Soft Range Information for Network Localization

Santiago Mazuelas, *Senior Member, IEEE,* Andrea Conti, *Senior Member, IEEE,* Jeffery C. Allen, *Member, IEEE,* and Moe Z. Win, *Fellow, IEEE*

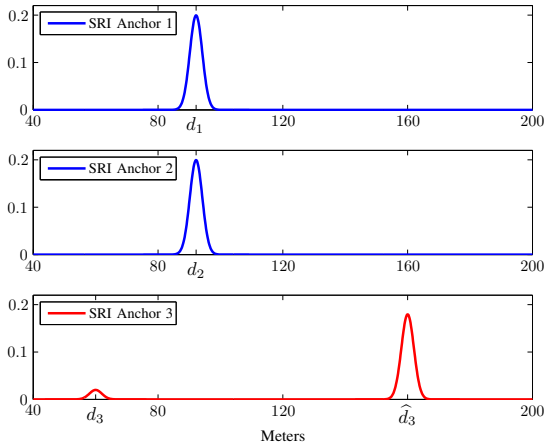
***Abstract*—The demand for accurate localization in complex environments continues to increase despite the difficulty in extracting positional information from measurements. Conventional range-based localization approaches rely on distance estimates obtained from measurements (e.g., delay or strength of received waveforms). This paper goes one step further and develops localization techniques that rely on all probable range values rather than on a single estimate of each distance. In particular, the concept of soft range information (SRI) is introduced, showing its essential role for network localization. We then establish a general framework for SRI-based localization and develop algorithms for obtaining the SRI using machine learning techniques. The performance of the proposed approach is quantified via network experimentation in indoor environments. The results show that SRI-based localization techniques can achieve performance approaching the Cramér–Rao lower bound and significantly outperform the conventional techniques especially in harsh wireless environments.**

***Index Terms*—Soft range information, network localization, wireless propagation, machine learning.**

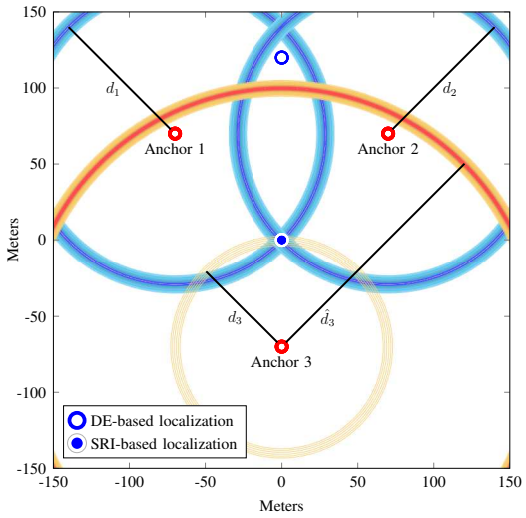
I. Introduction

NETWORK LOCALIZATION [1] is a key enabler for numerous emerging applications–including autonomous vehicles [2], logistics [3], smart cities [4], distributed sensing [5], environmental monitoring [6], public safety [7], medical services [8], and social networks [9]–that require highly accurate positional information [10]–[25]. However, harsh propagation environments such as indoors or urban canyons hinder accurate localization [26]–[34]. In particular, non-line-of-sight (NLOS) and multipath propagation prevent the extraction of reliable positional information from wireless signals using classical techniques [35]–[45].

In range-based network localization, the agents' positions are determined from measurements related to pair-wise distances and prior knowledge such as anchors' positions (see e.g., Fig. 1). Conventional approaches typically process these measurements to obtain distance estimates (DEs), which are



(a) Range-related measurements from anchors 1 and 2 accurately disambiguate the distances d_1 and d_2, while the range-related measurement from anchor 3 leads to a highly biased DE $\widehat{d}_3$.



(b) Localization based on DEs is inaccurate since the third anchor provides a highly biased DE. On the other hand, localization based on SRI is highly accurate since the information provided by the third anchor is fully utilized.

Fig. 1. Examples of conventional localization based on DE and proposed approach based on SRI.

Manuscript received October 17, 2016; revised May 17, 2017, August 25, 2017, and October 30, 2017; accepted November 13, 2017. Date of publication Month Day, 2018; date of current version Month Day, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ruixin Niu. This research was supported, in part, by the Office of Naval Research under Grants N00014-11-1-0397, N00014-16-1-2141, and N62909-18-1-2017, the National Science Foundation under Grant CCF-1116501, the Basque Government through the BERC 2014-2017, the Spanish Ministry of Economy and Competitiveness MINECO under Ramon y Cajal Grant RYC-2016-19383. The material in this paper was presented, in part, at the 2013 IEEE International Conference on Communications. *(Corresponding author: Santiago Mazuelas).*

S. Mazuelas was with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA, and he is now with the BCAM-Basque Center for Applied Mathematics, Spain (e-mail: smazuelas@bcamath.org).

A. Conti is with the Department of Engineering and CNIT, University of Ferrara, Italy (e-mail: a.conti@ieee.org).

J. C. Allen is with Space and Naval Warfare Systems Center Pacific (SSC Pacific), USA (e-mail: jeffery.allen@navy.mil).

M. Z. Win is with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, USA (e-mail: moewin@mit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.YYYY.ZZZZZZZ

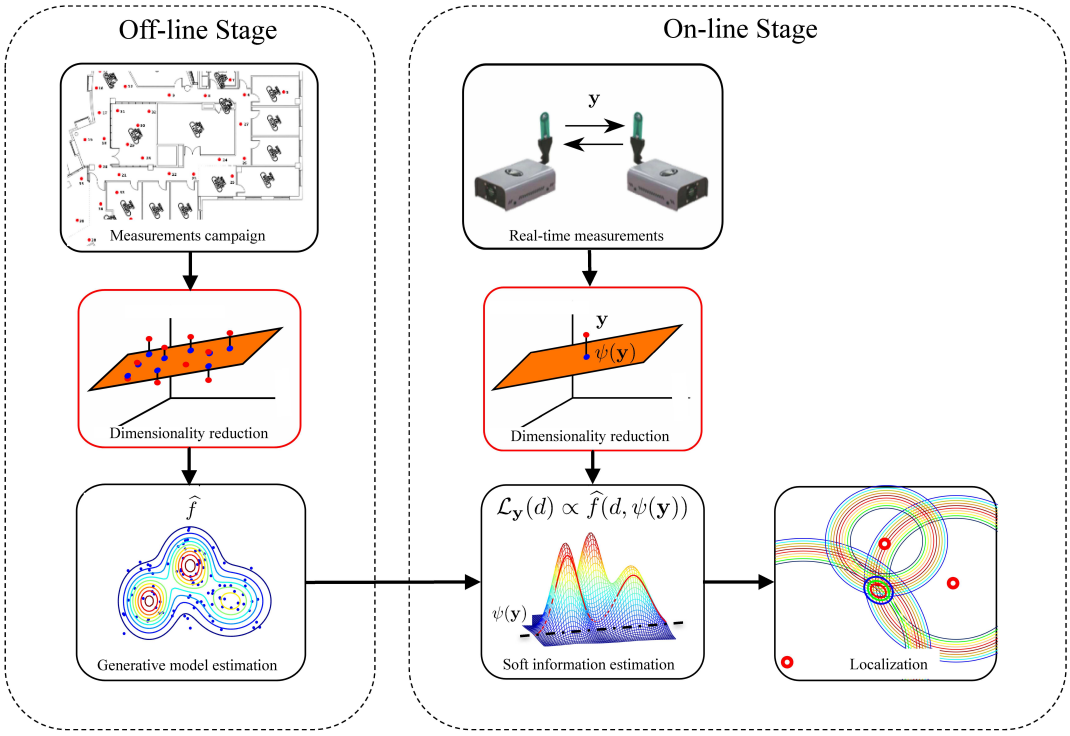


Fig. 2. Methodology for SRI-based localization.

subsequently used to determine the agents' positions [16]–[19]. As the harsh propagation conditions directly affect the measurements, these approaches focus on improving the DE to enhance localization performance. Most of the effort is dedicated to estimate the time-of-arrival (TOA) of direct-paths or to correct TOA of indirect-paths. Specifically, these approaches first identify NLOS conditions and then mitigate their effects. Such mitigation is accomplished by removing the positive bias due to NLOS propagation [34]–[37] or assigning different weights to line-of-sight (LOS) and NLOS DEs [39].

Most of the aforementioned approaches assume a fixed, often inaccurate, model for the relationship between the inter-node measurement and the distance (e.g., Gaussian distribution with a mean equal to the distance and a given variance [26], [35]). Extensions of such techniques adapt the model by varying the standard deviations of the Gaussian distributions [40], by using a few recent measurements [31], and by assigning samples of the power dispersion profile with different direct-path probabilities [46]. The limitations of conventional approaches based on DEs can be observed in the scenario depicted in Fig. 1, which describes the localization of an agent using range-related measurements with respect to 3 anchors. This example depicts a scenario in which the inter-node measurements from anchor 3 provide a high likelihood for distances much larger than the actual distance (see Fig. 1(a)). Therefore, any technique based on DEs would lead to inaccurate localization (see Fig. 1(b)).

We envision a new paradigm for high-accuracy localization that relies on the statistical characterization of the relationship between the inter-node measurements and ranges, hereafter referred to as soft range information (SRI). The main goal of this work is to design localization techniques that exploit SRI and to quantify their performance gain with respect to conventional localization techniques based on DEs. In the scenario depicted in Fig. 1, SRI-based localization provides accurate agent's position as the range information is fully utilized (see Fig. 1(b)). In particular, the SRI-based approach can account for the small but non-negligible likelihood of the actual distance from anchor 3 (see Fig. 1(a)).

In this paper, we propose a framework and develop algorithms for SRI-based network localization as depicted in Fig. 2. The main contributions of the paper can be summarized as follows:

- establishment of a general framework for SRI-based network localization;
- development of algorithms for determining the SRI via machine learning; and
- quantification through network experimentation of the benefits offered by SRI-based algorithms.

The remainder of the paper is organized as follows. Section II introduces the general methodology of SRI-based network localization and illustrates SRI benefits in a simple scenario. Section III develops algorithms for SRI estimation using machine learning. Section IV presents the performance of SRI-based localization via network experimentation. Finally, conclusions are given in Section V.

Notations: random variables (RVs) are displayed in sans serif, upright fonts and their realizations in serif, italic fonts; vectors are denoted by bold lowercase letters; a RV and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$; $\boldsymbol{x}[j]$ denotes the jth component of the vector $\boldsymbol{x}$; $f_{\mathbf{x}}(\boldsymbol{x})$ and, for brevity when

possible, $f(\boldsymbol{x})$ denote the Radon–Nikodym derivative of a RV x with respect to the base measure (Lebesgue for continuous RVs and counting for discrete RVs), e.g., $f(\boldsymbol{x})$ denotes a probability density function (PDF) in case of a continuous RV x; $f(\boldsymbol{x}|\boldsymbol{z})$ denotes either the conditional distribution of x given $\mathsf{z} = \boldsymbol{z}$ for a RV z or the distribution of x parametrized by $\boldsymbol{z}$ for a parameter $\boldsymbol{z}$; $\varphi(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the PDF of a Gaussian RV x with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathbb{E}\{\cdot\}$, $\mathbb{V}\{\cdot\}$, and $\mathbb{P}\{\cdot\}$ denote, respectively, the expectation, variance, and probability of the argument; $[\,\cdot\,]^{\mathrm{T}}$ denotes the transpose of the argument.

II. Soft Range Information

In this section, we first define SRI and describe range-based localization relying on SRI rather than DE. Then, we propose a general methodology for network localization based on accurate SRI estimation, which is illustrated in Fig. 2.

In range-based localization, nodes in a network obtain positional information from measurements related to the distance between pairs of nodes. Non-cooperative approaches use measurements related to the distance from agents (nodes at positions to be estimated) to anchors (nodes at known positions), while cooperative approaches additionally use measurements between agents. Localization approaches can also be classified, based on agent position model, as non-Bayesian and Bayesian: the former directly obtain position estimates from measurements, while the latter first determine posterior distributions of positions and then use them to estimate the positions.

