2015 IEEE International Conference on Robotics and Automation (ICRA)
Washington State Convention Center
Seattle, Washington, May 26-30, 2015

Communication-Constrained Multi-AUV Cooperative SLAM

Liam Paull, Guoquan Huang, Mae Seto, and John J. Leonard

***Abstract*— Multi-robot deployments have the potential for completing tasks more efficiently. For example, in simultaneous localization and mapping (SLAM), robots can better localize themselves and the map if they can share measurements of each other (direct encounters) and of commonly observed parts of the map (indirect encounters). However, performance is contingent on the quality of the communications channel. In the underwater scenario, communicating over any appreciable distance is achieved using acoustics which is low-bandwidth, slow, and unreliable, making cooperative operations very challenging. In this paper, we present a framework for cooperative SLAM (C-SLAM) for multiple autonomous underwater vehicles (AUVs) communicating only through acoustics. We develop a novel graph-based C-SLAM algorithm that is able to (optimally) generate communication packets whose size scales *linearly* with the number of observed features *since the last successful transmission*, constantly with the number of vehicles in the collective, and does not grow with time even the case of dropped packets, which are common. As a result, AUVs can bound their localization error without the need for pre-installed beacons or surfacing for GPS fixes during navigation, leading to significant reduction in time required to complete missions. The proposed algorithm is validated through realistic marine vehicle and acoustic communication simulations.**

I. Introduction

Multiple independently working robots can complete tasks more quickly in many cases. However, there is potential for even greater efficiency gains if the robots can *cooperate*. The ability to cooperate is contingent on the robots' ability to *communicate*. In this work we consider the task of simultaneous localization and mapping (SLAM) in the underwater environment where inter-vehicle communication is very challenging.

Cooperative SLAM (C-SLAM) can yield better performance than single robot SLAM since additional constraints are added through information sharing of one of two kinds: (i) *direct* encounters, where robots are able to make relative measurements of one another [1], and (ii) *indirect* encounters, where robots are able to make relative measurements through mutual observation of parts of the environment [2].

A visual depiction of the autonomous underwater vehicle (AUV) C-SLAM scenario is shown in Fig. 1. In the figure, two AUVs are performing a seabed coverage mission for mine hunting on orthogonal lawnmower (up and back) survey tracks. In the non-cooperative case, the AUVs submerge and dead reckon using a combination of compass, inertial and Doppler velocity sensors and position uncertainty will grow without bound.

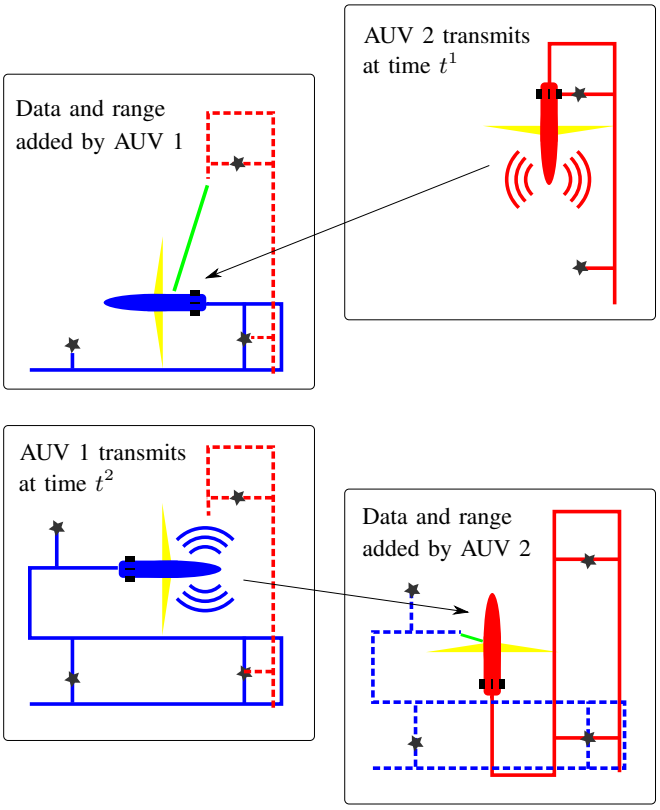

Fig. 1. Multi-AUV cooperative SLAM: Two AUVs navigate through an underwater environment while scanning the seabed with their sidescan sonar sensors (yellow). At time t^1, vehicle 2 broadcasts an acoustic packet (top right) that is received by vehicle 1. Vehicle 1 uses the transmitted packet to calculate a relative range to vehicle 2 (green) and also incorporates the transmitted features into its local feature map (top left). At some later time t^2 vehicle 1 makes a broadcast (bottom left) which is received by vehicle 2 (bottom right). Note that all communications go through an acoustic channel which is low-bandwidth, high-latency, and unreliable.

Historical methods to bound localization such as long baseline (LBL) and ultra short baseline (USBL) are costly and time-consuming to set up and also bound the operational area of the vehicles [3]. Another alternative is to periodically surface for GPS data; however, this can be very time consuming in deep water or under-ice applications and is sometimes undesirable due to covertness concerns.

In the proposed C-SLAM approach, AUVs can make direct measurements through acoustic ranging, and also communicate information about seabed features that they observe enabling indirect measurements resulting from mutually observed features. As a result, the AUV location uncertainty will become bounded.

This work was partially supported by Office of Naval Research (ONR) under grant no. N00014-13-1-0588, NSF Award IIS-1318392, and ONR Global.

L. Paull and J. Leonard are with the Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT, Cambridge, MA 02139, USA. Email: {lpaull,jleonard}@mit.edu

G. Huang is with the Dept. of Mechanical Engineering, University of Delaware, Newark, DE 19716, USA. Email: ghuang@udel.edu

M. Seto is with the Defense R&D Canada, Dartmouth, Nova Scotia, Canada. Email: mae.seto@drdc-rddc.gc.ca

Terrestrial [4] and aerial [2] applications of C-SLAM have demonstrated benefits in terms of efficiency and robustness; however, directly porting these algorithms to the undersea case is not straightforward. Sensing the environment is typically achieved with sonars which are of significantly lower fidelity than visible light cameras commonly used for terrestrial applications, and communicating though acoustics is challenging due to its high latency (signals travel at speed of sound $\approx$ 1500m/s), reduced throughput ($\approx$ 10-100 bytes/s), reduced bandwidth (channel sharing with time division multiple access) and low reliability ($\approx$ 20-50% dropout rate).

