Data-Driven Localization and Structure Learning in Reverberant Underwater Acoustic Environments

by

Toros Arikan

B. S., University of Illinois at Urbana-Champaign (2014)
M. S., University of Illinois at Urbana-Champaign (2017)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2023

Authored by: Toros Arikan
Department of Electrical Engineering and Computer Science
May 19, 2023

Certified by: Gregory W. Wornell
Sumitomo Electric Industries Professor of Engineering
Thesis Supervisor

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Data-Driven Localization and Structure Learning in Reverberant Underwater Acoustic Environments

by

Toros Arikan

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 2023, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Abstract

Passive localization and tracking of a mobile emitter, and the joint learning of its reverberant 3D environment, are important yet challenging tasks in the shallow-water underwater acoustic setting. A typical application is the monitoring of submarines or other man-made emitters with a small, surreptitiously-deployed receiver array. This task can be rendered more difficult by obstacles such as seamounts or piers, which can occlude the line of sight from the emitter to the receivers. Furthermore, the underwater acoustic domain is complex and difficult to model, and a good signal-to-noise ratio is not assured. We view these complexities as features that can be leveraged for improved localization performance, using global optimization and neural network methods. We develop a multi-stage optimization and tracking architecture that precisely maps the reflective boundaries in the environment, and thereby uses the non-line of sight reflected arrivals for robust and accurate localization. Each stage of this architecture establishes domain knowledge such as synchronization and occluder estimation, which are inputs for the following stages of more refined algorithms. Within this framework, we introduce a 2D neural network boundary estimation method that outperforms the existing methods in the literature, and is robust to the large time delay estimation errors that are common in the application domain. We analyze the performance and reliability of this holistic framework, both in simulation and in real-life reverberant watertank testbeds that model the shallow-water underwater acoustic setting. The results are encouraging for the future development of better-performing localization methods with novel capabilities, using data-driven learning algorithms.

Thesis Supervisor: Gregory W. Wornell
Title: Sumitomo Electric Industries Professor of Engineering

Acknowledgments

I am deeply indebted to a great number of excellent researchers for their generous support over the years of my doctoral studies.

I would like to firstly express my deepest gratitude for my advisor Professor Gregory Wornell, whose invaluable advice and supervision inculcated the values of high standards and rigor. He provided an excellent research environment with a wide range of collaborators, projects, and resources for me and my collaborators. My Masters advisor and thesis reader Professor Andrew Singer likewise provided essential support over many years that made our collaborative efforts and accomplishments possible. My thesis reader Professor Lizhong Zheng offered invaluable advice on my research as well. Of the world-class faculty members at MIT from whom I learned a great deal, I would like to specifically thank Professors Yury Polyanskiy and Guy Bresler for inspiring me to be a better researcher and opening my outlook beyond my specific field of signal processing.

I have had the great fortune of collaborating with many incredible researchers in the projects that I have been involved in. In their postdoctoral capacity, Dr. Amir Weiss, Dr. Hari Vishnu, and Dr. Christos Thrampoulidis provided support without which this research would not have been possible. Our mentor Dr. Grant Deane put together impressive experimental testbeds on his own initiative and gathered the underwater acoustic experimental results, without which we could not have progressed beyond a simulation environment. We likewise relied on his decades of research and his oceanographic expertise for our work. Dr. James Preisig gave critical support and advice for our projects. I would also like to thank Dariush Kari and Jae Won Choi for their efforts towards the accomplishment of our difficult experimental work. A key contribution for the development in this work was our collaboration with Lincoln Laboratory on indoor radio-frequency localization; and I would like to specifically thank Dr. Swaroop Appadwedula for his essential support and advice. I also had the honor of having dedicated and friendly graduate students in my research group, who without exception generously gave me both technical support and fellowship

throughout the years.

Throughout my research, I received generous financial support that enabled us to work with cutting-edge technological equipment. My work was supported, in part, by ONR under Grant N00014-19-1-2665, and by NSF under Grant CCF-1816209. I was also supported by a MathWorks Fellowship and an MIT Presidential Fellowship, for which I would like to express my gratitude for the respective organizations.

On a personal note, I would like to thank my dearest friends, Ege Iseri and Okan Koksal, for their warmth, generosity, honesty, and integrity over the many years that we have known each other. My parents Orhan and Feza Arikan, and my sister Lara Arikan, were the greatest support that one could ask for.

Finally, I would like to dedicate this thesis to my dear departed grandparents, Sabri and Sebahat Coskun. I miss them dearly, and while I wish they could have been present in person to see their grandson graduate, I believe that they can still feel warmth at my undying gratitude and love towards them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

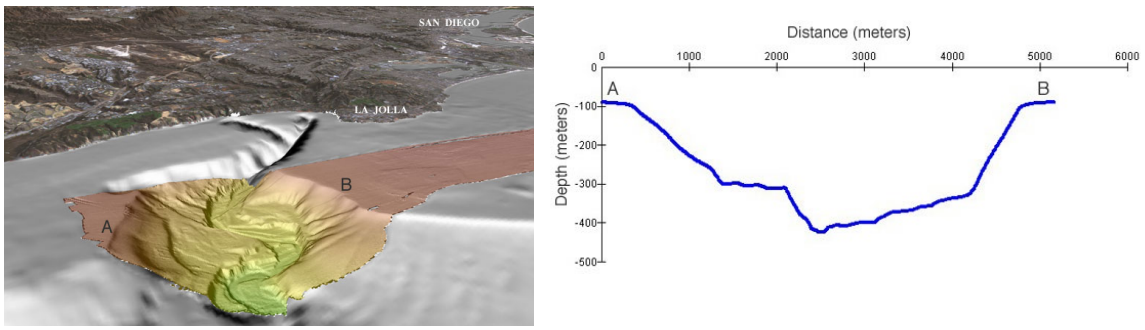
In many kinds of applications in electrical engineering, recent breakthroughs in computational resources and optimization methods has led to novel capabilities and large performance gains in well-established fields. There is a current trend towards automated, data-driven solutions to problems that were previously handled by handcrafted or user-supervised algorithms. Such solutions can offer superior robustness when expert knowledge is not available, or when there is no clear way to explicitly integrate domain knowledge into classical closed-form solutions. These developments encourage us to study underwater settings, specifically for sensing via acoustic modalities, as a field where new capabilities could be leveraged to approach various challenging localization and environment learning problems.

There is a long history of acoustic measurements being leveraged to reveal important information about underwater environments, with important civilian and defense-related applications [5]. Familiar examples of such methods are the passive time of arrival (TOA) localization and direction of arrival (DOA) estimation of acoustic sources, and side-scan sonar for the active imaging of the seabed. Traditionally, such methods have focused on comparatively simple propagation environments, and their associated analytically-tractable low-dimensional physical models. However, with the advent of modern data-driven artificial intelligence and machine learning methodologies, there is the potential to accommodate—and ultimately to exploit—the more complex structures present in difficult environments. Rather than explicitly

encoding domain knowledge into the calibration of localization and environment learning methods, which can be difficult due to complexity of the physical models involved, one can potentially use machine learning to implicitly incorporate this information into such methods in an automated manner.

Broadly, our focus has been on developing this potential in underwater acoustic sensing scenarios. In an illustrative example scenario that we consider throughout this study, there is a radiating source in the environment, whose location we cannot control, and a variety of unknown (non-radiating) structures in the environment, potentially including surface vessels and submarines, rigs, pilings, and buoys. While we would ideally like to have as many receivers as possible surrounding such an emitter, the recent development of inexpensive, autonomous underwater vehicles encourages us to accomplish with such a small receiver platform what could be done with a large array. Traditionally, one might be interested in the locations of the radiating sources, and view the non-radiating (but reflecting) structures in the environment as complicating the problem. However, the view we take is that this structure is also an important aspect of situational and environmental awareness, and thus it can be valuable to recover both the locations of sources and as much detail about the environment as possible. In essence, the source is effectively an acoustic 'flashlight' that illuminates the scene, allowing for a passive sensing of the broader environment. Thus, one could potentially leverage the capabilities of active, bistatic sonar methods within a passive setting.

To realize the application potentials of this example requires extensions to existing underwater acoustic computational modeling tools, and developing a suitable data-driven framework for signal processing and inference. Our development explicitly takes into account that in typical applications of interest, it is often unrealistic to have labeled data available for algorithm training. Even when the environment has been mapped ahead of time, as is the case for the bathymetric profile of Scripps Canyon as illustrated in Fig. 1-1, its inherent complexity poses major challenges for localization. For example, while environment estimation applications often approximate reflective surfaces as piecewise planar, the multipath setting that would be encountered is very



(a) (b)

Figure 1-1: Scripps Canyon: a shallow-water environment which exemplifies the complex seafloor profile that can be encountered in application settings (a); and a slice of its seafloor (bathymetric) profile for better illustration of the setting's scale (b) (images obtained from [1]). Note that while the zones labeled as A and B in (a) can roughly be approximated as planar surfaces, the rest of the environment could only be considered piecewise and locally planar depending on the localization application geometry.

different depending on where an emitter and receiver are situated. If we knew the emitter and receiver positions, we could roughly predict the multipath signal profile that we would encounter. However, when the emitter position is unknown in such a complex environment, simultaneously localizing it and leveraging the environmental structure becomes challenging. Ultimately, the goal is to develop data-driven and physics-informed methods that scale to very large numbers of parameters, and are thus capable of handling the kinds of natural complex structures that are frequently encountered in ocean acoustic applications.

The complexities of the ocean environment often lead to the need to rigorously train personnel who will fine-tune localization and estimation systems in real-time, or will interpret the estimation results accurately. Automation is important to reduce the operational burden and to streamline ocean operations. While the complexity of ocean-specific phenomena can be leveraged for improved performance, the difficulties posed by modeling such environments without trained human operators encourages the development of domain-agnostic algorithms. Such algorithms could then be adapted to work in different environments without expert knowledge.

Within the broad scope of underwater acoustic localization and environment es-

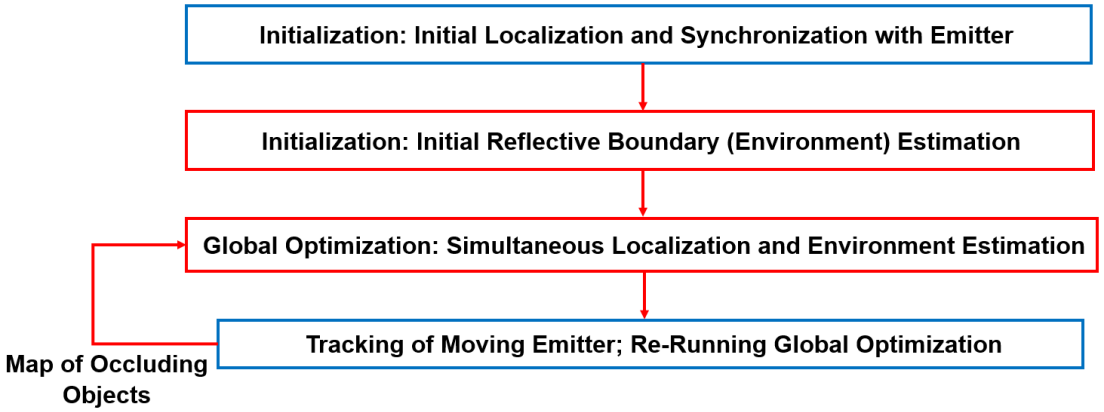


Figure 1-2: A conceptual summary of a potential simultaneous localization and environment estimation method. The sections with red outline the key steps where novel developments can be required to handle challenging environments. More advanced tracking algorithms could also be investigated in future work.

timation, we approach a more narrowly-defined problem that is amenable to the leveraging of the complex reverberant setting for improved performance. Presented with justifications further in the text are the following key features that are leveraged by our proposed methods:

- A known speed of sound which is constant throughout the environment, with the implicit assumption of short ranges and shallow seafloor depths for underwater acoustic localization.
- A pulsed, periodic, mobile man-made emitter that broadcasts a known high-frequency signal.
- Isotropic receiver hydrophones with accurately-known locations.

With the above features, we can conceptually approach the problem of simultaneous localization and environment estimation with a layered architecture such as illustrated in Fig. 1-2. This type of architecture proposes to solve a series of preliminary problems so that the overall ill-posed, non-linear and non-convex optimization task can be rigorously approached with global optimization methods. In this thesis, we will detail a specific architecture that we have designed to solve the localization and environment estimation task in simulation and in real-life settings.

Although we design and implement our methods for an underwater acoustic setting with such restrictive and simplifying features, we ultimately observe that there are novel capabilities and large performance gains that can be achieved over well-established and state-of-the-art methods even in this relatively basic setting. Hence, the algorithms that we modify or introduce could serve as building blocks of more complex systems that tackle increasingly difficult ocean environments and applications.

1.1 Challenges Posed by Underwater Acoustic Environments

The underwater acoustic setting is one of the most difficult and complex environments for passive localization applications, and indeed for other fields such as wireless acoustic communications. To understand how such a localization task can be accomplished with a realistically limited computational and experimental resources, we summarize here some key considerations for underwater acoustic environments [6]. We ultimately choose to design our methods to tackle a less challenging subset of underwater environments where straight-ray acoustic propagation models can hold. In our work, we ultimately observe that even this more amenable environment requires novel data-driven optimization methods for successful localization and environment learning.

1.1.1 Highly-Variable Environmental Parameters

Spatial and temporal (including seasonal) variability is very significant in underwater acoustics, and the basic channel parameters of a given setting depend on a wide range of such variable environmental conditions. For example, the attenuation constant of acoustic plane waves in the ocean can be modeled as a function of salinity, signal frequency, water temperature, and pressure. A key difficulty is that underwater acoustic channels are wideband in nature, due to the fact that the signal bandwidth is not negligible with respect to the center frequency. This is the cause of many practical is-

sues, such as when the Doppler shift can exceed the subcarrier frequencies of OFDM communication signals. Furthermore, phenomena such as a rough sea surface can cause significant frequency modulation. At the outset of our work, we therefore made the choice of basing our proposed methods on time delay estimation, rather than additional frequency-domain phenomena, due to the complexity of accurately modeling and replicating some of these features. For our localization and environment estimation methods, such issues as Doppler shifts are additional unmodeled sources of error in time delay estimation, rather than posing fundamental operational difficulties.

A key issue in deep ocean environments is that, as illustrated in Fig. 1-3, there are large variations in the speed of sound that can lead to the refraction of the emitted signals. This variability is due to factors such as the change in temperature, pressure, and salinity of the water column. Such variability of the speed of sound over large swathes of the ocean can lead to ducting phenomena, rather than the straight-ray propagation of acoustic waves. The realistic propagation models that incorporate such phenomena include normal mode theory, which can also model elastic seafloor layers, especially for low signal frequencies where such features have a more pronounced effect [7]. However, it is difficult to formulate a fully 3D propagation mode analysis model for general environments that can have additional reflective boundaries.

Since subtle changes in the soundspeed profile can cause large changes in the amplitudes of the received arrivals, it is unlikely that these amplitudes can be estimated accurately, especially in near-surface waters. The phase of the received signal is also difficult to predict due to scattering in shallow water. However, in comparison to these challenges, the TOAs and DOAs are more accurately-measurable characteristics of the received signal. Hence, a robust localization method can leverage such information in its core algorithms, and potentially leverage other features of the received signal to subsequently improve performance.

In practice, horizontal variations of the speed of sound are much smaller than vertical (i.e., along depth) variations, so over short ranges it is valid to approximate the horizontal speed of sound to be a constant. A variable soundspeed profile can dramatically affect the observed emitter signature, especially when monitored over

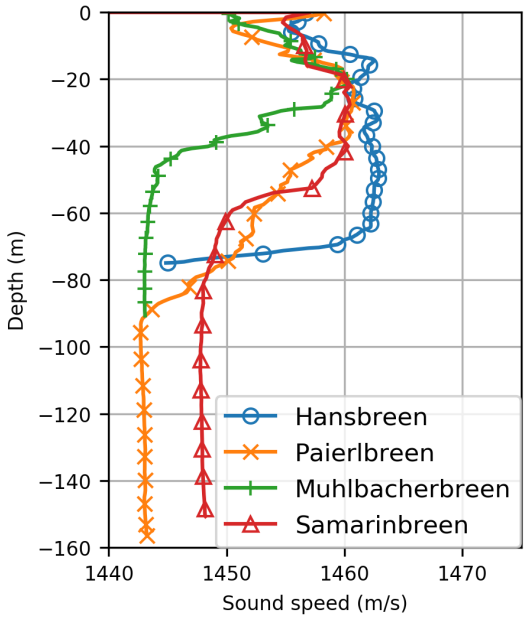


Figure 1-3: Representative depth-varying soundspeed profiles for underwater acoustic environments; in this case, different glacial bays in the same area (image reproduced with permission from [2]). Not only is the speed of sound highly variable with depth, but different bodies of water in the same geographical region can have large variations as well.

time and over geographic scale. Yet the spatially and temporally dense sounding of this underwater soundspeed profile over a wide ocean region and over different depths is a major operational challenge. In practice, if the ocean is sampled over a grid, interpolation has to be performed to obtain the ocean parameters that lie between the grid points. While such methods are active areas of oceanographic research, we focus on short ranges that side-step some of these domain-specific issues.

1.1.2 Complex Natural Features and Noise

Complicated underwater terrain profiles pose a major challenge for passive underwater environment estimation methods. The seabed is an acoustic medium, and the interactions of acoustic signals with the seafloor has to be treated as a propagation problem in a heterogeneous medium in order to obtain effective attenuation coefficients for the reflected waves. The reflective properties of the boundaries, such as the sediment soundspeed, have a strong effect on determining the magnitudes of different

multipath reflections; but their effects on modifying the TOAs are less pronounced. If we aim to leverage multipath for localization and environment learning, for example, muddy seafloors present a greater challenge than sandy seafloors, as they cause the reflections from the seafloor to be much weaker. In order to take such complexities into account over a wide range of operating frequencies and distances, one can solve a set of partial differential equations that describe acoustic propagation, and obtain the range- and depth-dependent Green's function [8]. When working with a narrowband signal, the effect of reflections will only be a constant factor in the received ray; but if the bandwidth is high, then the distortions due to the different absorption constants at different frequencies can distort the emitted signal.

In the ocean, the noise picked up by the receivers will be non-stationary. This is an additional complication in methods where parameters such as the noise power are required to be known or estimated, as these features can change very rapidly. For example, different multipath phenomena have different delay spreads which can evolve rapidly over time. Furthermore, there can be entanglements between different multipath arrivals, since they need not be uncorrelated due to various factors.

There are both natural and man-made sources of noise in the underwater environment. Interference from marine mammals, i.e., whale and dolphin songs can be significant in specific areas of the ocean. Schools of fish can also scatter or attenuate sound, and critical ocean regions can feature high biological interference. Near the shore, breaking waves can create a very noisy acoustic environment with high noise power. In fact, this wave-breaking is the dominant source of naturally-occurring ambient sound. Phenomena such as rainfall and strong winds can also contribute to producing a very different ambient soundscape. Rough seas lead to severe scattering phenomena and transmission losses. At lower frequencies, man-made sources of noise such as distant shipping can be dominant. In fact, passively listening to the soundscape of the open ocean, it is often difficult to correctly identify the source of the various ambient noises. The forecasting and accurate modeling of such features is very difficult.

This brief overview of the challenges of the ocean environment highlights the

need for expert knowledge in general underwater acoustic settings, and why trained operators are essential for many current localization and sensing systems. We now study some of the current methods used to deal with these issues, and outline our own approach to this problem.

1.2 Underwater Acoustic Localization and Environment Learning Methods

While we have limited our scope to the passive underwater acoustic localization and tracking of a mobile emitter in dynamic shallow-water settings [9], and the joint learning of its reverberant 3D acoustic environment, this task is still one of the most important and challenging localization problems [10], [11]. The emitter to be localized could be a surface vessel or underwater vehicle, and the receivers could be surreptitiously placed for passively monitoring a body of water. This is a different application setting compared to active localization modalities such as sonar, where multipath in the reverberant environment also poses difficulties that are met by a variety of methods [12]. The use of such active methods, however, could alert other receivers in the environment to the localization attempt, which encourages the deployment of passive localization methods in defense applications. The receivers in such applications are typically omnidirectional tethered hydrophones, but they could also be placed on a set of mobile underwater vehicles or suspended from them as an array, as illustrated in Fig. 1-4.

A representative scenario of our localization task is presented in Fig. 1-5, where the goal is to estimate the unknown source position at each emission point along a trajectory (red circles), in the presence of a large occluder, in a scaled model of the shallow-water setting. This is an instance of a difficult problem that also arises for indoor acoustic or radio-frequency localization [13], where the non-LOS (NLOS) arrivals due to multipath reflections must be judiciously exploited rather than mitigated, typically by using a suite of convex optimization techniques [14]. Similarly,

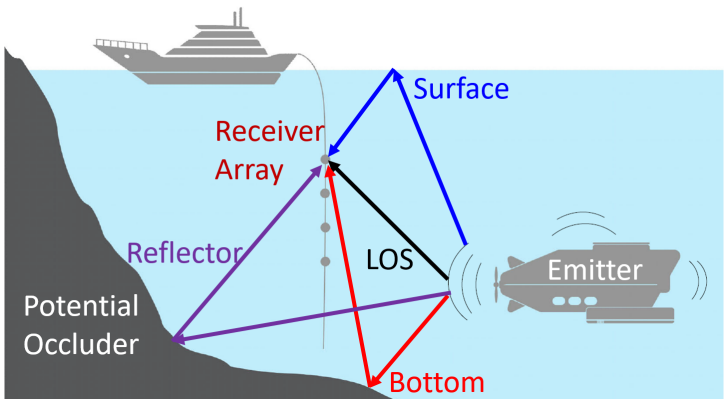
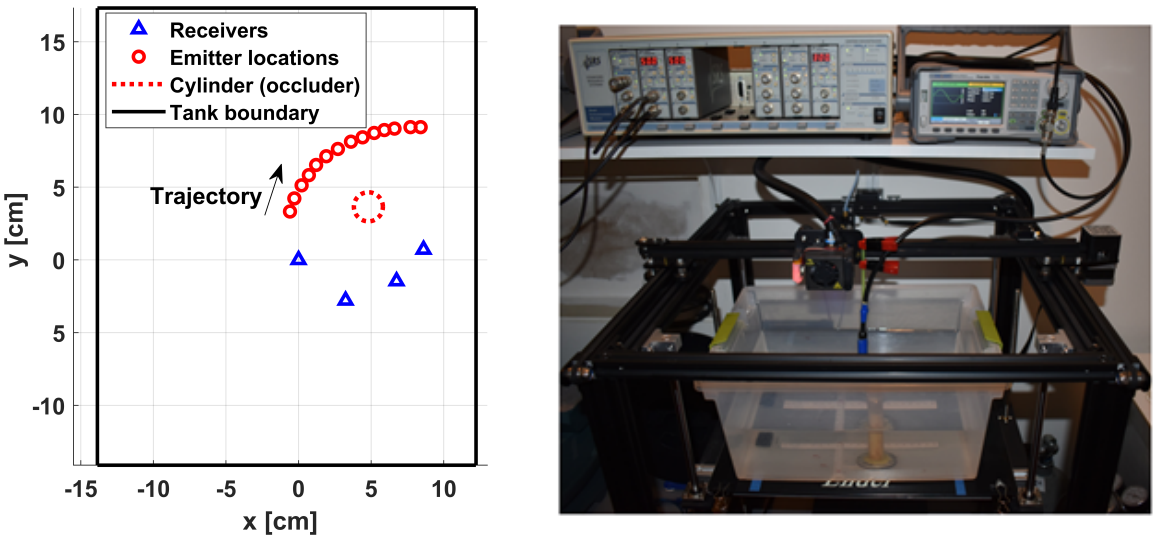


Figure 1-4: Typical shallow-water underwater acoustic localization problem setting; the image is modified from the original in [3]; licensed under a Creative Commons Attribution 4.0 International license.

the emitter's mobility must be leveraged in order to obtain an unambiguous solution of the unknown reverberant environment. Therefore, there is need for a robust passive localization method that can incorporate a wide range of side information from features of the task that would typically be viewed as challenges.

Without some restrictions on the passive localization setting, such problems are ill-posed and admit infinitely many solutions. Some of the important aspects of the localization setting are modeled as in Table 1.1; the limited information here leaves some key parameters undetermined, so that it is natural to adopt a multistage approach to first estimate them before using these estimates to solve the complete problem. Accordingly, as one method pursued to tackle this problem, we can deploy a series of localization and environment estimation steps sequentially. We call this architecture Passive End-to-End Localization (PEEL) [15]. Each stage of PEEL produces approximate solutions to different aspects of the overall localization and environment estimation problem, which then become inputs for the following stages.

PEEL is described as an end-to-end method because it starts from received signals that do not have a common clock with the emitter, and ultimately delivers a joint estimate of the full emitter trajectory and the environment. The static nature of the environment, and the spatial continuity in the trajectory of a moving emitter, are key to unambiguous localization and environment estimation. We demonstrate that PEEL can achieve strong performance, even when there is modeling mismatch that



(a) (b)

Figure 1-5: Top-down sketch of the test environment with emitter's trajectory (a), and the photograph of this experimental setup (b). This is a 3D localization and tracking problem with an occluding cylinder, and is more challenging than settings where the receivers surround the emitter [4].

Table 1.1: Model of the localization and environment learning setting.

Problem Feature	**Modeling Assumptions**
Speed of sound, v_s	Known, and constant within the environment
Environment	Static; only the emitter position is changing
Reflectors, $\boldsymbol{\eta}_j$	Planar, known number, unknown position and orientation
Transmissions	Periodic transmissions of deterministic pulses, with known period
Receivers, $\mathbf{p}_{r,i}$	At least 4 receivers deployed, with known positions

violates the assumptions in Table 1.1.

While there are many techniques for simultaneous localization and mapping in active settings, to the best of our knowledge there is currently no such passive localization method for omnidirectional receivers. Critically, we leverage motion and multipath, which are often seen as posing difficulties for localization, to improve performance and to disambiguate the reverberant environment.

One of the most challenging tasks handled by PEEL is finding an initial estimate of the reflective boundaries in the environment. This estimate is based on the known

positions of the receivers, the estimated position of the emitter, and the estimated times of arrival (TOAs) of the NLOS reflections. Note that we are not given which NLOS arrival is produced by which boundary, which is an issue known as the echo labeling problem. If this combinatorial-complexity problem were to be solved correctly, then existing algorithms such as least-squares (LS) or Euclidean distance matrices (EDM) localization may be employed to estimate the boundaries [16], [17]. However, these algorithms assume a high signal-to-noise ratio (SNR) to work well.

We propose a supervised learning method that performs image regression for 2D boundary estimation, incorporating convolutional neural networks (CNNs) that are trained on synthetic data, which we term COTANS-NN [18]. This CNN method outperforms the LS and EDM algorithms in our experiments, especially when operating outside of the high-SNR regime assumed by these alternative algorithms. A high-SNR regime can be unattainable in practical applications [19], so this robustness of COTANS-NN to noisy data is a critical advantage.

In PEEL, we deploy each of the algorithms discussed above in sequence, as depicted in Fig. 1-6. While we could conceptually solve the complete localization and environment estimation problem with a single overall optimization algorithm, in practice, it is computationally and algorithmically more feasible to tackle individual tasks as optimization sub-problems with well-justified initialization and re-iteration stages as outlined here.

1.2.1 Prior Work in Localization

When the ocean environment is accurately known, state-of-the-art matched-field processing (MFP) methods can be used for localization [20], [21], [22]. Such techniques can be augmented by incorporating additional prior probabilities on the environment [23], or by using adaptive algorithms [24]. The emission signatures of a variety of sources in the underwater environment can also be leveraged as domain knowledge that enhances MFP [25]. However, the standard MFP framework does not account for the presence of man-made features such as pier pilings or anchored vessels as illustrated in Fig. 1-7, or natural features such as unsurveyed reefs [26]. Depending

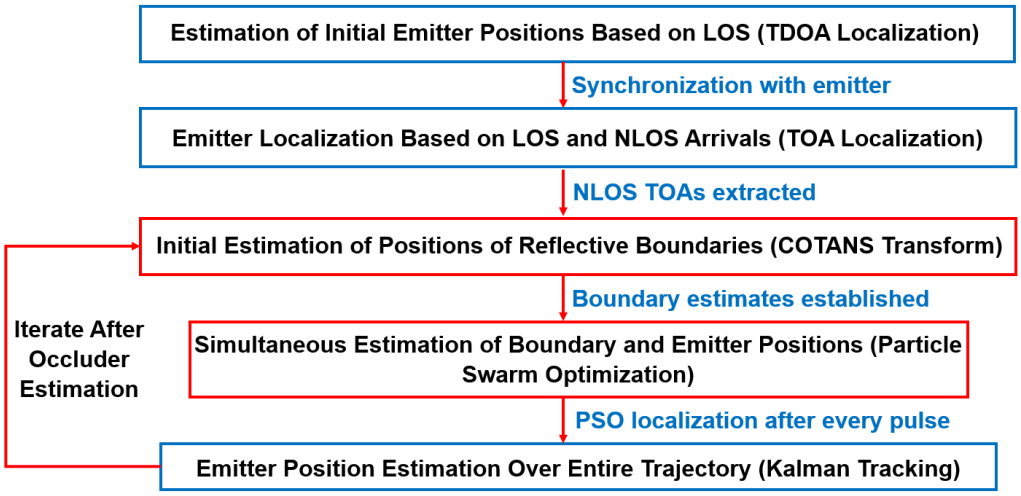


Figure 1-6: The sequence of algorithms in the PEEL method. The sections with red outlines incorporate novel developments or applications of those algorithms, which will be detailed at length in this thesis.

on their positions, such large and potentially unknown objects can act as additional reflective boundaries, or as occluders that block the line-of-sight (LOS) arrivals to receivers. MFP is sensitive to model mismatch [27], [28], so that additional algorithms may need to be implemented to ensure robust performance [29]. These mismatches may be caused by factors such as greater environmental variability [30], or source motion [31]. A mismatch due to the presence of unmodeled objects could lead to deteriorating MFP performance, which motivates our formulation of a localization framework that incorporates their presence.

While there is currently no end-to-end passive localization method in the literature such as our proposed PEEL architecture, similar frameworks have been designed for different settings. Joint acoustic tracking and environment learning problems can be solved using simultaneous localization and mapping (SLAM) methods [32], [33] in active settings, or when the emitter and receivers form a network [34], [35], [36]. In fact, there is a wide variety of probabilistic localization and tracking methods which may be employed with sensor networks [37], [38], [39], [40], [41]. When there is a large number of receivers in a regular array, simultaneous localization and environment estimation is possible without any prior estimates of the boundaries or the emitter location [42]. However, such methods may be unreliable in an environment with a few

(a) (b)

Figure 1-7: Scripps Pier: an example of an underwater acoustic localization setting with challenging natural and man-made features (a). Note the series of breaking waves along the beachline, which is a major source of acoustic noise. When a receiver array is suspended off the side of the pier (b), the pier pilings pictured here and the nearby civilian vessels create additional difficulties for localization and tracking applications.

irregularly-spaced isotropic receivers and potential occlusions of arrivals. While there are time-difference-of-arrival (TDOA) [43], [44] and time-of-arrival (TOA) [45], [46], [47] methods for localization, and the leveraging of reflective boundaries for improved localization performance [48], these methods do not account for unknown occluders as PEEL does.

In underwater acoustic applications, it is typically desirable to use as few receivers as possible for localization. To this end, single-receiver localization in underwater acoustic settings (a difficult challenge) has been the topic of ongoing research. This problem has been solved in non-underwater settings with known reflective boundaries [4]. Bearing estimates can be used in 2D underwater scenarios [49], and frequency-domain knowledge can be exploited in general 3D settings [50] with only a single receiver. While we have successfully performed single-receiver estimation with the techniques that we develop here, our framework handles a broader set of problems with an unknown environment and with unsynchronized emitters and receivers. For this more difficult setting, more than one receiver is typically needed unless accurate prior environment knowledge is available, and we have generally deployed 4-5 receivers in our experiments.

Neural network-based methods have been used in recent years for successful emitter localization in a range of environments, including the underwater acoustic setting [51], [52], and reverberant indoor environments [53]. These methods often involve generating simulated environments using tools such as Bellhop [54], which are then used to train NNs for localization. The successful application of NNs to the underwater setting builds on the success of other machine learning algorithms for localization tasks [55]. There are also studies that have been performed for NN-based localization in smaller model watertank settings, rather than the ocean [56]. An important application setting where NNs offer promising capabilities is single-hydrophone localization, where the LOS arrival and the NLOS surface reflection can be leveraged to allow for localization despite using the minimum number of receivers [57]. Direction of arrival estimation with NNs rather than conventional beamforming methods have also been demonstrated for the ocean environment [58].

One of the key challenges in NN-based localization is to devise techniques that are as agnostic to a specific setting as possible, to achieve broad applicability. A data-driven methodology is adopted in order to avoid the rigorous, explicit modeling of complex dynamic environments that can result in model mismatch. Collecting a large underwater acoustic dataset is time-consuming and expensive, so it is vital that realistic simulation results are employed for data augmentation in generating training datasets for NNs.

1.2.2 Prior Work in Environment Estimation

We have briefly mentioned that LS or EDM localization may be used to estimate the reflective boundary positions in the environment. An alternative boundary estimation methodology is developed in [34], where reflective boundaries are observed to be common tangents to ellipses (in 2D) or spheroids (in 3D) that are geometrically defined by the NLOS path distances. By fitting tangent planes to these ellipses, we can potentially estimate the boundaries while *avoiding* the error-prone echo labeling procedure. This methodology relies on a mathematical transformation that generates an image of overlapping curves from the NLOS TOAs. For convenience, we refer to the

domain to which the NLOS arrivals are mapped as the common tangents to spheroids (COTANS) domain, and refer to the transformation that maps to this domain as the COTANS transform. We envision assuming an estimated emitter position as a known ground truth, and using the NLOS arrivals to accurately estimate the boundaries, after which further joint localization and environment learning can be carried out.

