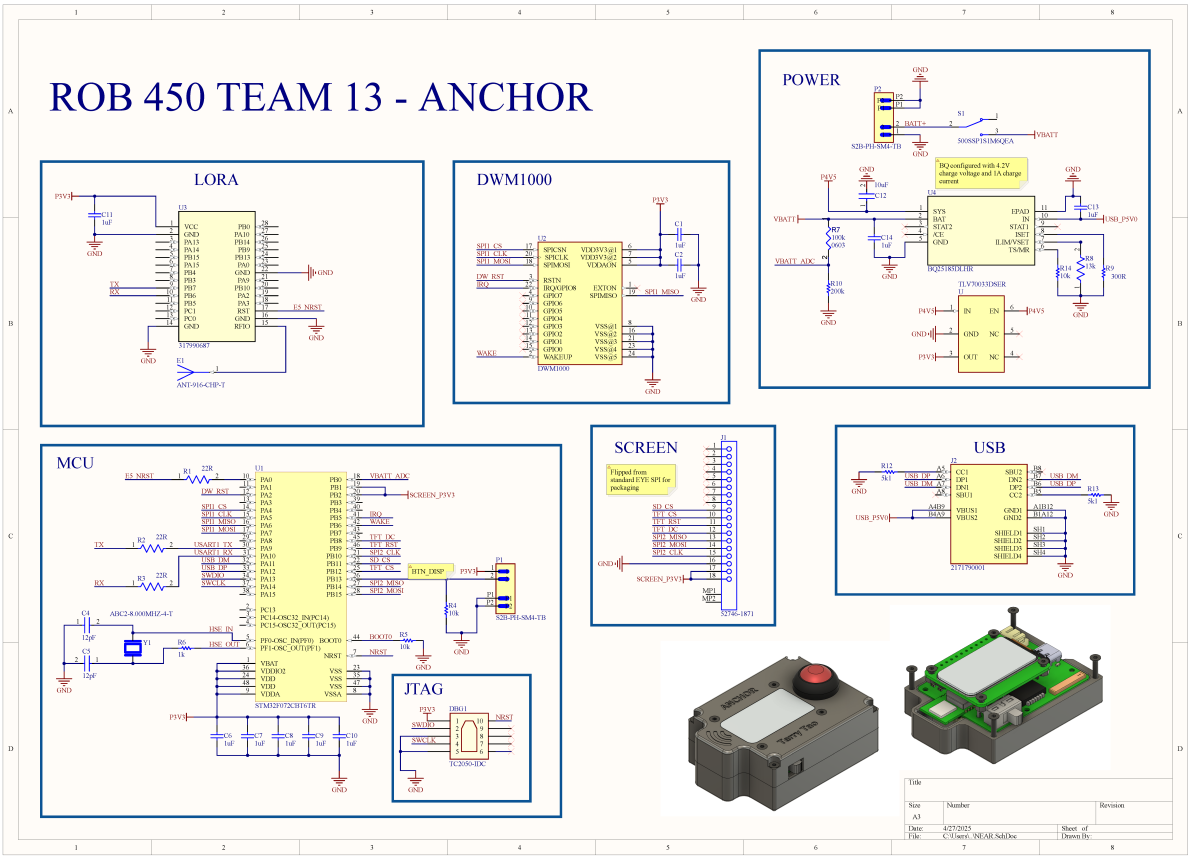


Fig 15: Electrical Schematic for the Anchor

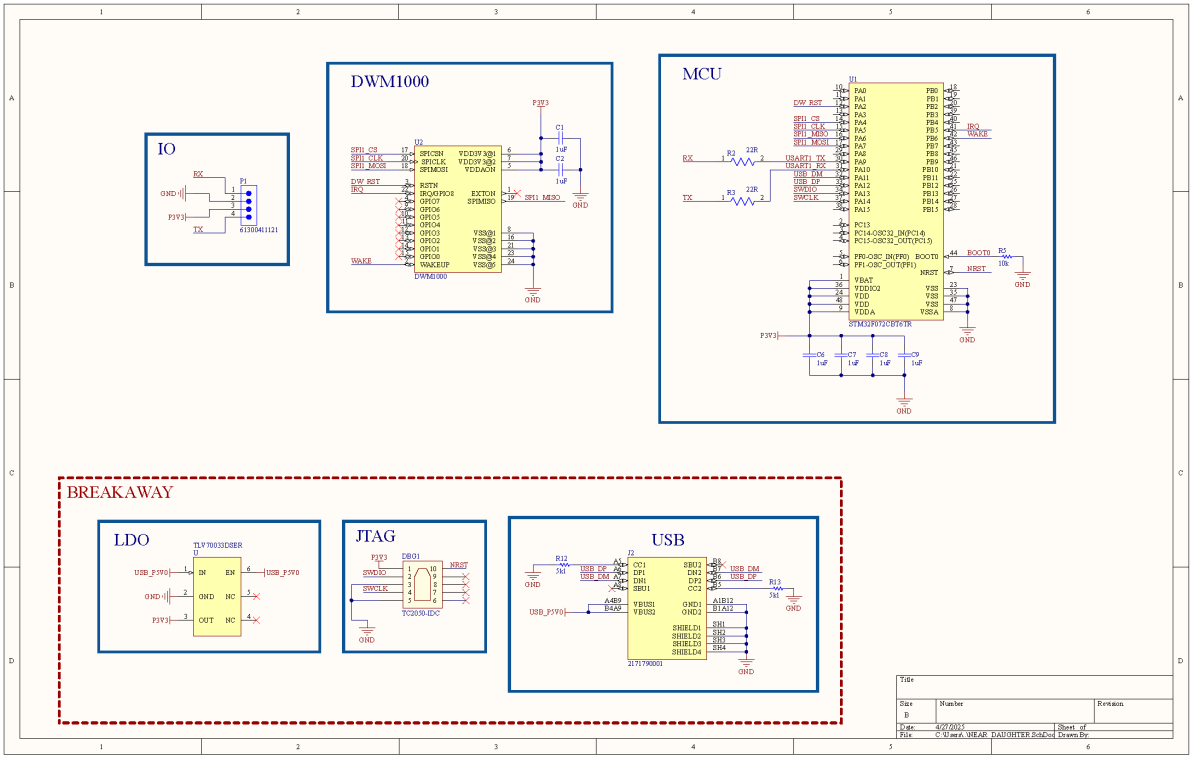