In this work, to the best of our knowledge, we develop the first multi-AUV C-SLAM algorithm that is specifically designed to operate *solely* with low-bandwidth acoustic communications.

We employ a graph-based approach that is able to *drastically* reduce communicated packet sizes by locally marginalizing out all vehicle pose estimates in between times of acoustic communications. We perform this marginalization *optimally* through Schur complement, rather than simply compounding measurements (which has been shown to be inconsistent [5]). This often results in a *dense* information matrix which has to be transmitted in order for consistent estimation and thus may still incur a significantly higher communication burden than the available bandwidth of the acoustic channel. To mitigate this issue, we perform *communication* reduction via graph sparsification inspired by recent work (e.g. [6], [7], [8], [5]) that instead seeks to reduce *computation* complexity. Specifically, we formulate a constrained convex optimization problem: minimize the Kullback-Leibler divergence (KLD) between the approximate and true distributions with a matrix inequality consistency constraint. We formulate the problem over the measurement information rather than the state information which leaves us free to design the new approximate measurements [8]. We show that an appealing choice is to design these new measurements to follow a predictable structure so that the measurement Jacobians will not have to be transmitted. Additionally, we design all measurements to be relative to the pose at the last known successful transmission time which has two key benefits: (i) the effect of linearization errors is reduced [6] and (ii) the algorithm is robust to partial or complete packet loss since the resulting factor graph will always be fully connected and consistent. Additionally, we show that for our choice of measurement structure, the optimization has a closed form solution. This extends the previous approach [8] by additionally considering the consistency constraint. The result is that we can achieve consistent C-SLAM by transmitting only: the new virtual measurement values (means) along with their associated block diagonal measurement information matrices. Consequently, total packet size is proportional to the number of observed measurements *since the last successful transmission* and does not grow with the number of vehicles in the team or the amount of time that has passed, even in the case of dropped packets. These packets can be generated on-demand in realtime onboard the AUVs even in the case of limited processing capabilities.

II. Related Work

An underwater SLAM solution provides a means of bounding localization errors without the need for pre-installed and localized beacon infrastructure. Approaches with different sensing modalities include: vision [9], sidescan sonar [10], forward-look sonar [11], and bathymetric sonar [12]. Our prior work [3] provides a review of recent research efforts in underwater navigation.

Recently, multi-AUV deployments have also been proposed that utilized the relative ranges derived from acoustic communications to perform cooperative localization (CL) underwater. For instance, Fallon et al. [13] propse a system for localization of a team of AUVs using acoustic communications from an autonomous surface craft. An alternative approach is to restrict underwater communications to be one-way to avoid information double-counting. For example, survey vs. aid teams [14], or client and server in the origin state method of Walls and Eustice [15]. Here, we allow fully bi-direction communication where every vehicle has equal access to the communications channel. This homogeneous topology has been used previously for CL [16]. In our recent work [17], we improve the scalability of [16] to allow larger team sizes and better robustness to failed communications.

Few if any works have proposed multi-AUV C-SLAM algorithms. Perhaps the closest known work is Fallon et al. which uses sidescan sonar landmarks and ranges from a surface vehicle over multiple sessions [16]. Here, we propose a method for full C-SLAM, which is more general than the underwater CL case previously considered.

A. C-SLAM

One issue in C-SLAM [18], [19] is how to find the initial relative poses of the robots. This can be achieved with "anchor nodes" in the pose graph formulation [2], directly from the maps themselves [20], or by requiring the robots to rendezvous [21], [18]. In this work, we assume such initialization is known before submergence, since vehicles start on the surface where they typically have GPS access.

The C-SLAM algorithm in the decentralized data fusion-smoothing and mapping (DDF-SAM) framework [22], [23] also marginalizes out the vehicles' poses. In particular, this is done in an overly conservative manner in [22], and later improved by using the concept of an "anti-factor" [23]. In our method we avoid double counting of information by removing all other vehicle poses from the factor graph during acoustic packet generation similar to [22], but restrict the AUVs only to transmit locally gathered data (as opposed to forward along data received from others). This may be overly restrictive in high-bandwidth systems, but it is necessary in underwater scenarios (where packets are often dropped), since it allows for reasonable scalability of packet sizes. The work of [24] uses the condensed measurement approach introduced in [25]. Specifically, each robot communicates the following information: (i) the last laser scan, (ii) the up-to-date estimates of the previous N nodes where N is the number of nodes since the last transmission, (iii) the indexes of the other robots' local maps that they have matched with their own, and (iv) part of the condensed graph computed

using the process in [25]. This approach can scale well with respect to the team size, however the packet size is on the order of 2.5 KBytes. Additionally, each communication is two-way where packet loss is not considered. Note that most terrestrial C-SLAM systems are built upon assumptions of communications throughput and bandwidth, which are unattainable underwater.

B. Graph Reduction for SLAM

Recently, graph-based SLAM approaches have become very popular [26]. Old states are not marginalized at every time step, leading to a sparse solution that can be solved incrementally and efficiently [27]. Nevertheless, computation and memory requirements will grow without bound, ultimately requiring some form of graph reduction to enable long-term operation.

In [28], pose-graph compression is proposed for laser-based SLAM by utilizing an approximate marginalization based on the Chow-Liu tree (CLT), while the method in [29] uses an information criterion to selectively remove uninformative loop closures. The work of [30] employs a Euclidean distance criterion for node removal to guarantee that state-space size grows only with the size of the mapped environment.