We adopted the COTANS methodology early in our work [34], but observed that the methods in the literature had implementation difficulties that could be improved on with a novel neural network (NN) approach. As originally proposed, COTANS applies a user-defined heuristic smoothing filter to COTANS images and picks the resulting maxima from overlapping smoothed curves. However, this method can be fragile when applied to real data with estimation errors.

Fitting tangent planes to spheroids for boundary estimation, as in our COTANS-NN method, is becoming increasingly widespread for indoor settings. This methodology was proposed in [59], [60] as the common tangent (COTA) algorithm. A similar approach [34] was used for joint localization and boundary estimation, using a method inspired by the Hough transform rather than leveraging an analytical cost function as in [59] (that assumed only small-scale errors). In [61], a Hough transform-inspired methodology was used to estimate an indoor environment and to perform echo labeling, with provisions to reject incorrectly-chosen second-order echoes. Though these techniques refer to the Hough domain, they do not fit planes to point clouds as the Hough transform does [62]. Instead, they sample points from a spheroid's surface, then deduce the corresponding tangent plane at each point. To avoid conflating this tangent plane-fitting method with the Hough transform, we refer to a COTANS transform and COTANS domain instead.

Plane-fitting boundary estimation methods typically apply a smoothing filter to COTANS images, followed by the extraction of maxima [34]. This local averaging approach aims to estimate the boundaries when errors are present, which can prevent the COTANS curves corresponding to the different NLOS arrivals from intersecting at the correct boundary parameters. However, this hand-crafted filtering operation is itself sub-optimal, and parameters such as filter sizes and kernels are manually

tuned to specific settings. A multi-scale filtering approach that can handle a range of different reverberant settings is needed. This difficulty is our main motivation for pursuing a NN-based method that can automate and combine these filtering and peak extraction tasks. If a NN is trained with a wide range of setting geometries and realistic estimation errors from different levels of noise, it can potentially learn the optimal inference rule, which can be viewed as joint (and implicit) filtering and peak extraction. The resulting network can then be efficiently re-trained for different environments. Thus, COTANS-NN is fundamentally different from other tangent-fitting methods, instead of being a straightforward NN extension of past work.

Hand-crafted filters make the implementation of alternative methods more difficult, since boundary estimation tasks have different physical scales or domains (e.g., indoor or underwater acoustic). It can also be challenging to make a fair performance comparison between these different methods, since implementation details such as filter sizes and kernels can be missing. In contrast, COTANS-NN can be re-trained in an automated manner without having to modify any implementation-specific hyperparameters. Furthermore, it is not straightforward to incorporate domain knowledge or prior information into a hand-crafted filter method, especially when considering that COTANS transform results are synthetic images where image processing heuristics are not necessarily applicable. Since COTANS-NN is a data-driven neural network method, it is able to learn and leverage such domain knowledge.

Oceanographic data is typically sparse in time and space, so that there are difficulties in using a sufficiently large dataset for a data-driven approach to localization and environment learning. Hence, it is typical to augment real-life datasets with simulation results, using tools such as Bellhop [54] or KRAKEN [63] for underwater acoustic settings [64], [65]. By leveraging such simulators, NNs can potentially also solve a range of associated estimation problems in environment learning. We have carried out similar studies for NN-based localization in the ocean [66]; however, the incorporation of tracking or occlusion to such a framework has not yet been attempted. The PEEL architecture that we propose incorporates NNs to solve the particular sub-problem of obtaining an initial estimate for the boundary locations, but it otherwise relies on

global optimization and tracking rather than being a NN method.

1.3 Implementation and Applicability Considerations

This work will address a holistic, architecture-based approach to the localization and environment learning tasks. Rather than considering various localization subproblems in isolation, we aim to perform a joint estimation so that our results are reflective of a practical system that is as widely-applicable as possible.

We deploy global optimization algorithms and NN-based methods that were prohibitively computationally expensive in the past. Even with the most advanced resources for personal use, such as 1 TB of RAM and a 64-core processor in a single computer, the prototyping stages of our methods were slow and featured day-long trials. It is therefore not surprising that some of the solutions which we have developed, do not seem to have been previously considered in the literature.

A running theme in our research has been solving localization and environment estimation problems with minimal model requirements. Because localization is a geometry-dependent task, it is possible to introduce unrealistic constraints that seemingly improve accuracy tremendously. Thus, we need to justify our choices of parameters at every step. This type of justification is provided by implementing a complete chain of estimation algorithms, which starts from a real raw received signal and estimates the necessary parameters. We also limit our target settings to situations where our models can be expected to hold; for example, we assume a short-range shallow underwater acoustic setting, so that we can expect an almost constant speed of sound through the water.

We present the signal and environment models that we use in our research in Chapter 2, including the simulations and preliminary experiments that were used to confirm the modeling assumptions. The design decisions, algorithms, and results for the PEEL and COTANS-NN methods are given in Chapters 3 and 4, respectively. We conclude with future directions in Chapter 5.

Chapter 2

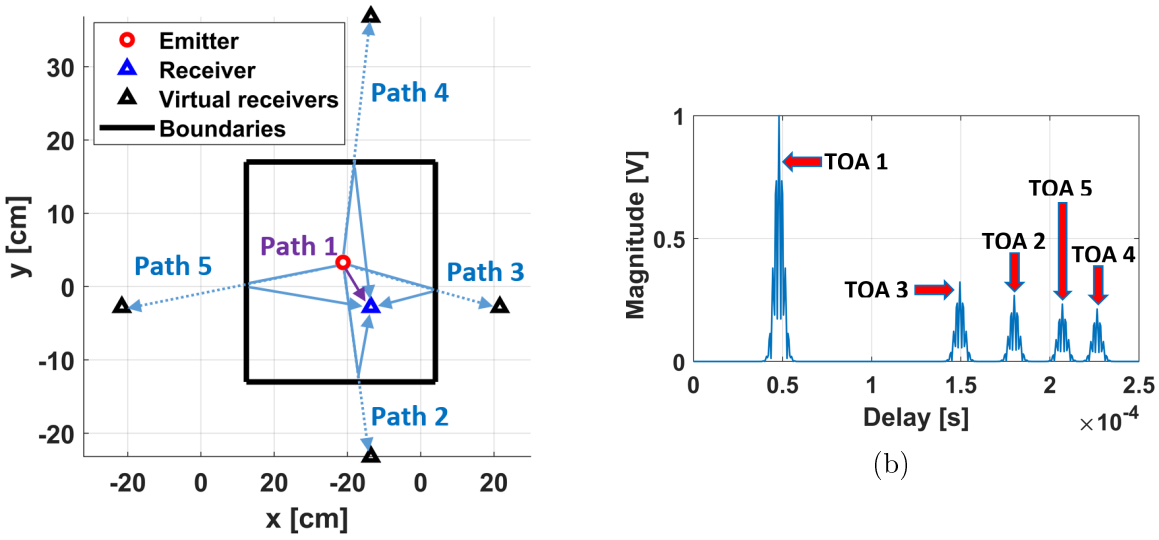
Signal and Environment Model

In this chapter, we introduce the notation and signal model, and present key geometrical considerations that are necessary to frame the localization and environment estimation problems. We specifically focus on how multipath arrivals can be modeled and used to improve localization performance. We present our signal model in Section 2.1, the planar 3D reflection model that underlines its development in Section 2.2.1, and the considerations for whether occluders could act as scattering objects in Section 2.2.2. Our real-life experiments were carried out with a scale watertank testbed, as detailed in Section 2.2.3; and in Section 2.2.4, we present the results of the simulation and calibration experiments that were performed in order to verify that our models were valid for our underwater acoustic application setting.

Note that in our subsequent mathematical notation, lowercase bold variables such as $\mathbf{p}$ denote vectors, uppercase bold variables such as $\mathbf{P}$ denote matrices, and uppercase calligraphic characters such as $\mathcal{P}$ denote sets.

2.1 Signal Model and Localization Principle: Exploiting Virtual Emitters and Receivers

The localization and boundary estimation methods which we develop generally rely on triangulation using time-domain statistics. Triangulation is a standard localization



(a) (b)

Figure 2-1: Illustration of virtual receivers produced by reflective planar boundaries in an occluder-free environment (a), and the corresponding noise-free simulated received signal for this scenario (b). A NLOS TOF to the real receiver is equivalent to the LOS TOF to a matching virtual receiver.

principle that models an emitter as the point of intersection of circles in 2D, or spheres in 3D, of equal time-of-flight (TOF) from the emitter to each receiver. Since the resulting TOF arcs do not perfectly intersect in practice (e.g., due to noise), an estimation method has to be used to find an approximate point of intersection; often, this method has to simultaneously solve an echo labeling problem [67]. When the locations of reflective boundaries in the environment are known, the multipath arrivals that they give rise to can be leveraged for better localization accuracy. These boundaries can be interpreted as producing 'virtual receivers' as illustrated in Fig. 2-1, which in this case corresponds to the structure of the scale watertank testbed setting which we will detail in this chapter.

We make the simplification that reflective structures are (piecewise) approximated as N static planar boundaries [68], where N is assumed to be known, and which are characterized by the range $\rho \in \mathbb{R}_+$, azimuth $\theta \in [0, 2\pi)$, and elevation $\phi \in [0, \pi)$ of their normal vector relative to the (arbitrarily-chosen) origin. Thus, the j-th boundary is parametrized as the vector $\boldsymbol{\eta}_j = [\rho_j, \theta_j, \phi_j]^{\mathrm{T}}$, for all boundaries $j \in \mathcal{S}_N$, where we denote $\mathcal{S}_K \triangleq \{1, \ldots, K\}$ for some $K \in \mathbb{N}$. We denote the Cartesian

coordinates of the emitter and the M isotropic receiver positions as $\mathbf{p}_{\mathrm{e}} \in \mathbb{R}^3$ and $\mathbf{p}_{\mathrm{r},i} \in \mathbb{R}^3$, respectively, with $i \in \mathcal{S}_M$.

The received signal at the i-th receiver, $r_i(t) \in \mathbb{R}$, is modeled as the sum of the LOS arrival and the single-reflection NLOS arrivals, delayed by their respective TOAs. We model the second-order reflections as heavily attenuated in the underwater acoustic setting [66], as compared to the first-order reflections from boundaries such as the sea surface and the seafloor. The LOS TOA $\tau_{i,0}$, is given by:

$$\tau_{i,0} = \frac{\|\mathbf{p}_{\mathrm{r},i} - \mathbf{p}_{\mathrm{e}}\|_2}{v_{\mathrm{s}}}, \quad \forall i \in \mathcal{S}_M, \tag{2.1}$$

where v_{s} is the speed of sound, approximated as a known constant. This approximation of v_{s} can be a valid assumption for well-mixed shallow-water settings at short ranges and high acoustic frequencies [69], [67].

Reflections can be interpreted as producing 'virtual emitters', and for the j-th boundary, we obtain the virtual emitter location $\mathbf{p}_{i,j}$ by finding the corresponding reflection of $\mathbf{p}_{\mathrm{e}}$. The NLOS TOA at the i-th receiver from the j-th boundary, $\tau_{i,j}$, is equal to the TOA from the i-th receiver ($\mathbf{p}_{\mathrm{r},i}$) to the corresponding j-th virtual emitter ($\mathbf{p}_j$):

$$\tau_{i,j} = \frac{\|\mathbf{p}_{\mathrm{r},i} - \mathbf{p}_{i,j}\|_2}{v_{\mathrm{s}}} \triangleq \frac{d_{i,j}}{v_{\mathrm{s}}}, \quad \forall i \in \mathcal{S}_M, \quad \forall j \in \mathcal{S}_N. \tag{2.2}$$

We denote a single copy of the periodically-emitted signal as $s(t)$, and we merge the effects of attenuation and reflection as the equivalent attenuation coefficient $\alpha_{i,j}$ for each path. The received signal $r_i(t)$ at the i-th receiver is thus modeled by:

$$r_i(t) = \sum_{j=0}^{N} \alpha_{i,j} s\left(t - \tau_{i,j}\right) + \xi_i(t), \tag{2.3}$$

where $j = 0$ corresponds to the LOS path, and $\xi_i(t)$ is an unknown noise signal that is a realization of a spectrally-flat Gaussian process. Note that $\alpha_{i,j}$ can be zero, as when an occluder blocks a particular arrival, so Eq. (2.3) incorporates the potential for occlusion. In practice, the environment can be reverberant, and can present issues such as higher-order reflections and noise that may not be Gaussian [69], [70].

Furthermore, the number of boundaries N that is assumed known may actually be inaccurate. A proposed localization method must therefore be robust to the resulting discrepancies that can be encountered in real-life settings.

It is important to note that, while localization approaches that rely on virtual emitters/receivers as we have discussed are useful and generalizable, there are other models and corresponding algorithms that can be used in more specific cases. For example, consider the case where we are given that in addition to an unoccluded LOS arrival, there is only a single NLOS arrival from the sea surface. This type of scenario can arise in the deep ocean, where reflections from the seafloor may be too attenuated or arrive much later than the surface reflection, yielding a less reverberant environment for localization. For such settings, cepstrum analysis is a useful tool [71], with promising recent research being performed to implement it on shallow-water settings as well [72]. We note this approach to highlight the potential for using different methodologies to leveraging domain knowledge in underwater acoustic localization.

As for the emitted signal in this localization setting, we model $s(t)$ as pulsed and periodic, which limits our estimation scenarios to such man-made sources. This model allows us to record $s(t)$ and to infer some key parameters: the approximate (truncated) channel impulse response length T_c [73], and the time period T_p between emissions. We require $s(t)$ to meet the following constraints in order to have accurate estimates $\{\hat{\tau}_{i,j}\}$ of the $\{\tau_{i,j}\}$. First, we define T_s as the temporal duration of $s(t)$, so that $s(t)$ is strictly zero outside the time interval $0 \leq t \leq T_s$. Next, we define T_A to be the full width at half maximum (FWHM) of the autocorrelation function of $s(t)$, which is the time resolution of the matched-filtered emitted signal. Finally, we define T_m as the minimum time difference between the main multipath arrivals. The operation modes for our localization methods are then as follows:

- If $T_s < T_m$, we have a pulse that is shorter than the multipath separation, as is the case with the Gaussian pulse in our experiments. Here, we do not need to know $s(t)$ for localization; T_s is short enough that the multipath arrivals are roughly separable.

- If $T_s > T_m$ and $T_A < T_m$, then after matched-filtering, the multipath peaks are roughly separable. Now, we do in fact need to know $s(t)$ for this matched-filtering step, so that we are able to identify the distinct arrivals and thereby solve for the environment.

We also need $T_p > T_c$ to avoid different periodic emissions being observed within the received signal's time interval. In practice, if $T_p > T_c$, then T_p can be estimated, but this is an environment-specific problem, as is an estimation of T_m.

Deployments in the ocean environment can introduce additional sources of error that are outside the scope of our methods, but which can be addressed by various algorithms in the literature. For example, we model the receivers as being perfectly synchronized with the same clock. While ensuring this condition was not an issue in our experiments, in ocean deployments with autonomous receivers there will be clock drift over time. There is a range of methods to address this difficulty using self-synchronization algorithms [74]. Likewise, while we have assumed the use of isotropic receivers in order to ensure the maximal applicability of our methods and the greatest coverage of potential NLOS arrivals, ocean applications often use directional receivers. Similar passive estimation methods that leverage ocean ambient noise can be employed to ensure a more reliable ground truth for localization purposes for such directional receivers [75].

With these introductory developments, we can rigorously develop our methods for reliable localization and environment estimation. Before such methods can be deployed, however, it is important to answer a number of questions that determine their applicability for a given setting. With a given signal of a certain bandwidth, and a given emitter and receiver placement, is it possible to resolve the individual multipath arrivals? Will the propagating acoustic waves penetrate significantly into seafloor and other boundaries, thereby being refracted and adding to the propagation distance, or will they be reflected mirror-like back into the water with negligible distortion? Will objects in the environment act as occluders, or scatterers? The answers to such questions determine whether it is possible to effectively leverage the NLOS arrivals for localization and boundary estimation. Hence, it is important to

set up an accurate simulation environment to obtain an approximate typical channel response, and to verify its validity with a series of real-life experiments. Such a simulator and test environment is also very useful for calibration purposes, and to verify the correct operation of the deployed equipment.

2.2 Verification of the Signal Model in Underwater Acoustic Propagation Environments

In order to develop acoustic systems capable of accommodating and exploiting complex ocean environments, it is important to have software tools that are capable of accurately simulating such environments, and to have test environments to confirm their operation. Indeed, packages such as Bellhop for simulating ocean acoustic propagation phenomena have played an important role in facilitating the development of existing systems; and the extensions of such packages to complex environments will play a corresponding role in future systems development. Accordingly, our initial investigations focused on the development of a simulation toolkit with which to test the feasibility of our methods, and the construction of a small testbed to replicate these simulation environments. These preliminary investigations ultimately verified the validity of the various simplifications in our signal and environment model.

The goal of our simulations is to be able to model reflective boundaries and occluding and scattering objects, and to produce the resulting predicted impulse responses for the given underwater acoustic channel. As discussed in Chapter 1.1, we assume an isovelocity environment in our simulations and algorithms, which is necessary for straightforward ray optics assumptions to hold. This requires the use of high frequency signals, so that the straight-ray prropagation approximation is valid. A modal solution that incorporates the derivation of the appropriate Green's function would be necessary at lower frequencies (<1000 Hz) [76], [77], but the more general environment geometries that we work with make the derivations of such functions difficult. We limit our work to relatively short ranges in the shallow water environment,

where we do not expect to observe complex waveguide propagation [78], [79].

Since we are operating at high frequencies, we adopt a ray propagation convention [80]. This convention allows easy visualization of the acoustic arrivals in terms of geometric raypaths. Volume absorption of the sound waves in the water is modeled using the Francois-Garrison model [81]. The absorption of sound upon reflection at the boundaries is modeled by making the reflection coefficients complex [80].

2.2.1 Planar 3D Environments

The modeling of 2D propagation in underwater environments has been studied extensively, and many propagation models for channels bounded by the sea-surface and the sediment bottom are available. Modeling of 3D environments also finds mention in the literature [82], [83], [84], [85], but it is less-explored than the 2D case. Most of these works focus on solving the well-studied wedge problem. To approach full 3D modeling capability, an $N \times 2D$ version of Bellhop is available commercially [54]; however, it is designed for forward propagation and cannot easily incorporate backscattering from reflective surfaces. This is a very important extension for localization in our applications, however. Although it was not featured in our publications, one of our first simulations was a simple set of Bellhop trials with a sea surface and seafloor at ranges of up to several hundred meters. These simulation results were used to verify that straight-ray propagation assumptions were being met for our chosen operating frequencies and ranges to the emitter, and that phenomena such as waveguide propagation were not being observed. Hence, we could proceed with tackling the task of introducing reflective boundaries to our simulated environments.

The modeling of environments whose geometry incorporates more than two reflecting boundaries is quite complex, compared to the cases with two or fewer boundaries (e.g., the water column or wedge geometries). This is because in the former case, even in an isovelocity environment, the rays from the emitter are often not visible at the receiver due to the geometry, thus creating the need for visibility checks for each pair of emitter/receiver locations. Furthermore, the number of multipath arrivals of significant amplitude increases considerably with the number of reflecting boundaries,

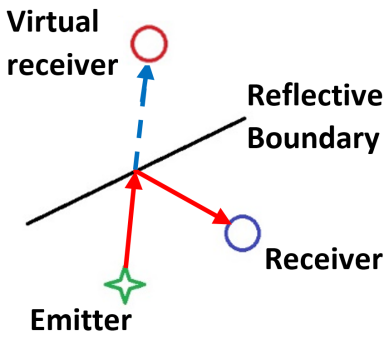


Figure 2-2: A reflected signal can be conceptualized as a direct path from the emitter to the mirror image of the receiver, or from the mirror image of the emitter to the receiver, with the same path length.

and is exponential in the number of reflections from boundaries.

In [68], a broad framework is provided for enumerating the multipath arrivals arising through different combinations of reflections at different boundaries. Here, a visibility check is used to determine which raypaths are visible at the receivers, and hence should be incorporated in modeling the recorded field. Once a simulated impulse response is obtained based on the visible raypaths and their corresponding path length-based attenuation coefficients, the emitted signal may be superimposed on these arrivals to generate a simulated received signal. While this framework is limited to the case where the reflecting boundaries are planar, it is a useful tool to understand even more complex environments. Thus, our modeling is undertaken based on [68]. We extend the visibility check approach to also account for reflections from finite boundaries in the environment, such as a finite corner placed inside a watertank. Visibility checks are more computationally expensive in these cases, but can still be carried out.

Throughout the work in this thesis, we have used vector geometry, with vectors expressed as arrays, to carry out operations such as reflections of points. Consider a point $\mathbf{p}_\mathrm{r} \in \mathbb{R}^3$ as being the location of a receiver, with Cartesian coordinates $[x_\mathrm{e}\ y_\mathrm{e}\ z_\mathrm{e}]^\mathrm{T}$. Recall that if we wish to find its mirror image, which produces a 'virtual receiver' as illustrated Fig. 2-2, we also need a vector geometric description of the plane of reflection.

It is convenient to define reflective boundaries as planes of infinite length, which

can be uniquely described by any point $\mathbf{p}_\mathrm{b}$ on the boundary (a convenient choice being a vertex or other such endpoint of the plane), and the unit normal vector $\mathbf{n}_\mathrm{b}$ that points into the halfspace which contains the emitter and receiver in question. With these two quantities, the plane is defined as the set of points $\mathbf{p}$ that satisfy the equation:

$$\mathbf{n}_\mathrm{b}^\mathrm{T}\left(\mathbf{p}-\mathbf{p}_\mathrm{b}\right)=0, \tag{2.4}$$

based on the fact that the normal vector is by definition orthogonal to any set of vectors $(\mathbf{p}-\mathbf{p}_\mathrm{b})$ that lie within the plane. Using this definition of reflective boundaries, the location of the virtual receiver $\mathbf{p}_\mathrm{v}$ produced by the particular boundary in Eq. (2.4) is given by:

$$\mathbf{p}_\mathrm{v}=\mathbf{p}_\mathrm{r}+2\left(\left(\mathbf{p}_\mathrm{b}-\mathbf{p}_\mathrm{r}\right)^\mathrm{T}\mathbf{n}_\mathrm{b}\right)\mathbf{n}_\mathrm{b}. \tag{2.5}$$

If a plane is of finite length, it is important to check whether it can give rise to a valid reflection, i.e., whether it can serve as a reflective boundary for a particular placement of an emitter and receiver. A geometric method (in 2D) for verifying the validity of a mirror image is to check whether the ray from the emitter to the virtual receiver (or from the virtual emitter to the receiver) intersects with the line segment that serves as the plane of reflection. In 3D, the same methodology may be used by first projecting the ray from the emitter to the virtual receiver onto the plane of reflection, thus producing an equivalent 2D scenario.

To find the point of intersection $\mathbf{p}_\mathrm{i}$ on the reflective boundary of the line segment from the emitter to the corresponding virtual receiver, we note that this point can be defined as:

$$\mathbf{p}_\mathrm{i}=\mathbf{p}_\mathrm{e}+\left(\mathbf{p}_\mathrm{v}-\mathbf{p}_\mathrm{e}\right)w, \tag{2.6}$$

where $w>0$ is a constant to be determined. Eq. (2.6) simply expresses the fact that the point of intersection lies somewhere in between the virtual receiver and the real emitter. Since this point of intersection must also lie on the reflective boundary, it must also satisfy the equation that defines the reflective plane (Eq. (2.4)), so that we get:

$$\mathbf{n}_\mathrm{b}^\mathrm{T}\left(\mathbf{p}_\mathrm{e}+\left(\mathbf{p}_\mathrm{v}-\mathbf{p}_\mathrm{e}\right)w\right)=\mathbf{n}_\mathrm{b}^\mathrm{T}\mathbf{p}_\mathrm{b}. \tag{2.7}$$

Therefore, solving this equation for w, we get:

$$w = \frac{\mathbf{n}_{\mathrm{b}}^{\mathrm{T}}\mathbf{p}_{\mathrm{b}} - \mathbf{n}_{\mathrm{b}}^{\mathrm{T}}\mathbf{p}_{\mathrm{e}}}{\mathbf{n}_{\mathrm{b}}^{\mathrm{T}}\left(\mathbf{p}_{\mathrm{v}} - \mathbf{p}_{\mathrm{e}}\right)}. \tag{2.8}$$

For a given localization and environment learning task, it is an important design choice to determine whether to use virtual emitters or receivers. Two key considerations in this regard are whether the emitter or the receiver positions are reliable and known ground truths, and whether either the emitter or receivers are mobile. For example, in our PEEL localization and environment estimation method, we have stationary known receivers and an unknown moving emitter to be localized, so we make use of virtual receivers. On the other hand, in our COTANS-NN boundary estimation method, we have emitter and receiver locations that are both (at least approximately) known, but we have only a single emitter and we have multiple receivers of potentially varying number, so we chose to use virtual emitters for algorithmic simplicity.

2.2.2 Scattering From 3D Objects

The PEEL localization and environment estimation method is designed to compensate for the presence of occluding objects in the environment. However, depending on their sizes, acoustic properties, and the emitted signal, such objects could potentially act as scattering objects rather than occluders. In this case, we would have an unmodeled source of error in our experiment, as our algorithms work with LOS and NLOS arrivals but do not currently account for strong scatterers. For example, in our PEEL experiments, we use a wooden cylinder as an occluder, as objects that are geometrically analogous to pier pilings. A key question that had to be answered was whether such wooden cylinders could act as operationally significant scatterers, for emitted Gaussian pulses with a high center frequency of 280 kHz.

Consider the case of vertical cylinders that extend from a flat bottom to a flat water surface, as will be the case for our watertank experiments. The cylinders scatter the incident field, generating a scattered field. In regions that are behind the cylinders (with reference to the source), a shadow is formed at high frequencies due

to destructive interference between the incident and the scattered field [86]. Cylinder scattering in such cases has been studied in the literature since the 1950s [87], for while it is rare to encounter an actual cylinder ocean, such an object serves as a reasonable and tractable mathematical model. It is simplest to model the acoustic scattering under the assumption that the incident waves have a planar wavefront, which is a far-field assumption. The scattering of a plane wave incident obliquely on an elastic cylindrical shell is discussed in [88]. Furthermore, the special cases of a rigid body cylinder and pressure-release cylinder are incorporated based on the expressions in [89], although these more complex models were ultimately not required in our watertank experiments.

More concretely, at a frequency f, the total acoustic field (including scattering effects) at a hydrophone at a radial distance of r_h and depth z_h from the point of incidence on the ellipse, and aligned at an azimuthal angle θ_h with respect to direction of propagation, is given by a summation over modes as [88]:

$$p_s(f, r_h, \theta_h, z_h) = p_i e^{ik\sin(\alpha)z_h} \sum_{n=0}^{\infty} A_n H_n(k\cos(\alpha) z_h r_h)\cos(n\theta_h), \tag{2.9}$$

where p_i is the complex incident wave amplitude, α is the elevation angle of incidence of the wave, H_n is the Hankel function, and $k = \frac{2\pi f}{v_\mathrm{s}}$ is the wavenumber, where v_s is the speed of sound. The expression for the modal coefficients A_n is given in detail in [88], and depends on the outer radius r_a and inner radius r_b of the cylindrical shell, and on its elastic properties, among other parameters. In practice, it is sufficient to restrict the summation to $M = kr_\mathrm{a} + 1$ modes to get fairly reliable results [90]. Using these basic building blocks of scattering theory, we can determine whether occluders will cause significant scattering.

2.2.3 Experimental Verification Platforms: Watertank Testbeds

Having developed the basic reflection and scattering theory that is necessary to evaluate the validity of our signal model, we now discuss the experimental platforms that allows us to verify that this body of theory is applicable to our localization setting.

Firstly, we discuss the high frequency acoustic testbed (abbreviated as HiFAAT for convenience) that we use for our PEEL experiments (Fig. 1-5(b)). This testbed provides high frequency (200-400 kHz) acoustic data from a reverberant environment, which is useful for carrying out experiments in a controlled setting and for testing algorithms before planning larger-scale experiments. Since a testbed such as HiFAAT uses reasonably accessible equipment, we provide its details here so that other users are able to assemble an equivalent test environment.

HiFAAT is designed around a modified Ender 5 Plus 3D (E5P) printer as more clearly seen in Fig. 2-3, which provides precise control of transducer placement and rapid mapping of tank volumes. The E5P has a relatively large print volume of 35 cm x 35 cm x 40 cm and precise, 0.1 mm carriage positioning. The tank volume (a plastic tub in our case), selected to fit on the Ender 5 plus print table, is roughly 25 cm x 32 cm x 15 cm. Acoustic signals are transmitted and received using ITC 1089D spherical ceramic crystals, which are useful in the 200-400 kHz frequency range. Signals for the ITC1089D transducer are designed using Matlab and uploaded to a networked Siglent SDG2042X arbitrary waveform generator. Received waveforms on the second ITC1089D are amplified and filtered with Stanford Research Systems SIM910 and SIM965 modules before acquisition with an Agilent Technologies DSO5034A oscilloscope. Signal transmission and acquisition is controlled with custom National Instrument LabVIEW software, which also positions the E5P carriage between transmission/acquisition cycles using standard G-code printer commands sent over a virtual serial interface.

The HiFAAT testbed is useful despite its small scale. The tank is roughly 70 wavelengths deep and 150 wavelengths wide in the horizontal direction. The most unrealistic element in this setup is excessive reverberation from the tank side walls, which will not be present in the ocean environment. Such reverberation can be time-gated out of the data, but imposes limits on source-receiver ranges that can be achieved without leading to unrealistically strong reverberation.

One of the important parameters to consider for our test environments is the frequency-dependent absorption loss coefficient which causes attenuation of the emit-

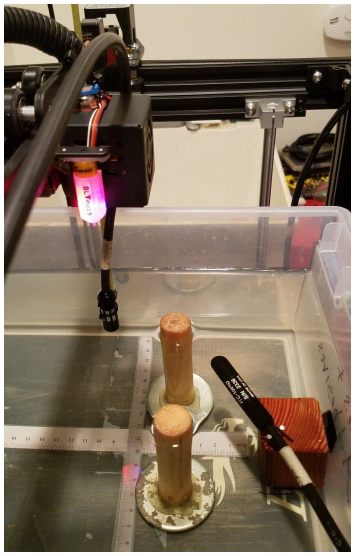
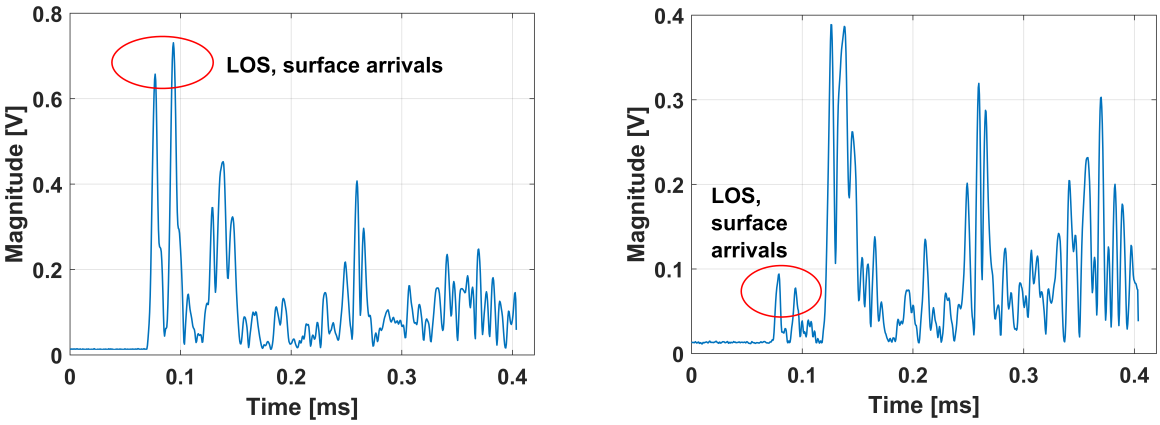


Figure 2-3: A picture of HiFAAT showing 2 ITC1089D transducers mounted between 2 vertical scattering cylinders. The center frequency of the pulse transmissions is 280 kHz with a wavelength of roughly 1.5 mm. The vertical source near the top tank wall is mounted to the carriage of an Ender 5 Plus 3D printer, which provides 3-axis computer-controlled movement for mapping out tank volumes.

ted signal, beyond the range-dependent free-space propagation attenuation. When comparing the modeled received signal strength to the experimental results, not accounting for this coefficient can lead to discrepancies. For our high-frequency pulses through freshwater watertank experiments, we calculated an absorption loss coefficient of 0.9986, and therefore neglected this term in our modeling. For more general emitted signals and over longer ranges, it is important to verify whether this term can be omitted in modeling, or must be incorporated.

The magnitude of a real underwater acoustic received signal from HiFAAT (obtained from an emitted Gaussian pulse) is given in Fig. 2-4(a); here, the most prominent peaks are the LOS and the first reflection from the water surface. Note that the LOS arrival has a smaller magnitude than the surface arrival for this particular geometric placement of the emitter and receiver, because these hydrophones are not perfectly isotropic. If we were given that there are 6 first-order reflections as is the case here, it would be difficult to determine the corresponding peaks. The situation is further complicated if an occluder blocks LOS as in Fig. 2-4(b). In this case, algorithms that count on a strong LOS arrival would fail. One of the key motivations



(a) (b)

Figure 2-4: Example of the magnitude of a received acoustic signal in the HiFAAT watertank testbed when no occluder is present (a), and when an occluder obstructs LOS (b).

of our proposed methods is to overcome such issues.

