

Non-Gaussian SLAM utilizing Synthetic Aperture Sonar

Mei Yi Cheung¹, Dehann Fourie², Nicholas R. Rypkema², Pedro Vaz Teixeira¹,
Henrik Schmidt¹, and John Leonard¹

Abstract—Synthetic Aperture Sonar (SAS) is a technique to improve the spatial resolution from a moving set of receivers by extending the array in time, increasing the effective array length and aperture. This technique is limited by the accuracy of the receiver position estimates, necessitating highly accurate, typically expensive aided-inertial navigation systems for submerged platforms. We leverage simultaneous localization and mapping to fuse acoustic and navigational measurements and obtain accurate pose estimates even without the benefit of absolute positioning for lengthy underwater missions. We demonstrate a method of formulating the well-known SAS problem in a SLAM framework, using acoustic data from hydrophones to simultaneously estimate platform and beacon position. An empirical probability distribution is computed from a conventional beamformer to correctly account for uncertainty in the acoustic measurements. The non-parametric method relieves the familiar Gaussian-only assumption currently used in the localization and mapping discipline and fits effectively into a factor graph formulation with conventional factors such as ground-truth priors and odometry. We present results from field experiments performed on the Charles River with an autonomous surface vehicle which demonstrate simultaneous localization of an unknown acoustic beacon and vehicle positioning, and provide comparison to GPS ground truths.

I. INTRODUCTION

Underwater acoustic beamforming typically uses phased arrays which consist of a fixed set of physically connected transducers that receive energy at varying times, depending on the geometry and travel speed of transmission. Increasing the physical aperture of the array can provide finer spatial resolution but is often constrained by the platform the array is mounted on. Synthetic aperture techniques operate by considering the array output over a segment of the platform’s trajectory, effectively creating a larger virtual array with increased aperture and thus resolution. In addition, distributing the beamforming task over a network of platforms (vehicles)—rather than one large expensive sensor—increases resilience and exploration rate, necessitating an accurate relative localization, mapping, and planning solution.

Conventional synthetic beamforming operations resolve the location of features in view by assuming the locations of transducer poses to be fixed and true – generally using some method of accurate odometry [1]. The inverse problem of resolving the transducer pose positions given a static scene assumption is more difficult. We believe that this work is the

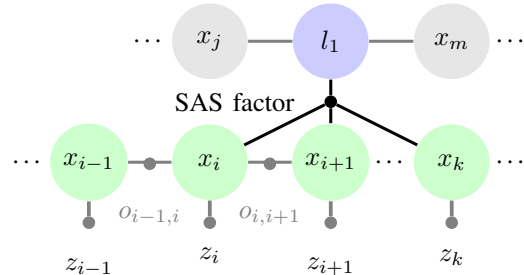


Fig. 1: Illustration of a SAS-factor (black) that can be used in combination with other factors in a factor graph (gray). Each consecutive vehicle pose x_i corresponds to a single element of the synthetic array. The measurement associated with each pose is an acoustic waveform, z_i . In our formulation, the synthetic aperture can be beamformed to inform the location of the landmark (conventional usage) and simultaneously, the correlation between waveforms can be used to inform the location of every pose (SLAM usage).

first to leverage single hydrophone acoustic waveforms in a common simultaneous localization and mapping solution, via the synthetic aperture process, to improve the accuracy of the vehicle’s pose estimate without prior knowledge of scene.

The main algorithmic contribution of this paper is incorporating conventional synthetic aperture beamforming into a common factor graph inference operation, with two immediate consequences: (i) recorded acoustic waveforms made by the vehicle in the field are used to simultaneously estimate the position of the acoustic source and each of the receiver positions associated with the synthetic array, and (ii) data-fusion inference with other data sources is performed by non-Gaussian smoothing using a factor graph.

II. LITERATURE REVIEW

A. Synthetic Aperture Sonar

Synthetic aperture radar (SAR) is a well-established technique. The underwater equivalent, synthetic aperture sonar (SAS), faces many challenges unique to the underwater and acoustic environment [2], but as we will establish, simultaneous localization and mapping (SLAM) methods can tackle these difficulties by providing high quality navigational information to an underwater vehicle using low-cost sensors. Conventionally, SAS is a method of processing sonar data in order to produce images of significantly higher azimuthal resolution than the physical array would normally permit. If a single receiver (hydrophone) or array travels in a straight line at a known and constant speed, the sampled acoustic data at each position in time can be seen as having being collected from sequential elements of a much larger, synthetic array. If the receiver’s position and velocity are known, the sampled signals can be phase-shifted into coherence, forming an image from the synthetic array.