More closely related, recent work has used a convex optimization formulation for sparsification, where the KLD is minimized with a consistency constraint [7]. In particular, Carlevaris and Eustice [5] present the formulation of generic linear constraints (GLCs), where marginalization induces a fully connected constraint over the Markov blanket of the marginalized node. This fully connected constraint is then sparsified using a CLT approximation. Our prior work [6] formulates the marginalization process in a similar way while sparsifying edges uses ℓ_1-regularization, which is appealing in its flexibility as it does not commit to the CLT graph structure. However, one challenge with this approach is that direct control over the structure of the resulting sparsified matrix is lost. Most recently, the work of [8] improves previous results by allowing non-linear measurements to approximate the dense constraint with "virtual" measurements which can be defined arbitrarily and then insightfully formulating the convex optimization over the *measurement*, rather than *state*, information matrix and proving that it remains convex. However, designing these virtual measurements is non-trivial and task specific. In all of these cases [5], [6], [8], the motivation for variable removal is complexity reduction. Methods of selecting variables to remove are not described. In this work, we design a variable removal strategy and design measurements to address the bandwidth constrained C-SLAM problem.

III. Problem Formulation

Consider that AUVs are equipped with the following sensor suite: (i) 3-axis compass, (ii) GPS receiver (only functional at surface), (iii) Doppler velocity log (DVL) for fore (u) and starboard (v) speed relative to the seabed, (iv) sidescan sonar (SSS), and (v) acoustic modem. The pose of vehicle i at time t is denoted by $x^i(t)$. Each vehicle has a slot in the time-division multiple access (TDMA) cycle with which to make an acoustic transmission. Denote the time of transmission $k = 1, \ldots, K$ as t^k. We use the shorthand notation x^i_k to denote the pose of vehicle i at either transmission or reception of transmission k. Vehicles can make relative range estimates of one another ("direct" encounters) by maintaining synchronized clocks onboard and calculating the time-of-flight of communication packets [31]. The range measurement made by vehicle i of vehicle j as a result of transmission k is denoted by $z^i_{rj}(t^k)$. As vehicles are operating underwater, they are also gathering sonar seabed imagery (presumably this is the purpose of the AUV mission). An automatic feature recognition system is used onboard to extract features from the imagery. The measurement, $z^i_{lm}(t)$, represents an observation by vehicle i of landmark m at time t.

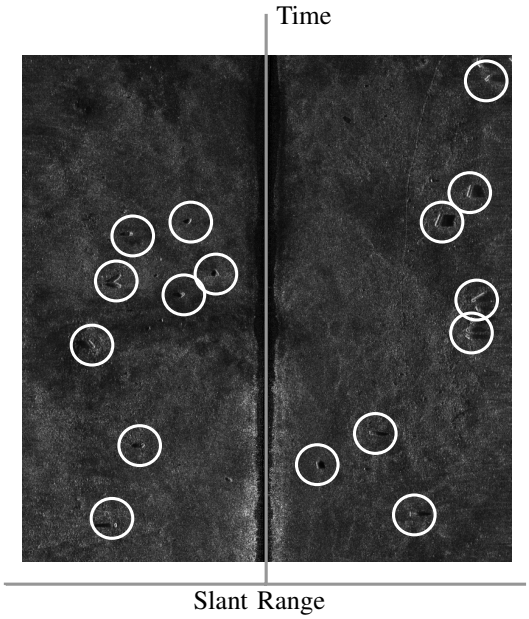

Fig. 2. Seabed mapping with sidescan sonar. Pings are registered into images. The vertical axis in the image represents the time of the ping and the x axis represents the slant range to the seafloor. Using the custom built automatic target recognition system, features of interest are detected in the image (highlighted with white circles). These feature observations are then added to the correct time-stamped pose in the trajectory based on the location of the object within the image.

A. Feature Detection from Sidescan Sonar

SSS returns are geo-registered and composed into images (see Fig. 2). The automatic target recognition (ATR) suite of tools used was developed at Defense R&D Canada. The matched filter in the ATR filters for geometric shapes with shadows (from the sonar insonification) based on templates. The ATR filters are similar to the existing algorithms [32] but optimized for mine-like object geometries. The AUV buffers the SSS data that it collects and then is able to generate the image and process the targets *in situ* using a single embedded processor on the vehicle. This ability to move the full sonar feature detection pipeline onboard the vehicle is essential for underwater SLAM.

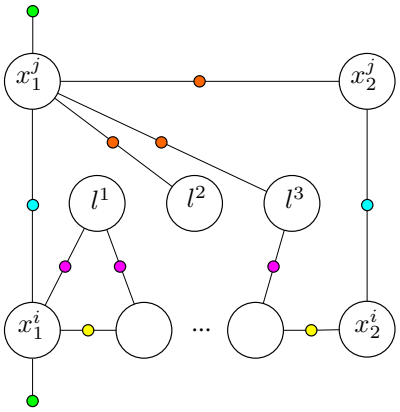

Fig. 3. Multi-vehicle factor graph maintained on vehicle i and used for navigation. In this graph, factors are color coded as follows: yellow: odometry, cyan: relative range measurements from TOF, magenta: landmark observations, green: occasional GPS measurements obtained at surface, and orange: "virtual" measurement resulting from marginalization/sparsification process presented in Section IV.

B. Multi-AUV C-SLAM with Factor Graphs

In this work, we follow a decentralized estimation paradigm which is robust to single-node failures. Each vehicle maintains in its state vector, X_C^i, (i) its own-vehicle trajectory, (ii) other vehicles' poses at communication/measurement times, and (iii) all detected features:

$$X_C^i(1:t) \triangleq [x^i(1:t), \{x^j(t^{1..K})\}_{j\in\{1..N\}\setminus i}, \mathcal{L}] \quad (1)$$

where $\mathcal{L} \triangleq \{l_m\}_{m=1..|\mathcal{L}|}$ is the set of all features; For visualization, we represent this as a factor graph as shown in Fig. 3. With mild assumptions (such as additive Gaussian noise of measurements), we often can convert the SLAM problem to a nonlinear least-squares (NLS) problem [26], which can be solved efficiently [27].

C. Bookkeeping with Confirmed Contact Points

To overcome the fact that broadcast packet reception is unknown until an acknowledgment is successfully received, a system of bookkeeping using confirmed ingoing and outgoing contact points is used [17]. Bookkeeping is required for vehicles to know which local factors should be generated to guarantee consistency of the multi-vehicle estimates maintained by others. Each vehicle i maintains a set of $N-1$ incoming and outgoing confirmed contact points. These contact points are the times of most recent confirmed successful communications to and from each other vehicle in the team. Incoming contact points are easily detectable based on the times at which communications are received. Outgoing contact points necessitate the use of communicated acknowledgment bits that are sent in subsequent data packet transmissions. In the case that an acknowledgement communication fails, the contact point time will not be updated, in essence assuming that the previous outgoing communication had failed. One of the advantages of the proposed measurement structure (see Sec. IV-B) is that it allows robustness to outgoing packets being dropped and also successful outgoing packets with dropped acknowledgements.