Definition 1 (Soft range information): Let $f(\boldsymbol{y}|d)$ be the distribution of range-related measurements set y conditioned on or parametrized by the distance between a pair of nodes. The SRI of a measurements set[1] $\mathsf{y} = \boldsymbol{y}$, denoted $\mathcal{L}_{\boldsymbol{y}}(d)$, is any function of distance d proportional to $f(\boldsymbol{y}|d)$, i.e., $\mathcal{L}_{\boldsymbol{y}}(d) \propto f(\boldsymbol{y}|d)$.

In contrast to conventional approaches for range-based localization relying on DE *values*, this paper proposes to first obtain SRIs from range-related measurements and then use those distance *functions* to determine node positions.[2] In fact, SRI provides richer information than DE by quantifying the odds of all possible distances, thus enabling *soft-decision localization* instead of conventional hard-decision localization.[3] This paper focuses on range-related measurements, however the methodology introduced here can analogously be used for measurements related to other positional features including angle, velocity, and acceleration. For a general positional feature θ, the corresponding soft information (SI) of a θ-related measurements set $\boldsymbol{y}$ would be a function of θ proportional to $f(\boldsymbol{y}|\theta)$ [47].

[1]In this paper, measurements set refers to a collection of observations, possibly of different types.

[2]Certain DE-based approaches including [26], [35], [40] characterize the relationship between the distance d and its estimate $\widehat{d}$ with a likelihood function of d centered at $\widehat{d}$. Those approaches differ from the SRI-based approach as the latter directly estimates the function $\mathcal{L}_{\boldsymbol{y}}(d)$ from measurements.

[3]The SRI can be used in both Bayesian and non-Bayesian formulations. In the latter case, the SRI coincides with the distance likelihood function. Note also that the SRI is defined up to a proportionality constant as this is sufficient for localization purposes.

Consider a network formed by N_{a} agents and N_{b} anchors with index sets denoted by $\mathcal{N}_{\mathrm{a}} = \{1, 2, \ldots, N_{\mathrm{a}}\}$ and $\mathcal{N}_{\mathrm{b}} = \{N_{\mathrm{a}} + 1, N_{\mathrm{a}} + 2, \ldots, N_{\mathrm{a}} + N_{\mathrm{b}}\}$, respectively. The network acquires measurements sets, where each set is composed of M scalars related to the distance between a pair of nodes. Such values can include TOA, received signal strength (RSS), and waveform samples or any combination thereof. Let $\boldsymbol{p}_i$ denote the position of node i; $d_{i,j} \in \mathbb{R}$ and $\boldsymbol{y}_{i,j} \in \mathbb{R}^M$ denote, respectively, the distance and a measurements set between nodes i and j; $d^{(k)}$ and $\boldsymbol{y}^{(k)}$ denote, respectively, the kth distance and the kth measurements set in a collection of distances and measurements sets;[4] and, d and $\boldsymbol{y}$ denote, respectively, the distance and a measurements set between an unspecified pair of nodes. Range-related measurements are considered, i.e., $f(\boldsymbol{y}_{i,j}|\{\boldsymbol{p}_k\}_{k\in\mathcal{N}_{\mathrm{a}}}) = f(\boldsymbol{y}_{i,j}|d_{i,j})$ for each pair of nodes i and j, and mutually independent measurements given distances, i.e., $f(\boldsymbol{y}_{i,j}, \boldsymbol{y}_{k,r}|d_{i,j}, d_{k,r}) = f(\boldsymbol{y}_{i,j}|d_{i,j}) f(\boldsymbol{y}_{k,r}|d_{k,r})$ for any two pairs of nodes $\{i, j\} \neq \{k, r\}$.

A. Benefits of SRI-based localization

A simple case study is presented to provide insights into how SRI offers richer information than DE for localization. Specifically, a localization system is considered, in which each measurements set $\boldsymbol{y} = [r, \delta]^{\mathrm{T}}$ is a collection of two values (a distance measurement r and a NLOS indicator δ). In particular, the distance measurement r is an instantiation of

$$\mathsf{r} = d + \mathsf{n} \tag{1}$$

where d is the distance between a pair of nodes and n is the measurement noise with PDF given by

$$f_{\mathsf{n}}(n) = \begin{cases} \varphi(n; 0, \sigma^2_{\mathrm{LOS}}) & \text{for LOS cases} \\ \varphi(n; b, \sigma^2_{\mathrm{NLOS}}) & \text{for NLOS cases} \end{cases} \tag{2}$$

where b is a positive bias due to NLOS conditions. The other component, δ, of the measurements set is the NLOS detector outcome with $\delta = 0$ and 1 corresponding to detected LOS and NLOS conditions, respectively. The error in detecting the propagation condition is accounted for by means of posterior probabilities of error

$$\epsilon_{\mathrm{NLOS}} = \mathbb{P}\{\mathrm{NLOS}|\delta = 0\} \tag{3a}$$

$$\epsilon_{\mathrm{LOS}} = \mathbb{P}\{\mathrm{LOS}|\delta = 1\}\,. \tag{3b}$$

The detector output and its errors are considered independent of the distance.

The SRI corresponding to a measurements set $\boldsymbol{y} = [r, \delta]^{\mathrm{T}}$, described in (1)–(3), is (see Appendix A)

$$\mathcal{L}_{\boldsymbol{y}}(d) \propto \begin{cases} (1 - \epsilon_{\mathrm{NLOS}})\mathcal{L}_{\mathrm{LOS}}(d) + \epsilon_{\mathrm{NLOS}}\mathcal{L}_{\mathrm{NLOS}}(d) & \text{for } \delta = 0 \\ \epsilon_{\mathrm{LOS}}\mathcal{L}_{\mathrm{LOS}}(d) + (1 - \epsilon_{\mathrm{LOS}})\mathcal{L}_{\mathrm{NLOS}}(d) & \text{for } \delta = 1 \end{cases} \tag{4}$$

with $\mathcal{L}_{\mathrm{LOS}}(d) = \varphi(r; d, \sigma^2_{\mathrm{LOS}})$ and $\mathcal{L}_{\mathrm{NLOS}}(d) = \varphi(r; d + b, \sigma^2_{\mathrm{NLOS}})$. Note that when the NLOS detector is highly reliable ($\epsilon_{\mathrm{LOS}} \simeq 0$ and $\epsilon_{\mathrm{NLOS}} \simeq 0$), the SRI is concentrated around the actual distance. Moreover, the SRI is robust to unreliable

[4]Negative values of k indicate values obtained during a training phase.

NLOS detectors; that is, SRI has positive values for both the actual distance and the biased distance weighted appropriately by the error probabilities of the NLOS detector given in (3).

The minimum mean square error (MMSE) distance estimator is obtained by modeling distances as RVs. Assuming a constant reference prior [48] for distances, such DE corresponding to a measurements set $\boldsymbol{y} = [r, \delta]^{\mathrm{T}}$, described in (1)–(3), is (see Appendix A)

$$\widehat{d} = \mathbb{E}\{\mathsf{d}|r,\delta\} = \begin{cases} (1-\epsilon_{\mathrm{NLOS}})r + \epsilon_{\mathrm{NLOS}}(r-b) & \text{for } \delta = 0 \\ \epsilon_{\mathrm{LOS}} r + (1-\epsilon_{\mathrm{LOS}})(r-b) & \text{for } \delta = 1\,. \end{cases} \tag{5}$$

Note that when the NLOS detector is highly reliable ($\epsilon_{\mathrm{LOS}} \simeq 0$ and $\epsilon_{\mathrm{NLOS}} \simeq 0$), the bias of NLOS cases is correctly subtracted to refine ranging [34]. However, in the presence of a NLOS detector error, the DEs are biased by $(1-\epsilon_{\mathrm{NLOS}})b$ in NLOS cases and by $-(1-\epsilon_{\mathrm{LOS}})b$ in LOS cases.[5] In addition, the distance estimator mean squared error (MSE) for scenarios with $\sigma_{\mathrm{LOS}} = \sigma_{\mathrm{NLOS}} = \sigma$ and $\epsilon_{\mathrm{LOS}} = \epsilon_{\mathrm{NLOS}} = \epsilon$ is (see Appendix A)

$$\mathrm{MSE}(\widehat{d}) = \epsilon\,(1-\epsilon)\,b^2 + \sigma^2 \tag{6}$$

which reduces to the MSE of LOS scenarios when the NLOS detector is totally reliable ($\epsilon = 0$).

The Fisher information inequality, also known as Cramér–Rao lower bound (CRLB), for a network with one agent at position $\boldsymbol{p}$ is

$$\mathbb{E}\{\|\widehat{\mathbf{p}} - \boldsymbol{p}\|^2\} \geqslant \mathrm{Tr}\{\boldsymbol{J}^{-1}\}$$

where $\widehat{\mathbf{p}}$ is a position estimator, $\mathrm{Tr}\{\cdot\}$ denotes the matrix trace, and $\boldsymbol{J}$ is the Fisher information matrix (FIM) given by

$$\boldsymbol{J} = \sum_{j\in\mathcal{N}_{\mathrm{b}}} \frac{\lambda_j}{d_j^2}(\boldsymbol{p}-\boldsymbol{p}_j)(\boldsymbol{p}-\boldsymbol{p}_j)^{\mathrm{T}}$$

in which d_j is the distance between agent and anchor j, and λ_j is the ranging information intensity (RII) of the measurements set $\mathbf{y}_j$ related to d_j [10], [30]. The RII determines the CRLB and can be obtained as shown in the following.

Proposition 1: The RII of a measurements set $\mathbf{y} = [\mathsf{r}, \boldsymbol{\delta}]$ described in (1)–(3) with $\sigma_{\mathrm{LOS}} = \sigma_{\mathrm{NLOS}} = \sigma$ and $\epsilon_{\mathrm{LOS}} = \epsilon_{\mathrm{NLOS}} = \epsilon$ is

$$\lambda = \frac{1}{\sigma^4}\left[\mathbb{E}\left\{(\mathsf{r}-d-\chi_0)^2|\boldsymbol{\delta}=0\right\}\mathbb{P}\{\boldsymbol{\delta}=0\} + \mathbb{E}\left\{(\mathsf{r}-d-\chi_1)^2|\boldsymbol{\delta}=1\right\}\mathbb{P}\{\boldsymbol{\delta}=1\}\right] \tag{7}$$

where

$$\chi_0 = \frac{b\,\epsilon\,\varphi(r;d+b,\sigma^2)}{(1-\epsilon)\,\varphi(r;d,\sigma^2) + \epsilon\,\varphi(r;d+b,\sigma^2)} \tag{8}$$

$$\chi_1 = \frac{b\,(1-\epsilon)\,\varphi(r;d+b,\sigma^2)}{\epsilon\,\varphi(r;d,\sigma^2) + (1-\epsilon)\,\varphi(r;d+b,\sigma^2)}\,. \tag{9}$$

Proof: See Appendix B. □

Remark 1: As expected, if $\mathbb{P}\{\boldsymbol{\delta}=0\} = 1$ and $\epsilon = 0$ (LOS scenario with totally reliable NLOS detector), the RII reduces to $\lambda = 1/\sigma^2$. For values $0 < \epsilon < 1$, the RII in (7), and, hence, the CRLB does not have closed-form expressions; however, expectations in (7) can be evaluated through numerical integration.

The analytical expression in Proposition 1 also reveals that NLOS detector errors do not fundamentally restrain performance as described in the following.