While the HiFAAT testbed allows us to conduct very precise localization experiments, we also worked with a much larger-scale testbed: the Scripps Ocean-Atmosphere Research Simulator (SOARS) wave tank facility in the Scripps Institution of Oceanography, which is depicted in Fig. 2-5. The length of this wave tank is 36 m, and its width and depth are 2.4 m. Hence, it can be used to approximate a shallow-water ocean environment better than HiFAAT can. The temperature of the water in the wave tank is modifiable as well in an automated manner. As the PEEL method for localization and tracking had already been implemented on data from HiFAAT by the time SOARS was accessible, the data that was collected at SOARS was instead primarily used to test the performance of our neural network-based boundary estimation method.

SOARS can be used to generate both winds and waves in a controlled manner, as depicted in Fig. 2-6. The relevant aspect of this feature for our particular localization scenarios is that, in future work, different sea surface conditions can be generated in a precise manner, to test our methods in a wide range of simulated environmental conditions.

The watertank experiments which we conduct take place in a controlled environ-

Figure 2-5: Overview of the SOARS watertank facility at the Scripps Institution of Oceanography.

ment with minimal ambient noise; when testing our algorithms, white Gaussian noise was synthetically added to our received signals to simulate a range of noisy conditions. In a real open-ocean environment, it is important to measure the ambient noise over time, and to test localization methods under such conditions instead of potential noise power levels that may be too low or too high than is realistic. There is a wide range of algorithmic and experimental methods in the literature for such ocean ambient noise measurement [91]. For future work, we have collected ambient noise and interference measurements from Scripps Pier, as illustrated in Fig. 2-7, which could be synthetically added to our experimental signals in this fashion.

2.2.4 Watertank Testbed Experiments for Model Verification

We first discuss the verification of the validity of our 3D straight-ray propagation and planar reflection model. Implementing the virtual emitter methodology of [68], we benchmarked our code against ray-modeling results available in the literature for the case of a 3D wedge [84]. Our simulated TOAs broadly agreed with the results in [84], agreeing on the number of visible virtual emitters and differing only slightly from the

Figure 2-6: The SOARS wave tank can be used to generate rough water surface conditions with user-controlled wind and wave parameters. A transparent window in the side of the tank allows us to monitor the experimental setup and the precise height of the waves.

literature results. These slight differences in TOAs can be attributed to the fact that [84] accounts for the more complex beam displacement effect that occurs when rays are totally internally refracted at surface discontinuities [80]. However, for the sake of simplicity, we have not included this effect, as it creates only minute differences in TOAs for our experiments. Since the results from our model match those in the literature within a reasonable degree of accuracy, we can proceed with further 3D modeling in complex environments with multiple planar boundaries.

As a proof-of-concept test of the validity of this model, we carried out an experiment in the HiFAAT watertank that is outlined in Fig. 2-8. Here, a corner-shaped reflector made of metallic sheets is placed in the interior, and impulse responses are recorded for different orientations of this wedge. The tank walls are modeled as plastic (polyethylene/polypropylene mixture), with a density and compressive speed of sound of 900 kg/m^3 and 2000 m/s, respectively. The fact that the tank walls are thin surfaces that border on air (even for the bottom, which is elevated) makes the penetration of sound waves into the material negligible. If this were not the case, we would have had to employ a more advanced model [83], and would have had to account for phenomena such as beam displacement [92].

Figure 2-7: Preparing for the deployment of an emitter and a vertical receiver array at Scripps Pier. Note the weights being tied to the ends of the lines on which the emitter and receivers have been attached, intended to prevent excessive swaying in the wavy and windy ocean environment so close to the surf zone.

We compare the results from the watertank experiment to results obtained by our simulator in Fig. 2-9. The simulation is carried out for rotation angles between 11° and 80°, as our model predicts that there is no LOS ray observed for angles lower than 11°. We are able to capture some major features of the received signal using our simulator, and good estimates of the TOAs. These include the first 2 stable arrivals, the arc-like pattern in the data (corresponding to reflections from the corner pieces, which are sensitive to the angle of the corner), and even some subsequent arrivals. However, there seems to be some mismatch in the relative amplitudes predicted by our model. This may have to do with the assumed acoustic parameters of the materials not matching their real properties. Our simulator also shows the reverberation dying down rapidly after 3.5×10^{-4} seconds, whereas this is not observed in the data. The later arrivals are stronger than expected, indicating that attenuation due to reflections is less than what the model indicates. Note that, if the angle of the corner is roughly known, there is enough information in the NLOS arrivals at a single sensor for

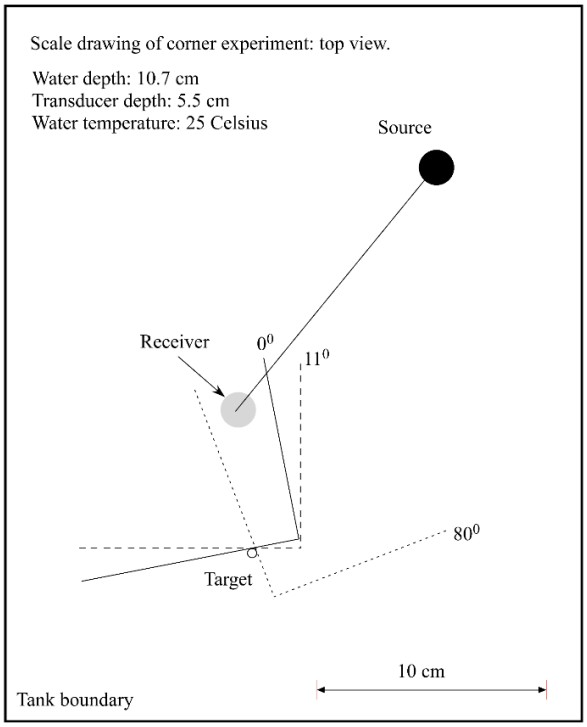
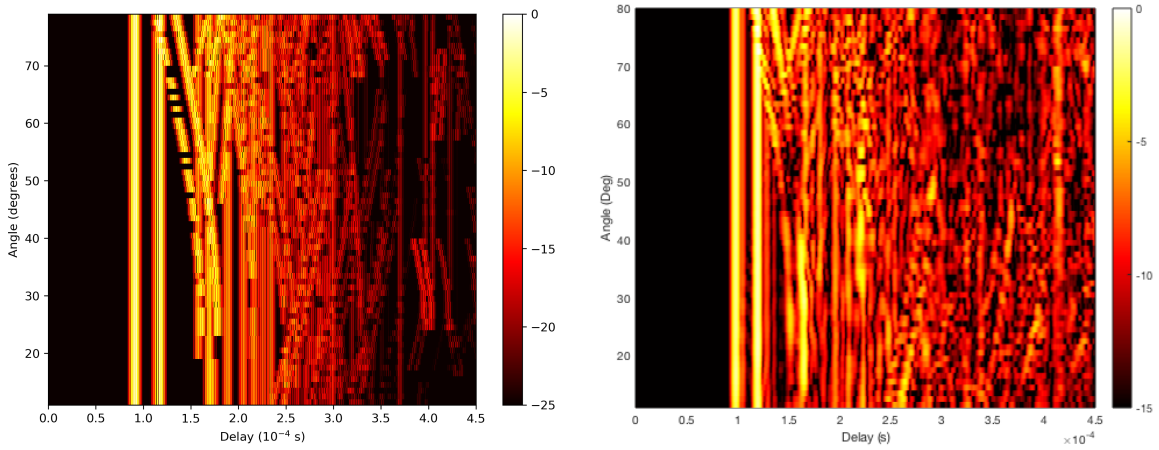


Figure 2-8: Experimental setup in the HiFAAT watertank to test whether reflections from boundaries are being accurately modeled.

localization. Likewise, from multiple known emitter or receiver locations, unknown corner boundaries could be mapped.

Next, we carry out a set of scattering experiments, where theoretical and practical results demonstrate that this is not ultimately an issue for the HiFAAT testbed. For the solid wooden cylinder used in our watertank experiments, we have $r_{\mathrm{a}} = 0.01$ m (a diameter of 2 cm), and $r_{\mathrm{b}} = 0$. The density, compressive acoustic speed and shear acoustic speed of the wooden cylinder are assumed to be 600 kg/m^3, 4150 m/s and 1050 m/s based on [93, 94]. In our particular experiments, the emitted signal was a Gaussian pulse with a nominal center frequency of 280 kHz. Various real-life calibration runs and simulations based on implementing Eq. (2.9) indicated that scattering was negligible at such high frequencies. The occlusion of the cylinder and the multipath from the reverberant environment were found to dominate the results, so that we did not pursue more detailed scattering modeling. However, at lower frequencies, it is strongly encouraged to confirm that scattering effects can



(a) (b)

Figure 2-9: Normalized and model-predicted scattered field (in dB) for the watertank experiment illustrated in Fig. 2-8 (a), compared against the real data recorded from HiFAAT (b).

be neglected for a given set of occluder dimensions and material properties, with modeling based on the reference literature provided here.

The closeness of the simulated and experimental results, and the ability to discriminate the different NLOS TOAs, served as the critical proof-of-concept for our subsequent localization and environment-learning experiments. The mismatches between these results would be even more pronounced in larger underwater acoustic environments, especially in the dynamic ocean setting [95]. This encourages the development of methods which are able to function robustly despite such discrepancies, and which rely more on extracted TOA data than the amplitudes of the arrivals in the received signals. Similarly, the discrepancies that were observed were more pronounced for the predicted phases of the received arrivals than their magnitudes. This feature of the experimental data encouraged us to pursue incoherent rather than coherent detection methods. These were general design directions that influenced the implementation of our methods on underwater acoustic settings; hence, our approach was to propose general modular frameworks to solve localization and environmental estimation problems. For other application settings where modeling can be more accurate and features like phase information can be reliably extracted, the proposed frameworks could be augmented with additional modules for enhanced performance.

Chapter 3

Joint Localization and Structure Learning

In this chapter, we outline the PEEL method for joint localization and structure learning. We present the theoretical and algorithmic development of this method in Section 3.1, and detail the corresponding results that are achieved in the scale watertank environment of Fig. 1-5 in Section 3.3. The main contribution which we present here is an architecture for joint localization and environment learning. Among other tools, PEEL makes use of:

- A global optimization method for joint boundary estimation and localization, which does *not* need to be tuned to a specific problem.
- The mapping of unknown occluders after tracking, followed by iterations with a refined optimization metric, which has not previously been implemented for passive localization.
- Closed-form derivations for some key steps of a boundary estimation technique [34], that improve computational efficiency and accuracy.

With the signal and environment model presented in Chapter 2, the localization problem where the emitter and reflector locations are both unknown can theoretically be disambiguated if there are multiple non-collinear emitter and receiver locations

in 2D, and non-coplanar locations in 3D. In practice, however, obtaining a robust fine-resolution solution for these emitter and receiver positions can be challenging. The solution critically depends on the setting geometry, and interference between multipath arrivals or their attenuation can lead to a poor localization result. The impact of reflective boundary positions on overall localization performance is hard to express in explicit functional form, and optimization algorithms are typically used to obtain approximations to the true solution. Based on these caveats for the design of practical localization methods, the sequence of algorithms that we deploy in PEEL leads to a tightly constrained search space for the emitter position, and thereby eases the computational load. This allows us to obtain accurate solutions for the localization task.

3.1 PEEL Method for End-to-End Localization and Structure Learning

In the PEEL method, we first use TDOA and TOA localization for initialization; then obtain an initial estimate of the reflective boundary locations; and finally track a moving emitter as we simultaneously learn the environment, as summarized in Fig. 3-1. Each step in Fig. 3-1 is presented in sequential order.

3.1.1 Stage 1: Coarse Localization of the Emitter

We first describe a standard TDOA localization algorithm (which does not require knowledge of $\{\boldsymbol{\eta}_j\}$) for synchronization with the emitter [96], and then a TOA algorithm that leverages multipath from known boundaries $\{\boldsymbol{\eta}_j\}$ after synchronization [97], [98]. In general terms, the TDOA and TOA algorithms that are outlined here produce ambiguity surfaces, which are plots of the likelihood of the emitter's presence at a certain location. Conceptually, TDOA localizes the emitter as the intersection point of hyperbolae defined by equal time differences, as opposed to TOA which examines the intersection of circles/spheres [99]. Both algorithms use a grid-search

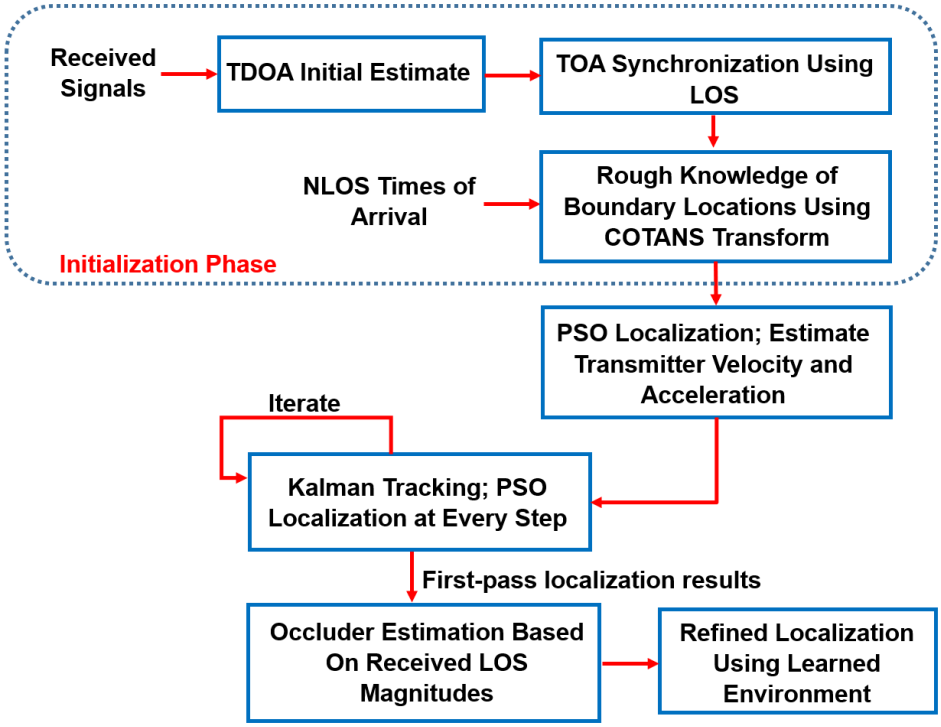


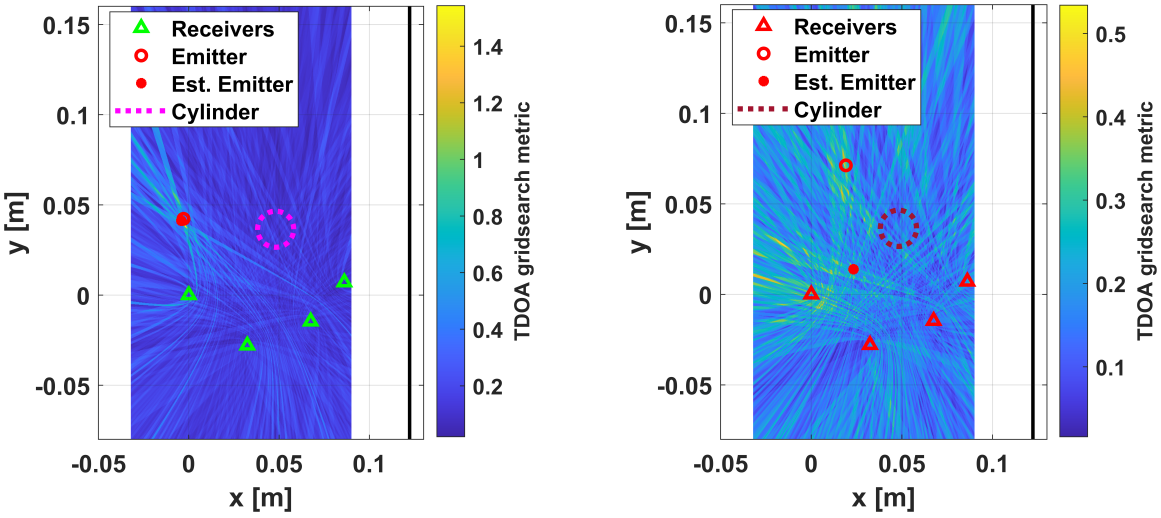
Figure 3-1: Conceptual flowchart for the proposed PEEL method.

for the emitter to obtain $\hat{\mathbf{p}}_{\mathrm{e}}$, where a set $\mathcal{P}$ of grid points $\mathbf{p}_{\mathrm{cand}}$ is defined over the target area. For example, in our HiFAAT watertank setting, the regularly-spaced grid-squares were chosen to have lengths of 1 mm. Each $\mathbf{p}_{\mathrm{cand}} \in \mathcal{P}$ corresponds to a potential emitter location, with an associated set of $\{\tau_{i,j,\,\mathrm{cand}}\}$ to the receivers $\mathbf{p}_{\mathrm{r},i}$ and the corresponding virtual receivers $\mathbf{p}_{i,j}$. Thus, each $\tau_{i,j,\,\mathrm{cand}}$ is a function of $\mathbf{p}_{\mathrm{r},i}$ and of $\boldsymbol{\eta}_j$. The results of the TDOA and TOA stages are used to initialize more refined techniques in the later stages of PEEL.

TDOA does not require a common clock with the emitter, and is suitable to establish synchronization [100], [101]. Denoting the calculated cross-correlation between the signals at receivers i_1 and i_2 as $R_{i_1,i_2}(\tau)$, the TDOA metric that we use is:

$$C_{\mathrm{TDOA}}\left(\tau_{\mathrm{cand}}\left(\mathbf{p}_{\mathrm{cand}}\right)\right) \triangleq \sum_{i_2 \neq i_1} \sum_{i_1=1}^{M} \left| R_{i_1,i_2}\left(\tau_{i_1,\mathrm{cand}}\left(\mathbf{p}_{\mathrm{r},i_1}\right) - \tau_{i_2,\mathrm{cand}}\left(\mathbf{p}_{\mathrm{r},i_2}\right)\right)\right|. \tag{3.1}$$

Using the generalized cross-correlation with phase transform (GCC-PHAT) method for better performance [102], which is a weighting of the cross-correlations used



(a) (b)

Figure 3-2: A TDOA grid-search that localizes the emitter on real data (a), and a grid-search that fails at a different location due to occlusion (b).

in TDOA to produce a sharper correlation peak, we estimate the emitter location $\hat{\mathbf{p}}_{\mathrm{e,TDOA}}$ as:

$$\hat{\mathbf{p}}_{\mathrm{e,TDOA}} = \underset{\mathbf{p}_{\mathrm{cand}} \in \mathcal{P}}{\operatorname{argmax}} \, C_{\mathrm{TDOA}} \left(\tau_{\mathrm{cand}} \left(\mathbf{p}_{\mathrm{cand}} \right) \right). \tag{3.2}$$

The average difference between the LOS peak delays and the expected TOFs from TDOA is then used to establish synchronization; though we note that more sophisticated methods for extracting the LOS peak or discriminating it from the NLOS peaks could also be used [103], [104]. If there are no occluders in the environment, TDOA could be sufficient to solve the main localization task, as demonstrated in Fig. 3-2(a). However, since this method relies on the LOS being unoccluded, the occluder can defeat this algorithm as in Fig. 3-2(b). Hence, while TDOA is important for time synchronization in our framework, the subsequent algorithms are necessary to ensure robust performance.

It is important to note a simple yet important caveat for localization illustrations such as Fig. 3-2 that is easy to overlook: the equalization of the scales of plot axes. Localization methods yield a variety of intersecting geometric objects such as circles, spheres, ellipses, spheroids, and hyperbolae, depending on the specific algorithm and whether it is implemented in 2D or 3D. Unless the plot axes have equal scales, such

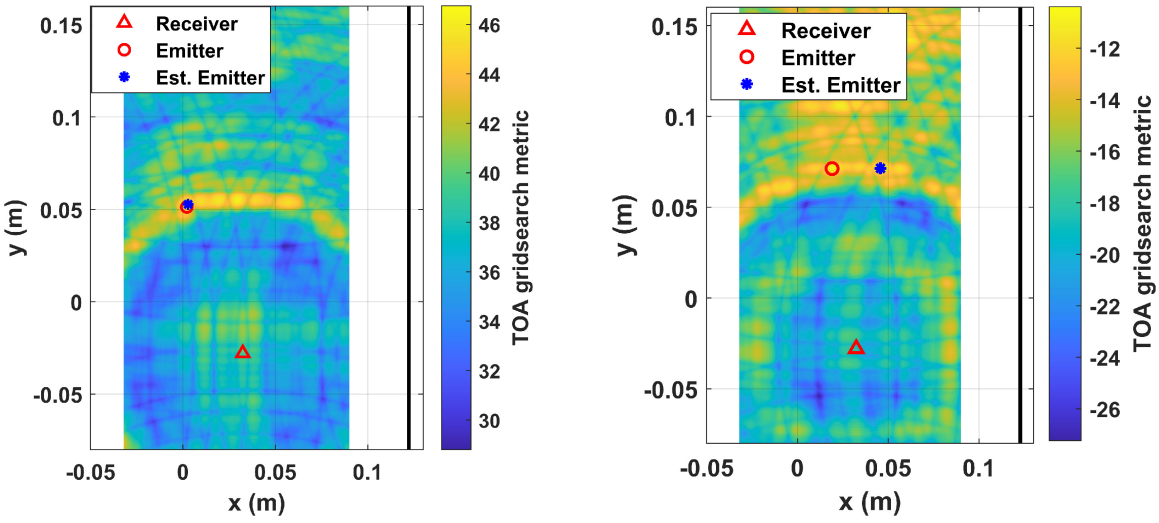


Figure 3-3: A TOA grid-search that localizes the emitter on real data (a), and a grid-search that fails at a different location due to the alternative arrival arc crossings (b), at ∼36 dB SNR.

shapes will be distorted, making the results difficult to interpret and making errors more difficult to debug.

Once synchronization has been established using TDOA localization, we can use the more accurate TOA localization methodology. We define the additive TOA metric $C_{\text{TOA}}(\tau_{\text{cand}}(\mathbf{p}_{\text{cand}}))$ as:

$$C_{\text{TOA}}\left(\tau_{\text{cand}}\left(\mathbf{p}_{\text{cand}}\right)\right) \triangleq \sum_{i=1}^{M}\sum_{j=0}^{N}\left|r_i\left(t-\tau_{i,j,\text{cand}}\left(\mathbf{p}_{\text{r},i},\boldsymbol{\eta}_j\right)\right)\right|, \tag{3.3}$$

where the time indices $\left(t-\tau_{i,j,\text{cand}}\left(\mathbf{p}_{\text{r},i},\boldsymbol{\eta}_j\right)\right)$ are rounded to the nearest sample in a discrete-time implementation. The estimate $\hat{\mathbf{p}}_{\text{e,TOA}}$ is then obtained as:

$$\hat{\mathbf{p}}_{\text{e,TOA}} = \underset{\mathbf{p}_{\text{cand}}\in\mathcal{P}}{\text{argmax}}\, C_{\text{TOA}}\left(\tau_{\text{cand}}\left(\mathbf{p}_{\text{cand}}\right)\right). \tag{3.4}$$

The heuristic metric in Eq. (3.3) is maximized when each $\tau_{i,j}$ corresponds to a peak location; such locations are the intersections of TOF arcs, as seen in Fig. 3-3(a). When the $\{\boldsymbol{\eta}_j\}$ are known, and therefore the virtual receivers are labeled correctly, $\hat{\mathbf{p}}_{\text{e}}$ is obtained accurately.

TDOA and TOA localization are more robust if the search space $\mathcal{P}$ can be limited to the vicinity of the correct location, $\mathbf{p}_{\mathrm{e}}$. Consider the results in Fig. 3-3(a); although $\mathbf{p}_{\mathrm{e}}$ is estimated accurately, there are many local maxima in the search space. These can lead to large errors, as in Fig. 3-3(b). Confining $\mathcal{P}$ to the vicinity of $\mathbf{p}_{\mathrm{e}}$ avoids such ambiguities. With these theoretical and practical considerations in mind, we summarize our TDOA and TOA localization algorithms in Algorithms 1 and 2.

Algorithm 1: TDOA localization for initialization.

Require: Unoccluded LOS arrivals at each receiver.
Input: $\mathcal{P}$, $\{r_i(t)\}_{i=1}^{M}$, $\{\mathbf{p}_{\mathrm{r},i}\}_{i=1}^{M}$
for $\mathbf{p}_{\mathbf{cand}} \in \mathcal{P}$ **do**
  Calculate C_{TDOA} (Eq. (3.1))
Get $\hat{\mathbf{p}}_{\mathrm{e,TDOA}}$(Eq. (3.2))
Result: $\hat{\mathbf{p}}_{\mathrm{e,TDOA}}$

Algorithm 2: TOA localization after initialization.

Require: Time synchronization after applying Algorithm 1; number of boundaries N is known.
Input: $\mathcal{P}$, $\{r_i(t)\}_{i=1}^{M}$, $\{\mathbf{p}_{\mathrm{r},i}\}_{i=1}^{M}$, $\{\boldsymbol{\eta}_j\}_{j=1}^{N}$
for $\mathbf{p}_{cand} \in \mathcal{P}$ **do**
  Calculate C_{TOA} (Eq. (3.3))
Get $\hat{\mathbf{p}}_{\mathrm{e,TOA}}$ (Eq. (3.4))
Result: $\hat{\mathbf{p}}_{\mathrm{e,TOA}}$

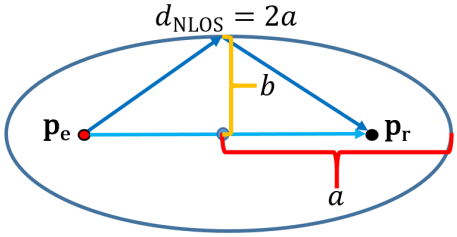
One advantage of Algorithm 2 is that it is a heuristic weighting of the arrivals based on their signal strengths. A localization method that is solely based on $\{\hat{\tau}_{i,j}\}$ will assign equal importance to each arrival, so that there will be no automatic compensation for faulty TOF estimates. A TOF estimate is likelier to have high error if the corresponding arrival is weak. In such cases, its contribution to the additive metric C_{TOA} in Algorithm 2 will be smaller, and its impact on $\hat{\mathbf{p}}_{\mathrm{e,TOA}}$ in Eq. (3.4) will be reduced. Note that Algorithm 2 does not require leveraging all of the boundaries to yield an accurate result. The additional NLOS arrivals provide spatial diversity that helps to increase localization accuracy and mitigate occlusion that causes loss of LOS (pun intended).

Algorithm 2 also allows us to compare different hypothesized environments. Given candidate boundaries $\{\boldsymbol{\eta}_j\}_{\text{cand}}$, we can run Algorithm 2 for each $\{\boldsymbol{\eta}_j\}_{\text{cand}}$, and pick the maximum metric solution as our joint estimate of $\hat{\mathbf{p}}_{\text{e,TOA}}$ and $\{\boldsymbol{\eta}_j\}$. C_{TOA} is therefore a heuristic measure of the estimate's goodness. Recall that the $\{\boldsymbol{\eta}_j\}$ cannot generally be assumed known; thus, we next propose a boundary estimation algorithm to produce a suitable set of $\{\boldsymbol{\eta}_j\}_{\text{cand}}$.

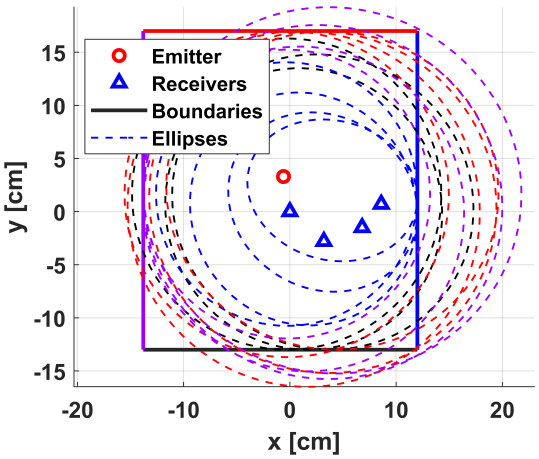
3.1.2 Stage 2: Preliminary Boundary Estimation

After TDOA and TOA localization, we have an initial estimate of the emitter's position and have approximately established synchronization. To jointly estimate the emitter and the environment in a computationally feasible manner, we need a good initial estimate of reflective boundary positions, which ultimately yields a tight search space for $\{\boldsymbol{\eta}_j\}$. As will be discussed in detail in Chapter 4, we first estimate a set of unlabeled $\{\hat{\tau}_{i,j}\}$, though some of these estimated arrival times might have large errors. We now develop a Hough transform-inspired method to estimate $\{\boldsymbol{\eta}_j\}$ which is robust to errors in $\{\hat{\tau}_{i,j}\}$, building on the algorithms presented in [34]. Our novel contribution of the neural network-based COTANS-NN method to solve this problem is detailed in Chapter 4; here, we present the techniques that were already established in the literature, which we applied to data that was obtained from the HiFAAT testbed detailed in Chapter 2. Since the COTANS-NN method is limited to 2D settings, we implemented the existing COTANS transform method for 3D boundary estimation.

As some background context, Euclidean distance matrices (EDM) [13] and convex optimization methods [14] have been used to solve for the virtual emitter positions and thereby obtain boundary positions. These methods can suffer from large errors if an arrival's $\hat{\tau}_{i,j}$ has been labeled incorrectly, and their performances are initialization-sensitive. An alternative approach relies on the fact that in 2D, an NLOS arrival (with an estimated time delay $\hat{\tau}_{i,j}$) corresponds to a path distance of d_{NLOS} that defines an ellipse whose foci are the emitter and receiver location, denoted as $\mathbf{p}_{\text{e}}$ and $\mathbf{p}_{\text{r}}$, respectively, as in Fig. 3-4(a). In 3D, the NLOS arrival equivalently defines a spheroid. By definition, points on the ellipse/spheroid have a total distance of



(a)



(b)

Figure 3-4: The NLOS arrivals define ellipses/spheroids of equidistance, with an emitter and receiver as their foci (a). In (b), we illustrate the ellipses defined by $\{\tau_{i,j}\}$ for our HiFAAT experimental setting (depicted in 2D for convenience). Each boundary is a common tangent to a single ellipse due to each receiver, highlighted here by matching colors. In such a rich multipath setting, solving the echo labeling problem is difficult.

d_{NLOS} to the emitter and receiver, and the reflective boundary itself is a tangent line/plane of this ellipse/spheroid. With multiple receivers, multiple spheroids are defined by such NLOS arrivals, and the reflective boundary is their common tangent. Therefore, by fitting common tangents to ellipses or spheroids, the boundaries can be estimated while avoiding the echo labeling problem in multipath environments [34], as illustrated in Fig. 3-4(b). Assigning ellipses to tangents is a problem of combinatorial complexity, and is error-prone for inaccurate time-delay estimates [105]. Missing or spurious arrivals in the received signals complicate the echo labeling, which motivates the development of an estimation algorithm that bypasses this task altogether.

It is convenient to define a given reflective boundary $\boldsymbol{\eta}_j$ by the spherical coordinates of a vector from the origin (arbitrarily selected) that is orthogonal to the boundary plane. Thus, each boundary is specified by a range ρ, azimuth θ, and elevation ϕ, where these local spherical coordinates correspond to a standard east-north-up Cartesian coordinate system.

The transform method for boundary localization that was outlined in [34], which for convenience we refer to as COTANS as previously explained in Section 1.2.2, is

derived as follows. A boundary defined by ρ, θ, and ϕ can be conceptualized as a point (ρ, θ, ϕ) in a COTANS transform domain; working out the (ρ, θ, ϕ) expression of a plane is to take the plane's COTANS transform [106]. When the space $\rho \times \theta \times \phi$ is discretized as a 3-mode tensor, each hypothetical tangent plane has a finite set of (ρ, θ, ϕ) on which it may lie. Incrementing this 'accumulator' tensor over every potential (ρ, θ, ϕ) entry yields a COTANS-domain image (as in Fig. 3-6(a)), with maxima at the true boundaries in the noiseless case.

Here, we extend the method of [34] to 3D, and we further derive a direct solution for the tangents in the COTANS domain, which *precludes* the need for randomly sampling points (ρ, θ, ϕ) on the surface of an ellipse/spheroid. Our approach is to first define a grid over the θ- and ϕ-coordinates, and then obtain the COTANS transform for the tangents at each θ and ϕ. In 2D, for a given θ, we obtain the COTANS transform's ρ in Appendix A.1 as:

$$\rho(\theta) = \sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}, \tag{3.5}$$

and in 3D, the result for a given θ and ϕ is calculated in Appendix A.2 as:

$$\rho(\theta, \phi) = \sqrt{c^2 \cos^2\phi + a^2 \sin^2\phi}, \tag{3.6}$$

where a, b and c are standard axes of spheroids calculated from the $\{\hat{\tau}_{i,j}\}$. This result is obtained for an origin-centered spheroid. To move a point cloud of (ρ, θ, ϕ) centered on the origin to the true emitter and receiver positions, we rotate the points to match the true spheroid's orientation (adjusting their θ and ϕ to be some θ_{rot} and ϕ_{rot}), and then translate the points to the true spheroid's position (yielding a final ρ_{COTANS}, θ_{COTANS}, and ϕ_{COTANS}). This sequence of COTANS-domain operations is illustrated in Fig. 3-5.