¹M. Cheung, P. V. Teixeira, H. Schmidt, J. Leonard are with the Mechanical Engineering Department, MIT, Cambridge, MA 02139, USA. {mc2922, pvt, henrik, leonard}@mit.edu

²D. Fourie, N. R. Rypkema are with the Electrical Engineering and Computer Science Department, MIT, Cambridge, MA 02139, USA. {fourie, rypkema}@mit.edu

The accuracy of SAS systems relies heavily on the positioning accuracy of the platform, typically slow-moving, very stable or ship-mounted for high-resolution imaging or object detection purposes [3]. AUVs can receive GPS localization when surfaced, but must rely on other methods while underwater. Unaided inertial navigation is highly susceptible to drift, and acoustic positioning systems such as ultra-short baseline (USBL) or long baseline (LBL) are conventionally expensive and/or require ship support. Doppler velocity logs (DVL) are accurate but expensive, and might require specific flying altitudes. Given the increasing affordability of small light-weight underwater vehicles [4], we formulate the problem using low-cost hydrophone sensors.

B. Simultaneous Localization and Mapping

SLAM is the simultaneous construction of a map of a robot's environment and estimation of its state (pose) [5]. The marine domain is an appealing application for SLAM due to the above-mentioned challenges of underwater navigation; however, another significant challenge lies in fusing measurements obtained from underwater sensors such as ranging and imaging sonars, optical modems, oceanographic (pressure, temperature, bathymetry) sensors, etc.

Mallios *et al.* [6] demonstrate a pose-based EKF-SLAM solution in an underwater cave system registering both imaging and profiling sonars. Single-beam sonar registration without the need for inertial/odometry measurements is shown with iSAM [7], or with forward-looking sonar and pose-graph SLAM in [8]. Acoustic structure from motion (ASFM) [9] presents a factor-graph formulation of fusing odometry and imaging sonar measurements to recover environmental point features in 3D.

Newman *et al.* [10] demonstrated incorporating range, azimuth, and elevation measurements from a SAS system into a constant-time SLAM (CTS) framework with manual data association, while indicating that a graph-theoretical perspective would likely resolve some of the limitations encountered by the CTS method. In contrast to CTS, our approach uses a non-Gaussian factor graph approach, and does not divide the problem into submaps. Instead, the Bayes (Junction) tree strategy is used to decompose the full graph problem into statistically independent branches, in a mechanism that allows previous computations to be recycled in a near-optimal manner [11]–[13].

III. PROBLEM STATEMENT AND METHOD

A. Conventional Beamformer (CBF)

An array of acoustic receivers can be used to estimate the angle between the array and a sound source through conventional beamforming (CBF) [14], as illustrated in Fig. 2. Assuming the wave of sound incident onto the array is planar (a valid assumption when the source is in the far-field), we use trigonometric principles and the known array geometry to apply time-delays (or phase-shifts in the frequency domain) to the received signals for a user-specified look-angle. The aim of beamforming is to estimate the angle to the sound source by varying this look-angle over a set of candidate

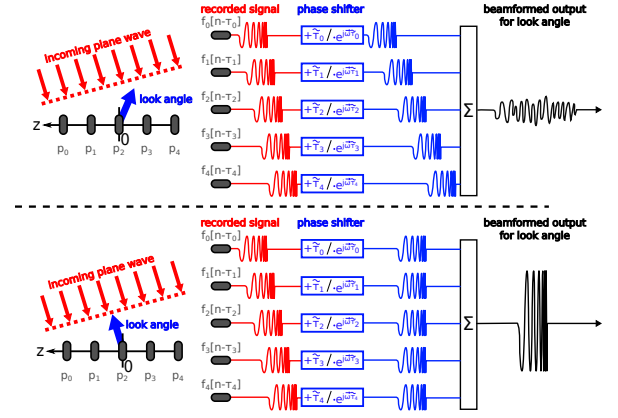


Fig. 2: Conceptual illustration of conventional beamforming. *Top*: The output power of the beamformer is low when the look-angle is not directed toward the incoming acoustic plane wave. *Bottom*: Signals received by each element constructively combine when the look-angle is directed toward the acoustic source, which is reflected by a high beamformer output power.

directions. The angle which results in maximum power (constructive interference on all received signals) usually corresponds to the look-angle is pointing towards the source.

For a given look-angle and set of array element positions \vec{x}_i , the time delay of a plane wave incident onto the array from that direction can be calculated as:

$$\tau_i = -\frac{\vec{u}^T \vec{x}_i}{c} \quad \text{where: } \vec{u} = [\cos(\vartheta) \quad \sin(\vartheta)]^T \quad (1)$$

where c is speed-of-sound in water and ϑ is the look-angle. This time-delay applied to the signal f_i from element i is identical to phase-shifts in frequency:

$$f_i[n - \tau_i] \xrightarrow{DFT/IDFT} F_i[\omega] \cdot e^{-j\vec{\omega}\vec{\tau}_i} \quad (2)$$

where $\vec{\omega}$ is the vector of frequencies of the n -point Discrete Fourier Transform (DFT). These time delays are induced by the array geometry, and beamforming works to negate these by using an element-wise spatial filter $e^{j\vec{\omega}\vec{\tau}_i}$ that applies opposing phase-shifts, followed by a summation over all signals as in Fig. 2:

$$Y[\omega; \vartheta] = \sum_{i=1}^m e^{j\vec{\omega}\vec{\tau}_i} \cdot F_i[\omega] \quad (3)$$

The output power averaged over all frequencies is:

$$|\tilde{Y}[\vartheta]|^2 = \frac{1}{n} \sum_n |Y[\omega_n; \vartheta]|^2 \quad (4)$$