Remark 2: For values of the bias significantly larger than the standard deviation of measurement noise n, i.e., $b \gg \sigma$, the RII in (7) and, hence, the CRLB is approximately the one of LOS scenario with totally reliable NLOS detector ($\lambda \simeq 1/\sigma^2$) independent of the detector's reliability ϵ. Note that when $b \gg \sigma$, the two Gaussian PDFs in (8) and (9) have negligible overlap. Hence $\chi_0 \simeq \chi_1 \simeq 0$ (resp. $\chi_0 \simeq \chi_1 \simeq b$) when r has mean d, i.e., LOS cases, (resp. $d+b$, i.e., NLOS cases) and standard deviation σ. Therefore, both expectations in (7) are approximately σ^2, for instance

$$\begin{aligned}\mathbb{E}\left\{(\mathsf{r}-d-\chi_0)^2|\boldsymbol{\delta}=0\right\} &= (1-\epsilon)\int (r-d-\chi_0)^2\varphi(r;d,\sigma^2)dr \\ &\quad + \epsilon\int (r-d-\chi_0)^2\varphi(r;d+b,\sigma^2)dr \\ &\simeq (1-\epsilon)\,\sigma^2 + \epsilon\,\sigma^2 = \sigma^2\,.\end{aligned}$$

The SRI encapsulates information not only about the most likely distance but also about other probable distances, and this richer information results in enhanced localization. In particular, in the system described above, a localization network based on SRIs would be resilient to NLOS detector errors that would significantly harm localization based on DEs. Fig. 3 shows the root mean square (RMS) error as a function of NLOS detector error probabilities for DE-based localization and SRI-based localization, along with the CRLB. In particular, we consider a localization network as depicted in Fig.1(b) with measurements sets between the agent and each anchor according to (1)–(3) with $b = 100\,\mathrm{m}$ and $\sigma = 2\,\mathrm{m}$. In the simulated scenario, $20\,\%$ of the measurements correspond to NLOS cases, and NLOS detectors for anchors 1 and 2 are totally reliable ($\epsilon = 0$) while that for anchor 3 has a varying probability of error ϵ. The agent position is estimated from DEs and SRIs using maximum likelihood (ML) criterion (see further details in Section II-B). Specifically, DE provides two numerical values (5) and (6) for the mean and variance of a Gaussian distribution that is used in ML positioning. On the other hand, SRI provides functions (4) that are directly used in ML positioning.[6]

Fig. 3 shows that SRI-based localization outperforms DE-based localization and approaches the performance given by the theoretical benchmark. It can be seen that the performance of DE-based localization matches that of SRI-based localization for cases with $\epsilon \simeq 0$, since in those cases the SRI is a Gaussian centered at the DE. It can also be observed that the performance of SRI-based localization attains the CRLB even in the presence of erroneous NLOS detections, while that of DE-based localization is highly harmed by those errors. Recall from Fig. 1(b) that every time a NLOS case is

[5]These kinds of biases would arise in any DE method since the distance is expected to be near r (resp. $r-b$) for cases detected as LOS (resp. NLOS).

[6]To ensure localization performance independent of specific algorithm implementations, the positions achieving maxima are obtained by exhaustive search over a regular grid.

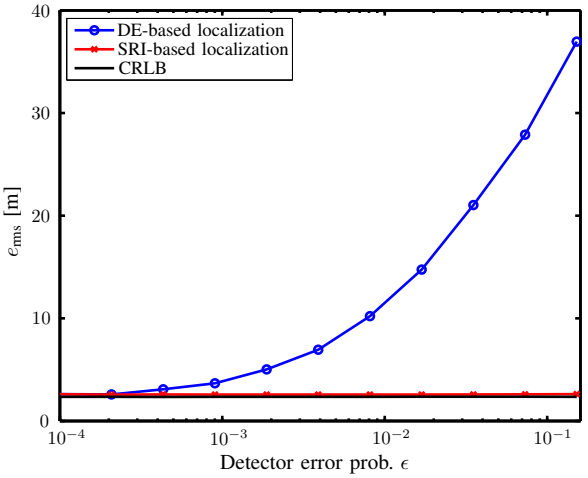


Fig. 3. The richer information provided by SRI results in improved accuracies. In the simulated case study, SRI-based localization achieves performances near the CRLB also with erroneous NLOS detection.

incorrectly detected as LOS, the localization error based on DE is around 120 m, while that based on SRI continues to be around 2 m. Both the CRLB and the SRI-based localization are insensitive to the detector's probability of error. For the CRLB, this fact is a consequence of Remark 2, while for SRI-based localization, it is a consequence of the expression of SRI in (4), which assigns non-negligible values to the actual distance for any detector's output and probability of error. Those non-negligible values get amplified after multiplication with the SRIs from anchors 1 and 2. Note that the benefits of SRI can also be achieved by using approximate SRIs. Even a rough approximation of the SRI without accurate values of b, ϵ_{LOS}, ϵ_{NLOS}, σ_{LOS}, and σ_{NLOS} would result in a more accurate localization than that based on DE. The final position estimate will be accurate as long as the approximated SRI from anchor 3 has non-negligible values around the actual distance (see Fig. 1).

The benefits of SRI-based localization come from the fact that the distance function offered by a SRI contains richer localization information than the distance value offered by a DE. Therefore, SRI-based localization is expected to outperform DE-based localization in all the situations where the information contained in the measurements cannot be fully encapsulated by a DE.

B. SRI in network localization

We now describe how SRI can be exploited to estimate the agents' positions in different localization settings.[7]

1) Non-cooperative non-Bayesian localization: Consider a scenario in which agents' positions are modeled as unknown parameters and are estimated based on measurements with respect to the anchors. The ML position estimate for the agent $i \in \mathcal{N}_{\text{a}}$ is given by

$$\widehat{\boldsymbol{p}}_i = \arg\max_{\boldsymbol{p}_i} f(\{\boldsymbol{y}_{i,j}\}_{j\in\mathcal{N}_{\text{b}}}|\boldsymbol{p}_i) = \arg\max_{\boldsymbol{p}_i} \prod_{j\in\mathcal{N}_{\text{b}}} f(\boldsymbol{y}_{i,j}|d_{i,j})\,. \tag{10}$$

If the distributions $f(\boldsymbol{y}_{i,j}|d_{i,j})$ are Gaussian with mean $d_{i,j}$, then the ML estimate leads to the least squares (LS) estimate or to the weighted least squares (WLS) estimate when the distributions have the same or different variances, respectively.

2) Non-cooperative Bayesian localization: Consider a scenario in which agents' positions are modeled as RVs and are estimated based on measurements with respect to the anchors. The MMSE and the maximum *a posteriori* probability (MAP) position estimates for the agent $i \in \mathcal{N}_{\text{a}}$ are given respectively by

$$\widehat{\boldsymbol{p}}_i = \int \boldsymbol{p}_i f(\boldsymbol{p}_i|\{\boldsymbol{y}_{i,j}\}_{j\in\mathcal{N}_{\text{b}}}) d\boldsymbol{p}_i \tag{11}$$

$$\widehat{\boldsymbol{p}}_i = \arg\max_{\boldsymbol{p}_i} f(\boldsymbol{p}_i|\{\boldsymbol{y}_{i,j}\}_{j\in\mathcal{N}_{\text{b}}}) \tag{12}$$

which compute the mean and mode of the posterior distribution

$$f(\boldsymbol{p}_i|\{\boldsymbol{y}_{i,j}\}_{j\in\mathcal{N}_{\text{b}}}) \propto f(\boldsymbol{p}_i) \prod_{j\in\mathcal{N}_{\text{b}}} f(\boldsymbol{y}_{i,j}|d_{i,j})\,. \tag{13}$$

If the prior distribution of $\mathbf{p}_i$ is constant, then the MAP estimate coincides with the ML estimate.

3) Cooperative non-Bayesian localization: Consider a scenario in which agents' positions are modeled as unknown parameters and are estimated based on measurements with respect to the anchors as well as to the neighboring agents. The ML positions estimates for the agents are given by

$$\begin{aligned}\{\widehat{\boldsymbol{p}}_i\}_{i\in\mathcal{N}_{\text{a}}} &= \arg\max_{\{\boldsymbol{p}_i\}} f(\{\boldsymbol{y}_{i,j}\}_{\substack{i\in\mathcal{N}_{\text{a}}\\ j\in\mathcal{N}_{\text{a}}\cup\mathcal{N}_{\text{b}}}}|\{\boldsymbol{p}_i\}_{i\in\mathcal{N}_{\text{a}}}) \\ &= \arg\max_{\{\boldsymbol{p}_i\}} \prod_{\substack{i\in\mathcal{N}_{\text{a}}\\ j\in\mathcal{N}_{\text{a}}\cup\mathcal{N}_{\text{b}}}} f(\boldsymbol{y}_{i,j}|d_{i,j})\,.\end{aligned} \tag{14}$$

If the distributions $f(\boldsymbol{y}_{i,j}|d_{i,j})$ are Gaussian with mean $d_{i,j}$, then the ML estimate leads to the LS estimate or to the WLS estimate when the distributions have the same or different variances, respectively.

4) Cooperative Bayesian localization: Consider a scenario in which agents' positions are modeled as RVs and are estimated based on measurements with respect to the anchors as well as to the neighboring agents. The MMSE and the MAP position estimates for the agents are given respectively by (11) and (12) using the posterior distribution[8]

$$f(\{\boldsymbol{p}_i\}_{i\in\mathcal{N}_{\text{a}}}|\{\boldsymbol{y}_{i,j}\}_{\substack{i\in\mathcal{N}_{\text{a}}\\ j\in\mathcal{N}_{\text{a}}\cup\mathcal{N}_{\text{b}}}}) \propto \prod_{i\in\mathcal{N}_{\text{a}}} f(\boldsymbol{p}_i) \prod_{j\in\mathcal{N}_{\text{a}}\cup\mathcal{N}_{\text{b}}} f(\boldsymbol{y}_{i,j}|d_{i,j}) \tag{15}$$

The centralized implementation of such estimates can be impractical due to the high dimensionality of all nodes positions, and various distributed implementations have been developed based on belief propagation (BP) techniques [49]–[51]. In

[7]For notational convenience, we consider the case in which a measurements set $\boldsymbol{y}_{i,j}$ is available for each pair of nodes i and j. The expressions for other cases can be obtained by removing the terms corresponding to node pairs with unavailable measurements.

[8]In this expression, the prior knowledge about each agent's position is considered independent of other agents' positions.

particular, each agent i approximates the posterior distribution of $\mathbf{p}_i$ by exchanging messages with neighboring nodes. For instance, if $\mathcal{Y}$ is a collection of measurements sets, agent i can update its approximated marginal posterior distribution $\widehat{f}(\boldsymbol{p}_i|\mathcal{Y})$ by using a new measurements set $\boldsymbol{y}_{i,j}$ as

$$\widehat{f}(\boldsymbol{p}_i|\boldsymbol{y}_{i,j},\mathcal{Y}) \propto \widehat{f}(\boldsymbol{p}_i|\mathcal{Y}) \int f(\boldsymbol{y}_{i,j}|d_{i,j})\widehat{f}(\boldsymbol{p}_j|\mathcal{Y})d\boldsymbol{p}_j \quad (16)$$

where the second term, which involves an integral, is known as the message from agent j [49]. BP techniques assume the information contained in messages from different neighbors are conditionally independent given the positions [51], and each node iteratively updates the approximated marginal posterior distribution in a process known as message passing.

In summary, a general range-based localization system can be described by the following steps:

1) Acquisition of range-related measurements sets $\boldsymbol{y}_{i,j}$ between different pairs of nodes.
2) Characterization of the SRI corresponding to each measurements set, $\mathcal{L}_{\boldsymbol{y}_{i,j}}(d_{i,j}) \propto f(\boldsymbol{y}_{i,j}|d_{i,j})$.
3) Localization of agents using one of the techniques described above relying on SRIs obtained in step 2).

Despite the popularity of approaches based on DEs, inference techniques for range-based localization rely on SRIs as shown above. In the following, we propose a methodology to obtain accurate SRI estimates from measurements, and then present machine learning techniques to obtain such estimates.

C. Localization via estimated SRI

The SRIs can be estimated directly from measurements sets without the need for DEs using a methodology with two phases (see Algorithm 1). In an offline phase, the data generative model is learned by using a measurement campaign. In an online phase, the SRI for each new measurements set is estimated based on the generative model learned in the previous phase. Specifically, during the offline phase, training data with measurements sets and corresponding distances are used to learn their joint distribution $f(\boldsymbol{y}, d)$, i.e., generative model. During the online phase, the SRI of a new measurements set $\boldsymbol{y}^{(k)}$ for $k > 0$ can be obtained directly from the generative model as $\mathcal{L}_{\boldsymbol{y}^{(k)}}(d) \propto f(\boldsymbol{y}^{(k)}, d)$ in the absence of prior information on the distance (using a constant reference prior [48]).[9]

Estimating the generative model from training data is challenging, especially for measurements sets with high dimensionality. In fact, to obtain detailed knowledge of the probability distribution of a RV with moderately high dimensions often requires a large number of instantiations [52]. For high-dimensional measurement sets, e.g., waveform samples with high delay resolution, an additional step for dimensionality reduction is needed. Such a dimensionality reduction step can be described as a function ψ transforming a measurements set $\boldsymbol{y} \in \mathbb{R}^M$ into features $\psi(\boldsymbol{y}) \in \mathbb{R}^{M'}$ where $M' \ll M$. Dimensionality reduction does not necessarily involve distance estimation, while distance estimation can be thought of as a specific type of dimensionality reduction. Algorithm 2 describes SRI estimation with dimensionality reduction.