Fig 16: Electrical Schematic for the MTag Daughterboard

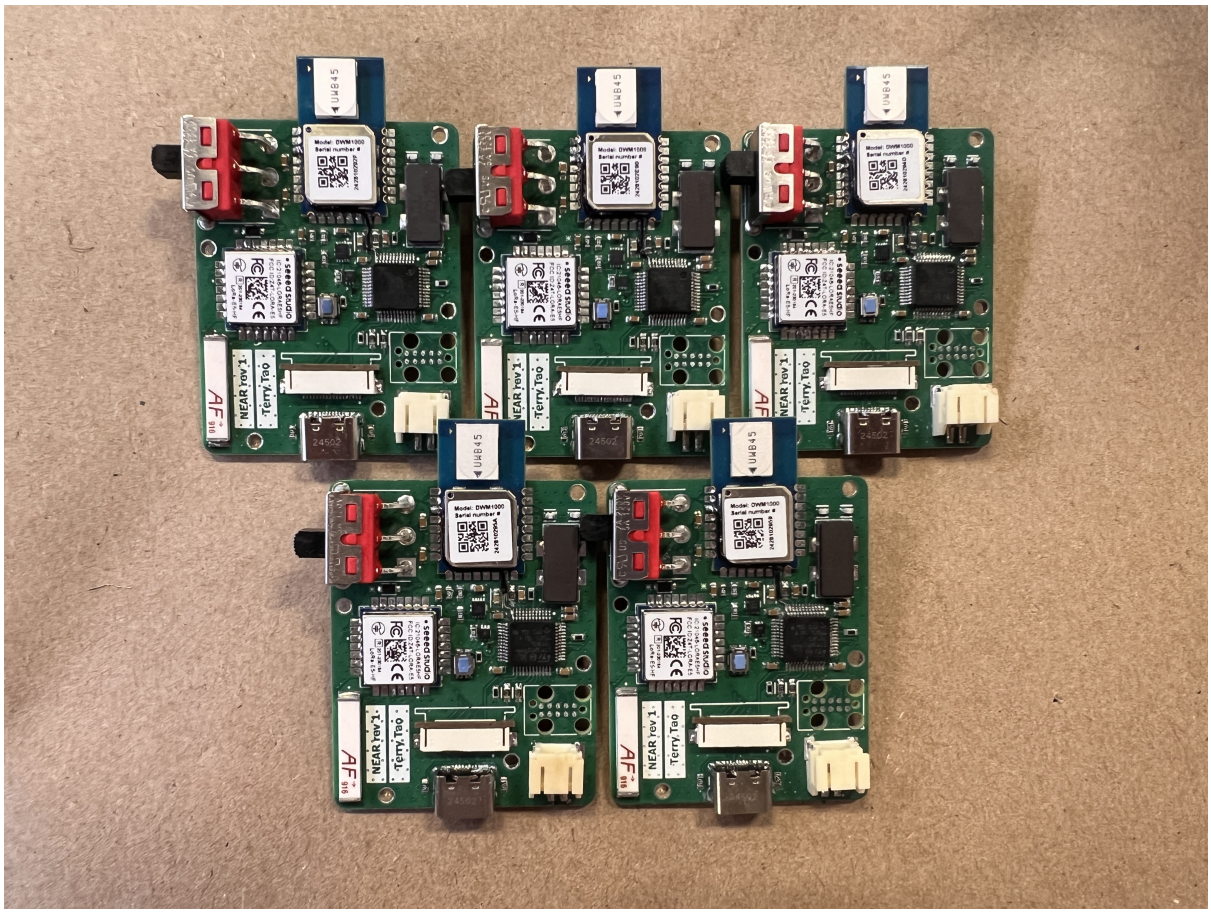


Fig 17: Fully soldered anchor PCBs