Algorithm 1 Acoustic Packet Generation Scheme

1: Transmission queue is empty
2: $t^{k_1} \leftarrow$ minimum contact point time
3: $t^{k_2} \leftarrow$ current time
4: Build local pose graph and solve for ${}^L\hat{X}_{k_{1,2}}$ (2) and H (3)
5: Marginalize states $t^{k_1}+1 : t^{k_2}-1$ using (4)-(7) to get H_R
6: Sparsify result with (13) and (15) to get $\hat{D}$ the diagonal consistent measurement information
7: Add local state estimates, ${}^L\hat{x}_2^i$, and ${}^L\hat{l}_{1:m}^i$ and measurement information $\hat{D}$ to transmission queue
8: Push data to the modem hardware for transmission

IV. Acoustic Packet Generation

In this section, we present in detail the core algorithm of packetizing local data that will be transmitted over the acoustic channel. The algorithm (see Algorithm 1) consists of two main components: (i) marginalization of nodes in between transmission times, and (ii) consistent sparsification of the resulting communication graph via convex optimization. In the sparsification step we enforce three design criteria: (i), we require that the solution be provably consistent (does not add information), (ii) we desire that the generated measurements follow a predetermined Jacobian structure (specifically the one shown on right panel of Fig. 4), and (iii) given the first two conditions we obtain the best possible approximation.

A. Marginalization of Intermediate Poses

Sending all intermediate poses between transmissions as well as all DVL, compass, and landmark observations is infeasible through the acoustic channel. To reduce data throughput we wish to relate all feature observations to states that coincide with acoustic transmissions. This is achieved through first removing all states corresponding to other vehicles (and their connected measurements) followed by a marginalization of own vehicle intermediate poses. Assume that the contact points times for the transmission as determined by the bookkeeping process previously described are t^{k_1} and t^{k_2}.

The states to be marginalized are $X_M \triangleq x^i(t^{k_1}+1 : t^{k_2}-1)$. The landmarks observed between times t^{k_1} and t^{k_2} are denoted by $\mathcal{L}_{k_{1,2}}^i \triangleq \{l_m | (\exists z_{l_m}^i(t) \in \mathcal{Z}^i) \wedge (t^{k_1} \le t \le t^{k_2})\}$, and similarly all the measurements (except relative range) obtained by vehicle i between t^{k_1} and t^{k_2} are $\mathcal{Z}_{k_{1,2}}^i$. As before we also define $X_{k_{1,2}}^i \triangleq [x^i(t^{k_1}+1 : t^{k_2}), \mathcal{L}_{k_{1,2}}^i]$.

To minimize world-frame linearization errors [5], [6], we transform all states to be relative to one of the remaining states. In this case we arbitrarily choose x_1^i, the own vehicle state at the last known point of contact. Note that this choice is arbitrary for the purposes of local marginalization, but it is important that it is consistent across vehicles such that incoming factors can be properly interpreted. We denote this local frame with the preceding superscript L. We formulate a NLS problem with respect to the desired section of the pose chain using only the local measurements available from

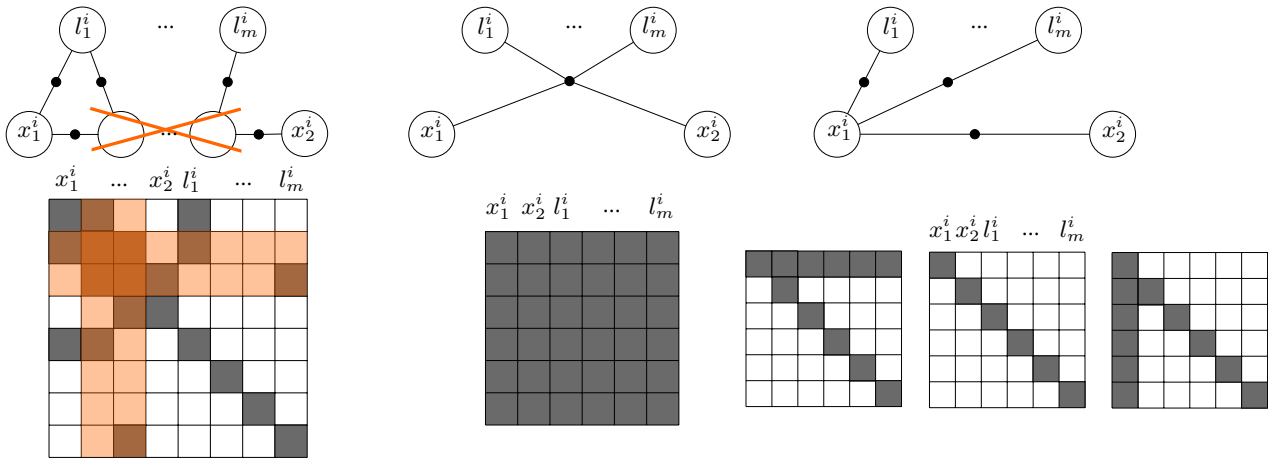

Fig. 4. *Left*: Intermediate nodes are marginalized from the information matrix H to reduce the amount of data to be transmitted acoustically. *Middle*: Marginalization results in a fully dense information matrix H_R. *Right*: We enforce a predetermined sparsity pattern and solve a convex optimization problem to determine the information matrix $H_{sparse} = F_R^T \hat{D} F_R$ corresponding to these new measurements that guarantee estimation consistency.