[9]SRI can be analogously obtained in a scenario with available prior knowledge about distance $f(d)$ as $\mathcal{L}_{\boldsymbol{y}^{(k)}}(d) \propto f(\boldsymbol{y}^{(k)}, d)/f(d)$.

Algorithm 1 – SRI estimation

Offline Phase

1: Acquire training data $\mathcal{T} = \{(\boldsymbol{y}^{(k)}, d^{(k)})\}_{k=-N}^{-1}$ through a measurement campaign.
2: Estimate the generative model as $\widehat{f}(\boldsymbol{y}, d)$.
3: Store the estimated generative model.

Online Phase

1: **for** $k \geqslant 0$ **do**
2: Acquire a new measurements set $\boldsymbol{y}^{(k)}$ at time t_k.
3: Estimate the SRI of the measurements set $\boldsymbol{y}^{(k)}$ from the stored generative model as $\mathcal{L}_{\boldsymbol{y}^{(k)}}(d) \propto \widehat{f}(\boldsymbol{y}^{(k)}, d)$.
4: **end for**

The proposed methodology can be used for any kind of range-related measurements, while the specific dimensionality reduction and generative model are technology-dependent as the RVs $\mathbf{y}$ corresponding to different technologies are not necessarily the same. In the following section we describe several techniques for both dimensionality reduction and generative model estimation based on machine learning. This diversity of techniques, each with different strengths and weaknesses, can offer the most adequate alternative for the specific choice of technology.

III. SRI ESTIMATION VIA MACHINE LEARNING

This section presents machine learning techniques for dimensionality reduction and generative model estimation. We first introduce three dimensionality reduction techniques: physical features (PFs), principal component analysis (PCA), and Laplacian eigenmap (LEM). Then, we present two density estimation techniques: the Fisher–Wald (FW) setting and kernel density estimation (KDE).

A. Dimensionality reduction

The purpose of dimensionality reduction is to find a mapping $\psi : \mathbb{R}^M \to \mathbb{R}^{M'}$ that transforms high-dimensional range-related measurements in $\mathbb{R}^M$ to low-dimensional features in $\mathbb{R}^{M'}$ with $M' \ll M$. In what follows, three dimensionality reduction techniques for SRI estimation are presented.

1) Physical features: PFs account for the intrinsic properties of the wireless link such as its strength, delay, and waveform shape. The PFs have been used to obtain DEs and to mitigate the effects of harsh propagation conditions [34], [36], [37], [39]–[41]. In this paper, the PFs are used to form a low-dimensional representation of range-related measurements. In addition to RSS and TOA, other PFs are considered such as the maximum amplitude (MA) ν_{MA}, rise time (RT) ν_{RT}, mean excess delay (MED) ν_{MED}, delay spread (DS) ν_{DS}, and kurtosis ν_{kurtosis}. These PFs are evaluated as follows

$$\nu_{\text{MA}} = \max_t\{|v(t)|\} \quad (17a)$$

$$\nu_{\text{RT}} = \arg\min_t\{|v(t)| \geqslant \beta_2\nu_{\text{MA}}\} - \arg\min_t\{|v(t)| \geqslant \beta_1\sigma_n\} \quad (17b)$$

Algorithm 2 – SRI estimation with dimensionality reduction

Offline Phase

1: Acquire training data $\mathcal{T} = \{(\boldsymbol{y}^{(k)}, d^{(k)})\}_{k=-N}^{-1}$ through a measurement campaign.
2: Perform dimensionality reduction to the training data:

$$\{(\boldsymbol{y}^{(k)}, d^{(k)})\}_{k=-N}^{-1} \rightarrow \{(\psi(\boldsymbol{y}^{(k)}), d^{(k)})\}_{k=-N}^{-1} .$$

3: Estimate the generative model as $\widehat{f}(\psi(\boldsymbol{y}), d)$.
4: Store the estimated generative model.

Online Phase

1: **for** $k \geqslant 0$ **do**
2: Acquire a new measurements set $\boldsymbol{y}^{(k)}$ at time t_k .
3: Perform dimensionality reduction to the new measurements set:

$$\boldsymbol{y}^{(k)} \rightarrow \psi(\boldsymbol{y}^{(k)}) .$$

4: Estimate the SRI using the reduced measurements set $\psi(\boldsymbol{y}^{(k)})$ from the stored generative model as

$$\mathcal{L}_{\boldsymbol{y}^{(k)}}(d) \propto \widehat{f}(\psi(\boldsymbol{y}^{(k)}), d) .$$

5: **end for**

$$\nu_{\text{MED}} = \int_{-\infty}^{+\infty} t \frac{|v(t)|^2}{\nu_{\text{RSS}}} dt \tag{17c}$$

$$\nu_{\text{DS}} = \int_{-\infty}^{+\infty} (t - \nu_{\text{MED}})^2 \frac{|v(t)|^2}{\nu_{\text{RSS}}} dt \tag{17d}$$

$$\nu_{\text{kurtosis}} = \frac{1}{\sigma_{|v|}^4 T} \int_T [|v(t)| - \mu_{|v|}]^4 dt \tag{17e}$$

where $v(t)$ is the received waveform at time t, σ_n is the standard deviation of the thermal noise, the values of β_1 and β_2 are chosen empirically,[10] and

$$\mu_{|v|} = \frac{1}{T} \int_T |v(t)| dt \tag{18a}$$

$$\sigma_{|v|}^2 = \frac{1}{T} \int_T [|v(t)| - \mu_{|v|}]^2 dt . \tag{18b}$$

The main advantages of PFs are as follows: i) they have a simple and intuitive meaning; ii) they do not require a training phase; and iii) they can often be obtained efficiently since many commercial devices are designed to compute some of them. However, PFs are not able to encapsulate all the localization information provided by rich measurement sets, e.g., waveform samples with high delay resolution.

2) Principal component analysis: PCA is a prevalent technique for dimensionality reduction [53]. PCA projects the data into a lower-dimensional linear subspace. Let $\boldsymbol{\Sigma}$ be the $M \times M$ empirical covariance matrix of measurements sets $\{\boldsymbol{y}^{(k)}\}_{k=-N}^{-1}$ obtained during the offline phase and $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{M'}$ be the M' eigenvectors of $\boldsymbol{\Sigma}$ corresponding to the largest eigenvalues. A new measurements set $\boldsymbol{y} = [y_1, y_2, \ldots, y_M]^{\mathrm{T}} \in \mathbb{R}^M$ can be projected onto the linear subspace generated by such eigenvectors to obtain its low dimensional representation, also known as principal components, that is

$$\psi(\boldsymbol{y}) = \left[\boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}_1, \boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}_2, \ldots, \boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}_{M'}\right]^{\mathrm{T}}. \tag{19}$$

The main advantages of PCA are as follows: i) it is the linear transformation that results in the lowest MSE; ii) it is easy to implement and only requires the choice of the number of principal components, which can be guided by the relative size of the eigenvalues, see e.g., [53, Chapter 6]; and iii) it can be used in the online phase by directly employing the eigenvectors of the empirical covariance. However, PCA is not able to encapsulate the localization information provided by measurement sets with strong nonlinear relationships since PCA projects the measurements into a lower-dimensional linear subspace.

3) Laplacian eigenmap: LEM provides a low-dimensional representation by approximating the support of the measurements set with a graph embedded in a low dimensional nonlinear space. In particular, it represents measurements sets using the eigenvectors of the Laplacian of a graph [54] obtained from the measurements sets. The graph has vertices corresponding to the measurements sets $\{\boldsymbol{y}^{(k)}\}_{k=-N}^{-1}$ and edges corresponding to pairs of similar measurements sets as follows. The ϵ-neighborhoods graph has an edge connecting each pair of measurements sets $(\boldsymbol{y}^{(k)}, \boldsymbol{y}^{(l)})$ with $\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^{(l)}\| \leqslant \epsilon$ [54], while the K-nearest neighborhoods graph has an edge connecting each pair of measurements $(\boldsymbol{y}^{(k)}, \boldsymbol{y}^{(l)})$ when both $\boldsymbol{y}^{(k)}$ is among the K-nearest neighbors of $\boldsymbol{y}^{(l)}$ and $\boldsymbol{y}^{(l)}$ is among the K-nearest neighbors of $\boldsymbol{y}^{(k)}$ [54].

Let $\boldsymbol{W}$ be a weighted adjacency matrix of the so formed graph; each component $W_{k,l}$ of matrix $\boldsymbol{W}$ quantifies the similarity between the pair of measurements $\boldsymbol{y}^{(k)}$ and $\boldsymbol{y}^{(l)}$ with

$$W_{k,l} = \begin{cases} e^{-\frac{\|\boldsymbol{y}^{(k)} - \boldsymbol{y}^{(l)}\|}{t}} & \text{if } \boldsymbol{y}^{(k)} \text{ and } \boldsymbol{y}^{(l)} \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $t \in \mathbb{R}$ is a parameter. Then, the unnormalized Laplacian matrix is $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is the diagonal matrix with entries $D_{k,l} = \sum_{l=1}^{N} W_{k,l}$. Let $\breve{\boldsymbol{e}}_1, \breve{\boldsymbol{e}}_2, \ldots, \breve{\boldsymbol{e}}_{M'}$ be the M' eigenvectors corresponding to the smallest non-zero eigenvalues of the generalized eigenvalue problem

$$\boldsymbol{L}\breve{\boldsymbol{e}} = \breve{\lambda}\boldsymbol{D}\breve{\boldsymbol{e}} \tag{21}$$

then a low dimensional representation of a measurement set $\boldsymbol{y}^{(\ell)} \in \{\boldsymbol{y}^{(k)}\}_{k=-N}^{-1}$ is obtained as the ℓth components of such eigenvectors, that is

$$\psi(\boldsymbol{y}^{(\ell)}) = \left[\breve{\boldsymbol{e}}_1[\ell], \breve{\boldsymbol{e}}_2[\ell], \ldots, \breve{\boldsymbol{e}}_{M'}[\ell]\right]^{\mathrm{T}} . \tag{22}$$

The main advantage of LEM is that it can efficiently reduce the dimensionality of measurements sets with strong nonlinear relationships. However, LEM requires the choice of several parameters such as t, K, and M', and its usage in the online phase requires an out-of-sample extension as described in [55].

[10] In the numerical results shown in Section IV $\beta_1 = 6$ and $\beta_2 = 0.6$.

B. Generative model estimation

In this section, techniques are provided to estimate the generative model, i.e., the joint distribution of measurements and distances, which can be cast as density estimation.[11] First a measurement preprocessing technique called "data sphering" is described, and then efficient techniques for density estimation based on the FW setting and KDE are presented. For notational convenience, in the following $\boldsymbol{x} = [\psi(\boldsymbol{y})^{\mathrm{T}}, d]^{\mathrm{T}}$, so that the goal is to estimate the generative model $f(\boldsymbol{x}) = f(\psi(\boldsymbol{y}), d)$.

