Radio Localization and Sensing—Part I: Fundamentals

Henk Wymeersch, *Senior Member, IEEE*, and Gonzalo Seco-Granados, *Senior Member, IEEE*
(Invited Paper)

***Abstract*— This letter is part of a two-letter tutorial on radio localization and sensing, with focus on mobile radio systems in 5G mmWave and beyond. Part I introduces the fundamentals, covering an overview of the relevant literature, as well as the different aspects of localization and sensing problems. Then, different performance metrics are presented, which are important in the evaluation of methods. Methods are detailed in the last part of this letter. Part I thus provides the necessary background to delve into more forward-looking problems in Part II.**

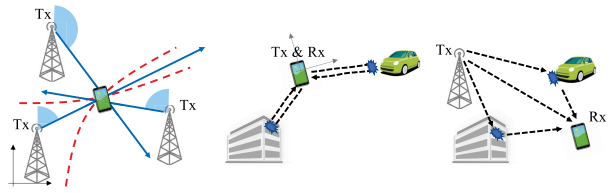
***Index Terms*— Localization, sensing, orientation estimation, synchronization.**

I. Introduction

LOCALIZATION (a term from robotics [1]) and positioning (a term from navigation and radio communication [2]) will be interchangeably used for estimation of the state (position, orientation) of a connected device in a global frame of reference (see Fig. 1-(a)). *Sensing* is broader and covers everything from channel parameter estimation and carrier sensing to presence detection [3]. In this letter, sensing will refer to state estimation of a passive object in the frame of reference of the sensor (see Fig. 1-(b,c)), and thus includes radar [4] and device-free localization [5].

Radio localization stems from military satellite-based navigation systems, most notably the global positioning system (GPS) [6]. When a receiver estimates pseudo-ranges from at least 4 synchronized satellites with a favorable geometric configuration, it can determine its 3D position and clock bias. Performance is mainly limited by signal blockages and multipath reflections. Positioning has also been part of evolving cellular standards [7]. Modern communication systems rely on a combination of time and angle measurements from several base stations (BSs) based on dedicated pilot resources to determine the 3D position of a user equipment (UE) [8].

Radar sensing also has roots in the military, with surveillance radar systems during World War II providing early warning of incoming bombers [9]. Due to its myriad of applications, radar has seen enormous developments, e.g., in automotive applications, where a modern radar can detect and track tens of moving objects and determine their distance/range, angle, and radial velocity, in the frame of reference of the radar, with very high accuracy [10]. In contrast to localization, monostatic radar is a local process, and can thus rely on a tailored, highly specialized, and hardware friendly waveforms, without strict standardization constraints. Bistatic sensing, on the other hand, is similar to communication and localization (i.e., when the transmitter or receiver have an unknown position). With radar systems and communication systems expected to operate in similar frequency bands, there is a potential convergence, both in terms of hardware and signals, of sensing and communication systems [11], [12]. Such integrated sensing and communication (ISAC), in addition to highly accurate 6D positioning, is expected to be among the main features of 6G [3].



(a) localization/positioning *(b) monostatic sensing* *(c) bistatic sensing*

Fig. 1. (a) Downlink localization of a connected UE; (b) Monostatic sensing where a UE acts as a radar sensor, with co-located transmitter and receiver; (c) bistatic sensing example with BS transmitter and UE receiver.

This letter introduces the fundamentals of model-based radio localization and sensing, and is organized as follows. First, the problem definitions, signal and channel models are detailed. Second, relevant performance metrics and bounds are described. Finally, an overview of the typical methods for localization and sensing is detailed. In Part II, a complementary literature review is provided, focusing on 6G and its challenges.

II. Models and Problem Definitions

In this section, we provide the basic formulations for the localization and sensing problem, within a mobile radio communication context. We start with a generic channel model, focusing on a frequency domain representation with N samples spaced Δ_f apart, spanning a total bandwidth of $W = N\Delta f$. This representation appears naturally with orthogonal frequency-division multiplexing (OFDM) signals, but it is not limited to them.