$t^{k_1} \leq t \leq t^{k_2}$:

$$
\begin{aligned}
{}^L\hat{X}^i_{k_{1,2}} &= \underset{{}^L X^i_{k_{1,2}}}{\operatorname{argmax}}\ p({}^L X^i_{k_{1,2}} | \mathcal{Z}^i_{k_{1,2}}) \\
&= \underset{{}^L X^i_{k_{1,2}}}{\operatorname{argmin}} \sum_{t'=t^{k_1}}^{t^{k_2}} ||z^i_g(t') - f^g({}^L x^i(t'))||^2_{\Sigma^i_g(t')} \\
&+ \sum_{t'=t^{k_1}}^{t^{k_2}} ||z^i_u(t') - f^u({}^L x^i(t'), {}^L x^i(t'-1))||^2_{\Sigma^i_u(t')} \\
&+ \sum_{t'=t^{k_1}}^{t^{k_2}} \sum_{l_m \in \mathcal{L}^i_{k_{1,2}}} ||z^i_{l_m}(t') - f^l({}^L x^i(t'), {}^L l_m)||^2_{\Sigma^i_l(t')}
\end{aligned}
\tag{2}
$$

where $f^u(\cdot)$ is the odometry function, $f^l(\cdot)$ is the robot-to-landmark measurement function, and $f^g(\cdot)$ is the global (GPS) function . This problem can be solved by Gauss-Newton iterations, in which we compute the (approximate) Hessian (information) matrix of the relative states (${}^L\hat{X}^i_{k_{1,2}}$ excluding x^i_1 [6]):

$$
H = \sum_{z \in \mathcal{Z}^i_{k_{1,2},}} (F_z)^T (\Sigma_z)^{-1} (F_z) \tag{3}
$$

F_z is the measurement Jacobian and Σ_z is the noise covariance for measurement z.

Once we have H we can proceed to marginalize out the intermediate nodes as desired through the Schur complement. Define the states to be marginalized out to be:

$$
{}^L X_M \triangleq [{}^L X^i(t^{k_1}+1 : t^{k_2}-1)] \tag{4}
$$

and the connected states to be retained to be:

$$
{}^L X_R \triangleq \begin{bmatrix} {}^L x^i_2 & {}^L \mathcal{L}^i_{k_{1,2}} \end{bmatrix} \tag{5}
$$

Then we can decompose the Hessian matrix accordingly:

$$
H = \begin{bmatrix} H_{MM} & H_{MR} \\ H_{RM} & H_{RR} \end{bmatrix} \tag{6}
$$

and write the expression for the information over the remaining states as:

$$
H_R = H_{RR} - (H_{RM})(H^{-1}_{MM})(H_{MR}) \tag{7}
$$

where H_R now encapsulates all of the information induced onto the connecting states through the marginalization process (subject to linearization error). Finally we can construct a new constraint that connects all of the connecting states (as shown in middle of Fig. 4):

$$
\begin{aligned}
z^i_R \triangleq {}^L\hat{X}^i_R &= {}^L X^i_R + \eta \\
&= r(X^i_R, x^i_1) + \eta \\
\eta &\sim \mathcal{N}(0, (H_R)^{-1})
\end{aligned}
\tag{8}
$$

where $r(\cdot)$ is a function that transforms the nodes X^i_R from the global from to the local frame relative to x^i_1 (similar to the root shift operation in GLC [5]) and ${}^L\hat{X}^i_R$ are the optimal estimates of the remaining states as determined by solving (2) which do not change through the marginalization process. It becomes clear from the middle panel of Fig. 4 that the information matrix is now fully populated. The number of entries in this matrix will scale $\mathcal{O}(|\mathcal{L}^i_{k_{1,2}}|^2)$. For even a very small number of measurements we will exceed the allowable packet size for acoustic transmission.

B. Consistent Sparsification of the Communication Graph

Since the dense matrix in the middle of Fig. 4 is still potentially too large for transmission through acoustics, we need to selectively *sparsify* it until the bandwidth considerations are met. We formulate the sparsification as a convex minimization problem whose cost function is the KLD between the original and the sparsified versions of the information matrices.

We seek to approximate the distribution $\mathcal{N}({}^L\hat{X}^i_R, H^{-1}_R)$ with a new distribution $\mathcal{N}({}^L\hat{X}^i_R, H^{-1}_{sparse})$. We enforce the following desired properties onto the new information matrix H_{sparse}.

1) Guaranteed consistency: This can be achieved by imposing the following constraint:

$$H_{sparse} \preceq H_R \tag{9}$$

where the $\preceq$ indicates that $H_R - H_{sparse}$ is positive definite.

2) Enforce predetermined sparsity pattern: We decompose the information matrix as follows:

$$H_{sparse} = F_R^T \hat{D} F_R \tag{10}$$

where F_R is the stacked Jacobian of the new measurements and the block diagonal matrix $\hat{D}$ contains all of the measurement informations (inverse covariances associated with these new measurements that we are defining). As such we have direct control over the structure of F_R and can enforce the desired measurement functions so long as we choose $\hat{D}$ such that the other constraints are satisfied [8]. Note that we have chosen a measurement pattern that induces a tree structure with tree depth equal to two in the resulting factor graph, where x_1^i is at the root. The result is that the Jacobian matrices follow a fixed a predetermined structure (Fig. 4-right) which is agreed upon by all vehicles and therefore not necessary to actual transmit.

3) Best approximation: The objective for minimization is the KLD between the original Gaussian distribution and the approximate one which can be formulated in terms of their information matrices assuming they have the same mean value [7]:

$$\begin{aligned} &D_{KL}(\mathcal{N}({}^L\hat{X}_R^i, H_R^{-1}) || \mathcal{N}({}^L\hat{X}_R^i, H_{sparse}^{-1})) \\ &= -\frac{1}{2}[\ln(\frac{|H_R|}{|H_{sparse}|})] + \text{trace}(H_{sparse} H_R^{-1}) - \dim({}^L\hat{X}_R^i) \end{aligned} \tag{11}$$