In practice, we first generate a set of points $\{(\rho, \theta, \phi)\}$ by defining an azimuth-elevation grid, and obtaining the COTANS transforms for each of the corresponding tangents. The resulting points $\{(\rho, \theta, \phi)\}_{\text{COTANS}}$ are then rounded to a desired accuracy. For these specific PEEL experiments, we use a resolution of 0.1 mm for ρ and

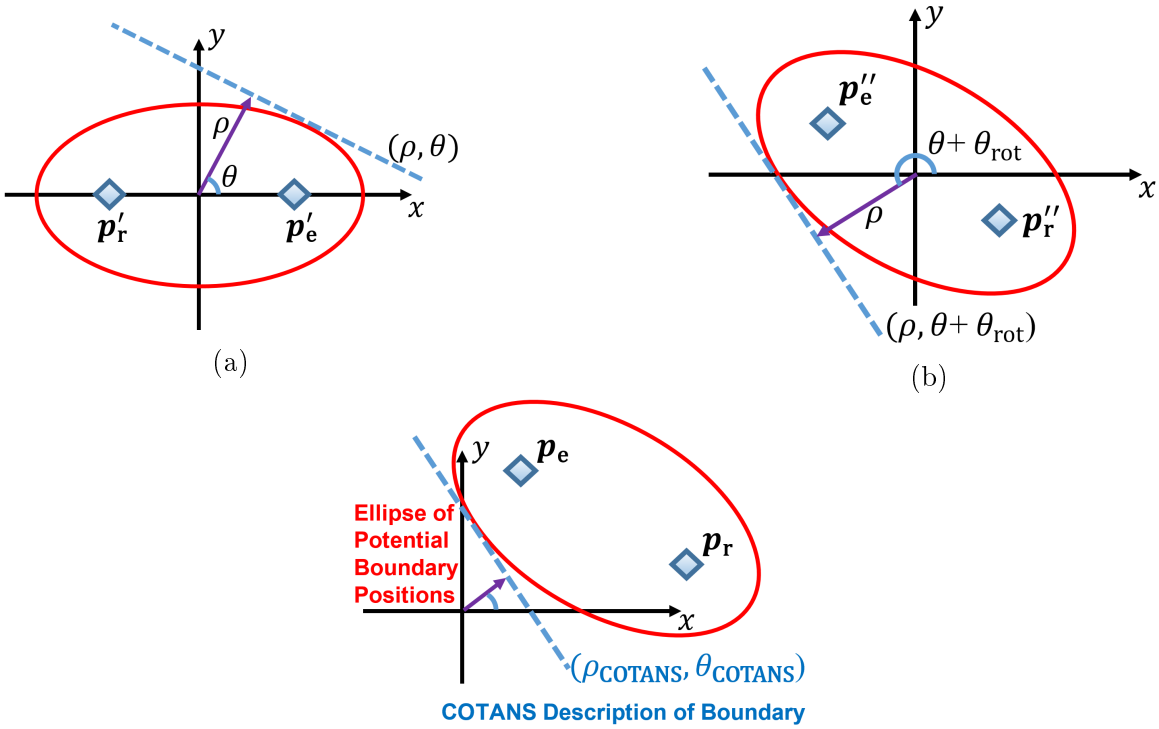


Figure 3-5: Illustration of the steps to obtain the COTANS transform of a particular tangent line (depicted in 2D). Description of one tangent line of a standard ellipse (a); the rotation of this origin-centered ellipse and its tangent (b); and the translation of this ellipse to its real position with the emitter and receiver as its foci (c).

0.1° for θ and ϕ. We define a tensor over ρ, θ, and ϕ with this resolution, and for each rounded $\{(\rho, \theta, \phi)\}_{\text{COTANS}}$ point, we increment the corresponding cell in the tensor by 1. We sum the tensors for each NLOS arrival at each receiver, to obtain a final tensor of superimposed COTANS curves such as in Fig. 3-6(a). Each ellipse contributes a curve; observe that the curves do not intersect at a single point because of errors in $\{\hat{\tau}_{i,j}\}$ and in the measured $\mathbf{p}_{\text{e}}$ and $\mathbf{p}_{\text{r}}$, so we have multiple maxima. We therefore use a smoothing filter to replace the tensor values with local averages. For our particular HiFAAT experiments, we heuristically use a uniform averaging filter that is 3 mm wide in ρ, and 3° wide in θ and ϕ. Then, we extract the locations of as many maxima as there are boundaries as in Fig. 3-6(b), where we set the neighborhood of each maximum in the tensor to zero, to avoid picking the same boundary multiple times. In this specific implementation, given a maximum of the accumulator result at some

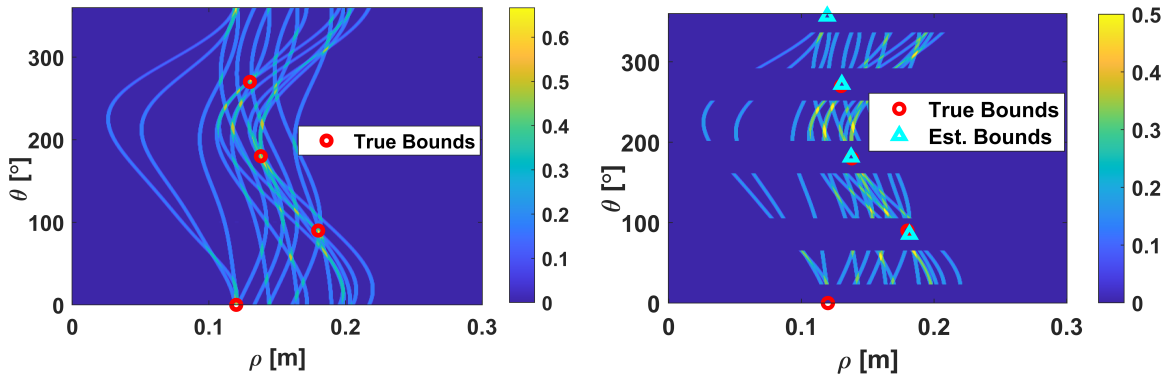


Figure 3-6: The COTANS accumulator for inconsistent $\{\hat{\tau}_{i,j}\}$ (a), and the resulting boundary estimates (b). The image is periodic in azimuth, so the maximum near 360° is close to the correct location at 0°

$(\rho_{\max}, \theta_{\max}, \phi_{\max})$, we zero out the θ and ϕ entries within ±15° of the maximum, and all ρ entries within this sector. The overall COTANS transform method for boundary estimation is summarized in Algorithm 3. In Chapter 4, we will discuss in detail an improved 2D neural network method for implementing Algorithm 3 without these heuristic smoothing and maxima-extracting steps.

Algorithm 3: COTANS transform-based boundary estimation.

Require: Time synchronization from Algorithm 2; number of boundaries N is known.

Input: $\{\hat{\tau}_{i,j}\}$, $\{\mathbf{p}_{\mathrm{r},i}\}_{i=1}^{M}$, $\hat{\mathbf{p}}_{\mathrm{e,TDOA}}$

for $\hat{\tau}_{i,j} \in \{\hat{\tau}_{i,j}\}$ **do**

  Calculate COTANS transform for spheroid, add to accumulator (Appendix A)

Result: Apply smoothing to COTANS accumulator; its N maxima are $\{\hat{\boldsymbol{\eta}}_j\}$.

Modeling the environment as consisting of planar boundaries makes environment learning more computationally tractable. Note that the virtual receiver model is not limited to planar boundaries: the planar boundary estimates can be tangents to a non-planar boundary. In this case, however, the boundary estimates will be inaccurate for localization purposes when the emitter moves to a different position. Therefore, rather than using the initial COTANS boundary estimates for the entire emitter trajectory, we need evolving boundary estimates over time, utilizing the global

optimization method which we now discuss.

3.1.3 Stage 3: Joint Boundary Estimation and Emitter Localization

At this stage, we have initial estimates of both the emitter position and the boundaries, allowing for simultaneous optimization over $\mathbf{p}_\mathrm{e}$ and $\{\hat{\boldsymbol{\eta}}_j\}$. We assume candidate boundaries $\{\boldsymbol{\eta}_j\}_\mathrm{cand}$ that are in the vicinity of the $\{\hat{\boldsymbol{\eta}}_j\}$, and then carry out TOA localization (Algorithm 2) for each $\{\boldsymbol{\eta}_j\}_\mathrm{cand}$. These $\{\boldsymbol{\eta}_j\}_\mathrm{cand}$ are drawn from a search space $\rho \in [\rho_j - \rho_0, \rho_j + \rho_0]$, $\theta \in [\theta_j - \theta_0, \theta_j + \theta_0]$, $\phi \in [\phi_j - \phi_0, \phi_j + \phi_0]$, where ρ_j, θ_j and ϕ_j are estimated from Algorithm 3, and ρ_0, θ_0 and ϕ_0 are error margins obtained as the typical maximum parameter estimation errors from COTANS, for a given setting. The $\{\hat{\boldsymbol{\eta}}_j\}$ associated with the highest maximum metric C_TOA (Eq. 3.3) yields the most likely $\mathbf{p}_\mathrm{e}$ and $\{\boldsymbol{\eta}_j\}$. Interlacing this algorithm with tracking over the emitter positions, we discover and map the occluders in the environment, and refine our estimates for joint localization and environment learning.

In this simultaneous optimization problem with measurement errors, it is advantageous to maintain multiple evolving estimates of the emitter and boundary locations, for which we use particle swarm optimization (PSO) [107]. In this iterative, cooperative, and partially random heuristic global optimization algorithm, each $\{\boldsymbol{\eta}_j\}_\mathrm{cand}$ is termed a 'particle'. Each particle is initialized at a random location and given an initial 'velocity' in the search space. Algorithm 2 is carried out for each particle, and the maximum C_TOA and corresponding $\hat{\mathbf{p}}_\mathrm{e,cand}$ is recorded. Particle velocities are modified at each PSO iteration towards the global maximum metric particle (with the maximum C_TOA), and also to the maximum metric particle in a random 'neighborhood' of particles, as detailed in Appendix B. Thus, particles explore the region around the current best approximation.

PSO was used in PEEL because its metaparameters do not need to be adapted to a specific physical setting, and because it can be run without re-initialization during tracking. There are alternative global optimization algorithms that can also be used

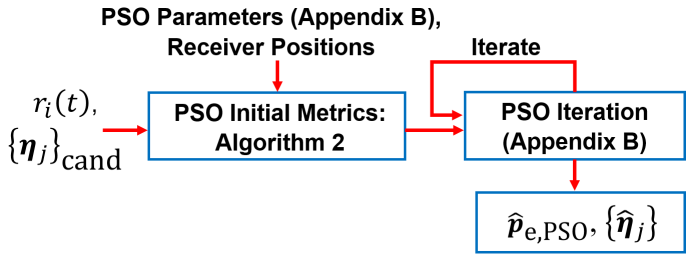


Figure 3-7: Flowchart for simultaneous boundary and emitter localization with PSO.

in PEEL, however, and it would be advantageous in an application setting to fine-tune the metaparameters of PSO or another algorithm for that particular environment.

In order to be able to accommodate for occlusion, we rely on a tracking procedure over the periodic transmissions of the moving emitter. Kalman tracking with a constant-acceleration model, for example, can be applied, as detailed in Appendix B.1. The $\hat{\mathbf{p}}_{\text{e}}$ obtained using PSO is used as the measurement vector to update the tracking. Critically, the grid-search is confined to a sufficiently small area $\mathcal{P}$ around the resulting prediction. This way, even if the emitter is temporarily occluded and the underlying C_{TOA} metric is erroneous, $\hat{\mathbf{p}}_{\text{e}}$ will still be close to the actual $\mathbf{p}_{\text{e}}$. Thus, if the emitter enters an area that is unfavorable for localization, PEEL can potentially mitigate the resulting estimation errors. Our algorithm for joint PSO localization and boundary estimation is given in Fig. 3-7, and our overall approach is summarized in Algorithm 4.

Algorithm 4: Joint PSO localization and Kalman tracking.

Input: $\{r_i(t)\}_{i=1}^{M}$, $\{\mathbf{p}_{\text{r},i}\}_{i=1}^{M}$, $\{\boldsymbol{\eta}_j\}_{\text{cand}}$
Require: Moving emitter, for tracking purposes.
  for each emitter position in trajectory **do**
    for pre-specified number of PSO iterations **do**
      Perform PSO, get $\hat{\mathbf{p}}_{\text{e,PSO}}$, $\{\hat{\boldsymbol{\eta}}_j\}$ (Fig. 3-7, Appendix B)
    end for
    Use $\hat{\mathbf{p}}_{\text{e,PSO}}$ to update Kalman tracking, get new $\mathcal{P}$ for next iteration.
  end for
  Result: $\hat{\mathbf{p}}_{\text{e,PSO}}$, $\{\hat{\boldsymbol{\eta}}_j\}$

From the results of Algorithm 4, we can determine whether each LOS path to receivers has been attenuated beyond expectation. Occlusion can be determined by

assuming a $1/\hat{R}$ attenuation factor in the received signal magnitude, where $\hat{R}$ is a given estimated LOS range. Unoccluded arrivals allow for the estimation of the emitted power, by scaling the corresponding LOS magnitudes by $\hat{R}$. If the received LOS magnitude is substantially below the magnitude predicted by geometric spreading (in our implementation, 70% below the predicted magnitude), we conclude that an occluder is present. Volumetric intersection methods can then be used to map the occluders [108]. With this type of approach, the volume of interest is discretized into parallel planes (lines in 2D). The cones of occlusion are then projected onto each of the parallel planes, followed by the occlusion-free cones of visibility being projected onto the same planes. The areas that are left occluded form a discretized approximation to the convex hulls of the occluders.

The accuracy of the occluder estimation procedure is dependent on the trajectory of the emitter and the time interval between its transmissions. If the emitter does not approximately move in an arc around the occluder and if the pulses are infrequent, the estimate of the occluder will be correspondingly coarse and blocky. In the HiFAAT watertank experiment, the trajectory, occluder position, and receiver positions have been arranged to make a good occluder estimate possible, and thus to demonstrate the full potential of this methodology.

After mapping the occluders, we re-run Algorithm 4. We now know which LOS arrivals are blocked by the occluder, and we change the additive metric C_{TOA} by leaving blocked paths out of the sum in Eq. (3.3). Our flexible cost function thereby allows us to compensate for model mismatch due to occlusion, and yields increasingly improved performance.

3.2 Simulation Experiments for Feasibility Analysis

Before PEEL was implemented on real experimental data from HiFAAT, a set of simulations was carried out to test the feasibility of joint boundary estimation and localization with PSO. In a noise-free environment similar to the shallow-water channel with a flat sea surface and seafloor, and two other boundaries forming a 90°

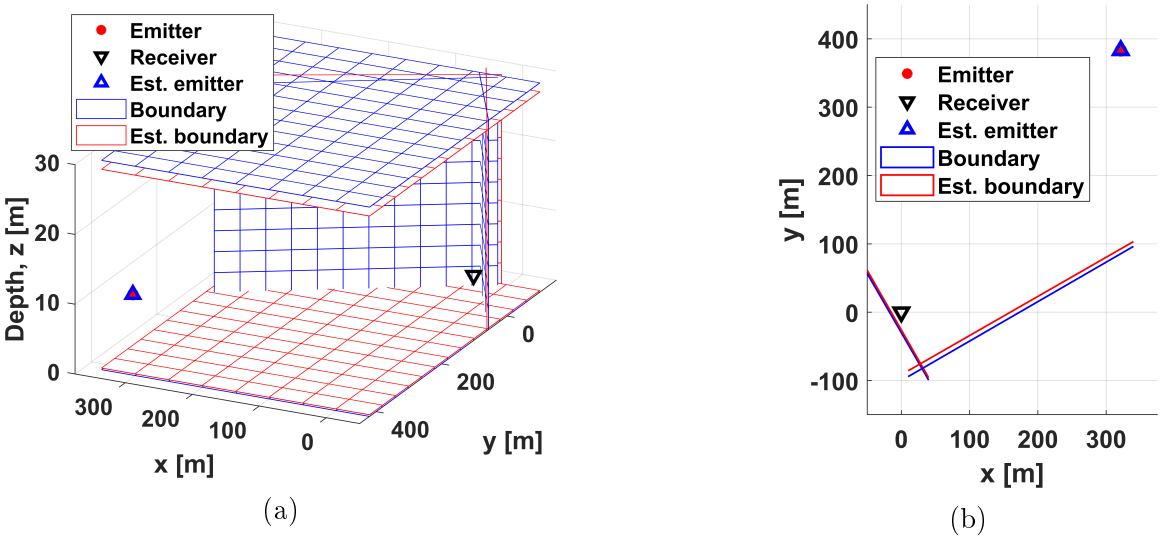


Figure 3-8: Top (a) and side (b) views of the simulated localization and environment learning experiment, and its estimated results.

wedge behind closely-spaced receivers, a synthetic signal was simulated using the mirror-image method detailed in Chapter 2. The simulation setting and the resulting emitter and boundary estimates are presented in Fig. 3-8. Here, the environment has a flat surface at depth $z_1 = 0$ m, and a flat bottom at depth $z_2 = 30$ m. The emitter is located at (500, 50, 10) m, and the reference receiver position is at (0, 0, 5) m. There are three other receiver positions at (2, 2, 6) m, (-1, 0.7, 4) m, and (-1.5, -1, 7) m, simulating a small receiver array with irregular spacing. Since the receiver positions are relatively close to each other, we only plot a single reference receiver position for convenience.

The search space for the boundaries is assumed to be $z_1 \in [-2, 2]$ m, $z_2 \in [-28, 32]$ m, $r_3 \in [20, 35]$ m, $\theta_3 \in [190, 225]^\circ$, and $r_4 \in [80, 110]$ m. These search spaces are chosen to reflect relatively good knowledge of the surface and bottom locations, and uncertainty in the distance and orientation of the corner behind the receiver. To localize the emitter, we run 20,000 iterations of PSO with 128 particles. This is a conservatively high number of computations, since we also achieved similarly good results with 48 particles and 7000 iterations.

Over 10 trials of PSO, we consistently localize the emitter at the correct location, while the boundary estimates can have small deviations from their correct positions,

as in Fig. 3-8. This is because, with four NLOS arrivals to work with, we have more than enough information for correct localization in this noise-free environment. Thus, the correct emitter estimate is achieved before PSO converges to a perfect estimate of the environment. This simulation result demonstrates the correct implementation of key modules of PEEL, before applying these algorithms to the real HiFAAT data.

We include these simple simulation results to outline the methodology for implementing and subsequently debugging the localization methods. To tackle the difficulties of calibrating real electronic equipment in a noisy environment, it is essential to first test the algorithms on simple simulated data. When modeling 3D scenarios with vector geometry, it is easy to make simple coding errors that are difficult to detect, such as degrees being used rather than radians, or sign errors when finding the mirror images of points in space. Verifying the correct operation of the algorithms in a noise-free simulation setting is very helpful in preparing for the challenges of real-life data.

3.3 Experimental Results With Watertank Testbed Data

We now evaluate our PEEL method for its passive localization and environment learning capabilities, seeking to demonstrate and understand its performance and robustness. For this purpose, we use the convenient setting of the HiFAAT watertank depicted in Fig. 1-5, as was detailed in Section 2.2.3. We first test PEEL on real data obtained from the watertank; then add synthetic noise to this data to analyze PEEL's performance as compared to other localization methods; and finally conduct trials in a simulation setting (still corresponding to the HiFAAT environment) with a wider range of localization scenarios.

In our experiments, we emit a high-frequency Gaussian pulse with a waveform given by:

$$s(t) = e^{-t^2 f_0^2} \cos\left(2\pi f_0 t\right), \tag{3.7}$$

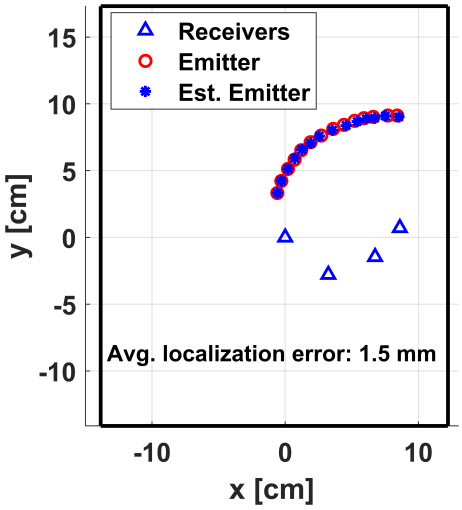


Figure 3-9: PEEL localization performance with no occluding cylinder present in the environment.

where the center frequency f_0 is set to 280 kHz, corresponding to a wavelength of 5 mm. The boundaries used for localization are the water surface, whose position is known since the receiver depths are known; and the bottom and three plastic sides of the tank, whose ranges to the origin and azimuths are respectively known to within ± 5 mm and $\pm 6^\circ$, as obtained using COTANS. One side-boundary is deliberately left unaccounted-for, which presents the difficulty of having an unmodeled set of arrivals in the data, so that we have $N = 5$. We use $M = 4$ receivers. The speed of sound through water in the tank was measured as 1485 m/s.

For our PEEL implementation, we use 128 particles and 5000 PSO iterations for each emitter position. At each PSO iteration, we use a grid $\mathcal{P}$ of 0.5 mm resolution and sides of 6 mm centered on the expected emitter location. We perform 10 experiments for each localization scenario, and obtain the root mean square (RMS) localization error. First, we run PEEL on a dataset where there is no occluding cylinder, to obtain the best possible performance as a basis for comparison. The results in Fig. 3-9 illustrate that the performance which is achieved in this case is within the margin of measurement error of emitter and receiver positioning for this experimental setup.

Implementing our method on the case of an occluding cylinder being present in the environment, the RMS error for the first-pass of PEEL is 3.9 mm, as seen in Fig. 3-10. To put this into perspective, the average distance of the emitter from the receivers

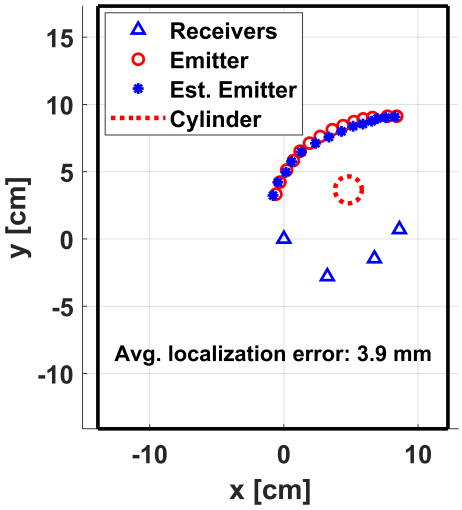
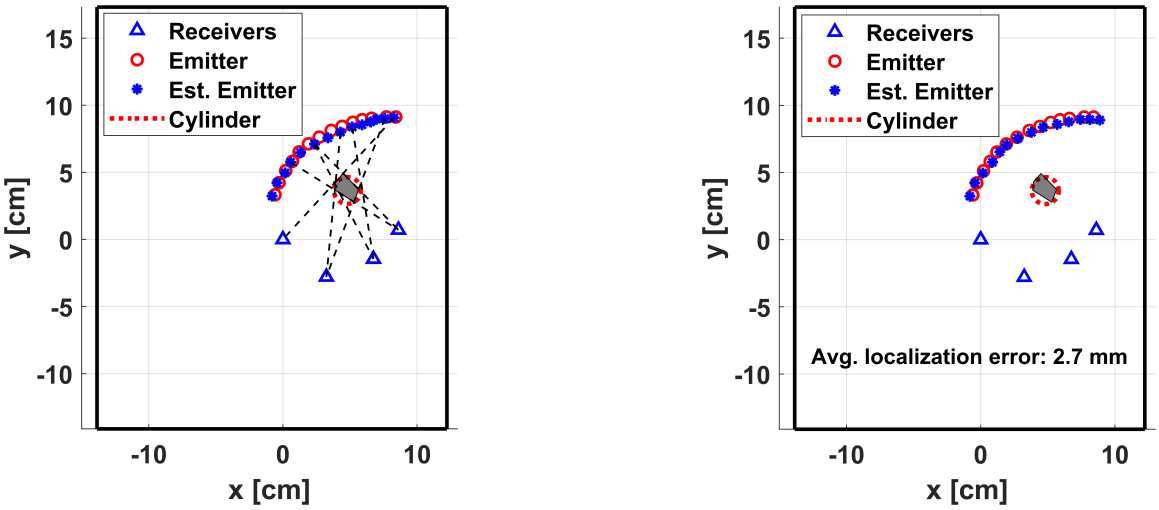


Figure 3-10: PEEL's first-pass localization performance, with an occluding cylinder present in the environment. The boundary that is unaccounted-for to introduce model mismatch is the bottom boundary in these top-view figures.

for this trajectory is 93.5 mm, and the RMS error in this scenario is 4.2% of the mean distance to the emitter. The final boundary estimates in Fig. 3-10 are obtained using the full estimated trajectory, which leads to a more accurate solution compared to the results obtained from using just one emitter position. Running PEEL's second-pass after estimating the occluder as in Fig. 3-11(a), we get the result in Fig. 3-11(b), where the RMS error is reduced to 2.7 mm (2.9% of the mean distance). The occluder is reconstructed with sufficient accuracy to improve our localization result. To understand how much the unknown occluder's presence affects the result, we note that the experiment with the same scenario but with no occluder (in Fig. 3-9) obtained an RMS error of 1.5 mm (1.6% of the mean distance), which is within our margin of measurement accuracy. In these trials, occlusion and interference due to reverberation are the key error factors; the SNR of unoccluded LOS arrivals is > 30 dB, so that noise was not a significant cause of localization errors.

If synchronization and an initial boundary estimate are given to us, single-receiver localization using Algorithm 4 becomes feasible. In an experiment with a single emitter at the origin and no occluder present, we get 2.1 mm of RMS error (2.2% of the mean distance). This strong performance is surprisingly only slightly worse than when we use 4 receivers.



(a) (b)

Figure 3-11: Estimation of the occluder (a), and PEEL's improved second-pass result, using the modified metric obtained by taking the occluder into account (b).

In order to analyze the performances of the localization algorithms that we have used in PEEL (TOA localization with known boundaries, and PSO), we carried out localization trials with synthetic Gaussian noise added to experimental signals. For the first position in the trajectory, where there is LOS to every receiver, the same noise power in units of dBm was added to each recorded signal. In Fig. 3-12(a), we compare PSO's performance for different numbers of PSO iterations, versus that of TOA localization with known boundaries. Because PSO optimizes for the boundary locations as well, its high runtime led to us conduct 150 trials for each noise level, while 5000 trials were carried out for TOA. There is a performance gap between PSO and the 'ground truth' TOA result, which is smaller for higher noise levels. Longer runtimes improve PSO's performance and help close this gap.

We next evaluate the performance of environment learning by adding synthetic Gaussian noise to the signals measured at the last position in the trajectory (where 2 LOS arrivals are occluded) and obtaining the RMS range error from the origin to all the boundaries, averaged over trials and the number of boundaries. We conducted 100 such trials for each noise level and for three scenarios: in the absence of the occluding cylinder as measured in a separate set of experiments, with the occluding cylinder present but not accounted for, and with the occluded arrivals removed in the revised

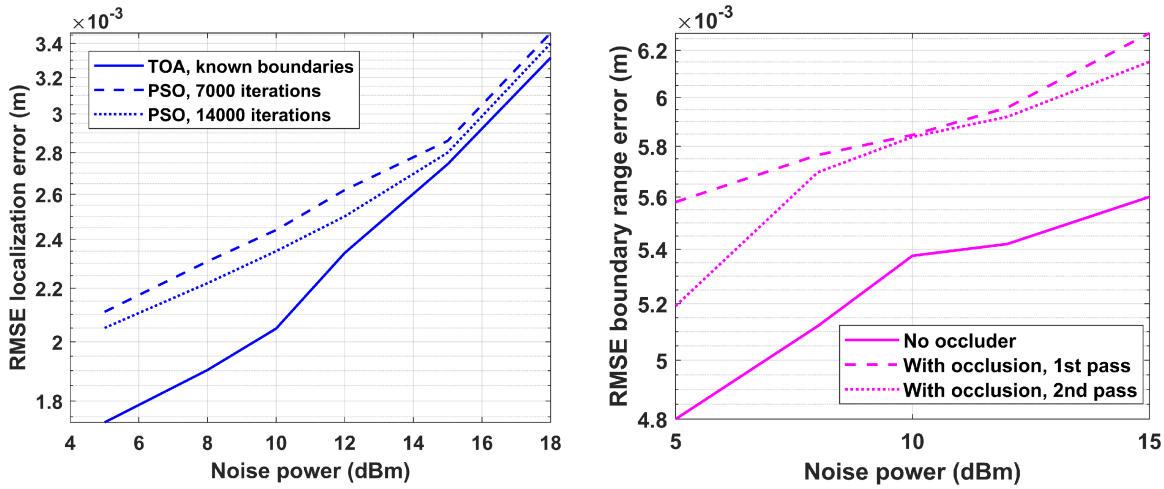


Figure 3-12: Performance of PSO localization versus TOA with known boundaries, highlighting better accuracy with more PSO iterations (a); and RMS boundary localization error performances for experimental results with synthetic added noise (b).

metric (modeling PEEL's second-pass). The results are given in Fig. 3-12(b), where we observe a gap in performance caused by the occlusion, beyond the performance deterioration caused by a lower effective SNR. Boundary estimation is improved by compensating for the occluder, though this improvement is less dramatic than the performance gains in emitter localization. Adding a noise power greater than 20 dBm causes much larger errors in the localization and boundary estimation tasks in this setting, as the noise peaks have much greater magnitude than the NLOS arrivals.

To test PEEL's performance on a wider range of experimental geometries, we generated synthetic data that modeled the HiFAAT watertank setting with randomized receiver locations. We implemented a simulation setting as per [68] that modeled a reverberant environment with higher-order reflections, producing synthetic received signals with the mirror image method. In these trials, the emitter's trajectory, the boundary locations, the occluder, and the position of the receiver at the origin were kept constant, and PEEL was run for 10 different deployments of the other 3 receivers. These receivers were distributed uniformly over the region bounded by the positions of receivers in the real experimental setting. We used the same Gaussian pulse that was used in the watertank experiment, though the only distortion that these pulses

underwent in simulation was attenuation. Ultimately, we obtained an RMS error of 2.6 mm over 10 trials, which was close to the watertank results.

Overall, we observed reliable performance across all of the trials for this particular demonstration. Once the unknown occluder is compensated-for, we only observe several millimeters of error, which is reasonable since there are already millimeter-scale errors in the measured receiver positions. These results are summarized in Table 3.1.

Table 3.1: Localization experiment results.

Scenario	**RMS Error (mm)**
4 receivers, no occluder	1.5
Single receiver, no occluder	2.1
4 receivers, occluder, first-pass	3.9
4 receivers, occluder, final result	2.7
Simulation, random receivers	2.6

3.4 Discussion of Experimental Results and Challenges

The trajectory to which PEEL was applied resulted in a difficult localization problem which nevertheless proved to be solvable. However, the strong results that we have presented from the HiFAAT watertank testbed were only obtained after intensive calibration work, and several initial trials had to be carried out to meet unexpected challenges. For future reference, it is important to illustrate some of the difficulties that can be encountered in such scaled-down underwater acoustics experiments.

In many localization applications, it is standard to consider the emitter/receiver's physical dimensions as negligible. This is often a reasonable assumption: if the bandwidth of the emitted signal only supports a distance estimation accuracy on the order of meters or tens of centimeters, a small centimeter-scale hydrophone can be conceptualized as a point source/receiver. However, when using wideband pulses that allow for millimeter-scale precision, this model no longer holds. The active element of a

hydrophone is a ceramic sphere embedded within it, with the plastic sheath around it being used to give the hydrophone structural integrity, and to protect the cables and other electronics connected to this sphere. In an emitter/receiver pair, the true distance traveled by a pulse is from the outer edge of one ceramic sphere to the outer edge of the other. Hence, the radii of these spheres must be taken into account and the nominal distances at which the hydrophones are placed must be modified to compensate for such an offset. Since we were not in a position to break open one of our cylindrical hydrophones to determine exactly where this ceramic sphere was positioned along its length, we carried out calibration experiments specifically to determine such parameters.

The millimeter-scale precision of the experiment setting also required care to setup details that are similarly easy to overlook. Just as it is typical to consider the emitter/receiver dimensions as negligible, it is typical to neglect the dimensions of the attachment points of transceivers. When attaching a hydrophone to a rod and then lowering it into the water, the resulting position offset must be taken into account in high-precision experiments.

Another unexpected precision-related difficulty was the watertank sides in HiFAAT not being perfect planar surfaces. Not only did this plastic watertank have sides that were slightly angled, but they were also slightly concave, only perceptible at a millimeter-scale that was difficult to observe without close examination. This concavity meant that our planar boundary assumption only held approximately, and that if a planar boundary ground truth were assumed, then the result would be large localization errors at some emitter positions. This is where PEEL's dynamic environment estimation capabilities proved critical for strong performance: at each position, it was imperative to re-calculate the planar boundary estimates for the watertank. The resulting boundary estimates differed from the trigonometrically-measured ground truth for the boundary positions, but led to better results. Until the complete PEEL method was implemented, however, this feature of HiFAAT was a major source of difficulty for our calibration experiments until it was identified, since it led to unexpected discrepancies in the TOAs of the NLOS arrivals.

An unexpected challenge encountered in the localization task was the directionality of the (approximately isotropic) emitter hydrophone. We have previously discussed in Chapter 2 that the predicted and observed magnitudes of the received pulses had large discrepancies due to environmental model mismatches, while their TOAs were accurate. However, we found that unaccounted-for directionality also played an important role in such discrepancies. Most significantly from an operational perspective, this directionality issue sometimes caused the surface-reflected NLOS arrival to have a larger magnitude than the LOS arrival. Since the LOS arrival has the shortest path length, it ought to have the smallest attenuation and thereby be the most significant arrival in the received signal in the case of no occlusion being present. Since the larger magnitude of the surface-reflection was an unexpected phenomenon, many tests had to be carried out to confirm that the directionality of the emitter was the culprit, and not another source of error or equipment failure.

Another complication caused by a NLOS arrival that is stronger than an unoccluded LOS arrival, is that a more complex algorithm has to be used for the identification of the LOS. Normally, one could simply pick the highest-magnitude peak in the matched-filtered received signals. When the LOS peak is not the highest, however, it will still be the earliest significant arrival, provided that it is unoccluded and the emitter is roughly isotropic, even if imperfectly so. Hence, a simple method in an application like HiFAAT (with relatively low noise power) is to set a threshold to e.g., 50% of the magnitude of the highest peak in the matched-filtered signal, and then to pick the earliest peak that exceeds this threshold as the LOS. Except for purposes of time synchronization for initialization, PEEL does not specifically require the identification the LOS; hence, we did not require more sophisticated LOS-identification algorithms at this stage.