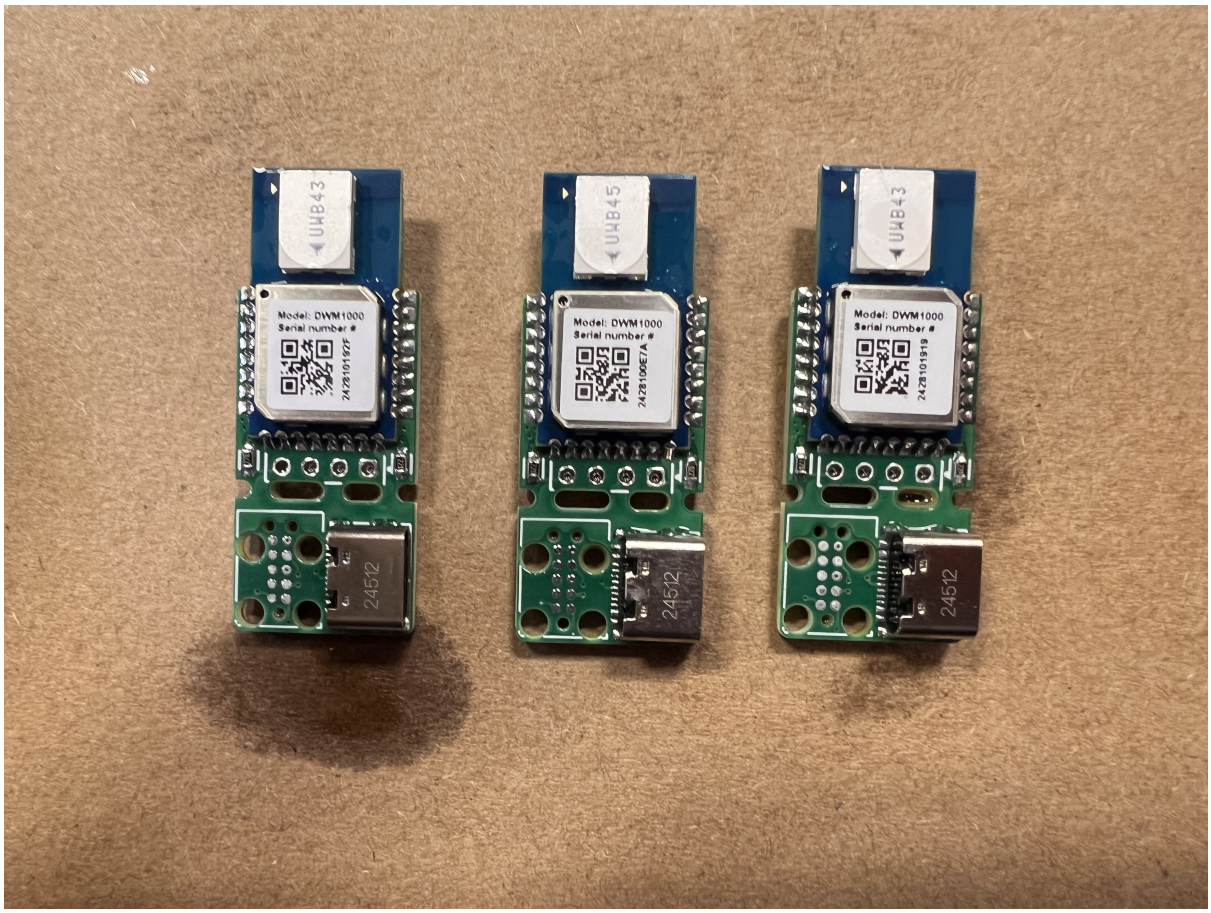


Fig 18: Fully soldered tag daughterboard PCBs

Fig 19: Fully assembled anchors