C. Convex Optimization Problem and Closed-Form Solution

We combine (9), (10), and (11) and the fact that we know H_R and $\dim({}^L\hat{X}_R^i)$ to formulate the following convex optimization over the block diagonal measurement information matrix D:

$$\begin{aligned} \min_{D \in \mathcal{D}} \quad & < F_R^T D F_R, H_R^{-1} > - \ln |F_R^T D F_R| \\ \text{subject to} \quad & F_R^T D F_R \preceq H_R \end{aligned} \tag{12}$$

where $\mathcal{D}$ is the set of block diagonal positive definite matrices. The Jacobian matrix that we have designed is both square and full rank (see the right-hand panel of Fig. 4). As a result, the *unconstrained* version of in (12) has a closed-form solution [8]:

$$D_i^* = (\{F_R H_R^{-1} F_R^T\}_i)^{-1} \tag{13}$$

where the subscript i denotes the i-th block diagonal component. However, this solution does not guarantee consistency so we project the solution onto the consistency constraint's domain:

$$\hat{D} = \underset{H_R^{-1} \preceq Y}{\text{argmin}} \; ||D^* - Y||_F^2 \tag{14}$$

where $||\cdot||_F$ is the matrix Frobenius norm. This can be analytically solved by using the eigendecomposition $D^* = V_D \text{diag}(\lambda_i^D) V_D^T$ and $H_R^{-1} = V_H \text{diag}(\lambda_i^H) V_H^T$, where λ_i^D and λ_i^H are the eigenvalues of D^* and H_R^{-1} respectively, and then computing:

$$\hat{D} = V_D \mathbf{Diag}\left(\max\{\lambda_i^D, \lambda_i^H\}\right) V_D^T \tag{15}$$

In our case the reduced Hessian has relatively small size, so this eigendecomposition can be performed efficiently even on an embedded processor.

D. Final Packet Contents

The solution to (12) is a new set of $|\mathcal{L}_{k_{1,2}}^i| + 1$ pairwise measurements, including one relative-pose constraint:

$$\begin{aligned} z_{\hat{D}_1}^i \triangleq {}^L\hat{x}_2^i &= r(x_2^i, x_1^i) + \eta_1 \\ \eta_1 &\sim \mathcal{N}(0, \hat{D}_1^{-1}) \end{aligned} \tag{16}$$

and $m = 1..|\mathcal{L}_{k_{1,2}}^i|$ relative pose-feature measurements:

$$\begin{aligned} z_{\hat{D}_m}^i \triangleq {}^L\hat{l}_m^i &= r(l_m^i, x_1^i) + \eta_m \\ \eta_m &\sim \mathcal{N}(0, \hat{D}_m^{-1}) \end{aligned} \tag{17}$$

where $\hat{D}_m$ is the block component of $\hat{D}$ corresponding to the m-th measurement.

In order for the packet recipient to be able to reconstruct the multi-vehicle pose graph shown in Fig. 3, we need to transmit the pose and feature estimates ${}^L\hat{x}_2^i$, ${}^L\hat{l}_{1:m}^i$ and the non-zero elements of the block diagonal measurement information matrix $\hat{D}$. In addition, we send the acknowledgement bits from the last cycle. Importantly, the size of the data packet scales *linearly* $\mathcal{O}(|\mathcal{L}_{k_{1,2}}^i|)$ with the number of features detected since the last transmission, *constantly* with time, even in the case of dropped packets, and *constantly* with the number of AUVs in the cooperative, making this approach applicable to very large-scale deployments.

V. Packet Reception

Upon reception of an acoustic packet from vehicle i on vehicle j, the received measurement values (state estimates) and associated information matrices are incorporated into the multi-vehicle factor graph shown in Fig. 3 allowing the construction of the full multi-vehicle NLS problem:

$$\begin{aligned} \hat{X}_C^j &= \underset{X_C^j}{\text{argmax}} \; p(X_C^j | \mathcal{Z}) \\ &= \underset{X_C^j}{\text{argmin}} \sum_{z_g^j(t) \in \mathcal{Z}} ||z_g^j(t) - f^g(x^j(t))||_{\Sigma_g^j(t)}^2 \\ &+ \sum_{z_u^j(t) \in \mathcal{Z}} ||z_u^j(t) - f^u(x^j(t), x^j(t-1))||_{\Sigma_u^j(t)}^2 \\ &+ \sum_{z_{l^m}^j(t) \in \mathcal{Z}} ||z_{l^m}^j(t) - f^l(x^j(t), l^m)||_{\Sigma_l^j(t)}^2 \\ &+ \sum_{z_{rj}^j(t^k) \in \mathcal{Z}} ||z_{rj}^j(t^k) - f^r(x_k^i, x_k^j)||_{\Sigma_r}^2 \\ &+ ||{}^L\hat{x}_2^i - r(x_2^i, x_1^i)||_{\hat{D}_1^{-1}}^2 \\ &+ \sum_{m=1}^{|\mathcal{L}_{k_{1,2}}^i|} ||{}^L\hat{l}_m^i - r(l_m^i, x_1^i)||_{\hat{D}_m^{-1}}^2 \end{aligned} \tag{18}$$

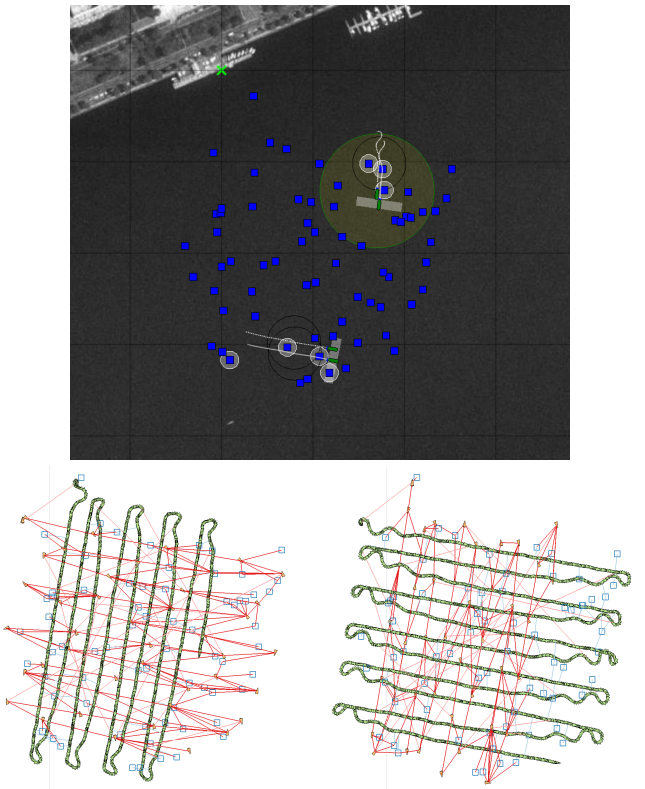

Fig. 5. *Top*: Shoreside viewer showing simulated features (blue squares) and ground truth and estimated locations of each vehicle. The ground truth vehicles have simulated sidescan sonar swaths that are being used to detect targets. At the current time, the vehicle on the North-South survey is transmitted as shown by the green circle. *Bottom*: Each of the two vehicles' decentralized factor graphs. Green triangles show own vehicle trajectory, other vehicle locations at times of communications are shown by orange triangles. The feature estimates are shown as blue squares. Red lines indicate virtual measurements (either pose to pose or pose to feature), blue lines are sidescan detection measurements, and pink lines are relative range measurements induced by acoustic communications with synchronized clocks.