Before performing the density estimation process, it is useful to preprocess the data to make the scales of different variables compatible [52]. We use the linear transformation called data sphering that maps the original data into a set with zero mean and identity covariance matrix. Specifically, if $\{\boldsymbol{x}^{(k)}\}_{k=-N}^{-1}$ is the set of original data, where $\boldsymbol{x}^{(k)} = [\psi(\boldsymbol{y}^{(k)})^{\mathrm{T}}, d^{(k)}]^{\mathrm{T}}$, the processed data are

$$\boldsymbol{z}^{(k)} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}), \text{ for } k = -N, -N+1, \ldots, -1 \tag{23}$$

where $\bar{\boldsymbol{x}}$ is the empirical mean of the original data, and $\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^{\mathrm{T}}$ is the spectral decomposition of the empirical covariance of the original data, i.e., $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements equal to the eigenvalues of the empirical covariance, and the columns of $\boldsymbol{A}$ are made of the eigenvectors of the empirical covariance. Then, the estimate of the density of the original data can be obtained from that of $f_{\mathbf{z}}(\boldsymbol{z})$ as

$$\widehat{f}_{\mathbf{x}}(\boldsymbol{x}) = \widehat{f}_{\mathbf{z}}(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}))\,|\det(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{A}^{\mathrm{T}})|\,. \tag{24}$$

1) Fisher–Wald setting: With this approach, the problem of density estimation is cast as an empirical risk minimization [56]. The risk of estimating the distribution $f(\boldsymbol{z})$ by another distribution $\widetilde{f}(\boldsymbol{z})$ is measured by

$$R(\widetilde{f}) = -\int_{\mathbb{R}^{M'+1}} f(\boldsymbol{z}) \log \widetilde{f}(\boldsymbol{z})\, d\boldsymbol{z}\,.$$

The minimum of such a risk is attained by distributions that differ from $f(\boldsymbol{z})$ only in a set of measure zero. Therefore, an estimate of $f(\boldsymbol{z})$ can be obtained from a measurement campaign as

$$\widehat{f} \in \arg\min_{\widetilde{f}\in\mathcal{F}} R_{\text{emp}}(\widetilde{f}) \tag{25}$$

where $R_{\text{emp}}(\widetilde{f}) = -\frac{1}{N}\sum_{k=-N}^{-1} \log \widetilde{f}(\boldsymbol{z}^{(k)})$ is the empirical risk of the density $\widetilde{f}$ for the N random instantiations $\{\boldsymbol{z}^{(k)}\}_{k=-N}^{-1}$, and $\mathcal{F}$ is a chosen family of distributions.

In this paper, we propose $\mathcal{F}$ to be the family formed by mixtures of m Gaussian distributions with parameters $\boldsymbol{\xi} = [\alpha_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \alpha_2, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \ldots, \alpha_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m]$. Each member of this family is given by

$$\widetilde{f}(\boldsymbol{z}; \boldsymbol{\xi}) = \sum_{i=1}^{m} \alpha_i \varphi(\boldsymbol{z}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{26}$$

[11]Note that, when some measurements are discrete, as those in Section II-A, the corresponding joint $f(\boldsymbol{y}, d)$ is a general Radon–Nikodym derivative. The approach presented in this section is still valid in those situations with small modifications such as estimating separately the cases corresponding to each discrete value.

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}^+$, and $\sum_{i=1}^m \alpha_i = 1$.

From empirical risk minimization, the density function is estimated using an ML estimator whose goal is to find the parameter $\boldsymbol{\xi}$ that solves (25). An approximation for the ML solution can be found via the expectation maximization (EM) algorithm [57], as follows:

- Let $\boldsymbol{\xi}^{[0]}$ be an initial solution.
- For each $j = 1, 2, \ldots, m$, iterate

$$\alpha_j^{[\ell+1]} = \frac{\alpha_j^{[\ell]}}{N} \sum_{k=-N}^{-1} \frac{\varphi(\boldsymbol{z}^{(k)}; \boldsymbol{\mu}_j^{[\ell]}, \boldsymbol{\Sigma}_j^{[\ell]})}{\widetilde{f}(\boldsymbol{z}^{(k)}; \boldsymbol{\xi}^{[\ell]})}$$

$$\boldsymbol{\mu}_j^{[\ell+1]} = \frac{\alpha_j^{[\ell]}}{N\alpha_j^{[\ell+1]}} \sum_{k=-N}^{-1} \boldsymbol{z}^{(k)} \frac{\varphi(\boldsymbol{z}^{(k)}; \boldsymbol{\mu}_j^{[\ell]}, \boldsymbol{\Sigma}_j^{[\ell]})}{\widetilde{f}(\boldsymbol{z}^{(k)}; \boldsymbol{\xi}^{[\ell]})}$$

$$\boldsymbol{\Sigma}_j^{[\ell+1]} = \frac{\alpha_j^{[\ell]}}{N\alpha_j^{[\ell+1]}} \sum_{k=-N}^{-1} \frac{\boldsymbol{z}^{(k)}(\boldsymbol{z}^{(k)})^{\mathrm{T}} \varphi(\boldsymbol{z}^{(k)}; \boldsymbol{\mu}_j^{[\ell]}, \boldsymbol{\Sigma}_j^{[\ell]})}{\widetilde{f}(\boldsymbol{z}^{(k)}; \boldsymbol{\xi}^{[\ell]})} - \boldsymbol{\mu}_j^{[\ell+1]}(\boldsymbol{\mu}_j^{[\ell+1]})^{\mathrm{T}}$$

in increasing ℓ until they convergence.

Each iteration of this algorithm reduces the empirical risk, i.e., increases the likelihood, while its performance highly depends on the initial solution [57]. In this paper, the initial solution $\boldsymbol{\xi}^{[0]}$ is found by first clustering the training data and then performing ML estimation for each cluster. Specifically,

- Cluster the training data $\{\boldsymbol{z}^{(k)}\}_{k=-N}^{-1}$ into m clusters, for instance, by using the k-means algorithm [58].
- For each cluster $j = 1, 2, \ldots, m$, $\alpha_j^{[0]}$ is the fraction of training data assigned to this cluster, and $\boldsymbol{\mu}_j^{[0]}$ and $\boldsymbol{\Sigma}_j^{[0]}$ are the empirical mean and covariance of the jth cluster, respectively.

The main advantages of density estimation in the FW setting described above are as follows: i) it can obtain a parsimonious parametric generative model characterized by α_j, $\boldsymbol{\mu}_j$, and $\boldsymbol{\Sigma}_j$ for $j = 1, 2, \ldots, m$; and ii) it only requires the choice of the number of components m in the mixture. However, generative models with a fixed number of parameters may lead to inaccuracies when the joint distribution of measurements sets and distances is highly complex.

2) Kernel density estimation: With this approach, the generative model is estimated using a sum of kernels centered at the $\{\boldsymbol{z}^{(k)}\}_{k=-N}^{-1}$. Specifically, $f(\boldsymbol{z})$ is approximated as

$$\widehat{f}(\boldsymbol{z}) = \frac{1}{N}\sum_{k=-N}^{-1} K(\boldsymbol{z} - \boldsymbol{z}^{(k)})$$

where $K(\cdot)$ is a positive kernel function. This paper uses the Gaussian kernel

$$K(\boldsymbol{z} - \boldsymbol{z}^{(k)}) = \varphi(\boldsymbol{z}; \boldsymbol{z}^{(k)}, \boldsymbol{H})$$

where $\boldsymbol{H}$ is a positive definite matrix called bandwidth matrix. Consider bandwidth matrices of the form $\boldsymbol{H} = \text{diag}([h_1, h_2, \ldots, h_{M'+1}])$, where the values of h_i can be calculated according to the normal reference rule as [59]

$$h_i = \left(\frac{4}{N(M'+3)}\right)^{\frac{1}{M'+5}} \sigma_i \tag{27}$$

where $M'+1$ is the dimension of $\boldsymbol{z}^{(k)}$ and σ_i^2 is the empirical variance of the ith components of the training data $\{\boldsymbol{z}^{(k)}\}_{k=-N}^{-1}$. For a Gaussian kernel, the approximating density is a mixture of Gaussians and the number of mixture components is equal to the number of elements N in the training data.

The main advantages of the KDE described above are as follows: i) it does not require an optimization process and the approximating density is obtained directly from the training data once a bandwidth matrix is chosen; and ii) it provides non-parametric estimates that can approximate highly complex distributions. However, the KDE results in a complexity increase since the approximating density has a large number of components, and its accuracy depends on the suitability of the bandwidth chosen.

IV. Case Study

The methodology presented for SRI-based localization is technology-agnostic since it is applicable to any technology capable of providing range-related measurements. The specific algorithm best suited for the dimensionality reduction and the generative model estimation steps are technology-dependent since they are contingent on the characteristics of the measurements. This section presents a case study in which ultra-wideband (UWB) signals are employed and shows the performance of SRI-based localization relative to conventional techniques.

A. UWB measurements

UWB technology offers the potential of high accuracy localization due to its ability to resolve multipath propagation and penetrate obstacles [60]–[66]. Commercial UWB impulse radios can obtain round-trip-time (RTT) measurements together with samples of the received waveform signal.[12] Each measurements set is given by

$$\boldsymbol{y} = \left[\tau, \nu_{\mathrm{MA}}, \breve{\boldsymbol{v}}[1], \breve{\boldsymbol{v}}[2], \ldots, \breve{\boldsymbol{v}}[M-2]\right]^{\mathrm{T}} \in \mathbb{R}^{M}$$

where τ and ν_{MA} are the RTT and MA of the received waveform, respectively, and $\breve{\boldsymbol{v}}[i]$ satisfies $0 \leqslant \breve{\boldsymbol{v}}[i] = |v(t_i)|/\nu_{\mathrm{MA}} \leqslant 1$ with $|v(t_i)|$ denoting the absolute value of the received waveform at each sampling time t_i. Note that the use of UWB technology can result in a large number of waveform samples and the dimensionality of $\boldsymbol{y}$ can be on the order of thousands. Therefore, the usage of such measured waveforms requires a step of dimensionality reduction as described in previous sections.[13]

B. Network experimentation

An extensive measurement campaign using Federal Communications Commission (FCC)-compliant UWB impulse radios was conducted in a typical office environment at the Massachusetts Institute of Technology (MIT) as described in [37]. In this network experimentation, 112 points in the monitored area were chosen, and 1024 pairs of nodes' position were used to collect measurements sets. For each pair, three elements were recorded: 1) a set of samples $v(t_i)$, $i = 1, 2, \ldots, M-2$ from the received waveform, where $M = 3503$ and $t_{i+1} - t_i = 41.3\,\mathrm{ps}$ for all i; 2) the delay estimate τ obtained from RTT measurements; and 3) the distance d between the two nodes. In addition, a label indicating LOS or NLOS condition was also recorded for each pair of nodes' positions; such labels were not used by the algorithms, they only enabled the performance results to be assessed in terms of LOS vs NLOS conditions. This campaign produced a database $\mathcal{D} = \{(\boldsymbol{y}^{(k)}, d^{(k)})\}_{k=-N_{\mathrm{d}}}^{-1}$ with $N_{\mathrm{d}} = 1024$.

C. SRI estimation

Algorithm 3 describes the specific approaches for SRI-based localization used in the case study.[14] The dimensionality reduction is achieved by the techniques described in Section III-A. Since the high dimensionality is due to the large number of waveform samples, dimensionality reduction is performed over the normalized samples $\breve{\boldsymbol{v}}[i]$ for $i = 1, 2, \ldots M-2$ while the measured delay and MA are left unmodified, that is

$$\psi(\boldsymbol{y}) = \left[\tau, \nu_{\mathrm{MA}}, \boldsymbol{\nu}[1], \boldsymbol{\nu}[2], \ldots, \boldsymbol{\nu}[M'-2]\right]^{\mathrm{T}} \in \mathbb{R}^{M'}$$

where $\boldsymbol{\nu} = \left[\boldsymbol{\nu}[1], \boldsymbol{\nu}[2], \ldots, \boldsymbol{\nu}[M'-2]\right]^{\mathrm{T}}$ is a low dimensional representation of $\breve{\boldsymbol{v}} = \left[\breve{\boldsymbol{v}}[1], \breve{\boldsymbol{v}}[2], \ldots, \breve{\boldsymbol{v}}[M-2]\right]^{\mathrm{T}}$.

The vector $\boldsymbol{\nu}$ is the RSS, RT, MED, DS, and kurtosis of $\breve{\boldsymbol{v}}$ for dimensionality reduction based on PFs; the $M'-2$ principal components of $\breve{\boldsymbol{v}}$ for dimensionality reduction based on PCA; and the components corresponding to $\breve{\boldsymbol{v}}$ of the first $M'-2$ eigenvectors of the generalized eigenvalue problem for dimensionality reduction based on LEM.[15] The generative model estimation is achieved by the two techniques described in Section III-B. Specifically, the joint density $f(\psi(\boldsymbol{y}), d)$ is estimated using a mixture of m Gaussians for the FW setting, and using a bandwidth matrix obtained from the normal reference rule for KDE.

D. Localization performance metrics

The performance of the proposed techniques is assessed via a semi-experimental approach using waveforms' measurements, collected at different receiver positions, as inputs to Monte Carlo simulation. In each Monte Carlo instantiation, the following steps are performed.