Note that we do not require a complete deduction of the occluders for improved localization performance. In this passive setting, a complete and accurate mapping of the occluders requires the emitter to move in an arc around them, as is the case in tomography applications. It is not guaranteed that occluders can be mapped in this manner, since the emitter's path is not under our control. In PEEL, however, as long

as occlusion is detected, the potential errors from summing over these non-existent arrivals in TOA localization are avoided.

Chapter 4

Environment Estimation with Machine Learning

In Chapter 3, we presented the COTANS transform method for boundary estimation that was featured in the literature, and noted that we had developed a NN method for 2D boundary estimation that was based on this transform. Focusing on reflective boundary estimation for shallow-water underwater acoustic settings, we specifically deal with problem geometries similar to that illustrated in Fig. 4-1. Our task is to estimate the reflective boundaries using both known emitter and receiver locations; the method that is designed for this task can then be embedded into the PEEL method which has a larger scope. While there is typically some prior knowledge of boundaries such as the sea surface and seafloor, we require more accurate knowledge of their positions in order to make use of the corresponding NLOS arrivals. Recall that over short ranges, we can approximate boundaries as planar, and can model them as producing mirror images of the emitter as 'virtual' emitters, as per Snell's Law [68]. Euclidean distance matrices (EDM) [13] or other methods [14] can then be used for boundary estimation through the localization of these virtual emitters. For indoor settings, there is theoretical work and experimental demonstrations of whether various arrangements of receivers can estimate rooms of varying numbers of walls [109]. In underwater acoustics, however, we often have to deal with low signal-to-noise ratios (SNRs) [19] and the model mismatch issues arising from a dynamic

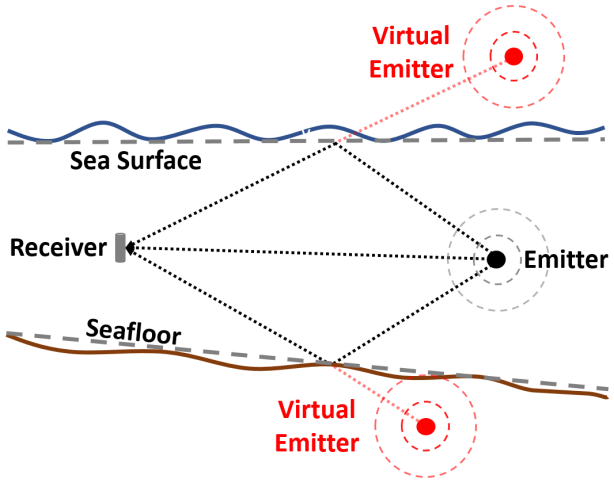


Figure 4-1: A general underwater acoustic setting, highlighting the typical reflective boundaries, their NLOS arrivals, and the corresponding virtual emitters. Note that for the COTANS boundary estimation task, the emitter location is known, and it is the virtual emitter locations that are unknown since the problem is the estimation of the unknown reflective boundary locations.

environment. To the best of our knowledge, existing methods generally do not address these difficulties.

In light of these challenges, we propose a convolutional neural network-based (CNN) regression method for boundary estimation through supervised learning. Our data-driven method operates by parametrizing tangents to ellipses by the range (ρ) and azimuth (θ) values of their normal vectors [34], calling this (ρ, θ) space the common tangents to spheroids (COTANS) domain. This COTANS transform maps the environment geometry and time-delay estimates to images in the (ρ, θ) space, transforming the data into a more natural input representation for the CNN. The proposed COTANS neural network (NN) method, termed COTANS-NN, incorporates a modified AlexNet [110] architecture. It is trained on a simulated dataset to estimate reflective boundaries from unlabeled NLOS arrivals over a wide range of SNRs. The resulting NN-based method can be used with both simulated and recorded data. A key influence for our work was the successful recent use of NNs for emitter localization and environment learning, including the underwater acoustic setting [51], [52] and reverberant indoor environments [53]. Although we target the short-range shallow-water underwater acoustic setting, we propose a general-purpose boundary

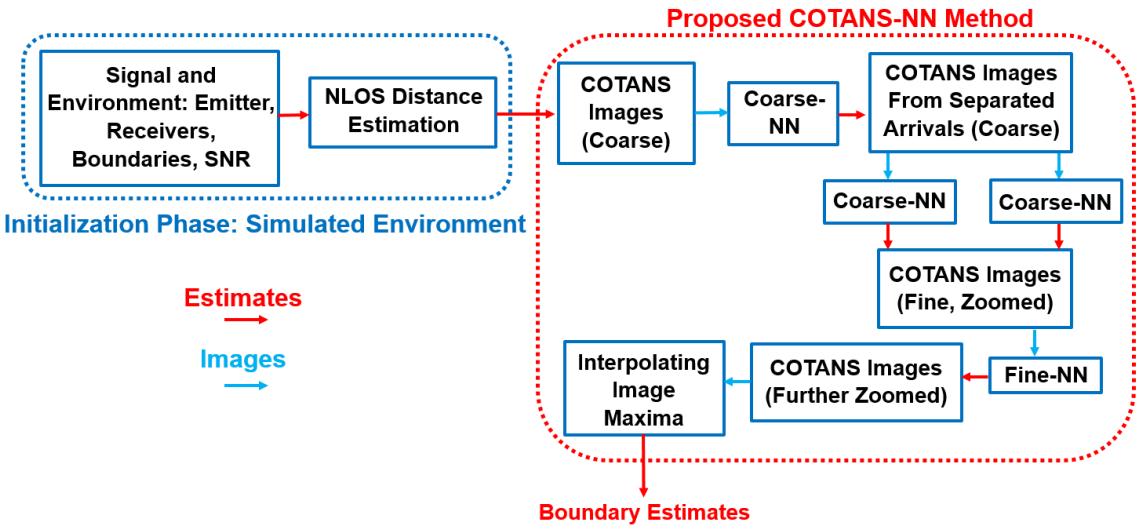


Figure 4-2: Summary of the COTANS-NN method's presentation.

estimation method for any setting where straight-ray propagation holds, and which outperforms its alternatives in simulation and in real-life experiments.

The main contributions detailed in this chapter are the following:

- A robust NN method for 2D boundary estimation that is superior to state-of-the-art alternatives, and is straightforward to re-train for different environments.
- A study of the performance and stability of alternative boundary estimation methods.
- A Cramér-Rao lower bound (CRLB) for boundary range estimation, which allows us to verify that the boundary estimation methods have been implemented correctly.

In the literature, the problems of time-delay estimation and localization are often treated separately, with different error models. Here, we provide a unified framework that establishes a common setting for these tasks, allowing for the study of COTANS-NN and its alternatives under realistic conditions. Note that the method which we develop considers the emitter location to be known; as a module of PEEL, the 'known' emitter location would be an estimate provided by previously-applied TDOA and TOA algorithms.

We develop COTANS-NN through the conceptual framework illustrated in Fig. 4-2. In Section 4.2, we present an error model for distance estimates that is important for generating a large realistic dataset for training and testing COTANS-NN. The COTANS methodology and our NN method are formulated in Section 4.3. Alternative methods are outlined in Section 4.4, and the CRLB for boundary range estimation is derived in Section 4.5. We present our results in Section 4.6.

4.1 Problem Formulation for Reflective Boundary Estimation

We presented the notation, signal model, and environment geometry that frames the 3D boundary estimation problem in Chapter 2. In the particular case of a 2D environment, the N planar boundaries are tangent lines, which are described by the range $\rho \in \mathbb{R}_+$ and azimuth $\theta \in [0, 2\pi)$ of their normal vector relative to the origin. Thus, the j-th boundary is parametrized as the vector $\boldsymbol{\eta}_j = [\rho_j \; \theta_j]^{\mathrm{T}}$. We assume a single isotropic emitter in the environment at a known location $\mathbf{p}_{\mathrm{e}} = [x_{\mathrm{e}} \; y_{\mathrm{e}}]^{\mathrm{T}}$, and M isotropic receivers at known $\mathbf{p}_{\mathrm{r},i} = [x_i \; y_i]^{\mathrm{T}}$, $i \in \mathcal{S}_M$, with isotropicity only needed for every boundary to produce a NLOS arrival at every receiver.

Recall that given the known emitted waveform as $s(t)$, which will ultimately be used to matched-filter the received signals, and equivalent attenuation coefficient $\alpha_{i,j}$ for each path, the received signal at the i-th receiver is modeled by:

$$r_i(t) = \sum_{j=0}^{N} \alpha_{i,j} s\left(t - \tau_{i,j}\right) + \xi_i(t), \tag{4.1}$$

where $j = 0$ corresponds to the LOS path, and $\xi_i(t)$ is a noise signal that is a realization of a spectrally-flat Gaussian process. We work with a discrete-time sampled version of Eq. (4.1) as $r_i[n] \triangleq \{r_i(t)|_{t=nT_{\mathrm{s}}}\}_{n\in\mathbb{Z}}$, where T_{s} is the sampling period, and the sampling frequency f_{s} is assumed to be greater than twice the Nyquist rate for the signals considered.

Table 4.1: Model of the boundary estimation problem.

Problem Feature	**Modeling Assumptions**
Speed of sound, v_{s}	Known, and constant within the environment
Environment	Static
Reflectors, $\boldsymbol{\eta}_j$	Planar, known number, unknown position
Emitter, $\mathbf{p}_{\mathrm{e}}$	Known position
Transmissions	Known pulses (i.e., known waveforms for matched-filtering)
Receivers, $\mathbf{p}_{\mathrm{r},i}$	Known positions

The geometric information for boundary estimation consists of the known $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$, and the unknown $\{\tau_{i,j}\}$. Hereafter, we assume that the NLOS TOAs are estimated using an (at least asymptotically) optimal estimator. For example, these estimates can be obtained by matched-filtering $r_i[n]$ with $s[n] \triangleq s(nT_{\mathrm{s}})$ and picking the TOAs corresponding to the N largest peaks in the result, after removing the LOS arrival, as $\{\hat{\tau}_{i,j}\}$. The distance estimates $\{\hat{d}_{i,j} \triangleq v_{\mathrm{s}}\hat{\tau}_{i,j}\}$ (from Eq. (2.2)) are then used for estimating the boundaries as $\{\hat{\boldsymbol{\eta}}_j\}_{j=1}^{N}$, recalling that the environment is modeled as per Table 4.1.

In many environment estimation methods, it is standard to model the $\{\hat{d}_{i,j}\}$ as corrupted by Gaussian noise [16]. However, it is the received signals of Eq. (2.3) that are in fact subject to Gaussian noise, and we thus require a more realistic error model for $\{\hat{d}_{i,j}\}$, as presented next.

4.2 NLOS Time-Delay Estimation Methods

Time-delay estimation is a topic that has been extensively studied in the literature [111], [112]; here, we describe our measurement model for the time-delay estimates $\hat{\tau}_{i,j}$. We generate a wide range of time-delay errors in our dataset, modeling operational conditions under both high and low SNR. For a given value of the SNR as S, we obtain a corresponding error $\epsilon_{i,j}(S)$ that is not necessarily additive or Gaussian.

In the boundary estimation problem, the NLOS arrivals from each boundary will produce a received signal with multiple peaks as illustrated in Fig. 4-3, and we will be picking a given number of the highest peaks from the matched-filtered $\{r_i(t)\}$ to ob-

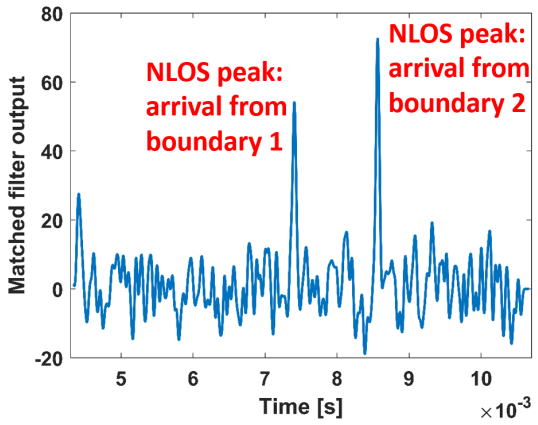


Figure 4-3: Example of a simulated received signal with two NLOS arrivals that has been matched-filtered, with the two peaks with highest magnitudes yielding the NLOS time-delay estimates. The emitted signal was a Gaussian pulse of 15.4 kHz bandwidth, and synthetic additive white Gaussian noise was introduced to produce this simulated signal.

tain the NLOS time-delays. This method yields the ML estimator for the time-delay estimation of a single arrival. We assume that the multipath arrivals are typically sufficiently separated, and the signal is of high enough bandwidth, so that after matched-filtering they do not affect each other's time-delay estimation performances through interference. This model was observed to be valid for our particular experiments, with noise and model mismatches posing much greater difficulties for performance than interference. However, for settings where time delay estimation must take interference into account, we present the more advanced expectation-maximization (EM) [113] and space-alternating generalized expectation-maximization (SAGE) TOA estimation algorithms [114], [115], which are able to handle overlapping arrivals.

Time-delay estimation performance is specific to a given emitted signal, such as a standard Gaussian pulse that we employ throughout:

$$p\left(t\right) = e^{-\frac{2\pi t^2}{\tau_p^2}}, \tag{4.2}$$

with $\tau_p = 1/B$, where B is the 3 dB bandwidth (in Hz). This pulse has energy $E_p = \int p^2\left(t\right) \mathrm{d}t$. While Eq. (4.2)'s infinite-length signal is truncated in practice to some finite length τ_d, the pulse exponentially decays to negligible magnitudes and the

finite-length signal is functionally equivalent to its infinite-length formulation. We assume that we have real additive white Gaussian noise[1], with the one-sided power spectral density of the noise $\xi_i(t)$ equal to N_0. Thus, the SNR is defined as E_p/N_0. For the sake of deriving consistent time-delay statistics, we assume that S is the same at each receiver, so that we can define an average error versus S for assessing boundary estimation performance. In practice, S varies across receivers. Given a desired SNR S in dB, we obtain $E_p/N_0 = 10^{S/10}$. For a sampling rate of f_s, the average noise power N_avg is $N_\mathrm{avg} = N_0 f_\mathrm{s}$. Thus, for a desired S, the required variance σ^2 of the synthetic additive Gaussian noise is:

$$\sigma^2 = \frac{f_\mathrm{s} E_p}{10^{S/10}}. \tag{4.3}$$

4.2.1 Cramér-Rao Lower Bound for Time Delay Estimation

It is well-known that time-delay estimation has a performance profile that is characterized by a transition from a non-informative region at low SNRs, through a threshold phenomenon, to a 'small-errors' regime at high SNRs. At high SNR, the Gaussian noise added to the emitted pulse to produce $r(t)$ results in a Gaussian, small-scale perturbation of $\hat{\tau}$, and a matched-filtering mean squared error (MSE) estimation performance that asymptotically coincides with the CRLB for time-delay estimation, which for a given S has variance $\sigma^2_\mathrm{C}(S)$ [116]. The resulting estimation error is called a 'local error', with $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2_\mathrm{C}(S))$.

4.2.2 Ziv-Zakai Lower Bound for Time Delay Estimation

At lower SNRs, peaks of noise can have a greater magnitude than that of the true arrival. Picking one of these spurious peaks results in a 'global error' that leads to a drastic performance reduction; since the noise peaks are distributed uniformly in the received signal, global errors cause $\hat{\tau}$ to be distributed uniformly on the observation time interval. Thus, if we examine $r(t)$ in the time interval $[\delta, T_p + \delta]$, we have

[1] If complex signals are used, some constants are modified in the following results.

$\hat{\tau}_{i,j} \sim \mathrm{U}(\delta, T_p + \delta)$, where $\mathrm{U}(a, b)$ denotes a uniform random distribution within the limits [a, b]. As SNR is progressively reduced below a heuristic threshold, there is a transition region where such global errors increasingly dominate the MSE. Note that time-delay estimation performance is specific to a given channel, pulse, and observation window.

Although we did not employ the ZZLB in our published research, it was an important step in designing our method. For a given emitted signal, the ZZLB helps to rigorously identify the SNR threshold for global errors. Hence, it is a key first step in studying the SNR performance in a particular localization or boundary estimation setting, and it is an aid for verifying the correct implementation of time delay estimators. Empirical matched-filtering performances are expected to have an SNR threshold below which they deteriorate significantly, that roughly corresponds to the threshold indicated by the ZZLB. A discrepancy between the theoretical and empirical thresholds could therefore highlight an implementation error.

4.2.3 Matched-Filtering Performance

In Fig. 4-4, we conduct a range estimation simulation with 10,000 realizations of the simulated noise added to $p(t)$ in Eq. 4.2, per SNR value considered, and compare it to the CRLB for time-delay estimation. These results illustrate the need for a boundary estimation method that performs accurately when errors are small (local errors) and robustly when they are large (global errors).

4.2.4 EM and SAGE Algorithms for Time Delay Estimation

The methodology of using matched-filtering to obtain TOA estimates fits the established framework in localization for interpreting local and global errors. When multipath interference is a greater issue for TOA estimation than noise, however, matched-filtering is a suboptimal algorithm. Therefore, for conditions where the multipath arrivals can have significant overlap, we briefly summarize the EM and SAGE algorithms that are designed to tackle this problem and can be deployed instead of

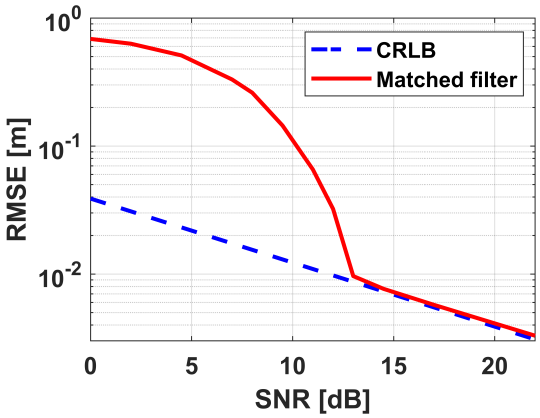


Figure 4-4: The CRLB on the range estimation root-mean squared error (RMSE) for a Gaussian pulse of 15.4 kHz bandwidth, and the simulated empirical matched-filtering performance. The global error threshold for this particular signal is observed to be 13.5 dB SNR, indicating that large errors can be encountered even at seemingly high signal strengths.

matched-filtering.

The EM algorithm for time delay estimation has been demonstrated to achieve the maximum likelihood estimation of the delay and amplitude parameters of NLOS echoes, even when the pulses all overlap [113]. Recall that in our multipath setting, a received signal given by:

$$r_i(t) = \sum_{j=0}^{N} \alpha_{i,j} s\left(t - \tau_{i,j}\right) + \xi_i(t), \quad \forall i \in \mathcal{S}_M, \quad \forall j \in \mathcal{S}_N, \tag{4.4}$$

where the attenuation coefficients $\{\alpha_{i,j}\}$, additive Gaussian noise realization $\xi_i(t)$, and number of reflective surfaces N are as detailed in Chapter 2. Given a copy of the emitted signal $s(t)$ (which starts at time T_i and lasts until time T_f), and the fact that there are $N+1$ arrivals, the EM algorithm identifies the amplitudes and delays of the pulse copies in the received signal. Although we specifically focus on time delay estimation, the amplitudes (and therefore the corresponding attenuations) of the received arrivals is also potentially useful information.

At the initialization stage (iteration $m = 0$), the EM algorithm randomly initializes $\{\alpha_{i,j}^{(0)}\}$ and $\{\tau_{i,j}^{(0)}\}$. Until a user-defined convergence criterion (which can simply be a fixed number of steps), it first iterates through an E-step, where for a given i

and for each j, it computes:

$$x_{i,j}^{(m)} = \alpha_{i,j}^{(m)} s\left(t - \tau_{i,j}^{(m)}\right) + \frac{1}{N+1}\left[r_i(t) - \sum_{l=0}^{N} \alpha_{i,l}^{(m)} s\left(t - \tau_{i,l}^{(m)}\right)\right]. \tag{4.5}$$

Then, it carries out an M-step, where it computes:

$$\{\tau_{i,j}^{(m+1)}\} = \underset{\tau}{\operatorname{argmax}} \left| g_j^{(m)}(\tau) \right|, \tag{4.6}$$

$$\{\alpha_{i,j}^{(m+1)}\} = \frac{g_j^{(m)}\left(\tau_{i,j}^{(m+1)}\right)}{E_{\mathrm{s}}}, \tag{4.7}$$

where:

$$E_{\mathrm{s}} \triangleq \int_{T_{\mathrm{i}}}^{T_{\mathrm{f}}} |s(t)|^2, \tag{4.8}$$

is the emitted signal energy, and:

$$g_j^{(m)}(\tau) \triangleq \int_{T_{\mathrm{i}}}^{T_{\mathrm{f}}} x_{i,j}^{m}(t)\, s^*(t-\tau)\, \mathrm{d}\tau, \tag{4.9}$$

is the matched-filtering result. With these iterative computations, we progressively bring a best-guess synthetic signal $x_{i,j}$ to match the real observed signal more closely.

One of the issues with this EM algorithm is that it does not prevent the selection of the same pulse multiple times. For example, a pulse of amplitude 1 can be interpreted as two identical and superimposed pulses with the same time delay, each with an amplitude of 0.5. To counteract this problem, we use the SAGE algorithm [115]. Rather than carry out the E-step for all echoes and then the M-step for all echoes separately, we simply compute the E- and M-step for a single echo, for each echo, and repeat. This decoupling of the echo estimations leads to the algorithm's convergence to different TOAs, as desired.

Note that while we have detailed several TOA estimation algorithms with the ultimate goal of boundary estimation, it is also possible to estimate the TDOAs from the received signals [117]. In the PEEL framework, we envision using TDOA localization

to obtain an estimate of the emitter location, followed by TOA localization. Hence, when a boundary estimation is employed as part of a larger system, the TOA estimation stage which we have presented here may potentially be preceded by a TDOA estimation stage.

4.3 COTANS-NN for Boundary Estimation

Having defined the boundary estimation problem in Section 4.1, we present the COTANS-NN method. Suppose that a set of $\{\hat{\tau}_{i,j}\}$ has been estimated from $\{r_i(t)\}$, and are unlabeled with respect to the corresponding boundaries. The goal is to estimate $\{\boldsymbol{\eta}_j\}$ in a way that is robust to errors in $\{\hat{\tau}_{i,j}\}$. We first discuss how the COTANS transform is used to generate images for a given geometry and set of $\{\hat{\tau}_{i,j}\}$ [34]. We then detail how COTANS-NN estimates the $\{\boldsymbol{\eta}_j\}$ from these COTANS images.

4.3.1 Additional Details of COTANS Image Generation

We previously summarized how COTANS images could be generated from time delay estimates in Chapter 3. Here, we briefly summarize the notation for COTANS image generation specifically for the 2D case, and present some additional details. Our focus is on the generation of COTANS images over a wide range of SNRs and boundary positions to create a large dataset, rather than just a specific instance from a specific experiment.

In 2D, for a given geometry and $\{\hat{\tau}_{i,j}\}$, a boundary defined by ρ and θ can be conceptualized as a point (ρ, θ) in a COTANS transform domain; working out the (ρ, θ) expression of a line is done by computing its COTANS transform [106]. Discretizing the space $\rho \times \theta$ as a matrix and incrementing this accumulator array (instead of a tensor as in 3D) over every potential (ρ, θ) for NLOS ellipses yields a COTANS-domain image (e.g., as in Fig. 4-5), whose maxima are expected to be at the true boundaries $\{(\rho_j, \theta_j)\}$ in the absence of errors. However, the time delay estimation errors that can arise as discussed in Section 4.2 pose a major challenge for this approach.

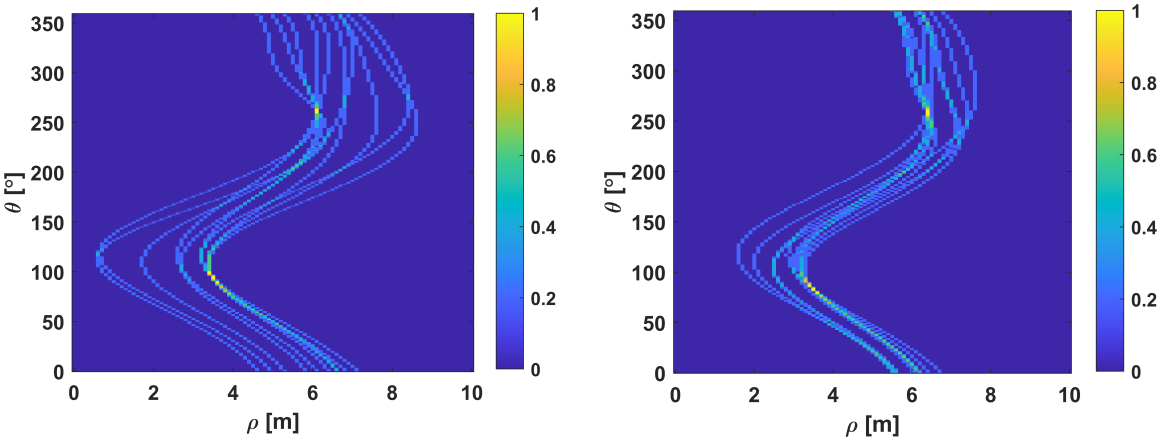


Figure 4-5: Examples of COTANS images for the transition region at 12 dB SNR (a); and for the high SNR region at 20 dB SNR (b), for an environment with two boundaries at $(\rho, \theta) = (3.5, 84)$ and $(6.4, 258)$. The images are colored for convenience; the original images are in grayscale.

We previously derived a mathematical solution for the COTANS transform which precludes the need for randomly sampling points (ρ, θ) on the surface of an ellipse [34]. Recall that for a given θ, we use vector geometry to obtain the COTANS transform's ρ for an origin-centered ellipse as:

$$\rho\left(\theta\right) = \sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}, \tag{4.10}$$

where $a = d_{\mathrm{NLOS}}/2$, and:

$$b = \sqrt{d_{\mathrm{NLOS}}^2 - d_{\mathrm{LOS}}^2}/2, \tag{4.11}$$

are the standard ellipse axes calculated from the $\{\hat{\tau}_{i,j}\}$. To move a collection of $\{(\rho, \theta)\}$ centered on the origin to $\mathbf{p}_{\mathrm{e}}$ and $\mathbf{p}_{\mathrm{r}}$ in 2D, we rotate the points to match the true ellipse's orientation (adjusting θ to be some θ_{rot}), and then translate the points (yielding a final ρ_{COTANS} and θ_{COTANS}).

It is important to note that the COTANS image curves need not be continuous between 0° and 360°. We observe such continuous curves in the vast majority of the experiments that we performed in simulation and in real-life experiments. However, this is mainly a result of our choice of the origin as located roughly in the center of

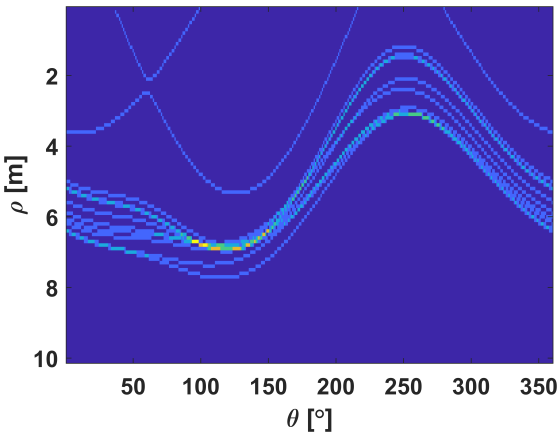


Figure 4-6: Example of a COTANS image where the NLOS curves are not necessarily continuous across the full range of azimuths.

the experimental setup, with the emitter on one side of the origin and the cluster of receivers positioned on the other side, and the reflective boundaries located relatively close by. In this case, the transformation steps in Fig. 3-5 typically yield a NLOS ellipse that still contains the origin within its interior. Hence, the set of potential tangent planes to this ellipse contain all azimuths in the interval [0°, 360°]. However, the ellipse need not contain the origin; this can happen due to global errors in our particular scenarios, but can happen in general for non-erroneous TOAs. In such cases, not all azimuths will be represented at the end of the COTANS transform, as exemplified in Fig. 4-6.

Discontinuous COTANS curves such as those observed in Fig. 4-6 are not an issue for our COTANS-NN method; these are a natural feature of the COTANS transform. However, they do rule out some algorithms that we considered for potentially enhanced performance, which traced the curves and exploited their continuity, or that naively included a for-loop over each value of θ. These discontinuities are also important to take into consideration when implementing a function that zooms into the vicinity of a COTANS image sector in ρ and θ via lookup tables, as we detail later.

When time-delay estimation errors are present, the COTANS curves do not exactly intersect at the correct boundary locations, as seen in Fig. 3-6(a). In the literature, a heuristic, hand-crafted smoothing filter is typically used for local averaging of the image [34], followed by the selection of as many maxima as there are boundaries as in

Fig. 3-6(b), where the neighborhood of each maximum is set to zero to avoid picking the same boundary multiple times. Note that if we were given constraints such as the environment being convex, we can rigorously select only the single-reflections as COTANS inputs [61]. This smoothing methodology can introduce estimation errors since it distorts the original COTANS image, and it only uses the information in a small part of the image for estimation rather than exploiting other potential patterns in the full image. Furthermore, the smoothing filter's dimensions and kernel are heuristically fine-tuned to specific environments and to specific COTANS image resolutions, making this smoothing process difficult to generalize. We therefore introduce a NN method for multi-scale filtering and peak extraction from COTANS images, that is not constrained by the above limitations.

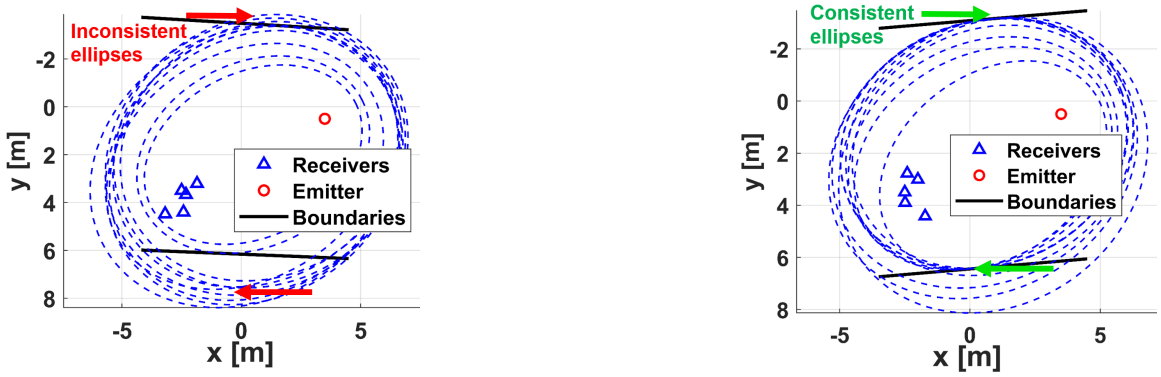
4.3.2 Detailed Explanation of the COTANS-NN Method

COTANS-NN uses CNNs for multi-output regression from COTANS images. We re-purpose the 8-layer and 2-GPU AlexNet architecture [110] by replacing the final classification layer with a regression layer, with MSE used as the cost function. Our approach is guided by previous similar approaches that re-purpose AlexNet for regression [118]. To work with color images, AlexNet has 3 channels; however, the COTANS images are like grayscale images in that they only have a single value for each pixel scaled to be within [0,1], so we modify AlexNet to only have 1 channel. Our network inputs are COTANS images, and outputs are the boundary parameter estimates $[\hat{\rho}_1 \cdots \hat{\rho}_N \ \hat{\theta}_1 \cdots \hat{\theta}_N]^{\mathrm{T}}$. We use the correct parameters $[\rho_1 \cdots \rho_N \, \theta_1 \cdots \theta_N]^{\mathrm{T}}$ as the targets for network training. Thus, our output layer size is $2N$, and the NN implements a function that projects COTANS images onto this $2N$-dimensional space.

We use AlexNet as a building block for COTANS-NN because it does not incorporate any specific features for image classification, as we have an image regression task instead. Thus, we replace the final classification layer with a regression layer and modify the input image dimensions, thereby not extensively re-designing the NN. Our NN training hyperparameters are given in Table 4.2; note that L2 regularization is set to zero because the COTANS transform images themselves are not corrupted by

Table 4.2: Training specifications for COTANS-NN.

Parameter	Value
Optimizer	Adam
Number of epochs	25
Mini batch size	50
L2 regularization	0
Initial learn rate	0.001



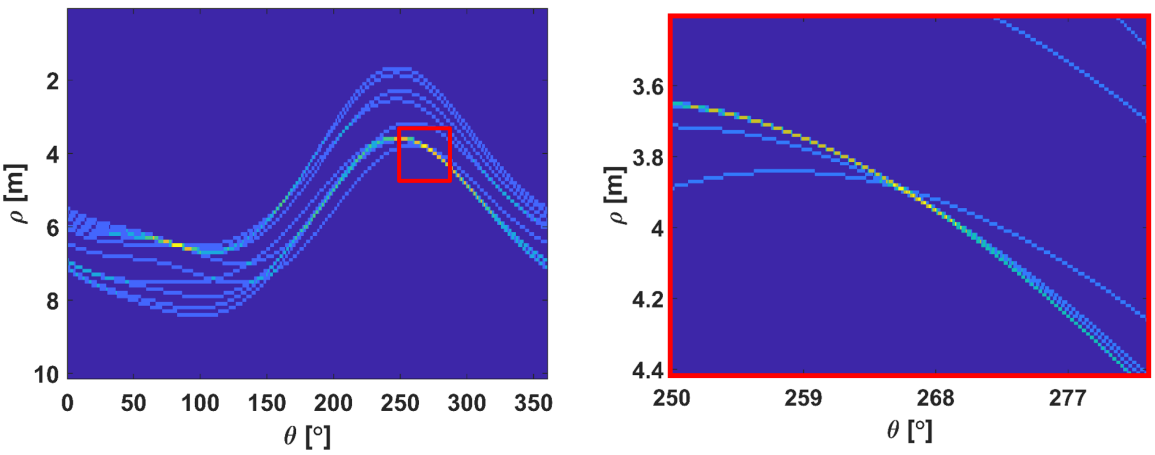
(a) (b)

Figure 4-7: Random geometries and the corresponding NLOS ellipses: the transition region at 8 dB SNR (a); and the high SNR region at 20 dB SNR (b).

any noise, though the underlying acoustic data may be.