Typically, conventional beamforming is performed using a geometrically fixed set of acoustic receivers. Synthetic aperture beamforming/sonar (SAS) extends this concept: if a single element coherently receives acoustic energy over time (either by actively ensonifying an object in the environment at every timestep, or in our case, by firing an acoustic beacon in sync with the element recording), then identical beamforming equations can be applied. There is only a minor change to Eq. 1, to reflect this dependency over the most recent n timesteps:

$$\tau_i = -\frac{\vec{u}^T \vec{x}_{t-i}}{c} \quad \text{where: } i = 0, 1, 2, \dots, n \quad (5)$$

Accurate estimation of the positions of the receiver(s), \vec{x}_{t-i} —the *micronavigation* problem—is critical to the performance of the SAS system, and must be performed using the few measurements available: received signals, odometry estimates, and occasional absolute positioning measurements.

B. Factor Graph Representation

The underlying inference problem that estimates state variables Θ , (such as vehicle poses or beacon locations) from non-ideal sensor measurements \mathbf{Z} is described by a bi-partite factor graph language [15]. The joint probability belief over all variables, $[\Theta | \mathbf{Z}] \in \mathcal{P}$, is a combination of factors (measurements \mathbf{z}_k^* and residual prediction models $\delta(\theta, z)$), where \mathcal{P} is set of all probability functions. By the chain rule, the product of independent measurements in the form of likelihoods $[\mathbf{Z}_i | \Theta_i]$ and variable prior potentials $[\Theta_j]$ —each over an alphabet/domain Ξ —represents the theoretical posterior joint probability density:

$$[\Theta | \mathbf{Z}] \propto \prod_i [\mathbf{Z}_i | \Theta_i] \prod_j [\Theta_j]. \quad (6)$$

The true measurements \mathbf{z}_i^* are assumed to originate from the true hidden variable states θ_i^* and the associated measurement process $\mathbf{z}_i^* \sim [\mathbf{Z}_i | \Theta_i = \theta_i^*]$. Factors (measurement models) are modeled by means of algebraic structure along with meaningful stochastic models. As an example, a basic Gaussian odometry measurement likelihood could be:

$$\mathbf{z}_i \sim [\mathbf{Z}_i | \Theta_i = \theta_i] = \mathcal{N}(\mu = \delta(\theta_i, \mathbf{z}_i^*), \Sigma_i). \quad (7)$$

Fig. 1 illustrates a synthetic aperture factor, containing hydrophone acoustic waveform recordings and required likelihood model, that can be used as part of a larger graph containing a variety of sensor data.

While much of the existing state estimation, localization, and mapping work in autonomy relies on point estimates and/or unimodal Gaussian approximations these yield less computational intensive algorithms, but they can over simplify the representation of uncertainty, miss important mission-relevant information, and yield brittle performance. In this work, the measurement likelihood model shown in Fig. 1 is completely non-Gaussian and used along with multimodal incremental smoothing and mapping (mm-iSAM) to perform belief space inference over all sensor data without having to resort to unimodal Gaussian assumptions.

C. Inference: Multimodal-iSAM

The multimodal-incremental smoothing and mapping inference algorithm (mm-iSAM) [11], [16] (and implementation [17]) infers the belief over state variables $\Theta_i = \hat{\Theta}_i$ that produced the measurements \mathbf{z}_k^* . The inference is also known as finding the posterior $[\hat{\Theta} | \mathbf{Z}]$, although mm-iSAM avoids resolving the full joint posterior, at the expense of more intensive computations. Mm-iSAM uses nine principles to reduce the computational complexity [11]. One major feature of mm-iSAM is that the full marginal posterior belief of any or all variables in the factor graph can be estimated.

The marginal belief of each variable then contains all the information introduced by the factors in the graph:

$$[\hat{\Theta}_i | \mathbf{Z}] = \int [\Theta | \mathbf{Z}] d\theta_i. \quad (8)$$

The marginal beliefs are approximate functions, and the current mm-iSAM strategy uses kernel density estimation with N evenly weighted, common leave-one-out likelihood cross validation for bandwidth selection Λ_i , Gaussian kernels [18] to represent the marginal beliefs:

$$[\hat{\Theta}_i | \mathbf{Z}] = \sum_n^N \frac{1}{N} \mathcal{N}(\theta_i; \mu_n, \Lambda_i). \quad (9)$$

Mm-iSAM requires a residual function for the probabilistic convolution of factors with current belief estimates of the dependent variables. We define the forward projection (akin to conventional SAS) as using receiver positions and recorded waveforms to resolve the likely position of the acoustic source. Similarly, the ‘backward’ projection estimates a single receiver position from the beacon’s position and neighboring poses.