which is solved using a similar technique as the relative case (2). In the above $f^r(x_k^i, x_k^j) = ||x_k^i - x_k^j||$ is the range measurement model and the last two terms are the virtual measurements received in the broadcast packet.

VI. Experimental Results

The system is implemented with a combination of open source projects developed at MIT: the mission-oriented operating suite (MOOS) with interval programming [33], Goby-Acomms [34], lightweight communications and marshaling (LCM) [35], and iSAM [27].

MOOS is a middleware and marine simulation software framework. A single MOOS community resides on each vehicle and a third as a shoreside for monitoring (see the top panel in Fig. 5). MOOS applications were written for simulating sensors, feature detection, the TDMA cycle, and processing and packaging acoustic packets to and from the simulated modem. In Goby, packets are defined as Google Protocol Buffers which are encoded using the dynamic compact control language at runtime. There is also functionality within Goby for simulating the acoustic channel and enforcing packet size restrictions. Data is parsed in MOOS and then passed through LCM channels to the backend iSAM

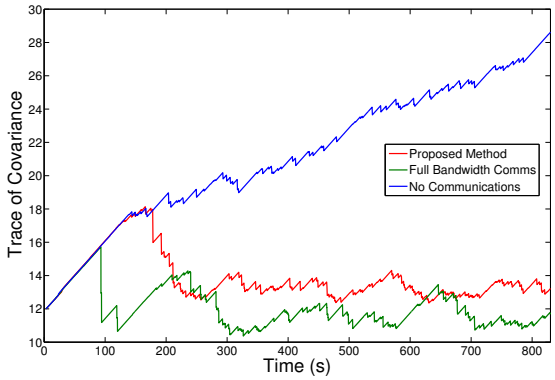

Fig. 6. Trace of the covariance matrix for three different simulated scenarios. As expected, the uncertainty grows unbounded for no communications. The performance of our proposed method is close to that of the full-bandwidth case. In our approach, packets are transmitted every 10s with a drop rate of 50% and maximum size of 192 bytes.

solver. This backend solver manages the multi-vehicle factor graph and uses built-in and custom factor definitions to maintain estimates of the entire vehicle trajectory and feature locations. Upon receipt of a packet request from MOOS, the packet relative factor-graph is built, optimized and then the convex optimization is solved and the packet contents are sent back to MOOS for transmission.

For a demonstration of the full system please refer to the attached video [1].

A two-vehicle simulation is shown in Fig. 5. In the top panel is the shoreside viewer that shows the ground truth locations of the discovered features (blue squares) as well as ground truth and estimated locations for each of the two vehicles. The decentralized multi-vehicle estimates maintained onboard each vehicle are shown in the LCM viewers at the bottom.

Fig. 6 shows the trace of the multi-AUV covariance matrix for three cases: (i) no communications, (ii) the proposed communication strategy, and (iii) a full bandwidth simulation which would be unachievable in reality. As evident, the position uncertainty is higher than the full bandwidth solution, but is nevertheless bounded, which is not the case for the no communications scenario. Note that the covariances in all cases grow until about 100s at which time vehicles start to observe mutual features. The fluctuations in the covariances are in general caused by the random placement of the features. During periods of few feature detections, uncertainties grow. Furthermore, Fig. 7 depicts the root mean square error (RMSE) of the position estimate as well as the 3σ bound which clearly shows to be bounded and consistent, thus satisfying the stated objectives.

VII. Conclusions and Future Work

We have presented an underwater C-SLAM framework for multiple AUVs communicating only through the unreliable and low bandwidth acoustic channel. Transmitted packet sizes are reduced by marginalization of unnecessarily

[1]Video also available at `http://people.csail.mit.edu/lpaull/projects/AUVCSLAM.xhtml`

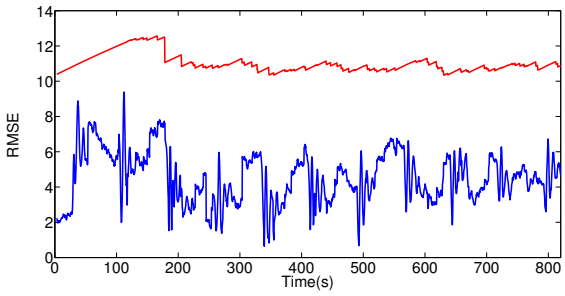

Fig. 7. Root mean squared error (RMSE) vs. time (blue) and 3σ bounds (red). It is clear that the proposed C-SLAM is able to bound the position errors and maintains consistency.

local variables followed by sparsification formulated as a constrained convex optimization. This enables us to bound the vehicles' position uncertainties without any pre-installed beacon infrastructure or having to surface for GPS data. This framework enables long-term deployment of teams of AUVs into a variety of less accessible environments such under ice and in the deep ocean.

REFERENCES

[1] S. Roumeliotis and G. Bekey, "Distributed multirobot localization," *Robotics and Automation, IEEE Transactions on*, vol. 18, no. 5, pp. 781–795, Oct. 2002.

[2] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, may 2010, pp. 3185–3192.

[3] L. Paull, S. Saeedi, M. Seto, and H. Li, "AUV navigation and localization: A review," *Oceanic Engineering, IEEE Journal of*, vol. 39, no. 1, pp. 131–149, Jan. 2014.