To use this architecture on COTANS images, we generate training image datasets by simulating scenarios with randomized $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$, and randomized boundary positions, as in Fig. 4-7. We train COTANS-NN on 14 different SNRs in the 10 to 30 dB SNR range, fully covering the transition region of global errors. We generate 50,000 training and 3,000 validation images for each SNR. Since COTANS curves for all receivers are all summed up into a single image, COTANS-NN does not need modification to handle variable numbers of receivers.

Limited NN training times and computational memory lead to practical constraints on COTANS image dimensions, necessitating a trade-off between complexity and bin resolutions. Thus, in order to plot the complete COTANS curves, we adopt input dimensions of 101 pixels in ρ and 360 pixels in θ. The θ-resolution is therefore 1°, while we use a ρ-resolution of 0.1 m, leading to a ρ-axis of 0 m to 10 m. This range of ρ is appropriate for our simulation and real experiment settings, but can be



(a) (b)

Figure 4-8: A coarse-resolution COTANS image, with the region highlighted in red centered on one of the Coarse-NN boundary estimates (a); and the resulting zoomed-in image in stage 2 (b).

scaled or translated to a different interval for different applications, while keeping the same image dimensions.

In order to surpass the performance limitations imposed by the resolution constraints on COTANS images, we design successive stages of COTANS-NN with finer resolutions for refined performance. This leads to the multi-stage, multi-resolution COTANS-NN method, as summarized in Fig. 4-2. A single NN, termed Coarse-NN, forms the first stage which is trained on coarse-resolution, complete COTANS images. While this NN is a good overall first-pass estimator, its performance saturates at high SNRs where the limited resolution can constrain performance. To overcome this limitation, we zoom into the vicinities of Coarse-NN boundary estimates on the full images, and perform further stages of estimation on these high-resolution images. At each stage of zooming, we increase the ρ- and θ-resolutions by a factor of ten, so that the second stage in our particular implementation yields images with a resolution of (0.01 m, 0.1°) in (ρ, θ). The image dimensions are retained to be the same at 101×360 pixels, so that the pre-training procedures used for the first stage can be used for successive stages as well. This zooming approach is illustrated in Fig. 4-8, where we have highlighted the vicinity of one of the two boundary estimates from Coarse-NN.

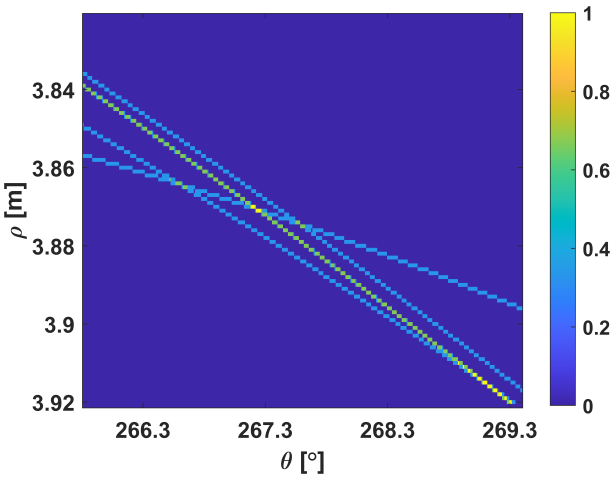


Figure 4-9: Example of a second-stage zoomed COTANS image, where finding the center of mass of the crossing lines is preferred to training a new NN.

To train a NN for fine-resolution images, we first generate a new set of 25,000 training and 1,500 validation coarse-resolution COTANS images for each SNR. We then zoom into the 1 m $\times$ 3.6° image region around the Coarse-NN estimates of each of the two boundaries, and generate two images within these regions, to obtain a zoomed dataset of 50,000 training and 3,000 validation images. Training the same NN architecture with this dataset yields a new NN which we call Fine-NN. Using Coarse-NN and Fine-NN in sequence leads to better performance.

After using Fine-NN for estimation, further stages of zooming have diminishing performance returns. These stages yield images featuring crisscrossing lines as in Fig. 4-9, rather than intersecting curves. Therefore, we employ a basic weighted averaging procedure for interpolating the estimates for these stages, instead of training any new stages of NNs. We threshold the image pixels to only retain the ones with values ≥ 0.5, indicating the intersections of two or more lines. Recall that the COTANS images are scaled so that the maximum intensity pixel, corresponding to the maximum number of crossing curves, has a value of 1. Weighting every remaining pixel by its value and finding the average pixel coordinates yields a refined boundary estimate. This stage's performance gain only becomes relevant at high SNRs.

COTANS-NN performs better when the image inputs to Coarse-NN are the echo-labeled curves from a single boundary, rather than the sum of unlabeled curves from

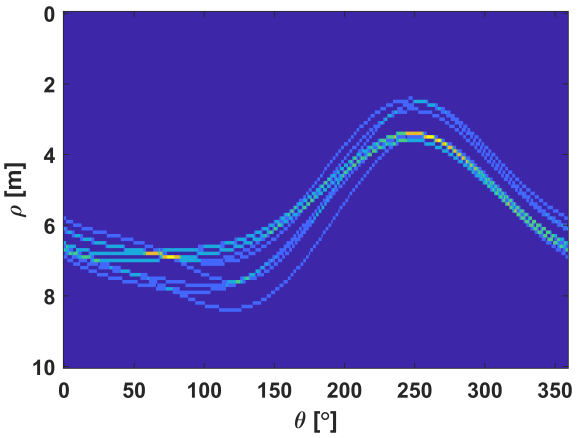


Figure 4-10: A COTANS image with curves from all the unlabeled NLOS arrivals, to be used as an input to Coarse-NN.

all the boundaries. Such single-boundary images can have less-distorted maxima, and the accuracy of successive stages of refinement is contingent on this Coarse-NN performance. However, we do not attempt to solve the complete, combinatorial echo labeling problem. Instead, we first carry out boundary estimation with Coarse-NN, then use the resulting estimates to estimate the correct assignment of echoes. Each boundary estimate corresponds to a set of NLOS TOAs $\tilde{\tau}_{i,j'}$ to each receiver, which differ from the NLOS TOAs $\hat{\tau}_{i,j}$ that were obtained by time-delay estimation. At each receiver i, we make the echo assignment of j to j'' that minimizes $\sum_{j=1}^{N} \|\tilde{\tau}_{i,j'} - \hat{\tau}_{i,j}\|$. We then generate separate new COTANS images for each j'' from the sorted time-delays $\hat{\tau}_{i,j''}$, and use Coarse-NN for estimation. An example of this procedure is given in the COTANS image of Fig. 4-10, which is used to generate the two COTANS images in Fig. 4-11, which correspond to NLOS curves assigned to different boundaries. Note that this procedure does not lead to a completely correct labeling when global errors have been made; but the performance of Coarse-NN is strong enough that the labeling of echoes with only local errors is generally accurate. The correct assignment of most of the echoes to separate images is found to be sufficient to yield better performance.

In implementation, the zooming operations that we have outlined do not involve simply generating higher-resolution complete COTANS-NN images and taking specific regions of these images into consideration. For a large dataset, the memory requirements of progressively higher-resolution images become prohibitive. Instead, we main-

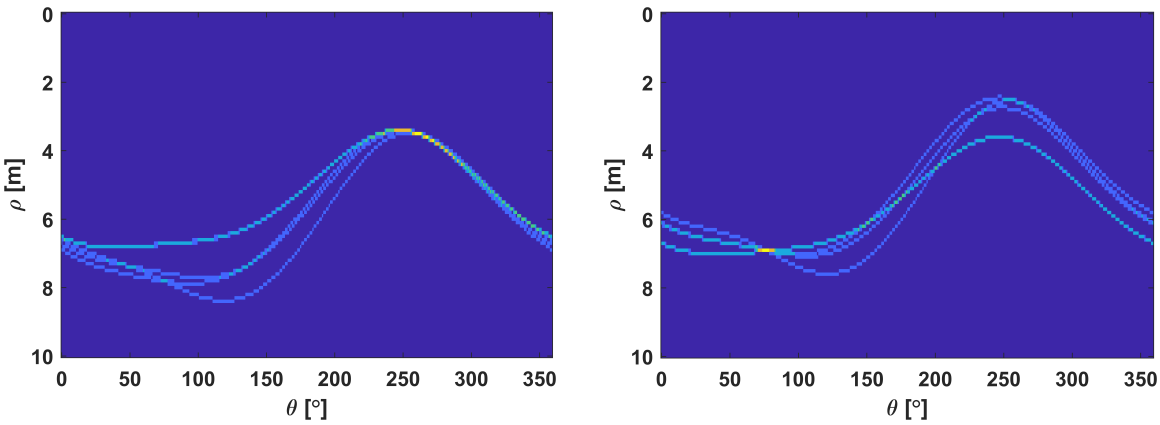


Figure 4-11: The NLOS arrivals in the COTANS image of Fig. 4-10 are partitioned into those estimated to come from boundary 1 (a), and boundary 2 (b). The maxima in the original image are distorted from the superposition of many NLOS curves, while the cleaner images belonging to the different boundaries serve better for COTANS-NN estimation.

tain a lookup table of the coarse-resolution azimuth transformations $\theta \to \theta_{\text{COTANS}}$. We generate the higher-resolution images by running the COTANS transformation only on the relevant θ-interval that yields the desired θ_{COTANS}-interval.

4.4 Existing Boundary Estimation Methods

In this section, we summarize the LS [16] and EDM [17] algorithms, which are the state-of-the-art alternatives to COTANS-NN. LS and EDM are used to localize emitters using LOS arrivals; they are similarly used for boundary estimation by localizing virtual emitters using NLOS arrivals. While COTANS-NN is currently limited to 2D, LS and EDM have the advantage of being 3D estimation methods; but we will observe that they assume a small-scale error regime, and require solving the echo labeling problem that COTANS-NN bypasses.

4.4.1 Least-Squares Solution for Boundary Estimation

In LS localization, the objective is to solve for the variables x_{e} and y_{e} that define the emitter's location $\mathbf{p}_{\text{e}}$ in Cartesian coordinates. Assuming a high-SNR regime, we

have the Gaussian noise-corrupted range estimates $\{r_i\}_{i=1}^{M}$ from the emitter to each receiver as:

$$r_i = d_i + \xi_i = \sqrt{(x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2} + \xi_i, \quad \xi_i \sim \mathcal{N}\left(0,\, \sigma_r^2\right). \tag{4.12}$$

Introducing the variable $r_\mathrm{e} \triangleq \sqrt{x_\mathrm{e}^2 + y_\mathrm{e}^2}$ for the emitter's range to the origin, the resulting set of linear equations to be solved is given by:

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & -0.5 \\ \vdots & \vdots & \vdots \\ x_M & y_M & -0.5 \end{bmatrix},\ \mathbf{x} = \begin{bmatrix} x_\mathrm{e} \\ y_\mathrm{e} \\ r_\mathrm{e}^2 \end{bmatrix},\ \mathbf{b} = \frac{1}{2}\begin{bmatrix} x_1^2 + y_1^2 - r_1^2 \\ \vdots \\ x_M^2 + y_M^2 - r_M^2 \end{bmatrix}. \tag{4.13}$$

In the presence of additive white Gaussian noise, the LS solution for Eq. (4.13) is given by:

$$\hat{\mathbf{x}}_\mathrm{e} = \arg\min_{\tilde{\mathbf{x}}} \left(\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\right)^\mathrm{T} \left(\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\right), \tag{4.14}$$

where $\tilde{\mathbf{x}}_\mathrm{e} = \left[\tilde{x}_\mathrm{e}\ \tilde{y}_\mathrm{e}\ \tilde{r}_\mathrm{e}^2\right]^\mathrm{T}$ is the optimization variable vector. Note that in the presence of range estimate errors (i.e., time delay estimation errors), it is critical to also introduce the nonlinear constraint:

$$\tilde{x}_\mathrm{e}^2 + \tilde{y}_\mathrm{e}^2 - \tilde{r}_\mathrm{e}^2 = 0, \tag{4.15}$$

and solving Eq. (4.14) constrained by Eq. (4.15) yields the LS algorithm for localization.

To localize virtual emitters rather than the real emitter, we only need to re-conceptualize our optimization variable vector as $\tilde{\mathbf{x}}_{\mathrm{v},j} = \left[\tilde{x}_{\mathrm{v},j}\ \tilde{y}_{\mathrm{v},j}\ \tilde{r}_{\mathrm{v},j}^2\right]^\mathrm{T}$ for each boundary $j = 1, \ldots N$, and to use the NLOS range estimates $r_{i,j}$. It is therefore implicit in this formulation that the echo labeling problem has been solved correctly. Going beyond the literature, inequality constraints on $\tilde{x}_\mathrm{e}$ or $\tilde{y}_\mathrm{e}$ can also be added, thus confining LS solutions to a region in which the emitter is known to be present, and thereby exploiting prior information on the environment, and in particular on

the area of interest.

4.4.2 Boundary Estimation with Euclidian Distance Matrices

EDM was designed with the insight that when matrices of squared distances between nodes (receivers and virtual emitters) are given, the matrix with the correct permutation of entries (i.e., with proper echo labeling) will have the lowest rank. Thus, if localization is carried out on each combination of echoes, we would simultaneously obtain the correct echo labeling and the correct virtual emitters [17]. In the presence of noise, however, it is difficult to use this (sensitive) rank criterion, and a heuristic metric and optimization method is applied instead.

EDM uses the 's-stress criterion', where given the measured $\{r_{i,j}\}$, the objective function to be minimized over $\tilde{x}_{\mathrm{v},j}$ and $\tilde{y}_{\mathrm{v},j}$ is:

$$s\left(\tilde{x}_{\mathrm{v},j},\, \tilde{y}_{\mathrm{v},j}\right) = \sum_{i} \left[\left(\tilde{x}_{\mathrm{v},j} - x_i\right)^2 + \left(\tilde{y}_{\mathrm{v},j} - y_i\right)^2 - r_{i,j}^2\right]^2 . \tag{4.16}$$

Ideally, Eq. (4.16) is minimized by the correct echo labeling. In our simulations, however, we had difficulty in using the s-stress criterion for noisy echo labeling, with the correct set of echoes not necessarily having the smallest s-stress. Therefore, in our implementation of EDM, we chose to use the correctly echo-sorted results for boundary estimation, as with LS. Thus, we give these methods an inherent advantage here, of access to knowledge that needs to be inferred from the data in reality.

4.5 CRLB for Boundary Range Estimation

We now derive a theoretical benchmark for the performances of boundary estimation algorithms given noisy range measurements as in Eq. (4.12), which enables us to verify their correct implementation, as well as assess the fundamental limitation of the asymptotic accuracy. Our starting point is the CRLB for emitter localization performance [16], which we modify to obtain the CRLB for boundary estimation.

The CRLB for estimating $\mathbf{p}_\mathrm{e} = [x_\mathrm{e}\, y_\mathrm{e}]^\mathrm{T}$ yields the Fisher information matrix:

$$\mathbf{I}(\mathbf{p}_\mathrm{e}) = \begin{bmatrix} \mathrm{I}_{1,1} & \mathrm{I}_{1,2} \\ \mathrm{I}_{1,2} & \mathrm{I}_{2,2} \end{bmatrix}, \tag{4.17}$$

where:

$$\mathrm{I}_{1,1} = \sum_{i=1}^{M} \frac{(x_\mathrm{e} - x_i)^2}{\sigma^2 \left((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2\right)}, \tag{4.18}$$

$$\mathrm{I}_{1,2} = \sum_{i=1}^{M} \frac{(x_\mathrm{e} - x_i)(y_\mathrm{e} - y_i)}{\sigma^2 \left((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2\right)}, \tag{4.19}$$

$$\mathrm{I}_{2,2} = \sum_{i=1}^{M} \frac{(y_\mathrm{e} - y_i)^2}{\sigma^2 \left((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2\right)}. \tag{4.20}$$

Inverting Eq. (4.17) yields a matrix with terms $\mathrm{J}_{1,1}$, $\mathrm{J}_{1,2}$, and $\mathrm{J}_{2,2}$. The CRLBs for the coordinates of $\mathbf{p}_\mathrm{e}$ are:

$$\mathrm{CRLB}\,(\hat{x}_\mathrm{e}) = \mathrm{J}_{1,1} = \frac{\mathrm{I}_{2,2}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}, \tag{4.21}$$

$$\mathrm{CRLB}\,(\hat{y}_\mathrm{e}) = \mathrm{J}_{2,2} = \frac{\mathrm{I}_{1,1}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}. \tag{4.22}$$

We also have the cross-term:

$$\mathrm{J}_{1,2} = -\frac{\mathrm{I}_{1,2}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}. \tag{4.23}$$

The CRLB on the range estimation error is then $\mathrm{CRLB}\,(\hat{x}_\mathrm{e}) + \mathrm{CRLB}\,(\hat{y}_\mathrm{e})$ [119]. For virtual emitters, $\mathbf{p}_\mathrm{v} = [x_\mathrm{v}\, y_\mathrm{v}]^\mathrm{T}$ can be substituted into Eqs. (4.21) and (4.22).

We specify a given boundary by its ρ and θ, and use the RMSE range error as the measure of performance of boundary estimation. Thus, we transform the CRLB for $\hat{\mathbf{p}}_\mathrm{v}$ into a CRLB for boundary range estimation.

Given $\mathbf{p}_\mathrm{v}$, we obtain the orthogonal vector $\boldsymbol{\rho}_\mathrm{v}$ to the boundary, referring to Fig. 4-12. First, note that $\mathbf{n}_\mathrm{v}$ pointing from the emitter to the virtual emitter has the same direction as the unit normal $\hat{\mathbf{n}}_\mathrm{v}$ from the origin to the boundary:

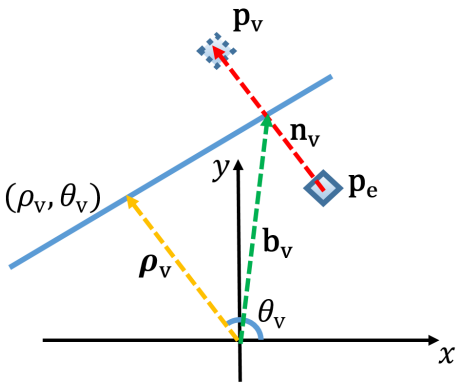


Figure 4-12: A geometric reference for the transformation of the CRLB for virtual emitter positions into the CRLB for boundary range estimation.

$$\hat{\mathbf{n}}_\mathrm{v} = \begin{bmatrix} \frac{x_\mathrm{v}-x_\mathrm{e}}{\sqrt{(x_\mathrm{v}-x_\mathrm{e})^2+(y_\mathrm{v}-y_\mathrm{e})^2}} \\ \frac{y_\mathrm{v}-y_\mathrm{e}}{\sqrt{(x_\mathrm{v}-x_\mathrm{e})^2+(y_\mathrm{v}-y_\mathrm{e})^2}} \end{bmatrix}. \tag{4.24}$$

The vector from the origin, to the intersection point of the boundary and $\mathbf{n}_\mathrm{v}$ is given by:

$$\mathbf{b}_\mathrm{v} = \begin{bmatrix} \frac{x_\mathrm{v}+x_\mathrm{e}}{2} & \frac{y_\mathrm{v}+y_\mathrm{e}}{2} \end{bmatrix}^\mathrm{T}. \tag{4.25}$$

The orthogonal vector from the origin to the boundary, is then given by the orthogonal projection of $\mathbf{b}_\mathrm{v}$ onto $\hat{\mathbf{n}}_\mathrm{v}$, i.e., by $\left(\mathbf{b}_\mathrm{v}^\mathrm{T}\hat{\mathbf{n}}_\mathrm{v}\right)\hat{\mathbf{n}}_\mathrm{v}$. We obtain this resulting vector as:

$$\boldsymbol{\rho}_\mathrm{v} = \begin{bmatrix} \frac{(x_\mathrm{v}-x_\mathrm{e})\left(x_\mathrm{v}^2-x_\mathrm{e}^2+y_\mathrm{v}^2-y_\mathrm{e}^2\right)}{2\left((x_\mathrm{v}-x_\mathrm{e})^2+(y_\mathrm{v}-y_\mathrm{e})^2\right)} \\ \frac{(y_\mathrm{v}-y_\mathrm{e})\left(x_\mathrm{v}^2-x_\mathrm{e}^2+y_\mathrm{v}^2-y_\mathrm{e}^2\right)}{2\left((x_\mathrm{v}-x_\mathrm{e})^2+(y_\mathrm{v}-y_\mathrm{e})^2\right)} \end{bmatrix}. \tag{4.26}$$

Finally, the range to the boundary is given by the magnitude of $\boldsymbol{\rho}_\mathrm{v}$, yielding the resulting scalar as:

$$\rho_\mathrm{v} = \frac{|x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2|}{2\sqrt{(x_\mathrm{v}-x_\mathrm{e})^2 + (y_\mathrm{v}-y_\mathrm{e})^2}}. \tag{4.27}$$

This set of operations is the same as those used to transform the LS and EDM virtual emitter estimates into boundary estimates. Without loss of generality, we will assume that $x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2 > 0$ for our subsequent derivations.

We have now obtained the range as a function of x_v and y_v, with x_e and y_e being known constants. The CRLBs for x_v and y_v can now be transformed into a CRLB

for ρ_{v}. We calculate the derivatives of ρ_{v} with respect to x_{v} and y_{v} as:

$$\frac{\partial \rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}} = \frac{2x_{\mathrm{v}}\left((x_{\mathrm{v}}-x_{\mathrm{e}})^2+(y_{\mathrm{v}}-y_{\mathrm{e}})^2\right)-(x_{\mathrm{v}}-x_{\mathrm{e}})\left(x_{\mathrm{v}}^2-x_{\mathrm{e}}^2+y_{\mathrm{v}}^2-y_{\mathrm{e}}^2\right)}{2\left((x_{\mathrm{v}}-x_{\mathrm{e}})^2+(y_{\mathrm{v}}-y_{\mathrm{e}})^2\right)^{3/2}}, \tag{4.28}$$

$$\frac{\partial \rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}} = \frac{2y_{\mathrm{v}}\left((x_{\mathrm{v}}-x_{\mathrm{e}})^2+(y_{\mathrm{v}}-y_{\mathrm{e}})^2\right)-(y_{\mathrm{v}}-y_{\mathrm{e}})\left(x_{\mathrm{v}}^2-x_{\mathrm{e}}^2+y_{\mathrm{v}}^2-y_{\mathrm{e}}^2\right)}{2\left((x_{\mathrm{v}}-x_{\mathrm{e}})^2+(y_{\mathrm{v}}-y_{\mathrm{e}})^2\right)^{3/2}}. \tag{4.29}$$

The resulting CRLB for ρ_{v} is obtained by the vector transformation of parameters as:

$$\begin{aligned}
\mathrm{CRLB}(\hat{\rho}_{\mathrm{v}}) &= \begin{bmatrix}\frac{\partial \rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}} & \frac{\partial \rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}}\end{bmatrix}\begin{bmatrix}\mathrm{I}_{1,1}^{-1} & \mathrm{I}_{1,2}^{-1}\\ \mathrm{I}_{1,2}^{-1} & \mathrm{I}_{2,2}^{-1}\end{bmatrix}\begin{bmatrix}\frac{\partial \rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}}\\ \frac{\partial \rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}}\end{bmatrix} && (4.30)\\
&= \mathrm{I}_{1,1}^{-1}\left(\frac{\partial \rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}}\right)^2+2\mathrm{I}_{1,2}^{-1}\frac{\partial \rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}}\frac{\partial \rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}}+\mathrm{I}_{2,2}^{-1}\left(\frac{\partial \rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}}\right)^2. && (4.31)
\end{aligned}$$

Hence, $\sqrt{\mathrm{CRLB}(\hat{\rho}_{\mathrm{v}})}$ is the lower bound on the RMSE boundary range estimation error.

4.6 Simulation and Experimental Results

We now study the performance of COTANS-NN as was presented in Section 4.3.2, and compare it to its alternatives which were presented in Section 4.4. We evaluate COTANS-NN in simulation, comparing its performance to LS and EDM, and also to the CRLB derived in Section 4.5. After re-training COTANS-NN, we apply it to a real-life underwater acoustic setting, where it outperforms LS. Finally, we conduct simulations that demonstrate the robustness of COTANS-NN to model mismatch and to reduced prior knowledge of the environment.

4.6.1 Simulated Performances

We test COTANS-NN on $K = 50{,}000$ COTANS images per SNR value. One boundary has its ρ and θ parameters uniformly drawn from the intervals [3, 3.5] m and [260, 280]°, respectively, while the other has parameters in [6, 6.5] m and [80, 100]°.

These boundaries model a sea surface and a shallow seafloor as in Fig. 4-7 The variations in range/angles of the boundaries could arise from either surface wave motion in the case of the sea surface, or bathymetric variations in the case of the seafloor. The $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$ are uniformly drawn from the vicinities of two fixed points, (3.5, 0.5) and $(-2.5, 3.5)$, respectively. While we have conducted simulation experiments with three boundaries as well, the resulting performance curves are qualitatively similar to the two-boundary case, only differing in the performances at specific SNRs. Hence, we only present the two-boundary results as being representative of the general performance of COTANS-NN.

We first compare COTANS-NN to an ideal LS implementation, which is initialized at the ground truth locations of the $\{\mathbf{p}_{\mathrm{v},j}\}$, with correct echo labeling, and with virtual emitter solutions constrained to lie within the same parameter space that COTANS-NN is trained on. Our performance metric is the range RMSE (in m) over all N reflective boundaries and all K environment realizations for each SNR S, defined as:

$$\rho_{\mathrm{RMSE}}(S) \triangleq \sqrt{\frac{\sum_{j=1}^{N}\sum_{k=1}^{K}\left(\rho_{j,k}^{(S)} - \hat{\rho}_{j,k}^{(S)}\right)^2}{NK}}. \tag{4.32}$$

Fig. 4-13(a) demonstrates that COTANS-NN and LS performances are nearly identical for SNR greater than 23 dB, a high-SNR operating regime where global errors are rare and the $\{\hat{\tau}_{i,j}\}$ are accurate due to small noise. Below 23 dB SNR, as global errors become increasingly common, COTANS-NN outperforms LS by up to 6 dB SNR. This performance advantage narrows at low SNRs, where accurate boundary estimation becomes infeasible using either method.

In Fig. 4-13(a), it appears that COTANS-NN merely outperforms LS at SNR less than 23 dB and has equivalent performance otherwise. In fact, in this plot, LS sometimes suffers failures which are constrained to lie within a relatively narrow parameter space, whenever global errors occur. We conduct the same experiment with unconstrained LS solutions, and also apply EDM to obtain Fig. 4-13(b). COTANS-NN outperforms LS and EDM by up to 9 dB SNR, and also marginally outperforms them in the high-SNR regime as well. LS and EDM have similar performances,

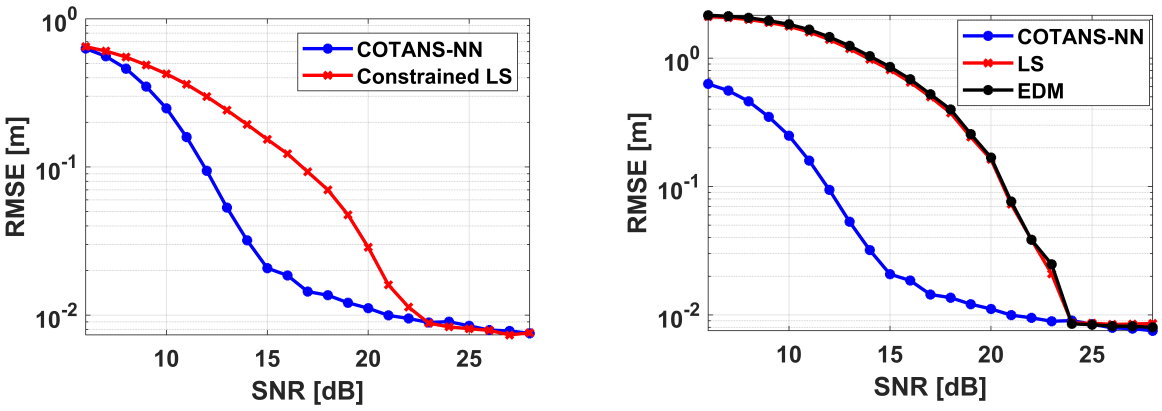


Figure 4-13: COTANS-NN performance comparison to constrained LS (a), and to LS and EDM (b).

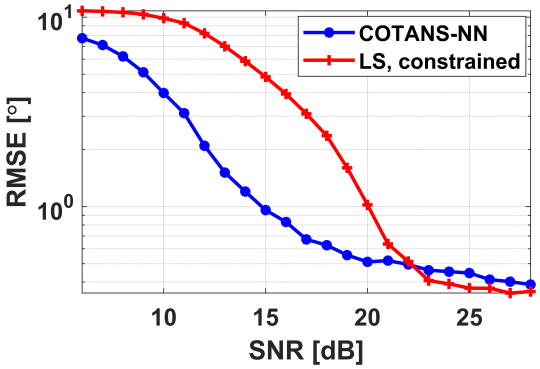


Figure 4-14: Azimuth estimation performance of COTANS-NN performance, as compared to constrained LS.

which arises from how they both minimize the squared error between measured and estimated distances.

We present in Fig. 4-14 the azimuth estimates from the simulation results corresponding to the scenario in Fig. 4-13(a). While the range estimation performances were very close to each other in the high-SNR regime, the azimuth estimation performances demonstrate a 2 dB advantage for LS over COTANS-NN. The absolute performance advantage that this difference corresponds to is very small, however, with azimuth estimation RMSE of 0.39°for LS as compared to 0.45°for COTANS-NN at 24 dB SNR, for example. In the intermediate SNR regime, COTANS-NN outperforms LS by up to 6 dB SNR as was the case for the range estimates, indicating that these estimation accuracies are consistent.

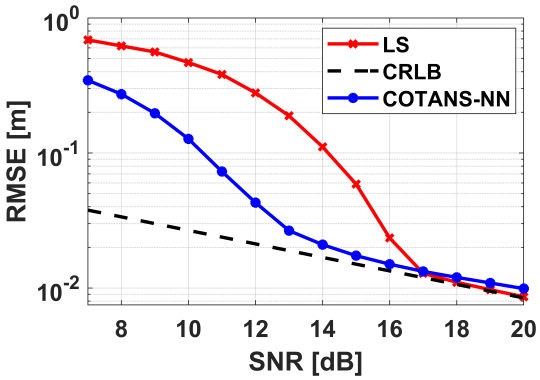


Figure 4-15: The CRLB for single boundary range estimation, calculated for a fixed scenario, and compared against the COTANS-NN and LS performances for the same scenario.

Finally, we conduct simulations with different noise realizations and a fixed environment, allowing us to compare against the CRLB in Section 4.5 for a single boundary present in the environment. Since LS localization performance approaches the CRLB for emitter localization at high SNR, LS virtual emitter localization should similarly approach the boundary range estimation CRLB. We observe that this is the case in Fig. 4-15, confirming that the CRLB has been formulated correctly, and is a benchmark for performance in the high-SNR regime as intended.

It is important to observe how the different stages of COTANS-NN contribute to overall performance. In Fig. 4-16, we compare the performances of the Coarse-NN, Fine-NN, and interpolation stages as previously outlined in Fig. 4-2. The performance of Coarse-NN saturates at high SNR due to limited image resolutions; the subsequent stage of Fine-NN improves this performance and avoids the saturation phenomenon in this selected SNR range. The large performance improvement from Fine-NN, even at intermediate SNRs, demonstrates that it is advantageous to always carry out this refinement step. For this particular simulation, the interpolation stage does provide a noticeable improvement in performance at high SNRs, albeit the gains of this subsequent stage are limited with diminishing returns. Hence, whether to carry out further refinements of COTANS-NN results after Fine-NN depends on particular application scenarios and requirements.

Another key stage to analyze is the grouping of curves into separate images and

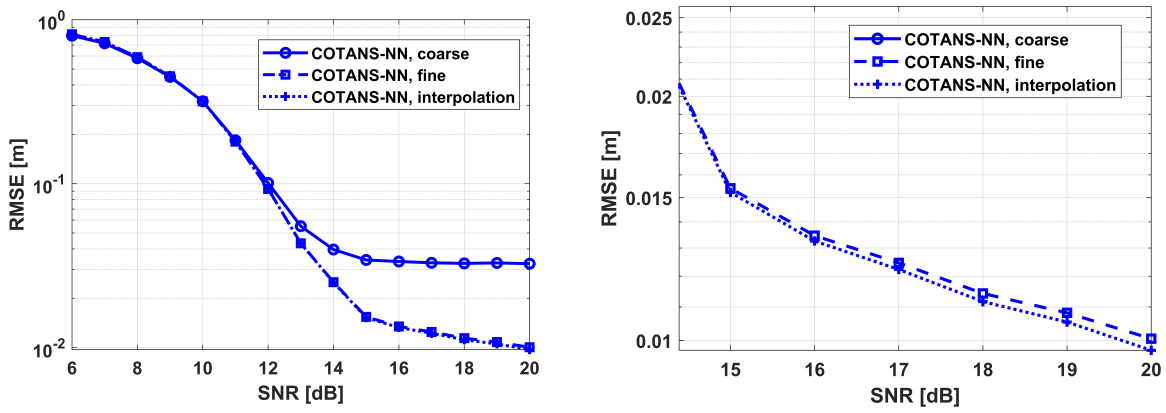


Figure 4-16: The performances of different stages of COTANS-NN (a), and a zoomed-in comparison of the fine NN performance and the subsequent interpolation refinement (b).

the application of Coarse-NN to each image, based on an initial first-pass estimate using Coarse-NN. In Fig. 4-17, we observe a large performance improvement of up to 4 dB SNR. A noticeable improvement takes place even in low- and medium-SNR regimes, where the rough echo labeling carried out here is not completely accurate due to global errors being present. Thus, the approximate grouping operation that we implement is robust in delivering performance gains. Not only is this result a major improvement in its own right over the first-pass application of Coarse-NN, but this enhanced performance is essential for Fine-NN and the subsequent interpolation stage to deliver large improvements as well, as opposed to these stages only having a marginal effect. The improved performance carries over into the performance of each of these sequential stages.