1) *Forward Convolution*: Probabilistic convolution is performed ‘forward’ or towards the beacon over the current marginal beliefs of each pose $X_i \in 1..m$ where m is the length of the synthetic array and the beamforming computations using the recorded waveform measurements as in a conventional beamformer (Section III-A):

$$[\hat{\mathbf{L}} | \mathbf{X}_{1..m}, \mathbf{Z}_1, \cdot] \propto [\mathbf{Z}_1 | \mathbf{L}, \mathbf{X}_{1..m}] \prod_{j=1}^m [\hat{\mathbf{X}}_j | \cdot] \quad (10)$$

Mm-iSAM numerically calculates this convolution by means of a residual function $\delta(\mathbf{l}_1, \mathbf{x}_{1..m}; \mathbf{z}_1)$ where \mathbf{z}_1 represents the recorded acoustic waveforms associated with the SAS likelihood factor. For the forward case, a simplification can be used whereby the azimuth angle beam pattern is converted into a representative probability density over $[0, 2\pi)$, i.e. $[\tilde{\Phi} | \mathbf{X} = \mathbf{x}_{1..m}, \mathbf{Z} = \mathbf{z}_1]$ as illustrated by the magenta traces in Figs. 3 and 6. The density is re-calculated once for each unique set of pose positions. The beamformer output over azimuth (eq. (4)) is used as a pseudo-probability density function after noise floor truncation and renormalization of the beam pattern integral

$$[\tilde{\Phi} | \cdot] = \frac{1}{Q} \int |\tilde{Y}[\theta, \phi]|^2 d\phi, \quad (11)$$

such that $\iint |\tilde{Y}[\theta, \phi]|^2 d\phi d\theta = Q$. Here Q is a hypothetical partition function that normalizes the beamformer output so that it will sum to 1. Strictly speaking the beam pattern also includes reproducible structure dependent on the environment and physical setup which can be deconvolved from the calculated beam pattern [19]; that calibration is not in the scope of this work. The model in Eq. (11) can then be used to sample the azimuth angles $\phi^{\{n\}}$ for numerical approximation of the convolution,

$$\phi^{\{n\}} \sim [\Phi | \mathbf{X} = \mathbf{x}_{1..m}, \mathbf{Z} = \mathbf{z}_1]. \quad (12)$$

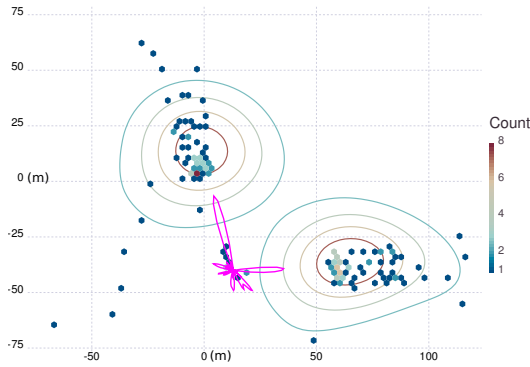


Fig. 3: Conventional SAS beamforming quantifies the distribution in received acoustic energy over azimuths for a series of vehicle poses and acoustic measurements. Range to the beacon is estimated from time-of-flight. All acoustic measurements are from field experiments in the Charles River. The output power of the CBF for a single array is drawn in magenta at the array's first pose. The range and azimuthal probability distribution is sampled to form a proposed posterior distribution for the beacon's location, here shown as a histogram of sampled locations and a contour plot of the resulting kernel density estimate representation. As typical for a CBF, the likelihood is symmetric about the linear array. Here, the beacon is located at (17.0, 1.78). Without any prior information, the CBF resolves two equally likely modes, for which a non-Gaussian representation is applicable.

For both time-synchronized one-way, or active SAS two-way time-of-flight operation, the range between beacon and transducer $\rho^{\{n\}}$ can be calculated via matched filtering. The correlator output of the matched filter is used to construct a pseudo-probability density:

$$\rho^{\{n\}} \sim [\mathbf{P} | \mathbf{X} = \mathbf{x}_{1..m}, \mathbf{Z} = \mathbf{z}_1] \quad (13)$$

The combination of range and bearing samples can then be used to perform a bearing, range, or bearing-range constraint between the SAS reference frame and beacon location

$$\delta_F \left(\mathbf{x}_{1..m}^{\{n\}}, l_1^{\{n\}}; \phi^{\{n\}}, \rho^{\{n\}} \right). \quad (14)$$

Using numerical root-finding, and given the independence of the kernels in each of the pose marginal estimates (as per eq. (9)), each of the pose values $\mu^{\{n\}}$ are used in conjunction with fresh samples $\phi^{\{n\}}, \rho^{\{n\}}$ to find the best value for each $l_1^{\{n\}}, \forall n \in [1, N]$

$$*l_1^{\{n\}} = \underset{l_1}{\text{argsolve}} \left(\mathbf{0} = \delta_F \left(\mathbf{x}_{1..m}^{\{n\}}, l_1^{\{n\}}; \phi^{\{n\}}, \rho^{\{n\}} \right) \right). \quad (15)$$