- Select a training set $\mathcal{T}$ randomly from the database $\mathcal{D}$ obtained by network experimentation, with $|\mathcal{T}| = N = N_{\mathrm{d}} - N_{\mathrm{b}}$ where $N_{\mathrm{d}} = |\mathcal{D}|$. Leave the remaining data sets $\mathcal{D} \setminus \mathcal{T} = \{(\boldsymbol{y}^{(k_j)}, d^{(k_j)})\}_{j=1}^{N_{\mathrm{b}}}$ for testing.[16]

[12]Note that RTT measurements enable the estimation of distances among asynchronized devices without a common time reference [38].

[13]In this section $M = 3503$ and we use $M' = 7$ for all methods to ensure fair comparison since 7 PFs are considered, which is also the number of PFs used in previous works [37], [40].

[14]The numerical results are obtained with $t = 0.8$ and $m = 5$ in equations (20) and (26).

[15]For LEM, we choose the K-nearest neighbors graph with $K = N$ and the online dimensionality reduction stage is performed by using its out-of-sample extension based on kernel eigenfunctions as described in [55].

[16]This methodology corresponds to leave-p-out cross-validation [67] with $p = N_{\mathrm{b}}$. The techniques for NLOS mitigation presented in [37] and [40] are also trained with the same data for a fair comparison.

Algorithm 3 – SRI estimation via network experimentation

Offline Phase

1: Acquire training data $\mathcal{T} = \left\{(\boldsymbol{y}^{(k)}, d^{(k)})\right\}_{k=-N}^{-1}$ from a network experimentation.

2: Perform dimensionality reduction of the training data for $k = -N, -N+1, \ldots, -1$

$$\left\{(\boldsymbol{y}^{(k)}, d^{(k)})\right\} \rightarrow \left\{([\tau^{(k)}, \nu_{\text{MA}}, \boldsymbol{\nu}^{(k)}], d^{(k)})\right\}$$

based on one of the following techniques described in Section III-A.

i) PF: Form the vector $\boldsymbol{\nu}^{(k)}$ using RSS, RT, MED, DS, and kurtosis obtained from $\breve{\boldsymbol{v}}^{(k)}$ via (17).

ii) PCA: Form the vector $\boldsymbol{\nu}^{(k)}$ using the $M'-2$ principal components of $\breve{\boldsymbol{v}}^{(k)}$ corresponding to the empirical covariance matrix of samples $\{\breve{\boldsymbol{v}}^{(l)}\}_{l=-N}^{-1}$ via (19).

iii) LEM: Form the vector $\boldsymbol{\nu}^{(k)}$ using the $M'-2$ eigenvectors with smallest non-zero eigenvalues of the generalized eigenvalue problem (21) for the graph with vertices $\{\breve{\boldsymbol{v}}^{(l)}\}_{l=-N}^{-1}$ via (22).

3: Preprocess the data through data sphering in (23).

4: Obtain the estimated generative model based on one of the following techniques.

i) FW setting: Obtain initial solution $\boldsymbol{\xi}^{[0]}$ using k-means algorithm with m clusters; and then obtain estimate for $\boldsymbol{\xi}$ using EM algorithm.

ii) KDE: Find the generative model estimation using $\boldsymbol{H}$ calculated according to the normal reference rule in (27).

5: Obtain the estimated generative model as $\widehat{f}(\psi(\boldsymbol{y}), d) = \widehat{f}_{\mathbf{x}}(\boldsymbol{x})$ using $\widehat{f}_{\mathbf{x}}(\cdot)$ from (24).

6: Store the estimated generative model.

Online Phase

1: **for** $k \geqslant 0$ **do**

2: Acquire a new set of normalized measurements at time t_k $\boldsymbol{y}^{(k)} = \left[\tau^{(k)}, \nu_{\text{MA}}^{(k)}, (\breve{\boldsymbol{v}}^{(k)})^{\text{T}}\right]^{\text{T}}$.

3: Perform dimensionality reduction to the new measurements set:

$$\boldsymbol{y}^{(k)} \in \mathbb{R}^{M+1} \rightarrow \left[\tau^{(k)}, \nu_{\text{MA}}^{(k)}, (\boldsymbol{\nu}^{(k)})^{\text{T}}\right]^{\text{T}} \in \mathbb{R}^{M'+1}.$$

4: Estimate the SRI of the measurements set $\left[\tau^{(k)}, \nu_{\text{MA}}^{(k)}, (\boldsymbol{\nu}^{(k)})^{\text{T}}\right]^{\text{T}}$ from the stored generative model as

$$\mathcal{L}_{\boldsymbol{y}^{(k)}}(d) \propto \widehat{f}(\tau^{(k)}, \nu_{\text{MA}}^{(k)}, \boldsymbol{\nu}^{(k)}, d).$$

5: **end for**

- Emulate a network with an agent at the origin and N_{b} anchors using $d^{(k_j)}$'s from $\mathcal{D} \setminus \mathcal{T}$, where the jth anchor is at known position

$$\boldsymbol{p}_j = [d^{(k_j)} \sin(\theta_j), d^{(k_j)} \cos(\theta_j)]^{\text{T}}. \tag{28}$$

- Infer the agent position for different localization techniques listed in Table I using $\boldsymbol{y}^{(k_j)}$'s from $\mathcal{D} \setminus \mathcal{T}$, and record the corresponding localization errors.[17]

TABLE I
COMPARISON OF DIFFERENT TECHNIQUES TO PROCESS RANGE-RELATED MEASUREMENTS

Technique	Dimensionality Reduction	Generative Model Est.	Localization Algorithm
LS	–	–	LS
WLS	–	–	WLS
SVM	PF	–	LS
GPR	PF	–	Bayesian
SRI Type I	PF	FW	Bayesian
SRI Type II	PCA	FW	Bayesian
SRI Type III	LEM	FW	Bayesian
SRI Type IV	PF	KDE	Bayesian
SRI Type V	PCA	KDE	Bayesian
SRI Type VI	LEM	KDE	Bayesian

The LS, WLS, support vector machine (SVM)-based, and Gaussian process regression (GPR)-based techniques use DEs obtained from the range-related measurements. Specifically, LS and WLS techniques use DEs obtained directly from RTTs (PFs are used in WLS to weight the DEs) [39], while SVM-based and GPR-based techniques use the DEs obtained from RTTs after subtracting a bias (PFs are used to estimate the RTTs bias) [37], [40]. Instead of DEs, SRI-based techniques use SRI estimates obtained directly from range-related measurements as described in Algorithm 3.

Localization error and localization error outage (LEO) are considered as metrics to provide insights into the behavior of different localization techniques. The localization error is defined as the Euclidean distance between the estimated position $\widehat{\boldsymbol{p}}$ and the actual position $\boldsymbol{p}$, namely $e(\boldsymbol{p}) = ||\widehat{\mathbf{p}} - \boldsymbol{p}||$. The LEO is given in terms of the outage probability based on the localization error, as

$$\begin{aligned} P_{\text{o}}(e_{\text{th}}) &= \mathbb{P}\left\{e(\boldsymbol{p}) > e_{\text{th}}\right\} \\ &= \mathbb{E}\left\{\mathbb{1}_{(e_{\text{th}}, +\infty)}(||\widehat{\mathbf{p}} - \boldsymbol{p}||)\right\} \end{aligned}$$

where e_{th} is the target (i.e., maximum allowable) localization error, and $\mathbb{1}_{\mathcal{A}}(x) = 1$ when $x \in \mathcal{A}$ and 0 otherwise. Here the statistical expectation $\mathbb{E}\{\cdot\}$ is over the ensemble of possible anchors positions and channel realizations. In addition, consider the RMS and median localization errors, which are respectively given by

$$\begin{aligned} e_{\text{rms}} &= \sqrt{\mathbb{E}\{||\widehat{\mathbf{p}} - \boldsymbol{p}||^2\}} \\ e_{\text{med}} &= P_{\text{o}}^{-1}(0.5). \end{aligned}$$

E. Localization performance results

In the following we show the performance of different localization techniques listed in Table I and its relationship with the measurement sets. Consider two scenarios: A) where $N_{\text{b}} = 3$ and each θ_j is uniformly distributed over $[0, 2\pi)$ in (28); and B) where $N_{\text{b}} = 5$ and each $\theta_j = j2\pi/5$ in (28).

[17] Bayesian localization is implemented by using MAP estimator and constant priors, where the maximum is obtained by exhaustive search over a regular grid with 0.12 m spacing in the region $[-10, 10] \times [-10, 10]\,\text{m}^2$.

TABLE II
PERFORMANCE OF DIFFERENT TECHNIQUES TO PROCESS RANGE-RELATED MEASUREMENTS

Technique	e_{rms} [m]	e_{med} [m]	$P_{\text{o}}(1\text{m})$	$P_{\text{o}}(3\text{m})$	T_{tr} [ms]
LS+LOS	1.96	0.56	27.2%	10.4%	–
LS	4.39	2.60	77.8%	44.1%	–
WLS	3.94	1.94	66.8%	35.5%	1.6
SVM	2.77	1.20	56.9%	18.1%	133
GPR	2.15	1.28	61.0%	12.9%	960
SRI Type I	1.97	1.07	52.5%	11.6%	177
SRI Type II	1.79	0.76	41.0%	9.1%	183
SRI Type III	1.81	0.77	41.7%	9.3%	1,230
SRI Type IV	2.06	1.16	56.4%	11.8%	870
SRI Type V	1.81	1.10	54.4%	11.0%	880
SRI Type VI	1.91	1.04	51.4%	10.6%	1,915

Table II shows the performance of the proposed SRI-based localization techniques in comparison with conventional techniques in terms of accuracy and training complexity. The accuracy of the LS estimator for the subset of cases where all the measurements are obtained in LOS (LS+LOS) is also shown as a benchmark for techniques that deal with harsh propagation conditions. Table II shows the accuracy for the localization techniques listed in Table I for scenario A) in terms of RMS localization error, median localization error, and LEO. It can be observed that SRI-based techniques achieve higher accuracy than conventional techniques. In particular, techniques based on SRI estimation can reduce the localization error of about 50% with respect to LS-based techniques (LS, WLS) and of about 30% with respect to regression-based techniques (SVM, GPR). Table II also shows the average processing time of the training stage, T_{tr} for the localization techniques listed in Table I. It can be observed that SRI techniques using simple dimensionality reduction together with parsimonious generative models (SRI Type I and SRI Type II) offer the best trade-off between accuracy and complexity in this case study. Hence, in the following the focus will be on these two SRI methods.

Fig. 4 shows the complementary LEO in scenario A) as a function of e_{th}. It can be observed that the SRI-based techniques achieve higher accuracy compared to conventional techniques and approach that of LS+LOS. In particular, the curves corresponding to SVM, GPR, and SRI Type I show a direct comparison between conventional approaches based on regression and the proposed approach based on SRI using the same waveform features. It can also be observed from the curve corresponding to SRI Type II that further improvement can be obtained by using more sophisticated dimensionality reduction techniques such as PCA.

Fig. 5 shows LEO in scenario B) as a function of NLOS probability p_{NLOS} for $e_{\text{th}} = 2\text{m}$ and $e_{\text{th}} = 1\text{m}$. It can be observed that SRI-based localization techniques can outperform conventional techniques regardless of the occurrence of NLOS situations.

Finally, the effects of different measurements types on localization accuracy are considered. This is of practical relevance since cost-effective devices may provide waveform features instead of waveform samples. Fig. 6 shows the complementary LEO as a function of e_{th} for LS localization technique in comparison with SRI-based techniques that use RTT measurements only (SRI+RTT), RTT and RSS measurements (SRI+RTT+RSS), or RTT and waveform samples (WS) measurements (SRI+RTT+WS). As expected, the use of richer measurements sets results in higher accuracy. This figure also shows that the addition of RSS or waveform samples to RTT measurements in SRI-based localization increases its localization accuracy of about 20% or 40%, respectively.