4.6.2 Underwater Acoustic Experiment Results

To verify that COTANS-NN performs well under realistic conditions, we perform experiments in a controlled underwater acoustic setting. We use the Scripps Ocean-Atmosphere Research Simulator (SOARS) wave tank facility in the Scripps Institution of Oceanography (Fig. 4-18(a)), which was previously discussed in Section 2.2.3, with the top-view of this particular experiment geometry set up as in Fig. 4-18(b). The

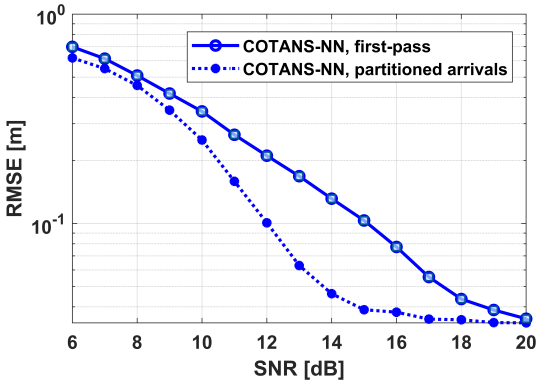


Figure 4-17: The first-pass Coarse-NN performance, compared to the average performance obtained by applying Coarse-NN to separate images in the second-pass.

hydrophones are suspended at the same depth so that we have a 2D estimation problem for the side walls, located at y = -1.235 m and y = 1.235 m.

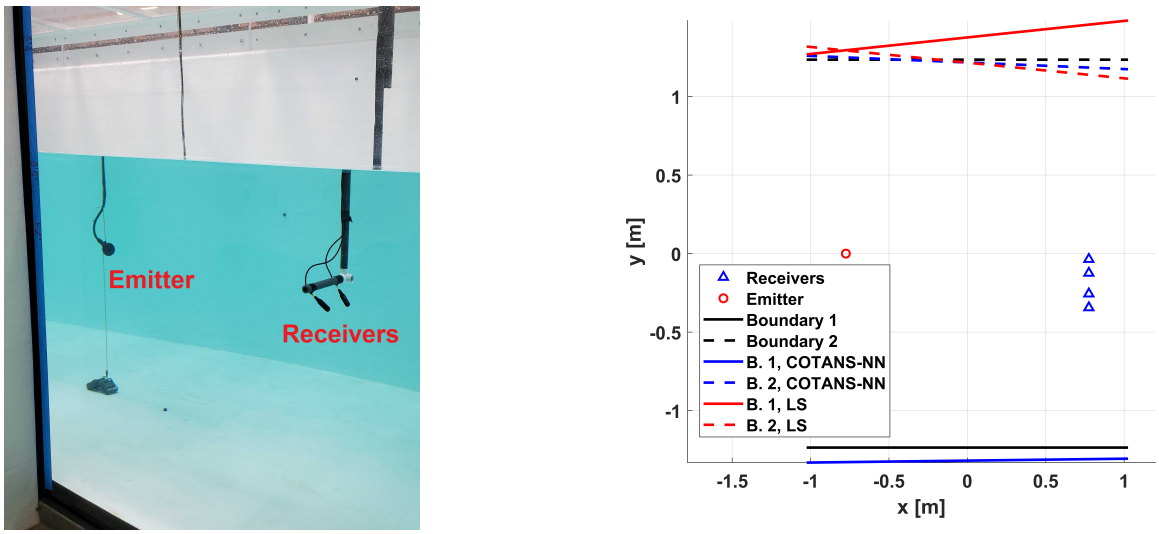
We re-train the COTANS-NN networks that were used to obtain the results in Fig. 4-13(b) on a dataset that is similar to the geometric scenario of SOARS. We then use the COTANS image generated from the SOARS experiment to estimate the boundaries. The results with COTANS-NN and with LS are given in Table 4.3. COTANS-NN achieves an accuracy on the order of centimeters in ρ and a few degrees in θ. LS suffers a large error for one boundary, and is consistently outperformed by COTANS-NN.

Table 4.3: SOARS estimation error magnitudes for boundary parameters, with drastic errors highlighted in red.

Parameter	COTANS-NN	LS
ρ_1	0.083 m	0.134 m
ρ_2	0.019 m	0.025 m
θ_1	0.7°	174°
θ_2	2.4°	5.7°

4.6.3 COTANS-NN Robustness Analysis

We now present simulation results that study COTANS-NN's robustness. A common pitfall in NN design is to over-train on a particular dataset, yielding a network that is fragile to model mismatch or one that only works with a narrow parameter space.



(a) (b)

Figure 4-18: Emitter deployment in the SOARS watertank (a), and the top-view schematic of the watertank illustrating the experiment geometry and the estimation results (b).

A robust method will have a gradual performance decline under model mismatch rather than abrupt deterioration, and will remain functional for difficult estimation scenarios.

First, we explore the effect of model mismatch in the ranges of $\mathbf{p}_e$ assumed in generating the training data. We train COTANS-NN with $\mathbf{p}_e$ drawn randomly from a square 0.25 m wide, centered at (3.5, 0.5). We progressively shift the center of this square by 0.25 m in the x- and y-directions, while continuing to use the original COTANS-NN trained on source locations within the first square region. We thereby obtain the results in Fig. 4-19, where COTANS-NN continues to be a stable method, despite increasingly worse performance as model mismatch creates unaccounted-for estimation biases.

In a different experiment, we relax the bounds on θ that the $\boldsymbol{\eta}_j$ can have, so that COTANS-NN handles a larger parameter space. We re-train COTANS-NN, originally having a $\pm 10^\circ$ θ-margin as in Sec. 4.6.1, with $\pm 20^\circ$ and $\pm 30^\circ$ θ-margins as well. A larger parameter space requires a correspondingly larger training set, but we instead use 50,000 training images per SNR as before, to assess COTANS-NN's robustness. Our results in Fig. 4-20(a) indicate that COTANS-NN remains stable, despite being

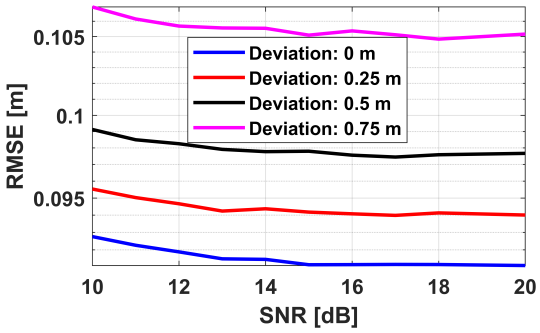


Figure 4-19: Performance of COTANS-NN as the average assumed emitter position increasingly deviates from the true one.

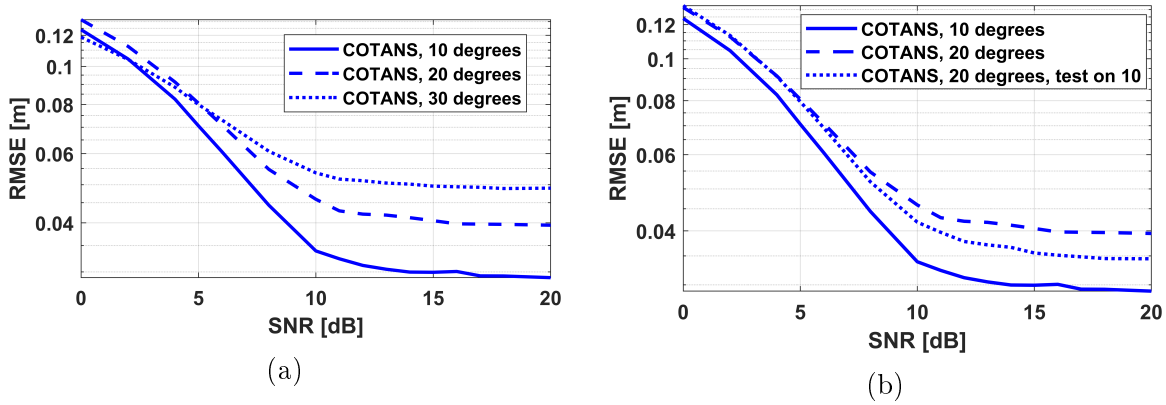


Figure 4-20: COTANS-NN performance on progressively larger θ-margins (a), and performance on the same margin after being trained on different margins (b).

trained on harder scenarios.

To analyze the deterioration due to a larger parameter space, we re-train COTANS-NN to operate on a $\pm 20^\circ$ θ-margin, then test it on the same $\pm 10^\circ$ θ-margin dataset of the original network. The resulting performances in Fig. 4-20(b) indicate that by sequentially using COTANS-NN modules on progressively smaller parameter spaces, we could achieve greater accuracy.

Chapter 5

Conclusion and Future Work

The overarching theme for this work has been the study of challenging problems in localization and environment estimation, for which no framework existed that provided the guarantee of a working computational approach. At the outset of this study, we had in our view a variety of passive localization and sensing methods that were individually well-defined, and for which theoretical guarantees of performance were available. However, we could not discover a study that approached the task of passive localization and environment learning in a general setting. Thus, although we had an intuition that various techniques such as PSO and COTANS-NN would work well enough, it was not obvious that they would yield good results on our particular data or function successfully within a single overall architecture. It was therefore a leap of faith to implement these methods on our real data, in an underwater acoustic setting that they had not been originally designed for. The successful performance of these methods in such challenging settings indicates that the tools which we have designed are applicable for a wide range of environments, which could be much less adversarial for localization tasks.

We developed a method for simultaneous localization and environment learning, which was tested in a 3D reverberant underwater environment with occluding objects. Our PEEL method performed iteratively refined localization, with each solution yielding results that served as assumed knowledge for the following more sophisticated algorithm stages. This multi-stage approach handled a complex problem

where successively establishing synchronization, boundary estimation, emitter tracking, and occluder estimation, made solving the overall task computationally feasible. Applying these stages separately is suboptimal as compared to a hypothetical joint optimization over all of the aspects of the problem. However, PEEL breaks down the localization task into sub-problems with well-posed solutions. We ultimately achieved accurate simulation and experimental results in end-to-end localization.

The proposed COTANS-NN image regression method for 2D reflective boundary estimation exploited the multi-scale filtering and domain adaptation capabilities of CNNs. Our method leveraged prior knowledge of the environment to deliver robust performance in both simulation and a real underwater acoustic setting despite model mismatch, in part by avoiding separated sub-optimal echo labeling and filtering steps that are fragile without high SNR. Simulations and a real watertank experiment demonstrated that COTANS-NN consistently delivers accurate estimates even when large errors are present in the time-delay estimates, outperforming alternative boundary estimation methods in the literature.

As a continuation of our work, the next key task which we aim to accomplish is to implement a NN approach that can perform 3D boundary estimation, now that we have demonstrated our capabilities in 2D. We cannot trivially scale our current architecture up to this task, however. COTANS-NN makes full use of the available memory and runtime resources, and going from 2D to 3D inputs would require at least an order of magnitude more computational power. While replacing the 2D convolutional layers with 3D layers is a first step, the key difficulty is that 3D data increases computational demands dramatically (the "curse of dimensionality"). Hence, it is imperative to redesign our NN architectures to deliver the same performance or better, while simultaneously reducing computational costs.

In order to scale up to 3D, we ultimately need to design a custom neural network that can solve the boundary estimation problem much more efficiently. We propose to first devise a new and efficient architecture in 2D, which should hint at how a more difficult 3D setting could be tackled. Perhaps the network could have two sub-branches, one for range and the other for azimuth estimation, which are then

combined in a set of fully-connected layers. There is a wide range of techniques that have been used to adapt NNs to specific image classification tasks. There is a smaller literature for image regression tasks, which makes this future study harder but also makes it more significant from a research perspective.

5.1 Near-Term Avenues for Research

We wish to present some of our ideas for improving PEEL and COTANS-NN in detail, as being relatively straightforward and near-term extensions of our current methods. Some of these research avenues may yield a large improvement in performance or computational efficiency, and would thereby enable the implementation of new capabilities such as robust and accurate NN-based 3D boundary estimation. Others would serve to provide important insights into the operation of our algorithms, and could thereby point to better conceptual methods for approaching the localization and boundary estimation tasks.

5.1.1 Ocean Implementation of PEEL

When deploying PEEL in an open water setting, there are improvements that must be made to cope with this more difficult environment. To use as general a model as possible, we have used the simplest versions of each algorithm, such as GCC-PHAT for TDOA, basic PSO for optimization, and a constant acceleration model for Kalman tracking. Each of these algorithms can be enhanced if more domain knowledge is available. Applying an algorithm that tracks the multipath arrivals [120], [121], for example, can help to further refine the boundary estimates. The underwater environment's bathymetric profile or the presence of occluding objects may already be known, and must then be taken into account. In a 2D slice of the ocean environment, there are performance bounds on estimation accuracy with multipath-aided localization [122]. The extension of such theoretical analysis to PEEL's 3D setting is an important future avenue of research as well.

5.1.2 Different Optimization Algorithms for PEEL

PEEL currently incorporates global optimization over the emitter position and the locations of planar boundaries. A more sophisticated environmental model would allow it to work robustly with the complex non-planar boundaries in real-life shallow-water environments. In the literature, global optimization with simulated annealing and genetic algorithms have been applied to MFP [123] for geoacoustic inversion, using known emitters. Optimization over not just the locations of their boundaries but their physical properties as well, while being more computationally expensive, could potentially lead to improved results in field deployments. There are different versions of the PSO optimization algorithm that could also be implemented, for potentially improved performance and faster (or more reliable) convergence [124], [125]. The incorporation of particle filtering for localization and tracking [126] may also be promising for PEEL's enhancement. While we have used PEEL for only a single emitter being present in the environment, other algorithms may also be incorporated to handle the case of tracking multiple sources [127].

5.1.3 Different Cost Functions for COTANS-NN

In COTANS-NN, we currently use the MSE between the real and estimated boundary ranges and azimuths (all scaled to the interval [0, 1]) as our cost function. This was initially chosen for expediency, as it was the default regression layer setting in Matlab, and was retained due to the achievement of strong performance. In practice, the mean absolute error (MAE) and mean relative error (MRE) are alternative cost functions that are also widely used in the literature [128]. We could also use a modified cost function specifically for azimuth estimation. We have worked with boundaries whose azimuths were not close to 0° or 360°; if a boundary is present near these extremes, then the cost function has to be cyclic in order to give accurate answers at either of these endpoints. These relatively simple changes could potentially lead to a large improvement in convergence and performance. If a new error function allows for faster and more reliable convergence in training, we could then reduce our training

set sizes, which is critical for the future extension of COTANS-NN to 3D settings.

5.1.4 Error Analysis for COTANS-NN

The analysis of the boundary estimation cost surface itself may yield important insights for our work. This type of analysis is best conducted for the case of only a single boundary being present in the environment, as in this case, the cost will only be a function of two parameters and can be easily visualized if the environment's geometry is fixed. For different cost functions, we could visualize how many local minima there are and how easy optimization is in general. This analysis could also rigorously guide a choice of cost function for our NNs, as compared to ad hoc experimentation with different functions.

5.1.5 Analysis of Neural Networks in COTANS-NN

The detailed analysis of the neural networks employed in COTANS-NN will be important for the future improvement of this method. There is a wide variety of tools for analyzing NNs at present, allowing us to study the individual layers in order to gain insight into what sort of function the NN actually learns over the course of training. Such an analysis could lead to an overhaul of the NN architecture, since it would provide insight into how the boundary estimation task could be approached more effectively.

5.2 Longer-Term Research Tasks for PEEL and for COTANS-NN

In the field of passive localization and environment learning, we encountered several difficult problems over the course of our work, whose solutions could greatly enhance estimation performance and provide novel capabilities. Ultimately, we wish to highlight that a data-driven learning paradigm can be used to not only approach our specific problem area, but in the broader field of sensing applications.

5.2.1 Future Work for PEEL

There are broader questions in TOA localization that pertain to PEEL. Currently, PEEL exploits the emission of deterministic, pulsed signals. In practice, other types of emissions may arise, as with communications sources or vessel signatures. It would be challenging to establish time-synchronization with such signals. Although we have experimented with performance in increasing noise, it would also be of interest to obtain a full bandwidth/noise characterization over different environments and signals.

An interesting direction would be combining our proposed methods with blind identification and channel estimation methods [129], [130]. Research has already been performed to utilize channel estimation methods in the shallow-ocean underwater acoustic domain [131], where the presence of significant surface waves is a major challenge. The multipath arrivals associated with scattering from such surface waves fluctuates rapidly over time [132], and the sea surface roughness is typically modeled in a probabilistic manner [133], [134]. Assessment of the delay-Doppler spread function is important in such cases, since the Doppler shifts associated with the time-varying surface can be approximately constant over an averaging time interval.

While we have focused on extracting the time delays from the received signal based on either a known emitted pulse, or a replica of the LOS arrival, in an application setting we may consider using a separate module for obtaining an estimate of the channel impulse response. This could provide prior knowledge for our methods, to help us to reject global errors and to improve our time delay estimation.

5.2.2 Future Work for COTANS-NN

One of the key tasks that can be investigated is that of model order selection. Currently, we have different NNs for different numbers of boundaries present in the environment. Ideally, we would have a single NN that simultaneously determines how many boundaries there are, and also where these boundaries are. How this sort of task is to be accomplished is an open question, and we propose first study how similar problems are currently tackled in NN design. Perhaps we would have a classification

network applied to the data first, whose output is the number of boundaries, followed by the sort of regression network that we have already implemented. Alternatively, we could have a regression network with a large number of outputs in its final layers (e.g., for 6 boundaries), but then have additional output parameters that assigns a likelihood score to whether the boundary is actually present. This type of network design would require a much larger dataset of environment geometry examples than at present, so we can only tackle this challenge once we succeed at greatly reducing our existing dataset sizes. While the final layer of the NN currently only provides the boundary location parameters, we could expand the network's capabilities to label the arrivals, estimate the number of boundaries present, or produce a metric of confidence in the estimation results. If the NLOS TOAs come from an underwater acoustic simulator such as Bellhop [54] or Watermark [135], the training dataset would be richer than our current signal model, potentially leading to better performance in ocean deployments.

Model order selection is an important problem for PEEL's robustness as well, and the augmentation of COTANS-NN can yield better overall performance. In cases where one is not confident about the assumption on the number N of reflective boundaries, we envisioned running several instances of PEEL in parallel with different values of N. A major discrepancy between the results of these instances would be indicative of errors. This parallelization strategy for PEEL is also important for handling adversarial scenarios that can defeat our method. The emitter may move behind an occluder and come to a halt, for example; or it may take a sharp turn which throws off the tracking algorithm. Parallelizing PEEL may help detect and handle such conditions; however, a NN-aided estimation and tracking framework could improve robustness much more reliably than such ad hoc algorithmic strategies.

The boundary estimation problem is amenable to further theoretical work. Our derivation of the CRLB for this task, highlights how this problem is sufficiently self-contained and well-defined, unlike the full joint localization and environment learning problem. The CRLB for boundary range estimates could be extended to boundary azimuth estimates as well, or could be derived for 3D settings as well as 2D. These

results would offer additional means of verify the correct operation of boundary estimation methods, and could help debug any potential implementation errors.

There are alternative ways of assembling the training data for boundary estimation, such as incorporating attenuation coefficients by scaling each NLOS curve by its magnitude in the COTANS images. It may also be possible to modify the network inputs, providing this data to the NN in other formats than COTANS images, to explore a wider range of estimation methods. COTANS-NN's demonstration of the feasibility and desirability of transform-based NN boundary estimation is encouraging for these future studies.

A significant enhancement of COTANS-NN's capabilities would be the incorporation of tracking in the presence of a moving emitter. At present, given multiple known emitter positions, COTANS-NN can estimate the boundary positions with greater accuracy, thanks to the spatial diversity provided by these different emitter placements. One could also envision having an evolving estimate of the boundaries, over different snapshots. There are methods in the literature for solving such problems, especially in SLAM, which do not require the use of NNs [136]. In the context of COTANS, as an emitter moves, the curves in the COTANS images would also shift, in a continuous way that incorporates information on the setting which is difficult to express analytically but which could be learned by a NN. Implementing a NN-based approach for tracking within COTANS-NN is an interesting open question, and one that could lead to a re-design of the overall PEEL method.

In future work, we also envision connecting physics and machine learning for an improved operational model of the ocean environment. Not all ocean variability is acoustically significant, especially considering that there are different acoustic regimes for different frequencies and environments. We may not know ahead of time which features of the ocean will be operationally useful. While there is a massive number of parameters to take into account when modeling the ocean, the number of parameters is not necessarily the same as the number of degrees of freedom of the resulting environment. For example, coupled parameters contribute to the same degree of freedom. Hence, a machine learning method that can infer the key parameters in an

automated manner can augment complex physical models, or implement them with reduced computational complexity. The integration of such environmental priors to machine learning is important for both improved performance and computational efficiency.

5.2.3 Markov Chain Monte Carlo Methods for Environment Estimation

In PEEL, we have focused on exploring passive localization and environment estimation. There is a Markov chain Monte Carlo (MCMC) method for active settings, however, which may be adaptable to our passive setting [137]. This methodology has been used for the reconstruction of unknown indoor environments from NLOS arrivals [138], and could potentially be used to work with random emitted signals and more complex environments. The 3D structure to be recovered is modeled as a point process [139], and reversible-jump MCMC (RJ-MCMC) moves are proposed to sample the posterior distribution of interest [140]. This approach is an extension of Metropolis-Hastings methods to a setting where the dimensionality is not known ahead of time, such as with an unknown number of boundaries or occluding objects in the environment. Here, we model the received data as drawn from a random process, and find a parametric best-fit to the received signals.

In lidar scene reconstruction, the spatial function to be fit to the data is a point cloud of partly-reflective voxels, or alternatively, a set of 2D rectangles that approximate planar surfaces. Observing the signal at each pixel of the camera, we obtain a histogram of photon counts, that is modeled by a Poisson process. Each point in this point cloud has an independent contribution to received signal. A probabilistic prior is defined on this point cloud so that points tend to cluster together. RJ-MCMC is then used to perturb the current configuration of points, by expanding or contracting a neighborhood of points, or by adding or removing points to or from the cloud, favoring their clustering. Thus, given a good initial guess of the scene, we iteratively improve the point cloud's conformity to objects in the scene.

The advantage of such a Monte Carlo approach is that the estimation accuracy is independent of the dimensions of the state space, and only depends on the number of particles. The MCMC method may therefore be better scalable to a much larger number of parameters, or to settings where there are multiple emitters [141].

PEEL currently deals with relatively low-dimensional data, through modeling the reflective boundaries as piecewise planar. A more thorough model of the environment which captures its real-life complexity, however, may have thousands of unknown parameters which must be determined. The key difficulty to be resolved before MCMC methods can be used for our underwater acoustic environment for passive sensing is the formulation of an appropriate probabilistic setting. While we have worked with deterministic high-frequency pulses as the emitted signals under consideration, such an MCMC method may be more amenable for estimation with probabilistic emitted signals that can be modeled as a given distribution whose parameters are unknown.

Appendix A

Derivation of the COTANS Transform

A.1 COTANS Transform for Tangent Lines to Ellipses in 2D

We present the derivation of the COTANS transform of tangent lines to an ellipse in 2D. First, we derive the tangent line to a standard ellipse centered around the origin, with a closed-form description in terms of the normal vector to the tangent. We then consider this ellipse to be rotated and translated to its actual position in 2D, and modify the normal-vector description of the corresponding rotated and translated tangent. The final parametrization of the tangent in terms of its normal vector is $(\rho_{\text{COTANS}}, \theta_{\text{COTANS}})$; this is a point in $\{(\rho, \theta)\}$-space, where ρ is the length of the vector and θ is its direction as measured with respect to the x-axis. This description is by definition the COTANS transform of the tangent line.

First, we obtain a description of tangent lines to standard ellipses (as in Fig. 3-4(a)). The equation for a standard ellipse centered on the origin is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where a and b are its major and minor axes, respectively. Consider a point (x_0, y_0) on the ellipse, and note that this point can be parametrized as $(a\cos\alpha, b\sin\alpha)$, where the angle α is an intermediate parametrization. Differentiating the ellipse's equation, the

slope of a tangent line is:

$$\frac{2x}{a^2} + \frac{2y}{b^2}\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = -\frac{b^2 x}{a^2 y}. \tag{A.1}$$

From the fact that (x_0, y_0) satisfies both $\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} = 1$ and the equation of the tangent's slope, we obtain the equation of the tangent line at (x_0, y_0) as $\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1$.

Now, we obtain the COTANS transform of this tangent line, by parametrizing it in terms of its Euclidean distance ρ from the origin and the azimuth θ of the normal line to the tangent going through the origin. Thus, a tangent to the ellipse at $(a\cos\alpha, b\sin\alpha)$ will be expressed as a point $(\rho_\alpha, \theta_\alpha)$ in the COTANS domain. One derivation of this COTANS transform result is as follows. Our method will obtain $\rho(\theta)$, the distance to the origin of a given tangent plane as a function of azimuth.

The distance ρ_α of the tangent line to the origin is the solution to:

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_2 \\ \text{s.t.} \quad & \mathbf{w}_\alpha^T \mathbf{x} = 1, \end{aligned} \tag{A.2}$$

where $\mathbf{x} = [x \ y]^T$ is the point of intersection of the tangent line and its normal through the origin, and $\mathbf{w}_\alpha^T = \left[\frac{x_0}{a^2} \ \frac{y_0}{b^2}\right] = \left[\frac{\cos\alpha}{a} \ \frac{\sin\alpha}{b}\right]$. ρ_α can be found by using Lagrange multipliers: technique, where:

$$\mathcal{L}(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^T\mathbf{x} - \lambda\left(\mathbf{w}_\alpha^T\mathbf{x} - 1\right),$$

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{0} \Rightarrow \mathbf{x} - \lambda\mathbf{w}_\alpha = \mathbf{0} \Rightarrow \mathbf{x} = \lambda\mathbf{w}_\alpha,$$

$$\nabla_\lambda\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{0} \Rightarrow \mathbf{w}_\alpha^T\mathbf{x} = 1 \Rightarrow \lambda\|\mathbf{w}_\alpha\|_2^2 = 1 \Rightarrow \lambda = \frac{1}{\|\mathbf{w}_\alpha\|_2^2},$$

$$\Rightarrow \mathbf{x} = \lambda\mathbf{w}_\alpha = \frac{\mathbf{w}_\alpha}{\|\mathbf{w}_\alpha\|_2^2} = \frac{a^2b^2}{a^2\sin^2\alpha + b^2\cos^2\alpha}\left[\frac{\cos\alpha}{a} \ \ \frac{\sin\alpha}{b}\right]^T,$$

$$\Rightarrow \rho_\alpha = \|\mathbf{x}\|_2 = \frac{1}{\|\mathbf{w}_\alpha\|_2} = \frac{ab}{\sqrt{a^2\sin^2\alpha + b^2\cos^2\alpha}}. \tag{A.3}$$

We now have ρ_α in Eq. (A.3), but our goal is to eliminate α and obtain $\rho(\theta)$. From

Eq. (A.1), we know that lines which are orthogonal to the tangent line have a slope of $\frac{a^2 y}{b^2 x}$, so:

$$\tan\theta = \frac{a\sin\alpha}{b\cos\alpha} = \frac{a}{b}\tan\alpha,$$

$$\Rightarrow \cos^2\alpha = \frac{1}{1+\tan^2\alpha} = \frac{1}{1+\frac{b^2}{a^2}\tan^2\theta}, \tag{A.4}$$

$$\sin^2\alpha = \frac{\frac{b^2}{a^2}\tan^2\theta}{1+\frac{b^2}{a^2}\tan^2\theta}. \tag{A.5}$$

Therefore, for a given θ, the corresponding $\rho(\theta)$ is obtained after substitution as:

$$\rho(\theta) = \frac{ab}{\sqrt{a^2\sin^2\alpha + b^2\cos^2\alpha}} = \sqrt{a^2\cos^2\theta + b^2\sin^2\theta}. \tag{A.6}$$

Eq. (A.6) allows us to obtain the set of points $\{(\rho,\theta)\}$ in the COTANS domain for a standard ellipse, but we wish to obtain such points for a general ellipse whose foci are $\mathbf{p}_\mathrm{e}$ and $\mathbf{p}_\mathrm{r}$. We are given that the LOS distance between the emitter and receiver is d_LOS and that the distance of the NLOS reflection from a given boundary is d_NLOS. Our method is to first obtain (ρ,θ) for a standard ellipse with foci at $\mathbf{p}'_\mathrm{r} = \left[-\frac{d_\mathrm{LOS}}{2}\ \ 0\right]^T$ and $\mathbf{p}'_\mathrm{e} = \left[\frac{d_\mathrm{LOS}}{2}\ \ 0\right]^T$, with axes $a = \frac{d_\mathrm{NLOS}}{2}$, $b = \frac{\sqrt{d_\mathrm{NLOS}^2 - d_\mathrm{LOS}^2}}{2}$. One example for such an ellipse is given in Fig. 3-5(a). The ellipse corresponding to the true localization scenario is obtained by rotating this standard ellipse as in Fig. 3-5(b), and then translating it to its correct position as in Fig. 3-5(c). The (ρ,θ) of the tangent line must be correspondingly transformed to match the new position of the rotated and translated tangent that yields the desired $(\rho_\mathrm{COTANS}, \theta_\mathrm{COTANS})$.

To modify (ρ,θ) as per Fig. 3-5, first consider the azimuth of the vector $\mathbf{p}_\mathrm{e} - \mathbf{p}_\mathrm{r}$, calling this angle θ_rot. To align the standard ellipse with the target ellipse, we replace each (ρ,θ) with $(\rho, (\theta+\theta_\mathrm{rot}) \bmod 2\pi)$ which we term (ρ,θ''). This rotation leaves the ρ-value of each tangent line unchanged, and only affects the azimuth, while the new foci are at $\mathbf{p}''_\mathrm{e}$ and $\mathbf{p}''_\mathrm{r}$. Next, we calculate a translation vector $\mathbf{p}_\mathrm{trans} = \mathbf{p}_\mathrm{r} - \mathbf{p}''_\mathrm{r}$, which would be added to any point on the rotated standard ellipse to obtain the target ellipse. To obtain the resulting $(\rho_\mathrm{COTANS}, \theta_\mathrm{COTANS})$ pairs, we first calculate the dot product $\rho_\mathrm{proj} = \mathbf{p}_\mathrm{trans} \cdot \hat{\rho}$, where $\hat{\rho} = [\cos\theta''\ \ \sin\theta'']^T$ is the unit vector pointing

towards the tangent line. Thus, we project the translation vector $\mathbf{p}_{\text{trans}}$ onto $\hat{\rho}$. If $\rho_{\text{proj}} \geq 0$, then we merely advance the tangent line in the same direction without changing its azimuth, so we replace (ρ, θ'') with $(\rho + \rho_{\text{proj}}, \theta'')$. If $\rho_{\text{proj}} < 0$, then we replace ρ with $|\rho - |\rho_{\text{proj}}||$. If $\rho_{\text{proj}} < 0$, then we subtract the projection result from ρ, thus replacing ρ with $|\rho - |\rho_{\text{proj}}||$, which ensures that ρ is positive as per definition. If $|\rho_{\text{proj}}| < \rho$ and $\rho_{\text{proj}} < 0$, then we do not modify the azimuth θ''; else, since the line has been translated past the origin and the direction of the $\hat{\rho}$-vector has been flipped, we replace θ'' with $(\theta'' + \pi) \bmod 2\pi$. Carrying out these operations for each of the starting $\{(\rho, \theta)\}$ points, we obtain a final transformed set of points as $\{(\rho_{\text{COTANS}}, \theta_{\text{COTANS}})\}$, rounded to a desired accuracy. We define an array over ρ and θ with this resolution, and for each rounded point, increment the corresponding array cell by 1.