2) *'Backwards' Convolution:* The 'backwards' probabilistic convolution is the process by which the position of the beacon informs the position of each element in the synthetic array. We note that it is not a forward-backward solve in the typical sense. The backward convolution is approximated by the mm-iSAM infrastructure using the factor residual function for all kernels $n \in [1, N]$ as in Eq. (9), however, in the backwards case we can directly optimize the correlation of one element at a time in a leave-one-out strategy:

$$*x_i^{\{n\}} = \underset{p_i}{\text{argmin}} \left(\delta_F \left(\mathbf{x}_{1..m}^{\{n\}}, l_1^{\{n\}}; \mathbf{z}_1 \right) \right)^2. \quad (16)$$

A leave-one-out strategy is used to ensure pose estimates are calculated independently of their previous estimates, such

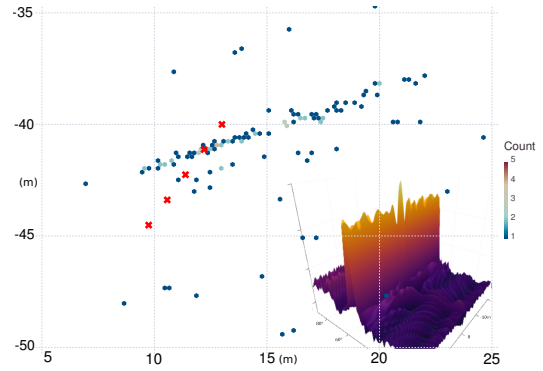


Fig. 4: Example of coherence during beamforming during 'backwards' convolution. GPS ground-truth for five synthetic array elements are shown as red crosses. Sampled solution locations for a single pose (2nd element) are shown, from which mm-iSAM constructs its posterior distribution. The clear 'line' of solutions perpendicular to the ray from array to beacon represents the far-field assumption, as moving linear array elements perpendicular to the correct look-angle does not change the phase-shift applied to the measurement waveform and its corresponding correlation. *Inset:* Correlation ridge plotted against azimuth and δx , z-axis is beamformer power. Note the large peak along the correct azimuth ($\approx 60^\circ$) and the additional peak at the correct position in x .

that $\vec{x}_j, j \in 1..m \setminus i$. When solving for the leave-out element \mathbf{x}_i , the leave-in array element waveforms are time shifted according to their positions $[\mathbf{X}_j = \hat{\mathbf{x}}_j^{\{n\}} | \cdot]$.

The origin of the synthetic array is arbitrarily chosen as one of the leave-in elements $\mathbf{x}_o, o \in j$, and the estimated position of the beacon is $[\mathbf{L}_1 | \cdot]$. The relative azimuth between pose and beacon is computed from the current belief over the beacon's location and the origin. The conventional beamformer filter is applied to all of the leave-in elements, and combined with the proposed phase-shifted waveform. This produces a residual function as shown in Fig 4 (inset), the largest peak of which is the position that most accurately resolves the current look-angle.

$$\vartheta = \tan^{-1} \left(\frac{y_b - y_o}{x_b - x_o} \right) \quad (17)$$

$$\hat{\tau}_i = -\frac{\vec{u}^T \hat{\mathbf{x}}_i}{c} \quad (18)$$

$$\tau_j = -\frac{\vec{u}^T \vec{x}_j}{c} \quad \text{where: } \vec{u} = [\cos(\vartheta) \quad \sin(\vartheta)]^T \quad (19)$$

$$Y[\omega; \vartheta] = \frac{1}{m-1} \sum_{j \in (1..m) \setminus i} e^{j\omega \hat{\tau}_j} \cdot F_j[\omega] + e^{j\omega \hat{\tau}_i} \cdot F_i[\omega] \quad (20)$$

IV. EXPERIMENTAL RESULTS

Field experiments were conducted from the MIT Sailing Pavilion on the Charles River, Cambridge, MA. The operational area is characterized acoustically by very shallow water depths (3-12m), concrete seawalls on both banks and acoustic noise mainly from boating activity, surface wind effects, and multi-path effects [20]. An acoustic beacon transmits a known, time-synced chirp waveform (250 – 2000 Hz) from two experimental locations (A, B): (Trial A) a stationary floating dock, and (Trial B) a drifting motor boat. Fig. 5 shows an FFT of the chirp waveform, and a representative spectrogram of the received signal.

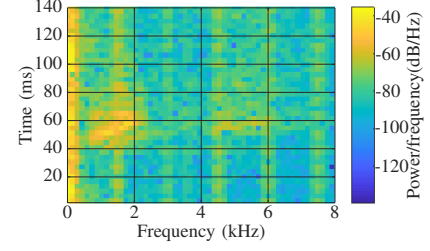
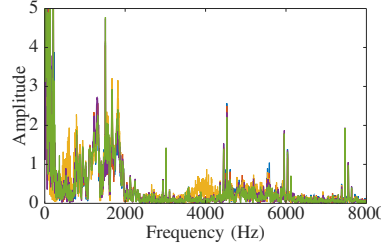


Fig. 5: *Left*: ASV *Nostromo* and acoustic beacon at the MIT Sailing Pavilion. *Center*: Frequency content of experiment waveform after windowed chirp z-transform. *Right*: Sample spectrogram of acoustic waveform recorded by a single hydrophone, with chirp at left.