[4] S. Saeedi, L. Paull, M. Trentini, M. Seto, and H. Li, "Group mapping: A topological approach to map merging for multiple robots," *Robotics Automation Magazine, IEEE*, vol. 21, no. 2, pp. 60–72, June 2014.

[5] N. Carlevaris-Bianco, M. Kaess, and R. Eustice, "Generic node removal for factor-graph SLAM," *Robotics, IEEE Transactions on*, vol. 30, no. 6, pp. 1371–1385, Dec 2014.

[6] G. Huang, M. Kaess, and J. Leonard, "Consistent sparsification for graph optimization," in *Mobile Robots (ECMR), 2013 European Conference on*, Sept 2013, pp. 150–157.

[7] J. Vial, H. Durrant-Whyte, and T. Bailey, "Conservative sparsification for efficient and consistent approximate estimation," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Sept 2011, pp. 886–893.

[8] M. Mazuran, G. D. Tipaldi, L. Spinello, and W. Burgard, "Nonlinear graph sparsification for SLAM," in *Robotics Science and Systems Conference*, 2014, pp. 1–8.

[9] R. M. Eustice, H. Singh, J. J. Leonard, and M. R. Walter, "Visually mapping the RMS Titanic: Conservative covariance estimates for SLAM information filters," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1223–1242, 2006.

[10] I. T. Ruiz, S. de Raucourt, Y. Petillot, and D. Lane, "Concurrent mapping and localization using sidescan sonar," *Oceanic Engineering, IEEE Journal of*, vol. 29, no. 2, pp. 442–456, Apr. 2004.

[11] F. S. Hover, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Int. J. of Robotics Research*, vol. 31, no. 12, pp. 1445–1464, 2012.

[12] S. Barkby, S. B. Williams, O. Pizarro, and M. V. Jakuba, "Bathymetric particle filter SLAM using trajectory maps," *International Journal of Robotics Research*, vol. 31, no. 12, pp. 1409–1430, 2012.

[13] M. F. Fallon, G. Papadopoulos, J. J. Leonard, and N. M. Patrikalakis, "Cooperative AUV navigation using a single maneuvering surface craft," *International Journal of Robotics Research*, vol. 29, no. 12, pp. 1461–1474, Aug. 2008.

[14] A. Bahr, M. Walter, and J. Leonard, "Consistent cooperative localization," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, May 2009, pp. 3415–3422.

[15] J. M. Walls and R. M. Eustice, "An exact decentralized cooperative navigation algorithm for acoustically networked underwater vehicles with robustness to faulty communication: Theory and experiment," in *Proceedings of the Robotics: Science & Systems Conference*, Berlin, Germany, June 2013.

[16] M. Fallon, G. Papadopoulos, and J. Leonard, "A measurement distribution framework for cooperative navigation using multiple AUVs," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, May 2010, pp. 4256–4263.

[17] L. Paull, M. Seto, and J. Leonard, "Decentralized cooperative trajectory estimation for large teams of autonomous underwater vehicles," in *Intelligent Robots and Systems (To Appear)*, 2014, pp. 1–8.

[18] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *International Journal of Robotics Research*, vol. 25, no. 12, pp. 1243–1256, December 2006.

[19] A. I. Mourikis and S. I. Roumeliotis, "Predicting the performance of cooperative simultaneous localization and mapping (C-SLAM)," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1273–1286, 2006.

[20] S. Saeedi, L. Paull, M. Trentini, and H. Li, "Neural network-based multiple robot simultaneous localization and mapping," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2376–2387, 2011.

[21] X. Zhou and S. Roumeliotis, "Multi-robot SLAM with unknown initial correspondence: The robot rendezvous case," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, Oct 2006, pp. 1785–1792.

[22] A. Cunningham, M. Paluri, and F. Dellaert, "DDF-SAM: Fully distributed SLAM using constrained factor graphs," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, Oct 2010, pp. 3025–3030.

[23] A. Cunningham, K. Wurm, W. Burgard, and F. Dellaert, "Fully distributed scalable smoothing and mapping with robust multi-robot data association," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 1093–1100.

[24] M. Lazaro, L. Paz, P. Pinies, J. Castellanos, and G. Grisetti, "Multi-robot SLAM using condensed measurements," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 1069–1076.

[25] G. Grisetti, R. Kummerle, and K. Ni, "Robust optimization of factor graphs by using condensed measurements," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 581–588.

[26] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous location and mapping via square root information smoothing," *Int. J. of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, 2006.

[27] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[28] H. Kretzschmar and C. Stachniss, "Information-theoretic compression of pose graphs for laser-based SLAM," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1219–1230, 2012. [Online]. Available: http://ijr.sagepub.com/content/31/11/1219.abstract

[29] V. Ila, J. Porta, and J. Andrade-Cetto, "Information-based compact pose SLAM," *Robotics, IEEE Transactions on*, vol. 26, no. 1, pp. 78–93, Feb 2010.

[30] H. Johannsson, M. Kaess, M. Fallon, and J. Leonard, "Temporally scalable visual SLAM using a reduced pose graph," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 54–61.

[31] S. E. Webster, R. M. Eustice, H. Singh, and L. L. Whitcomb, "Advances in single-beacon one-way-travel-time acoustic navigation for underwater vehicles," *The International Journal of Robotics Research*, vol. 31, no. 8, pp. 935–950, 2012.

[32] C. Sanderson, D. Gibbins, and S. Searle, "On statistical approaches to target silhouette classification in difficult conditions." *Digital Signal Processing*, vol. 18, no. 3, pp. 375–390, 2008.

[33] M. Benjamin, P. Newman, H. Schmidt, and J. Leonard, "An overview of MOOS-IvP and a brief users guide to the IvP Helm autonomy software," http:// dspace. mit. edu/ bitstream/ handle/ 1721.1/45569/ MIT-CSAIL-TR-2009-028.pdf, June 2009.

[34] T. Schneider and H. Schmidt, "Model-based adaptive behavior framework for optimal acoustic communication and sensing by marine robots," *Oceanic Engineering, IEEE Journal of*, vol. 38, no. 3, pp. 522–533, July 2013.

[35] A. Huang, E. Olson, and D. Moore, "LCM: Lightweight communications and marshalling," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, Oct 2010, pp. 4057–4062.