A.2 COTANS Transform for Tangent Planes to Spheroids in 3D

The equation of an origin-centered spheroid is $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$. A point (x_0, y_0, z_0) on its surface can be parametrized as $(a\cos\alpha\sin\beta, b\sin\alpha\sin\beta, c\cos\beta)$, where α and β are intermediate parameters. Similarly to 2D, the tangent plane at (x_0, y_0, z_0) is $\frac{xx_0}{a^2} + \frac{yy_0}{b^2} + \frac{zz_0}{c^2} = 1$, and we have an optimization similar to Eq. (A.2), this time using $\mathbf{w}_{\alpha,\beta}^T = \left[\frac{x_0}{a^2} \;\; \frac{y_0}{b^2} \;\; \frac{z_0}{c^2}\right] = \left[\frac{\cos\alpha\sin\beta}{a} \;\; \frac{\sin\alpha\sin\beta}{b} \;\; \frac{\cos\beta}{c}\right]$ instead of $\mathbf{w}_\alpha^T$. A Lagrangian method as in Eq. (A.3) yields:

$$\rho_{\alpha,\beta} = \frac{1}{\|\mathbf{w}_{\alpha,\beta}\|_2} = \frac{1}{\sqrt{\frac{\cos^2\alpha\sin^2\beta}{a^2} + \frac{\sin^2\alpha\sin^2\beta}{b^2} + \frac{\cos^2\beta}{c^2}}}, \tag{A.7}$$

$$\mathbf{x} = \frac{\mathbf{w}_{\alpha,\beta}}{\|\mathbf{w}_{\alpha,\beta}\|_2^2} = \rho_{\alpha,\beta}^2 \left[\frac{\cos\alpha\sin\beta}{a} \;\; \frac{\sin\alpha\sin\beta}{b} \;\; \frac{\cos\beta}{c}\right]^T. \tag{A.8}$$

To eliminate α and β and express $\mathbf{x}$ in terms of the azimuth θ and elevation ϕ, we use:

$$\mathbf{x} = \rho_{\alpha,\beta} \left[\cos\theta \sin\phi \quad \sin\theta \sin\phi \quad \cos\phi\right]^T. \tag{A.9}$$

Firstly, noting that $\tan\phi = \frac{a}{b}\tan\alpha$, we have the same results of Eq. (A.4) for $\cos^2\alpha$ and Eq. (A.5) for $\sin^2\alpha$. Note that for the case of a spheroid where $a = b$, we have $\alpha = \theta$, since the x-y cross-section of the ellipsoid is then a circle. To eliminate β, we observe that:

$$\tan^2\phi = \frac{c^2}{a^2}\frac{\cos^2\alpha}{\cos^2\theta}\tan^2\beta = \frac{c^2}{a^2}\tan^2\beta,$$

$$\Rightarrow \cos^2\beta = \frac{1}{1 + \frac{a^2}{c^2}\tan^2\phi}, \tag{A.10}$$

$$\sin^2\beta = \frac{\frac{a^2}{c^2}\tan^2\phi}{1 + \frac{a^2}{c^2}\tan^2\phi}. \tag{A.11}$$

We substitute into Eq. (A.7) our results in Eq. (A.4), (A.5), (A.10), and (A.11) to get:

$$\rho(\theta,\phi) = \sqrt{c^2\cos^2\phi + a^2\sin^2\phi}. \tag{A.12}$$

Note that because of symmetry around the z-axis, there is no θ-dependence at this stage.

Using Eq. (A.12), we can take the COTANS transform of tangent planes to a standard spheroid with $a = b$. We can rotate this spheroid and translate it, and correspondingly transform the points (ρ,θ,ϕ) that define the tangent planes. Our operations are similar to those in the 2D case. We first define our standard spheroid to have foci at $\mathbf{p}'_{\mathrm{r}} = \left[0 \;\; 0 \;\; -\frac{d_{\mathrm{LOS}}}{2}\right]^T$ and $\mathbf{p}'_{\mathrm{e}} = \left[0 \;\; 0 \;\; \frac{d_{\mathrm{LOS}}}{2}\right]^T$, with axes $a = b = \frac{\sqrt{d^2_{\mathrm{NLOS}} - d^2_{\mathrm{LOS}}}}{2}$ and $c = \frac{d_{\mathrm{NLOS}}}{2}$. We calculate the azimuth and elevation of the vector $\mathbf{p}_{\mathrm{e}} - \mathbf{p}_{\mathrm{r}}$, and call these angles θ_{rot} and ϕ_{rot}, respectively. We rotate the spheroid by ϕ_{rot} with the y-axis as the axis of rotation (thereby obtaining the correct elevation), then rotate the result by θ_{rot} with the z-axis as the axis of rotation (thereby obtaining the correct azimuth), and finally we translate it to its correct position.

Rotating the spheroid by ϕ_{rot} will not change the ρ of a given (ρ,θ,ϕ) tangent plane, but may affect its azimuth and elevation. The unit vector orthogonal to (ρ,θ,ϕ)

in Cartesian coordinates is given by $\hat{\rho} = [\cos\theta\sin\phi \;\; \sin\theta\sin\phi \;\; \cos\phi]^T$. To rotate this vector with the y-axis as the axis of rotation, we multiply it by the corresponding rotation matrix to obtain:

$$\begin{bmatrix} \cos\phi_{\text{rot}} & 0 & \sin\phi_{\text{rot}} \\ 0 & 1 & 0 \\ -\sin\phi_{\text{rot}} & 0 & \cos\phi_{\text{rot}} \end{bmatrix} \begin{bmatrix} \cos\theta\sin\phi \\ \sin\theta\sin\phi \\ \cos\phi \end{bmatrix} = \begin{bmatrix} \cos\phi_{\text{rot}}\cos\theta\sin\phi + \sin\phi_{\text{rot}}\cos\phi \\ \sin\theta\sin\phi \\ -\sin\phi_{\text{rot}}\cos\theta\sin\phi + \cos\phi_{\text{rot}}\cos\phi \end{bmatrix} \tag{A.13}$$

The azimuth and elevation of the vector obtained in Eq. (A.13) are then substituted as the new θ and ϕ values of the plane, θ'' and ϕ''. Next, we rotate the tangent plane by θ_{rot}, which does not change ρ or ϕ'' but replaces θ'' with $(\theta'' + \theta_{\text{rot}}) \bmod 2\pi = \theta'''$. At this stage, the point corresponding to the transformed value of $\mathbf{p}'_{\text{r}}$ is a point $\mathbf{p}''_{\text{r}}$. Finally, we calculate the translation vector $\mathbf{p}_{\text{trans}} = \mathbf{p}_{\text{r}} - \mathbf{p}''_{\text{r}}$, and obtain $\rho_{\text{proj}} = \mathbf{p}_{\text{trans}} \cdot \hat{\rho}$. The transformation of the tangent plane in the COTANS domain is similar to the 2D case. If $\rho_{\text{proj}} \geq 0$, we replace $(\rho, \theta''', \phi'')$ with $(\rho + \rho_{\text{proj}}, \theta''', \phi'')$. If $\rho_{\text{proj}} < 0$, then we subtract the projection result from ρ, thus replacing ρ with $|\rho - |\rho_{\text{proj}}||$. If $|\rho_{\text{proj}}| < \rho$ and $\rho_{\text{proj}} < 0$, then we do not modify the azimuth θ''' or the elevation ϕ''; else, we replace θ''' with $(\theta''' + \pi) \bmod 2\pi$ and ϕ'' with $\pi - \phi''$. Carrying out this procedure for each of the starting $\{(\rho, \theta, \phi)\}$ points, we obtain a final set of $\{(\rho_{\text{COTANS}}, \theta_{\text{COTANS}}, \phi_{\text{COTANS}})\}$ points in the 3D COTANS domain.

Appendix B

Implementation of PSO Localization and Environment Estimation

In the standard PSO algorithm [107], we define N particles $\mathbf{x}_i^{(k)}$, corresponding to points $\mathbf{x}_i^{(k)} = [x_{i,1}^{(k)} \dots x_{i,D}^{(k)}]^T$ in a D-dimensional search space at iteration k of the PSO algorithm. This search space is delimited by the boundaries $[x_{j,\min}, x_{j,\max}]$ for each of the $j \in \{1 \dots D\}$ parameters that we optimize over. Each particle has a velocity vector $\mathbf{v}_i^{(k)} = [v_{i,1}^{(k)} \dots v_{i,D}^{(k)}]^T$, assigned randomly at initialization. Particles remember their best locations $\mathbf{p}_i$ in the search space over iterations 1 through k, where their objective function $f(\mathbf{p}_i)$ was the largest. Finally, each particle is in a set of M randomly chosen neighbors, over which we determine the best location $\mathbf{p}_{ig}$ over past iterations. The PSO update step for the velocity and position of each particle is then given for each dimension d by:

$$v_{id}^{(k+1)} \leftarrow \alpha v_{id}^{(k)} + \mathrm{U}\left(0, \beta\right)\left(p_{id} - x_{id}^{(k)}\right) + \mathrm{U}\left(0, \beta\right)\left(p_{ig} - x_{id}^{(k)}\right), \tag{B.1}$$

$$x_{id}^{(k+1)} \leftarrow x_{id}^{(k)} + v_{id}^{(k+1)}, \tag{B.2}$$

where α and β that define a particle's inertia and acceleration for updating its velocity (standard values are 0.7298 and 1.4986, respectively), and $\mathrm{U}(0, \beta)$ is a sample of the uniform random variable. Particles that exit the boundaries of the search space are

reflected back. Note that in the notation in this appendix, we have used a conventional indexing from the PSO literature for ease of comparison, rather than the one which we have developed for the PEEL framework.

The key user-defined parameter for PSO is the number of particles, which is typically on the order of tens for relatively easy optimization problems and hundreds for more challenging ones. We have been consistently successful with 128 particles and 5000 iterations for each emitter position for our localization task.

B.1 Kalman Tracking Implementation Details

Kalman tracking is the only stage in our method that needs to be tuned to a particular setting. While we use a constant acceleration model that is valid for a wide variety of trajectories, the state error variance and the measurement noise variance parameters have to be adjusted for each application.

We have assumed a process noise variance of 0.0005^2 on the velocity in the x- and y-directions in m/s, and a measurement noise variance of 0.002^2, which takes into account that our expected estimation errors in position and velocity are on the order of millimeters. These parameters have to be tuned to the expected scale of emitter motion, and the tracking method can be improved if more knowledge is available about the trajectory. The extended Kalman filter (EKF) that assumes a constant acceleration model is a standard extension of the method that we have used and has been successful in underwater acoustic localization and tracking in the literature [142], [143]; but we have obtained better performance with a standard Kalman filter in our particular HiFAAT experiments. The position errors relative to the physical scale of the experiment are often large enough to cause more sensitive tracking methods to diverge, so care must be taken in using a robust tracker at this stage, before attempting to use an EKF or other algorithmic improvement.

Bibliography

[1] USGS Pacific Coastal and Marine Science Center, accessed at https://cmgds.marine.usgs.gov/data/pacmaps/sd-pers5.html, "San Diego Margin Perspective View," Last retrieved - 2023.

[2] H. Vishnu, G. B. Deane, M. Chitre, O. Glowacki, D. Stokes, and M. Moskalik, "Vertical directionality and spatial coherence of the sound field in glacial bays in Hornsund Fjord," *J. Acoustical Soc. America*, vol. 148, no. 6, pp. 3849–3862, 2020.

[3] J. Kim, J. Kim, L. T. Nguyen, B. Shim, and W. Hong, "Tonal signal detection in passive sonar systems using atomic norm minimization," *EURASIP J. Adv. Signal Process.*, vol. 43, no. 1, pp. 1–9, 2019.

[4] P. Setlur, G. E. Smith, F. Ahmad, and M. G. Amin, "Target localization with a single sensor via multipath exploitation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 1996–2014, 2012.

[5] M. Deffenbaugh, "Optimal ocean acoustic tomography and navigation with moving sources," Ph.D. dissertation, Massachusetts Institute of Technology, 1997.

[6] P. C. Etter, *Underwater acoustic modeling and simulation*. CRC Press, 2018.

[7] K. B. Smith, "A three-dimensional propagation algorithm using finite azimuthal aperture," *J. Acoustical Soc. America*, vol. 106, no. 6, pp. 3231–3239, 1999.

[8] R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," *J. Acoustical Soc. America*, vol. 74, no. 1, pp. 188–195, 1983.

[9] T. C. Yang, "Properties of underwater acoustic communication channels in shallow water," *J. Acoustical Soc. America*, vol. 131, no. 1, pp. 129–145, 2012.

[10] E. M. Fischell, A. R. Kroo, and B. W. O'Neill, "Single-hydrophone low-cost underwater vehicle swarming," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 354–361, 2019.

[11] J. Waterston, J. Rhea, S. Peterson, L. Bolick, J. Ayers, and J. Ellen, "Ocean of Things: Affordable maritime sensors with scalable analysis," in *OCEANS 2019-Marseille*, 2019, pp. 1–6.

[12] A. B. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," in *Thirty-Third Asilomar Conference on Signals, Systems, and Computers*, vol. 1, Oct 1999, pp. 103–108.

[13] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean Distance Matrices: Essential theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 12–30, Nov 2015.

[14] H. Naseri, M. Costa, and V. Koivunen, "Multipath-aided cooperative network localization using convex optimization," in *48th Asilomar Conf. Signals, Syst. Comput.*, 2014, pp. 1515–1520.

[15] T. Arikan, A. Weiss, H. Vishnu, G. B. Deane, A. C. Singer, and G. W. Wornell, "An architecture for passive joint localization and structure learning in reverberant environments," *J. Acoustical Soc. America*, vol. 153, no. 1, p. 665–677, 2023.

[16] K. W. Cheung, H. C. So, W. K. Ma, and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1121–1130, 2004.

[17] I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. National Academy of Sciences*, vol. 110, no. 30, pp. 12 186–12 191, 2013.

[18] T. Arikan, A. Weiss, H. Vishnu, G. B. Deane, A. C. Singer, and G. W. Wornell, "Learning environmental structure using acoustic probes with a deep neural network," accepted to be published in the proceedings of *ICASSP 2023-2023 IEEE Int. Conf. on Acoustics, Speech and Signal Processing.*

[19] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proc. IEEE*, vol. 97, no. 2, pp. 404–426, 2000.

[20] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE J. Oceanic Eng.*, vol. 18, no. 4, pp. 401–424, 1993.

[21] N. O. Booth, P. A. Baxley, J. A. Rice, P. W. Schey, W. S. Hodgkiss, G. L. D'Spain, and J. J. Murray, "Source localization with broad-band matched-field processing in shallow water," *IEEE J. Oceanic Eng.*, vol. 21, no. 4, pp. 402–412, 1996.

[22] Z. H. Michalopoulou, A. Pole, and A. Abdi, "Bayesian coherent and incoherent matched-field localization and detection in the ocean," *J. Acoustical Soc. America*, vol. 146, no. 6, pp. 4812–4820, 2019.

[23] P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of a posteriori probability distributions," *J. Acoustical Soc. America*, vol. 104, no. 2, pp. 808–819, 1998.

[24] J. C. Preisig, "Robust maximum energy adaptive matched field processing," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1585–1593, 1994.

[25] C. M. Verlinden, J. Sarkar, B. D. Cornuelle, and W. A. Kuperman, "Determination of acoustic waveguide invariant using ships as sources of opportunity in a

shallow water marine environment," *J. Acoustical Soc. America*, vol. 141, no. 2, pp. EL102–EL107, 2017.

[26] M. Chitre and K. Pelekanakis, "Channel variability measurements in an underwater acoustic network," in *UComms 2014*, 2014, pp. 1–4.

[27] Y. Le Gall, F. Socheleau, and J. Bonnel, "Matched-field performance prediction with model mismatch," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 409–413, 2016.

[28] W. Xu, A. B. Baggeroer, and H. Schmidt, "Performance analysis for matched-field source localization: Simulations and experimental results," *IEEE J. Oceanic Eng.*, vol. 31, no. 2, pp. 325–344, 2016.

[29] L. M. Zurl and B. H. Tracey, "Depth-shifting of shallow water guide source observations," *J. Acoustical Soc. America*, vol. 118, no. 4, pp. 2224–2233, 2005.

[30] L. T. Fialkowski, M. D. Collins, W. A. Kuperman, J. S. Perkins, L. J. Kelly, A. Larsson, J. A. Fawcett, and L. H. Hall, "Matched-field processing using measured replica fields," *J. Acoustical Soc. America*, vol. 107, no. 2, pp. 739–746, 2000.

[31] L. M. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," *J. Acoustical Soc. America*, vol. 113, no. 5, pp. 2719–2731, 2003.

[32] L. Li and J. Krolik, "Simultaneous target and multipath positioning," *IEEE J. Select. Topics Signal Process.*, vol. 8, no. 1, pp. 153–165, 2013.

[33] F. Ribeiro, C. Zhang, D. Florêncio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1781–1792, 2010.

[34] H. Naseri and V. Koivunen, "Cooperative simultaneous localization and mapping by exploiting multipath propagation," *IEEE Signal Process. Mag.*, vol. 65, no. 1, pp. 200–211, 2016.

[35] I. Shames, A. N. Bishop, and B. D. Anderson, "Analysis of noisy bearing-only network localization," *IEEE Trans. Automat. Contr.*, vol. 58, no. 1, pp. 247–252, 2012.

[36] B. D. Anderson, I. Shames, G. Mao, and B. Fidan, "Formal theory of noisy sensor network localization," *SIAM J. Discrete Mathematics*, vol. 24, no. 2, pp. 684–698, 2010.

[37] S. Oh, L. Schenato, P. Chen, and S. Sastry, "Tracking and coordination of multiple agents using sensor networks: System design, algorithms and experiments," *Proc. IEEE*, vol. 95, no. 1, pp. 234–254, 2007.

[38] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1389–1403, 2012.

[39] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, 2003.

[40] H. Naseri and V. Koivunen, "A bayesian algorithm for distributed network localization using distance and direction data," *IEEE Trans. Signal Inform. Process. over Networks*, vol. 5, no. 2, pp. 290–304, 2018.

[41] L. Wang, Y. Yang, and X. Liu, "A direct position determination approach for underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 033–13 044, 2020.

[42] Y. El Baba, A. Walther, and E. Habets, "3D room geometry inference based on room impulse response stacks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 5, pp. 857–872, 2017.

[43] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1479–1489, 2008.

[44] P. Annibale, J. Filos, P. A. Naylor, and R. Rabenstein, "TDOA-based speed of sound estimation for air temperature and room geometry inference," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 21, no. 2, pp. 234–246, 2012.

[45] S. Venkateswaran and U. Madhow, "Localizing multiple events using times of arrival: a parallelized, hierarchical approach to the association problem," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5464–5477, 2012.

[46] X. Alameda-Pineda and R. Horaud, "A geometric approach to sound source localization from time-delay estimates," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 6, pp. 1082–1095, 2014.

[47] O. Bialer, D. Raphaeli, and A. J. Weiss, "Maximum-likelihood direct position estimation in dense multipath," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2069–2079, 2012.

[48] F. Ribeiro, D. E. Ba, C. Zhang, and D. A. Florêncio, "Turning enemies into friends: Using reflections to improve sound source localization," in *2010 IEEE Int. Conf. Multimedia and Expo*, Jul 2010, pp. 731–736.

[49] J. Bonnel, A. M. Thode, S. B. Blackwell, K. Kim, and A. Michael Macrander, "Range estimation of bowhead whale (Balaena mysticetus) calls in the Arctic using a single hydrophone," *J. Acoustical Soc. America*, vol. 136, no. 1, pp. 145–155, 2014.

[50] C. O. Tiemann, A. M. Thode, J. Straley, V. O'Connell, and K. Folkert, "Three-dimensional localization of sperm whales using a single hydrophone," *J. Acoustical Soc. America*, vol. 120, no. 4, pp. 2355–2365, 2006.

[51] H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in Santa Barbara Channel using machine learning classifiers," *J. Acoustical Soc. America*, vol. 142, no. 5, pp. 455–460, 2017.

[52] H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, "Deep-learning source localization using multi-frequency magnitude-only data," *J. Acoustical Soc. America*, vol. 146, no. 1, pp. 211–222, 2019.

[53] Y. Wu, R. Ayyalasomayajula, M. J. Bianco, D. Bharadia, and P. Gerstoft, "Sslide: Sound source localization for indoors based on deep learning," in *ICASSP 2021-2021 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Jun 2021, pp. 4680–4684.

[54] M. B. Porter, "The BELLHOP Manual and User's Guide: PRELIMINARY DRAFT," Tech. Rep., 2011.

[55] H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," *J. Acoustical Soc. America*, vol. 142, no. 3, pp. 1176–1188, 2017.

[56] J. Yangzhou, Z. Ma, and X. Huang, "A deep neural network approach to acoustic source localization in a shallow water tank experiment," *J. Acoustical Soc. America*, vol. 146, no. 6, pp. 4802–4811, 2019.

[57] E. L. Ferguson, R. Ramakrishnan, S. B. Williams, and C. T. Jin, "Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor," in *ICASSP 2017-2017 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Mar 2017, pp. 2657–2661.

[58] E. Ozanich, P. Gerstoft, and H. Niu, "A feedforward neural network for direction-of-arrival estimation," *J. Acoustical Soc. America*, vol. 147, no. 3, pp. 2035–2048, 2020.

[59] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2010, pp. 2822–2825.

[60] F. Antonacci, J. Filos, M. R. Thomas, E. A. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses,"

IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 10, pp. 2683–2695, 2012.

[61] S. Park and J. Choi, "Iterative echo labeling algorithm with convex hull expansion for room geometry estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, no. 3, pp. 1463–1478, 2021.

[62] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[63] M. B. Porter, "The KRAKEN normal mode program," Naval Research Lab, Washington DC, Tech. Rep., 1992.

[64] Z. Lei, K. Yang, and Y. Ma, "Passive localization in the deep ocean based on cross-correlation function matching," *J. Acoustical Soc. America*, vol. 139, no. 6, pp. EL196–EL201, 2016.

[65] Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," *J. Acoustical Soc. America*, vol. 143, no. 4, pp. 2321–2331, 2018.

[66] A. Weiss, T. Arikan, H. Vishnu, G. B. Deane, A. C. Singer, and G. W. Wornell, "A semi-blind method for localization of underwater acoustic sources," *IEEE Trans. Signal Process.*, vol. 70, pp. 3090–3106, 2022.

[67] Y. M. Too, M. Chitre, G. Barbastathis, and V. Pallayil, "Localizing snapping shrimp noise using a small-aperture array," *IEEE J. Oceanic Eng.*, vol. 44, no. 1, pp. 207–219, 2017.

[68] G. B. Deane, "A three-dimensional analysis of sound propagation in facetted geometries," *J. Acoustical Soc. America*, vol. 96, no. 5, pp. 2897–2907, Nov 1994. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.411299

[69] M. Chitre, "A high-frequency warm shallow water acoustic communications channel model and measurements," *J. Acoustical Soc. America*, vol. 122, no. 5, pp. 2580–2586, 2007.

[70] G. B. Deane and M. D. Stokes, "Model calculations of the underwater noise of breaking waves and comparison with experiment," *J. Acoustical Soc. America*, vol. 127, no. 6, pp. 3394–3410, 2010.

[71] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.

[72] Y. Gao, M. Clark, and P. Cooper, "Time delay estimate using cepstrum analysis in a shallow littoral environment," in *Conf. Undersea Defence Techn.*, vol. 7, 2008, p. 8.

[73] W. Sun and Z. Wang, "Online modeling and prediction of the large-scale temporal variation in underwater acoustic communication channels," *IEEE Access*, vol. 6, pp. 73 984–74 002, 2018.

[74] K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," *IEEE J. Oceanic Eng.*, vol. 30, no. 2, pp. 338–347, 2005.

[75] Y. Zha, S. C. Webb, and W. Menke, "Determining the orientations of ocean bottom seismometers using ambient noise correlation," *Geophysical Research Lett.*, vol. 40, no. 14, pp. 3585–3590, 2013.

[76] H. Weinberg and R. Burridge, "Horizontal ray theory for ocean acoustics," *J. Acoustical Soc. America*, vol. 55, no. 1, pp. 63–79, 1974.

[77] E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," *J. Acoustical Soc. America*, vol. 100, no. 6, pp. 3631–3645, 1996.

[78] C. H. Harrison, "Three-dimensional ray paths in basins, troughs, and near seamounts by use of ray invariants," *J. Acoustical Soc. America*, vol. 62, no. 6, pp. 1382–1388, 1977.

[79] K. D. Heaney and J. J. Murray, "Measurements of three-dimensional propagation in a continental shelf environment," *J. Acoustical Soc. America*, vol. 125, no. 3, pp. 1394–1402, 2009.

[80] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*. New York, NY: Springer New York, 2011. [Online]. Available: http://www.springerlink.com/index/10.1007/978-1\-4419-8678-8

[81] R. E. Francois and G. R. Garrison, "Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption," *J. Acoustical Soc. America*, vol. 72, no. 6, pp. 1879–1890, Dec 1982. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.388673

[82] M. J. Buckingham, "Theory of three-dimensional acoustic propagation in a wedgelike ocean with a penetrable bottom," *J. Acoustical Soc. America*, vol. 82, no. 1, pp. 198–210, 1987.

[83] G. B. Deane and M. J. Buckingham, "An analysis of the three-dimensional sound field in a penetrable wedge with a stratified fluid or elastic basement," *J. Acoustical Soc. America*, vol. 93, no. 3, pp. 1319–1328, 1993.

[84] E. K. Westwood, "Broadband modeling of the three-dimensional penetrable wedge," *J. Acoustical Soc. America*, vol. 92, no. 4, pp. 2212–2222, 1992.

[85] P. S. Petrov and F. Sturm, "An explicit analytical solution for sound propagation in a three-dimensional penetrable wedge with small apex angle," *J. Acoustical Soc. America*, vol. 139, no. 3, pp. 1343–1352, 2016. [Online]. Available: http://dx.doi.org/10.1121/1.4944692

[86] H. Medwin and C. Clay, *Fundamentals of Acoustical Oceanography*. Academic Press, 1998.

[87] J. J. Faran, "Sound Scattering by Solid Cylinders and Spheres," *J. Acoustical Soc. America*, vol. 23, no. 4, pp. 405–418, 1951.

[88] N. D. Veksler, *Resonance Acoustic Spectroscopy*, ser. Springer Series on Wave Phenomena, H. Überall, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, vol. 11, no. 9. [Online]. Available: http://link.springer.com/10.1007/978-3-642-84795-0

[89] A. Nagl, H. Überall, P. Delsanto, J. Alemar, and E. Rosario, "Refraction effects in the generation of helical surface waves on a cylindrical obstacle," *Wave Motion*, vol. 5, no. 3, pp. 235–247, Jun 1983. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0165212583900148

[90] F. Honarvar, "Nondestructive evaluation of cylindrical components by resonance acoustic spectroscopy," Ph.D. dissertation, University of Toronto, 1997.

[91] J. H. Haxel, R. P. Dziak, and H. Matsumoto, "Obtaining baseline measurements of ocean ambient sound at a mobile test berth site for wave energy conversion off the central Oregon coast," in *OCEANS'11 MTS/IEEE KONA*, 2011, pp. 1–5.

[92] C. T. Tindle and G. E. J. Bold, "Improved ray calculations in shallow water," *J. Acoustical Soc. America*, vol. 70, no. 3, pp. 813–819, 1981.

[93] M. Roohnia, *Wood: Vibration and Acoustic Properties*. Elsevier Ltd., 2019, no. January 2016. [Online]. Available: http://dx.doi.org/10.1016/B978-0-12-803581-8.01996-2

[94] https://cedarstripkayak.wordpress.com/lumber-selection/162-2/, "Wood Density Chart," Last retrieved - 2020.

[95] J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoustical Soc. America*, vol. 116, no. 4, pp. 2067–2080, 2004.

[96] T. Korhonen, "Acoustic localization using reverberation with virtual microphones," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2008, pp. 211–223.

[97] F. Ribeiro, C. Zhang, D. A. Florêncio, and D. E. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1781–1792, 2010.

[98] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP J. Audio, Speech, Music Process.*, pp. 1–17, 2010.

[99] R. Kaune, "Accuracy studies for TDOA and TOA localization," in *IEEE 15th International Conference on Information Fusion*, Jul 2012, pp. 408–415.

[100] Z. Su, G. Shao, and H. Liu, "Semidefinite programming for NLOS error mitigation in TDOA localization," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1430–1433, 2017.

[101] T. Qiao, S. Redfield, A. Abbasi, Z. Su, and H. Liu, "Robust coarse position estimation for TDOA localization," *IEEE Wireless Commun. Lett.*, vol. 2, no. 6, pp. 623–626, 2013.

[102] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[103] I. Guvenc and Z. Sahinoglu, "Threshold selection for UWB TOA estimation based on kurtosis analysis," *IEEE Commun. Lett.*, vol. 9, no. 12, pp. 1025–1027, 2005.

[104] I. Guvenc, C. C. Chong, and F. Watanabe, "NLOS identification and mitigation for UWB localization systems," in *2007 IEEE Wireless Commun. Networking Conf.*, 2010, pp. 1571–1576.

[105] M. Crocco, A. Trucco, and A. D. Bue, "Uncalibrated 3D room geometry estimation from sound impulse responses," *Journal of the Franklin Institute*, vol. 354, no. 18, pp. 8678–8709, 2017.

[106] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3D Hough transform for plane detection in point clouds: A review and a new accumulator design," *3D Research*, vol. 2, no. 2, p. 3, 2011.

[107] M. Clerc, *Particle swarm optimization*. John Wiley & Sons, 2010.

[108] P. Srinivasan, P. Liang, and S. Hackwood, "Computational geometric methods in volumetric intersection for 3D reconstruction," *Pattern Recognition*, vol. 23, no. 8, pp. 843–857, 1990.

[109] M. Kreković, I. Dokmanić, and M. Vetterli, "Shapes from echoes: uniqueness from point-to-plane distance matrices," *IEEE Trans. Signal Process.*, vol. 68, pp. 2480–2498, 2020.

[110] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, vol. 25, 2012.

[111] A. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Proces.*, vol. 31, no. 2, pp. 472–486, 1983.

[112] ——, "Fundamental limitations in passive time delay estimation–Part II: Warrow-band systems," *IEEE Trans. Acoust., Speech, Signal Proces.*, vol. 32, no. 5, pp. 1064–1078, 1984.

[113] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Proces.*, vol. 36, no. 4, pp. 477–489, 1988.

[114] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, 1994.

[115] R. Demirli and J. Saniie, "Model-based estimation of ultrasonic echoes. Part I: Analysis and algorithms," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 3, pp. 787–802, 2001.

[116] D. Dardari, C. Chong, and M. Win, "Improved lower bounds on time-of-arrival estimation error in realistic UWB channels," in *2006 IEEE International Conference on Ultra-Wideband*, 2006, pp. 531–537.

[117] M. Cobos, F. Antonacci, L. Comanducci, and A. Sarti, "Frequency-sliding generalized cross-correlation: A sub-band time delay estimation approach," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1270–1281, 2020.

[118] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in neural information processing systems*, vol. 26, 2013.

[119] T. Jia and R. M. Buehrer, "A new Cramer-Rao lower bound for TOA-based localization," in *MILCOM 2008- IEEE Military Communications Conference*, 2008, pp. 1–5.

[120] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I.Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 434–450, 1999.

[121] J. Salmi, A. Richter, and V. Koivunen, "Detection and tracking of MIMO propagation path parameters using state-space approach," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1538–1550, 2008.

[122] B. Friedlander, "Accuracy of source localization using multipath delays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 346–359, 1988.

[123] J.-P. Hermand, "Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: Theory and experimental results," *IEEE J. Oceanic Eng.*, vol. 24, no. 1, pp. 41–66, 1999.

[124] S. K. S. Fan and E. Zahara, "A hybrid simplex search and particle swarm optimization for unconstrained optimization," *European Journal of Operational Research*, vol. 181, no. 2, pp. 527–548, 2007.

[125] R. Brits, A. P. Engelbrecht, and F. van den Bergh, "A niching particle swarm optimizer," in *Proc. 4th Asia-Pacific Conf. Simulated Evolution and Learning*, vol. 2, Nov 2002, pp. 692–696.

[126] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 826–836, 2003.

[127] X. Li, Y. Ban, L. Girin, X. Alameda-Pineda, and R. Horaud, "Online localization and tracking of multiple moving speakers in reverberant environments," *IEEE J. Select. Topics Signal Process.*, vol. 13, no. 1, pp. 88–103, 2019.

[128] Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan, "Source localization using deep neural networks in a shallow water environment," *J. Acoustical Soc. America*, vol. 143, no. 5, pp. 2922–2932, 2018.

[129] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[130] I. Kang, M. P. Fitz, and S. B. Gelfand, "Blind estimation of multipath channel parameters: a modal analysis approach," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1140–1150, 1999.

[131] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Oceanic Eng.*, vol. 32, no. 4, pp. 927–939, 2007.

[132] C. Eckart, "The scattering of sound from the sea surface," *J. Acoustical Soc. America*, vol. 25, no. 3, pp. 566–570, 1953.

[133] E. I. Thorsos, "Acoustic scattering from a "Pierson–Moskowitz" sea surface," *J. Acoustical Soc. America*, vol. 88, no. 1, pp. 335–349, 1990.

[134] M. Siderius and M. B. Porter, "Modeling broadband ocean acoustic transmissions with time-varying sea surfaces," *J. Acoustical Soc. America*, vol. 124, no. 1, pp. 137–150, 2008.

[135] P. A. van Walree, F. X. Socheleau, R. Otnes, and T. Jenserud, "The watermark benchmark for underwater acoustic modulation schemes," *IEEE J. Oceanic Eng.*, vol. 42, no. 4, pp. 1007–1018, 2017.

[136] J. Palacios, G. Bielsa, P. Casari, and J. Widmer, "Communication-driven localization and mapping for millimeter wave networks," in *IEEE INFOCOM 2018-IEEE Conf. Computer Communications*, Apr 2018, pp. 2402–2410.

[137] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. Mclaughlin, and J. Y. Tourneret, "Bayesian 3D reconstruction of complex scenes from single-photon lidar data." *SIAM J. Imaging Sciences*, vol. 12, no. 1, pp. 521–550, 2019.

[138] J. Rapp, C. Saunders, J. Tachella, J. Murray-Bruce, Y. Altmann, J. Y. Tourneret, S. McLaughlin, R. M. Dawson, F. N. Wong, and V. K. Goyal, "Seeing around corners with edge-resolved transient imaging," *Nature communications*, vol. 11, no. 1, p. 5929, 2020.

[139] D. L. Snyder and M. I. Miller, *Random point processes in time and space*. Springer Science & Business Media, 2012.

[140] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *3D Research*, vol. 82, no. 4, pp. 711–732, 1995.

[141] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for multi-target tracking," *IEEE Trans. Automat. Contr.*, vol. 54, no. 3, pp. 481–497, 2009.

[142] A. Shefi, C. Therrien, D. Kirk, R. Saez, and B. Friedlander, "Passive multipath target tracking in inhomogeneous acoustic medium," in *ICASSP'87: IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 12, Apr 1987, pp. 463–466.

[143] R. Duan, K. Yang, Y. Ma, Q. Yang, and H. Li, "Moving source localization with a single hydrophone using multipath time delays in the deep ocean," *J. Acoustical Soc. America*, vol. 136, no. 2, pp. EL159–EL165, 2014.