An autonomous surface vehicle (ASV) is used as an experimental proxy for an underwater platform, with the benefit of providing ground truth via RTK-GPS to an accuracy within 0.1m [21]. In two experiment trials, the ASV travels an elliptical or circular loiter pattern relative to the beacon at a constant forward speed of approximately 1m/s. We note that the experimental ASV is able to travel faster (up to 4m/s) and has a smaller turning radius than a typical AUV, thus the dynamics of trajectory-following are not the focus for this experiment. MOOS-IvP autonomy software is used to handle communication between mission-level trajectory behaviors and low-level vehicle speed (thruster) and heading (servo) controllers [22].

Acoustic signals are recorded with a stern-mounted (1.5m depth) tetrahedral hydrophone array [23]. While only measurements from a single element of the tetrahedral array are processed, the factor graph representation is suited to the incorporation of diverse factors, including potential long baseline (LBL) measurements from the array. The length of each mission is approximately 7 minutes, over which the vehicle travels about 470m. Vehicle acoustic waveform measurements and poses are recorded at the rate of 1Hz, sampled at a frequency of 37.5 kHz for a total of 8000 samples. The speed of sound is assumed to be 1481m/s.

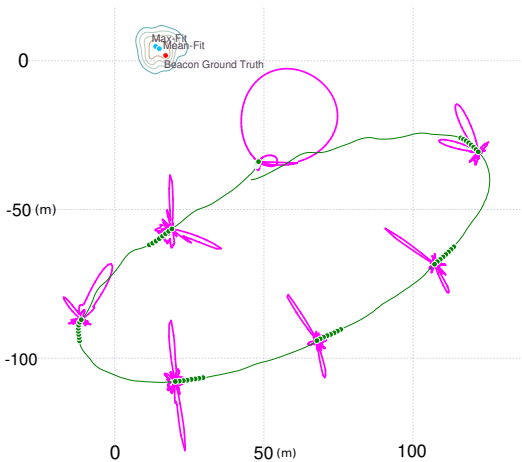


Fig. 6: SAS-SLAM result with GPS priors on every pose for Trial A. Ground truth GPS trajectory of the vehicle is traced in green, with poses used for SAS shown as points. Estimated beam patterns are shown as magenta traces at the origin of each array (arbitrarily selected as the earliest pose). The true, MAP, and mean-fit beacon locations are shown, underlaid by the full posterior distribution represented as a density contour plot, with each contour containing 20% of the probability mass.

A. Factor Graph Construction

For Trial A, a stationary beacon is located at the MIT Sailing Pavilion. For Trial B, a beacon is located on an anchored motorboat equipped with accurate differential GPS, with maximum drift under 20m. Both SAS-SLAM solutions assume the beacon to be stationary. No prior is given for the beacon’s position, assumed to be entirely unknown. Trial A has 300 poses and Trial B has 400 poses in total.¹

First, we show results with the assumption that every vehicle pose is supplied with an accurate GPS ground-truth prior (reasonable for a surface vehicle). This functions as an objective analysis of the forward convolution in the SLAM framework. Fig. 6 shows an instance of Trial A with six SAS factors, each comprising eight poses (synthetic array length of approximately 8m) selected at equal intervals around the mission’s elliptical trajectory. With accurate positioning, the beamformer strongly identifies the correct azimuth (and its equally likely symmetric second mode) for every synthetic array, with the exception of the very first one (closest to the beacon). The first circular pattern represents the case where no resolution in azimuth is obtained due to the very short effective array length (as the vehicle was relatively stationary at the start of the mission). The ability to resolve the correct look-angle is also affected by end-fire conditions, as shown by the third and seventh factor (counting counter-clockwise).

The 2-norm error $e_b = \sqrt{(\hat{x} - x_o)^2 + (\hat{y} - y_o)^2}$ from both a Mean-Fit (Mean) estimate and a MAP estimate over the posterior to ground-truth are tabulated in Table. I. The multi-modality of the posterior is evaluated by computing its Kullback-Leibler divergence (KLD) from a fitted multivariate normal distribution. We conclude that given even just three SAS-factors, inference over the joint distribution produced from the CBF resolves the posterior probability of the beacon’s location very accurately. The observed increase in prediction error from 7 factors to 16 can be attributed to the increasing number of SAS-factors with uninformative beamformer distributions (as discussed in Fig. 6)

TABLE I: Results (GPS Priors)

Trial	# SAS-F	Mean e_b (m)	MAP e_b (m)	KLD
A	3	5.50	6.22	0.018
A	7	3.12	4.52	0.025
A	16	7.42	7.05	0.033
B	3	5.51	8.05	0.359
B	8	4.10	6.70	0.027
B	14	4.53	4.40	0.048