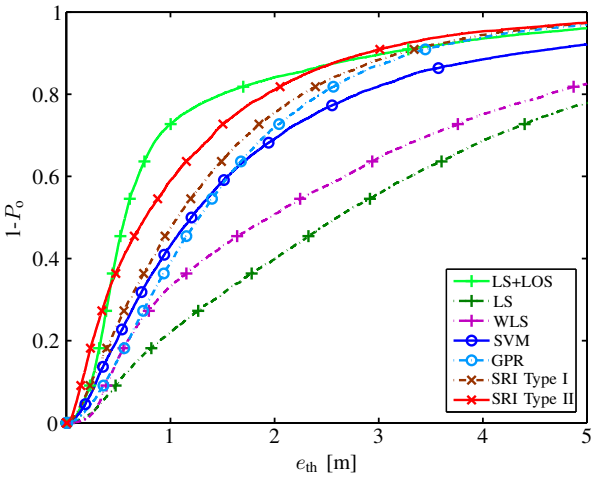


Fig. 4. Complementary LEO as a function of target localization error for different localization techniques.

V. FINAL REMARK

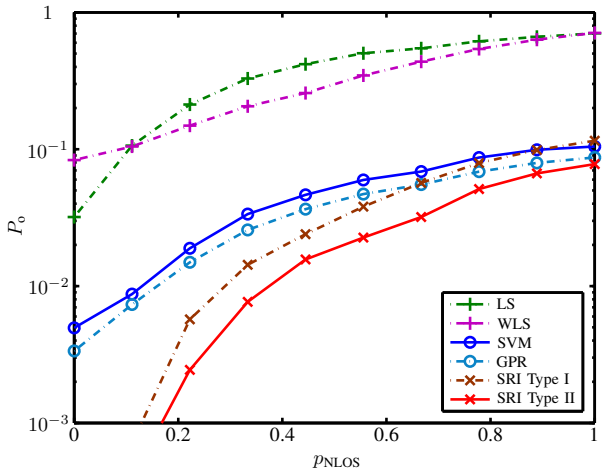
The paper introduces the new paradigm of soft range information (SRI)-based localization. SRI enables soft-decision localization by capturing the odds of all possible distances instead of a single most likely distance as in conventional techniques based on distance estimates. To obtain the SRI, core of the proposed methodology, algorithms are developed based on machine learning. The performance of SRI-based localization is evaluated using measurements obtained via network experimentation in wireless environments. The results reveal that the proposed method outperforms conventional localization techniques in harsh-propagation environments. This work shows that SRI encompasses richer information than conventional approaches, opening a way to a new level of location-awareness.

APPENDIX

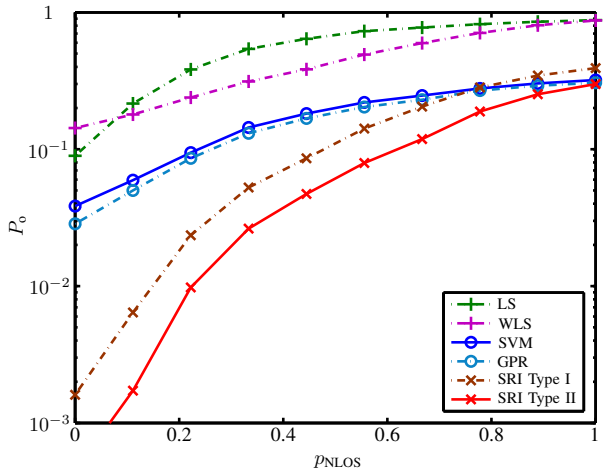
A. Derivations for DE and SRI in Section II-A

Consider measurements $\boldsymbol{y} = [r, \delta]^{\text{T}}$ where r and δ are defined as in (1) and (2), respectively. The SRI of $\boldsymbol{y}$ is

$$\mathcal{L}_{\boldsymbol{y}}(d) \propto f(\boldsymbol{y}|d) = \mathbb{P}\{\delta = \delta|d\} f(r|\delta, d).$$



(a) Case $e_{th} = 2$m.



(b) Case $e_{th} = 1$m.

Fig. 5. LEO as a function of NLOS occurrences for different localization techniques.

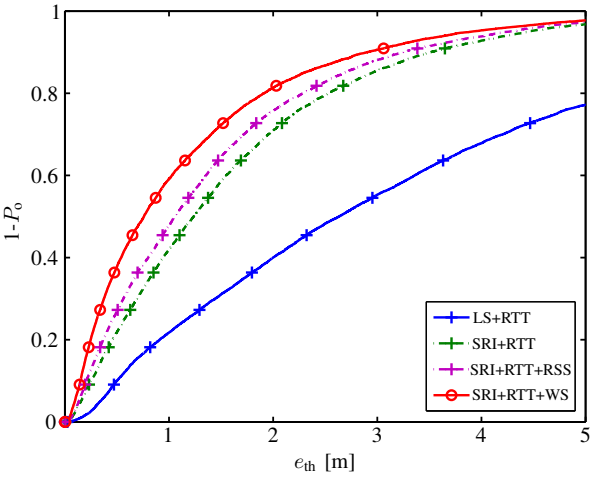


Fig. 6. Complementary LEO as a function of target localization error for different localization techniques.

Let s be the RV identifying the propagation scenario, i.e., $\mathsf{s} =$ LOS or $\mathsf{s} =$ NLOS; since the detector output is independent of the distance, we have

$$\begin{aligned}\mathcal{L}_{\boldsymbol{y}}(d) &\propto f(r, \mathsf{s}=\text{LOS}|\delta, d) + f(r, \mathsf{s}=\text{NLOS}|\delta, d)\\ &= \mathbb{P}\left\{\mathsf{s}=\text{LOS}|\delta, d\right\} f(r|\mathsf{s}=\text{LOS}, \delta, d)\\ &\quad + \mathbb{P}\left\{\mathsf{s}=\text{NLOS}|\delta, d\right\} f(r|\mathsf{s}=\text{NLOS}, \delta, d).\end{aligned}$$

This gives (4) as the distance measurement r is independent of the NLOS detector $\boldsymbol{\delta}$ given the LOS/NLOS condition s and the distance, and the detector errors do not depend on the distance.

The MMSE distance estimator is obtained by using a Bayesian formulation, which models the distance as a RV, and by computing the posterior distribution $f(d|\boldsymbol{y})$ [68]. Such posterior is proportional to the likelihood $f(\boldsymbol{y}|d)$ for a constant reference prior on the distance. Hence $f(d|\boldsymbol{y})$ equals the right hand side of (4) as it integrates to one with respect to d, which leads to the expression for the MMSE distance estimate in (5). In addition, its MSE is the expectation of $\mathbb{V}\{\mathsf{d}|\mathsf{r}, \boldsymbol{\delta}\}$, which becomes for $\epsilon_{\text{LOS}} = \epsilon_{\text{NLOS}} = \epsilon$

$$\begin{cases} (1-\epsilon)\left(((1-\epsilon)\,b)^2 + \sigma^2\right) + \epsilon\left((\epsilon\,b)^2 + \sigma^2\right) & \text{for } \delta = 0\\ \epsilon\left(((1-\epsilon)\,b)^2 + \sigma^2\right) + (1-\epsilon)\left((\epsilon\,b)^2 + \sigma^2\right) & \text{for } \delta = 1\end{cases}$$

using the expression for the variance of a mixture [69]. Then, (6) is obtained after some algebra.

B. Proof of Proposition 1

The FIM for a network with one agent and $\mathcal{N}_{\text{b}}$ anchors is

$$\boldsymbol{J} = \sum_{j\in\mathcal{N}_{\text{b}}} \mathbb{E}\left\{\frac{\partial \log f(\mathbf{y}_j|\boldsymbol{p})}{\partial \boldsymbol{p}}\left(\frac{\partial \log f(\mathbf{y}_j|\boldsymbol{p})}{\partial \boldsymbol{p}}\right)^{\text{T}}\right\}.$$

For $j \in \mathcal{N}_{\text{b}}$,

$$\frac{\partial \log f(\boldsymbol{y}_j|\boldsymbol{p})}{\partial \boldsymbol{p}} = \frac{\partial \log f(\boldsymbol{y}_j|\boldsymbol{p})}{\partial d_j}\frac{\partial d_j}{\partial \boldsymbol{p}}$$

and

$$\begin{aligned}\frac{\partial \log f(\boldsymbol{y}_j|\boldsymbol{p})}{\partial d_j} &= \frac{\partial \log f(\boldsymbol{y}_j|d_j)}{\partial d_j}\\ \frac{\partial d_j}{\partial \boldsymbol{p}} &= \frac{1}{d_j}(\boldsymbol{p} - \boldsymbol{p}_j)\,.\end{aligned}$$

Therefore,

$$\boldsymbol{J} = \sum_{j\in\mathcal{N}_{\text{b}}} \frac{\lambda_j}{d_j^2}(\boldsymbol{p} - \boldsymbol{p}_j)(\boldsymbol{p} - \boldsymbol{p}_j)^{\text{T}}$$

where, for $j \in \mathcal{N}_{\text{b}}$,

$$\lambda_j = \mathbb{E}\left\{\left(\frac{\partial \log f(\mathbf{y}_j|d_j)}{\partial d_j}\right)^2\right\}. \tag{29}$$

Finally, from (4) and after some algebra, for each measurement set $\boldsymbol{y} = [r, \delta]^{\text{T}}$ related to distance d we obtain

$$\frac{\partial \log f(\boldsymbol{y}|d)}{\partial d} = \frac{1}{\sigma^2}\begin{cases} r - d - \chi_0 & \text{for } \delta = 0\\ r - d - \chi_1 & \text{for } \delta = 1\end{cases}$$

with χ_0 and χ_1 given by (8) and (9), respectively. Hence, the result is obtained by using (29) and the fact that $f(\boldsymbol{y}|d) = \mathbb{P}\{\delta = \delta\} f(r|\delta, d)$ as the detector output is independent of the distance.

ACKNOWLEDGMENT

The authors wish to thank R. Cohen, W. Dai, Z. Liu, S. Mao, and X. Wang for their helpful suggestions and careful reading of the manuscript.

REFERENCES

[1] M. Z. Win, A. Conti, S. Mazuelas, Y. Shen, W. M. Gifford, D. Dardari, and M. Chiani, "Network localization and navigation via cooperation," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 56–62, May 2011.

[2] R. Karlsson and F. Gustafsson, "The future of automotive localization algorithms: Available, reliable, and scalable localization: Anywhere and anytime," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 60–69, Mar. 2017.

[3] X. Wang, S. Yuan, R. Laur, and W. Lang, "Dynamic localization based on spatial reasoning with RSSI in wireless sensor networks for transport logistics," *Sensors and Actuators A: Physical*, vol. 171, no. 2, pp. 421–428, Nov. 2011.

[4] G. Cardone, P. Bellavista, A. Corradi, C. Borcea, M. Talasila, and R. Curtmola, "Fostering participaction in smart cities: A geo-social crowdsensing platform," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 112–119, Jun. 2013.

[5] F. Zabini and A. Conti, "Inhomogeneous Poisson sampling of finite-energy signals with uncertainties in $\mathbb{R}^d$," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4679–4694, Sep. 2016.

[6] D. Dardari, A. Conti, C. Buratti, and R. Verdone, "Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 7, pp. 790–802, Jul. 2007.

[7] S. M. George, W. Zhou, H. Chenji, M. Won, Y. O. Lee, A. Pazarloglou, R. Stoleru, and P. Barooah, "Distressnet: A wireless Ad Hoc and sensor network architecture for situation management in disaster response," *IEEE Commun. Mag.*, vol. 48, no. 3, pp. 128–136, Mar. 2010.

[8] J. Ko, T. Gao, R. Rothman, and A. Terzis, "Wireless sensing systems in clinical environments: Improving the efficiency of the patient monitoring process," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 2, pp. 103–109, Mar. 2010.

[9] L. Evans, "Maps as deep: Reading the code of location-based social networks," *IEEE Technol. Soc. Mag.*, vol. 33, no. 1, pp. 73–80, Mar. 2014.

[10] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization – Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.

[11] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization – Part II: Cooperative networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4981–5000, Oct. 2010.

[12] Y. Shen, S. Mazuelas, and M. Z. Win, "Network navigation: Theory and interpretation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1823–1834, Oct. 2012.

[13] K. Pahlavan, X. Li, and J.-P. Mäkelä, "Indoor geolocation science and technology," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 112–118, Feb. 2002.

[14] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

[15] G. Han, H. Xu, T. Duong, J. Jiang, and T. Hara, "Localization algorithms of wireless sensor networks: a survey," *Telecommunication Systems*, vol. 52, no. 4, pp. 2419–2436, 2013.