¹Trial A SAS-SLAM video: https://youtu.be/_RfXLQ67N4o

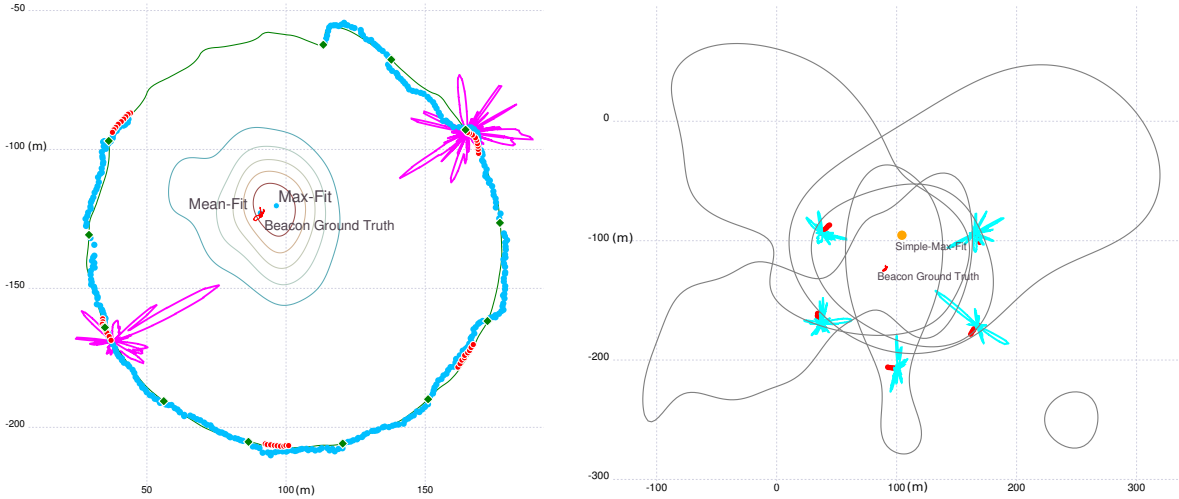


Fig. 7: *Left*: (Trial B) Full SAS-SLAM result with inference over GPS, odometry and SAS factors. GPS priors are shown as green diamonds and ground truth GPS trajectory is in green. Drifted position estimates (proxy inertial measurements) are shown as red points. Final MAP values of poses after inference is shown in blue circles. A single raw beam pattern is shown as magenta traces for two of the SAS-factors. The beacon locations are: ground truth (solid red line), MAP estimate and Mean-Fit estimate (blue points). The posterior distribution is represented as a density contour plot, with each contour containing 20% of the probability mass. Five SAS-factors are created, each with eight array elements. X-Y scale is shown in (m). *Right*: Comparison without SAS-SLAM: Array poses are shown in red. The CBF output of each synthetic array is shown in cyan. The overall beacon location predicted with a Max-Fit (orange point) over all five distributions after ‘forward’ convolution only. Each grey contour contains 50% of the probability mass.

Subsequently, GPS priors are provided on poses at different rates simulating dives. Navigational data is corrupted with a specified amount (0.2, 0.4 m/s) of cumulative navigation drift as a proxy for the accumulation of inertial error via odometry on an underwater vehicle. Drift is only accumulated between GPS priors, reflecting fixes obtained at the surface. Vehicle poses are initialized with corrupted position estimates, and consecutive vehicle poses are related with odometry factors ($\sigma = 0.5\text{m}$ in x, y). Fig. 7 (Left) shows an instance of Trial B with five SAS-factors, instantiated at equal intervals around a trajectory encircling the beacon from approximately 75m. For comparison, using the same dataset, Fig. 7 (Right) shows the density estimates from each of the five synthetic arrays and their beam patterns where each synthetic array is computed only from drifted position data (no SLAM). While ultimately the most probable location for the beacon lies on the interior of the trajectory, the fusion of the non-Gaussian acoustic measurements is key to identifying the posterior distribution of that probability given relevant symmetric modes.

TABLE II: Results (GPS Priors and Odometry)

Trial	Observations		$\sigma_{\dot{x}, \dot{y}}$	e_b (m)		KLD	e_p (m)	
	GPS	SAS		Mean	Max		Initial	Final
A	3	5	0.2	15.3	26.3	0.258	7.6	2.6
	3	11	0.2	4.48	5.16	0.173	16.9	4.8
	3	16	0.2	2.92	3.12	0.035	13.7	2.6
A	13	10	0.2	2.28	4.65	0.456	48.3	4.4
	13	10	0.4	3.38	4.30	0.037	72.9	6.4
	13	10	0.6	6.66	14.5	0.024	65.1	5.87
B	2	5	0.2	35.5	36.5	0.030	57.2	8.01
	2	12	0.2	6.77	11.0	0.044	22.4	5.7
	2	21	0.2	22.2	21.2	0.015	25.7	3.5
B	5	5	0.2	6.07	8.24	0.006	18.3	3.81
	18	5	0.2	6.27	5.56	0.006	3.08	2.29
	35	5	0.2	2.33	3.67	0.051	2.75	1.91

In addition to beacon localization errors, we compute the initial (drift) and final errors in vehicle positions e_p , and

evaluate estimation performance under different conditions: number of GPS priors, number of SAS factors, and variance of the velocities drift in the position data ($\sigma_{\dot{x}, \dot{y}}$). The MAP estimate was used to compute both e_b and e_p . The results, tabulated in Table II, show that the algorithm succeeds in estimating the position of both beacon and vehicle to comparable levels of accuracy, bringing e_b and e_p to the order of a few meters. Increasing the drift in odometry measurements appears to have a more noticeable effect on the accuracy of the beacon estimate rather than the vehicle’s. For Trial A, the increase in the number of observations quickly resolves the initially unknown beacon position estimate to within 10m of its true value. Platform position error, despite the low number of absolute measurements, remains low ($<5\text{m}$). For Trial B, even with only five SAS-factors, the localization performance was within 10m.

V. CONCLUSIONS

This work demonstrates that combining conventional beamforming with simultaneous localization and mapping is a desirable way to address the problem of micronavigation in synthetic aperture sonar applications, particularly when faced with inaccurate odometry and infrequent GPS updates. In the future, we anticipate that this method can extend to active SAS systems in two-way operation, either to build more dense map representations using a single transducer with moderate odometry in highly GPS-denied environments or, conversely, allow teams of collaborating underwater vehicles to map and localize more efficiently by communicating relevant acoustic SAS measurements.

ACKNOWLEDGMENTS

This work was partially supported by the Office of Naval Research under grants N00014-18-1-2832 and N00014-16-1-2628, and the *Systems That Learn Initiative* at CSAIL, MIT.

REFERENCES

- [1] Andrea Bellettini and Marc A Pinto. Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna. *IEEE journal of oceanic engineering*, 27(4):780–789, 2002.
- [2] Roy Edgar Hansen, Hayden John Callow, Torstein Olsmo Sabo, and Stig Asle Vaksvik Synnes. Challenges in seafloor imaging and mapping with synthetic aperture sonar. *IEEE Transactions on geoscience and Remote Sensing*, 49(10):3677–3687, 2011.
- [3] M. P. Hayes and P. T. Gough. Synthetic aperture sonar: A review of current status. *IEEE Journal of Oceanic Engineering*, 34(3):207–224, July 2009.
- [4] Oscar A Viquez, Erin M Fischell, Nicholas R Rypkema, and Henrik Schmidt. Design of a general autonomy payload for low-cost auv r&d. In *Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES*, pages 151–155. IEEE, 2016.
- [5] C. D. Cadena Lerma, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. Leonard. Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. *IEEE Trans. Robotics*, 2016.
- [6] Angelos Mallios, Pere Ridao, David Ribas, Marc Carreras, and Richard Camilli. Toward autonomous exploration in confined underwater environments. *Journal of Field Robotics*, 33(7):994–1012, 2016.
- [7] J. Wang, S. Bai, and B. Englot. Underwater localization and 3d mapping of submerged structures with a single-beam scanning sonar. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4898–4905, May 2017.
- [8] J. Li, M. Kaess, R. M. Eustice, and M. Johnson-Roberson. Pose-graph slam using forward-looking sonar. *IEEE Robotics and Automation Letters*, 3(3):2330–2337, July 2018.
- [9] Tiffany A Huang and Michael Kaess. Towards acoustic structure from motion for imaging sonar. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 758–765. IEEE, 2015.
- [10] P. M. Newman, J. J. Leonard, and R. J. Rikoski. Towards constant-time SLAM on an autonomous underwater vehicle using synthetic aperture sonar. *Robotics Research*, pages 409–420, 2005.
- [11] D. Fourie. *Multi-Modal and Inertial Sensor Solutions to Navigation-type Factor Graphs*. PhD thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2017.
- [12] M. Kaess, V. Ila, R. Roberts, and F. Dellaert. The Bayes tree: An algorithmic foundation for probabilistic robot mapping. In *Intl. Workshop on the Algorithmic Foundations of Robotics, WAFR*, Singapore, December 2010.
- [13] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31:217–236, February 2012.
- [14] H. A. Van Trees. *Optimum array processing: Part IV of detection, estimation and modulation theory*, volume 1. Wiley Online Library, 2002.
- [15] F.R. Kschischang, B.J. Frey, and H-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2), February 2001.
- [16] D. Fourie, J.J. Leonard, and M. Kaess. A nonparametric belief solution to the Bayes tree. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS*, Daejeon, Korea, Oct 2016.
- [17] Dehann Fourie and contributors. Caesar.jl, 2018. <https://github.com/JuliaRobotics/Caesar.jl>.
- [18] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.
- [19] P. V. Teixeira, M. Kaess, F. S. Hover, and J. J. Leonard. Multibeam data processing for underwater mapping. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018.
- [20] M. Stojanovic B. Reed, J. Leighton and F. Hover. *Multi-vehicle Dynamic Pursuit Using Underwater Acoustics*, pages 79–94. Springer International Publishing, Cham, 2016.
- [21] T. Takasu and A. Yasuda. Development of the low-cost rtk-gps receiver with an open source program package rtklib. In *International Symposium on GPS/GNSS*, 2009.
- [22] M. R. Benjamin, J. J. Leonard, H. Schmidt, and P. M. Newman. An Overview of MOOS-IvP and a brief users guide to the IvP Helm autonomy software. 2009.
- [23] H. Schmidt N. R. Rypkema, E. M. Fischell. One-way travel-time inverted ultra-short baseline localization for low-cost autonomous underwater vehicles. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages pre–print, 2017.