[16] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[17] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, Aug. 2001.

[18] U. A. Khan, S. Kar, and J. M. F. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2000–2016, May 2009.

[19] ——, "DILAND: An algorithm for distributed sensor localization with noisy distance measurements," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1940–1947, Mar. 2010.

[20] Z. Liu, W. Dai, and M. Z. Win, "Mercury: An infrastructure-free system for network localization and navigation," *IEEE Trans. Mobile Comput.*, vol. IEEE Xplore Early Access, 2017.

[21] E. Grossi and M. Lops, "Sequential along-track integration for early detection of moving targets," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3969–3982, Aug. 2008.

[22] Y. Yan and Y. Mostofi, "Impact of localization errors on wireless channel prediction in mobile robotic networks," in *Globecom*, Atlanta, GA, Dec. 2013, pp. 1374 – 1379.

[23] T. Wang, Y. Shen, A. Conti, and M. Z. Win, "Network navigation with scheduling: Error evolution," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7509–7534, Nov. 2017.

[24] W. Dai, Y. Shen, and M. Z. Win, "A computational geometry framework for efficient network localization," *IEEE Trans. Inf. Theory*, vol. IEEE Xplore Early Access, 2017.

[25] L. Lu, H. Zhang, and H.-C. Wu, "Novel energy-based localization technique for multiple sources," *IEEE Syst. J.*, vol. 8, no. 1, pp. 142–150, Mar. 2014.

[26] D. B. Jourdan, D. Dardari, and M. Z. Win, "Position error bound for UWB localization in dense cluttered environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 2, pp. 613–628, Apr. 2008.

[27] B. Alavi and K. Pahlavan, "Modeling of the TOA-based distance measurement error using UWB indoor radio measurements," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 275–277, Apr. 2006.

[28] L. Stoica, A. Rabbachin, and I. Oppermann, "A low-complexity non-coherent IR-UWB transceiver architecture with TOA estimation," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1637–1646, Jun. 2006.

[29] S. Bartoletti, W. Dai, A. Conti, and M. Z. Win, "A mathematical model for wideband ranging," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 216–228, Mar. 2015.

[30] S. Mazuelas, R. M. Lorenzo, A. Bahillo, P. Fernandez, J. Prieto, and E. J. Abril, "Topology assessment provided by weighted barycentric parameters in harsh environment wireless location systems," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3842–3857, Jul. 2010.

[31] J. Prieto, S. Mazuelas, A. Bahillo, P. Fernandez, R. M. Lorenzo, and E. J. Abril, "Adaptive data fusion for wireless localization in harsh environments," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1585–1596, Apr. 2012.

[32] S. Bartoletti, A. Giorgetti, M. Z. Win, and A. Conti, "Blind selection of representative observations for sensor radar networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1388–1400, Apr. 2015.

[33] A. Conti, D. Dardari, M. Guerra, L. Mucchi, and M. Z. Win, "Experimental characterization of diversity navigation," *IEEE Syst. J.*, vol. 8, no. 1, pp. 115–124, Mar. 2014.

[34] A. Conti, M. Guerra, D. Dardari, N. Decarli, and M. Z. Win, "Network experimentation for cooperative localization," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 467–475, Feb. 2012.

[35] S. Mazuelas, F. A. Lago, J. Blas, A. Bahillo, P. Fernandez, R. Lorenzo, and E. Abril, "Prior NLOS measurements correction for positioning in cellular wireless networks," *IEEE Trans. Vehicular Technology*, vol. 58, no. 5, pp. 2585–2591, Jun 2009.

[36] J. Khodjaev, Y. Park, and A. S. Malik, "Survey on NLOS identification and error mitigation problems in UWB-based positioning algorithms for dense environments," in *Annals of Telecommunications*, 2009.

[37] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010.

[38] D. Dardari, A. Conti, U. J. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proc. IEEE*, vol. 97, no. 2, pp. 404–426, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.

[39] I. Güvenç, C.-C. Chong, F. Watanabe, and H. Inamura, "NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistics," *EURASIP J. Adv. in Signal Process.*, vol. 2008, pp. 1–14, Article ID 271 984, 2008.

[40] H. Wymeersch, S. Maranò, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for UWB localization," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1719–1728, Jun. 2012.

[41] B. Denis and N. Daniele, "NLOS ranging error mitigation in a distributed positioning algorithm for indoor UWB ad-hoc networks," in *International Workshop on Wireless Ad-Hoc Networks*, Oulu, Finland, May/Jun. 2004, pp. 356–360.

[42] J.-Y. Lee and R. A. Scholtz, "Ranging in a dense multipath environment using an UWB radio link," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1677–1683, Dec. 2002.
[43] S. Mazuelas, A. Bahillo, R. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. Abril, "Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 821–831, Oct. 2009.
[44] R. Casas, A. Marco, J. J. Guerrero, and J. Falcó, "Robust estimator for non-line-of-sight error mitigation in indoor localization," *EURASIP J. Appl. Signal Process.*, no. 1, pp. 156–156, 2006.
[45] S. Venkatesh and R. M. Buehrer, "NLOS mitigation using linear programming in ultrawideband location-aware networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 3182–3198, Sep. 2007.
[46] M. Rydstrom, L. Reggiani, E. Strom, and A. Svensson, "Suboptimal soft range estimators with applications in UWB sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4856–4866, Oct. 2008.
[47] A. Conti, S. Mazuelas, S. Bartoletti, M. Z. Win, and W. C. Lindsey, "Soft information for localization of things," *Proc. IEEE*, 2018, special issue on *Foundations and Trends in Localization Technologies*.
[48] J. O. Berger, J. M. Bernardo, and D. Sun, "The formal definition of reference priors," *The Annals of Statistics*, vol. 37, no. 2, pp. 905–938, 2009.
[49] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.
[50] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 809–819, Apr. 2005.
[51] M. Cetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 42–55, Jul 2006.
[52] J. Klemelä, *Smoothing of Multivariate Data: Density Estimation and Visualization*. John Wiley & Sons, 2009.
[53] I. Jolliffe, *Principal Component Analysis*. John Wiley & Sons, 2005.
[54] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 585–591.
[55] Y. Bengio, J. F. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering," in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004, pp. 177–184.
[56] V. N. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.
[57] R. A. Rednerf and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, Apr. 1984.
[58] D. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
[59] D. Scott, *Multivariate density estimation: theory, practice, and visualization*. Wiley Com, 2009.
[60] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, Feb. 1998.
[61] ——, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, Apr. 2000.
[62] M. Z. Win, "A unified spectral analysis of generalized time-hopping spread-spectrum signals in the presence of timing jitter," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1664–1676, Dec. 2002.
[63] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.
[64] L. Yang, "Timing PPM-UWB signals in ad hoc multiaccess," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 4, pp. 794–800, Apr. 2006.
[65] A. F. Molisch, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. Schantz, K. Siwiak, and M. Z. Win, "A comprehensive standardized model for ultrawideband propagation channels," *IEEE Trans. Antennas Propag.*, vol. 54, no. 11, pp. 3151–3166, Nov. 2006, special issue on *Wireless Communications*.
[66] J. Zhang, P. V. Orlik, Z. Sahinoglu, A. F. Molisch, and P. Kinney, "UWB systems for wireless sensor networks," *Proc. IEEE*, vol. 97, no. 2, pp. 313–331, Feb. 2009.
[67] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40–79, 2010.
[68] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
[69] B. Lindsay, *Mixture Models: Theory, Geometry, and Applications*. Alexandria, Virginia: American Statistical Association, 1995.

Santiago Mazuelas (M'10-SM'17) received the Ph.D. in Mathematics and Ph.D in Telecommunications Engineering from the University of Valladolid, Spain, in 2009 and 2011, respectively.

Since 2017 he has been Ramon y Cajal Researcher at the Basque Center for Applied Mathematics (BCAM). Prior to joining BCAM, he was a Staff Engineer at Qualcomm Corporate Research and Development from 2014 to 2017. He previously worked from 2009 to 2014 as Postdoctoral Fellow and Associate in the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT). His general research interest is the application of mathematics to solve engineering problems, currently his work is primarily focused on statistical signal processing, machine learning, and data science.

Dr. Mazuelas is Associate Editor for the IEEE Communications Letters and served as Co-chair for the Symposiums on Wireless Communications at the 2014 IEEE Globecom and at the 2015 IEEE ICC. He has received the Best Student Prize in Telecommunications Technical Engineering from University of Valladolid in 2006, the Best Doctorate Thesis Award from University of Valladolid in 2011, and the Young Scientist Prize from the Union Radio-Scientifique Internationale (URSI) Symposium in 2007. His papers received the IEEE Communications Society Fred W. Ellersick Prize in 2012, and Best Paper Awards from the IEEE ICC in 2013, the IEEE ICUWB in 2011, and the IEEE Globecom in 2011.

Andrea Conti (S'99-M'01-SM'11) received the Laurea (*summa cum laude*) in telecommunications engineering and the Ph.D. in electronic engineering and computer science from the University of Bologna, Italy, in 1997 and 2001, respectively.

He is an Associate Professor at the University of Ferrara, Italy. Prior to joining the University of Ferrara, he was with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni and with the IEIIT-Consiglio Nazionale delle Ricerche. In Summer 2001, he was with the Wireless Systems Research Department at AT&T Research Laboratories. Since 2003, he has been a frequent visitor to the Wireless Information and Network Sciences Laboratory at the Massachusetts Institute of Technology, where he presently holds the Research Affiliate appointment. His research interests involve theory and experimentation of wireless systems and networks including network localization, distributed sensing, adaptive diversity communications, and network secrecy. He is recipient of the HTE Puskás Tivadar Medal and co-recipient of the IEEE Communications Society's Stephen O. Rice Prize in the field of Communications Theory and of the IEEE Communications Society's Fred W. Ellersick Prize.

Dr. Conti has served as editor for IEEE journals, as well as chaired international conferences. He has been elected Chair of the IEEE Communications Society's Radio Communications Technical Committee. He is a co-founder and elected Secretary of the IEEE Quantum Communications & Information Technology Emerging Technical Subcommittee. He is an elected Fellow of the IET and has been selected as an IEEE Distinguished Lecturer.

Jeffery C. Allen (M'94) received the Ph.D. degree in mathematics from the University of California, San Diego, in 1988.

Since 1989, he has been with the Space and Naval Warfare Systems Command (SPAWAR), San Diego. He serves as TM for the Office of Naval Research (ONR) and the Defense Advanced Research Projects Agency (DARPA). His current research interests are in the field of wireless communication and localization networks including: system-level performance and Pareto-front computations; radiowave propagation on stochastic sea surfaces; wireless channel modeling; as well as network localization and navigation.

Moe Z. Win (S'85-M'87-SM'97-F'04) is a professor at the Massachusetts Institute of Technology (MIT) and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories and NASA Jet Propulsion Laboratory.

His research encompasses fundamental theories, algorithm design, and network experimentation for a broad range of real-world problems. His current research topics include network localization and navigation, network interference exploitation, and quantum information science. He has served the IEEE Communications Society as an elected Member-at-Large on the Board of Governors, as elected Chair of the Radio Communications Committee, and as an IEEE Distinguished Lecturer. Over the last two decades, he held various Editorial posts for IEEE journals and organized numerous international conferences. Currently, he is serving on the SIAM Diversity Advisory Committee.

Dr. Win is an elected Fellow of the AAAS, the IEEE, and the IET. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). Together with students and colleagues, his papers have received numerous awards including the IEEE Communications Society's Stephen O. Rice Prize in the Field of Communications Theory (2012), the IEEE Aerospace and Electronic Systems Society's M. Barry Carlton Award (2011), the IEEE Communications Society's Guglielmo Marconi Prize Paper Award (2008), and the IEEE Antennas and Propagation Society's Sergei A. Schelkunoff Transactions Prize Paper Award (2003). Other recognitions include the IEEE Communications Society Edwin H. Armstrong Achievement Award (2016), the International Prize for Communications Cristoforo Colombo (2013), the Copernicus Fellowship (2011) and the *Laurea Honoris Causa* (2008) from the University of Ferrara, and the U.S. Presidential Early Career Award for Scientists and Engineers (2004). He is an ISI Highly Cited Researcher.