A. Generic Observation Model

1) Channel Model: The channel between a transmitter (Tx) with N_{Tx} antennas and a receiver (Rx) with N_{Rx} antennas over frequency $n \in \{0, \ldots, N-1\}$ and symbol $k \in \{0, \ldots, K-1\}$ can be approximated by [13]

$$\boldsymbol{H}_{n,k} = \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_{\text{rx}}(\boldsymbol{\theta}_l)\boldsymbol{a}_{\text{tx}}^{\top}(\boldsymbol{\phi}_l)e^{-\jmath 2\pi n \Delta_f \tau_l} e^{\jmath 2\pi k T_s \nu_l}, \quad (1)$$

Manuscript received 2 July 2022; revised 31 August 2022; accepted 10 September 2022. Date of publication 15 September 2022; date of current version 12 December 2022. This work was supported by the European Commission through the H2020 project Hexa-X (Grant Agreement no. 101015956), by the ICREA Academia Program, and by the Spanish R+D project PID2020-118984GB-I00. The associate editor coordinating the review of this letter and approving it for publication was A. Zappone. *(Corresponding author: Henk Wymeersch.)*

Henk Wymeersch is with the Department of Electrical Engineering, Chalmers University of Technology, 41258 Gothenburg, Sweden (e-mail: henkw@chalmers.se).

Gonzalo Seco-Granados is with the Department of Telecommunications and Systems Engineering, Universitat Autonoma de Barcelona, Bellaterra, 08193 Barcelona, Spain (e-mail: gonzalo.seco@uab.cat).

Digital Object Identifier 10.1109/LCOMM.2022.3206821

where L is the number of physical propagation paths (as, e.g., would be given by a ray-tracer), α is a complex channel gain, $\boldsymbol{a}_{\text{rx}}(\boldsymbol{\theta}) \in \mathbb{C}^{N_{\text{Rx}}}$ is the Rx array response as a function of the angle-of-arrival (AoA) $\boldsymbol{\theta} \in \mathbb{R}^2$ in azimuth and elevation, $\boldsymbol{a}_{\text{tx}}(\boldsymbol{\phi}) \in \mathbb{C}^{N_{\text{Tx}}}$ is the Tx array response as a function of the angle-of-departure (AoD) $\boldsymbol{\phi} \in \mathbb{R}^2$ in azimuth and elevation, τ is the time-of-arrival (ToA), ν is the Doppler shift, and T_s is the symbol duration. The AoA is defined in the reference frame of the Rx, the AoD in the reference frame of the Tx, and thus these angles depend on the respective orientations. Below 6 GHz, due to limited delay and angle resolution, combined with a weak connection of the paths to the environment geometry,[1] explicit geometric information in the channel is hard to harness. In contrast, at mmWave and above, paths are more closely related to the environment geometry and can be more easily resolved [14]. Hence, we will assume each path in (1) corresponds to a physical object.

2) Signal Model: The observation at the Rx is then of the form [13]

$$\boldsymbol{y}_{n,k} = \boldsymbol{W}_k^{\mathsf{H}} \boldsymbol{H}_{n,k} \boldsymbol{f}_{n,k} + \boldsymbol{n}_{n,k}, \tag{2}$$

where $\boldsymbol{W}_k \in \mathbb{C}^{N_{\text{Rx}} \times M_{\text{Rx}}}$ is an orthonormal analog Rx combiner, with $\boldsymbol{W}_k^{\mathsf{H}} \boldsymbol{W}_k = \boldsymbol{I}_{M_{\text{Rx}}}$ using $M_{\text{Rx}} \leq N_{\text{Rx}}$ RF chains, $\boldsymbol{f}_{n,k}$ is the k-th Tx signal across the Tx array, with $\mathbb{E}\{\|\boldsymbol{f}_{n,k}\|^2\} = P_{\text{tx}}/W$, and $\boldsymbol{n}_{n,k} \sim \mathcal{CN}(\boldsymbol{0}, N_0 \boldsymbol{I}_{M_{\text{Rx}}})$ is noise after the combining. Here, P_{tx} is the average transmit power and N_0 denotes the noise power spectral density. The transmit signals $\boldsymbol{f}_{n,k}$ are generally known (pilots in localization or bistatic sensing or known data in monostatic sensing), but may be partially unknown for semi-blind estimation [15].

B. The Localization Problem

In localization, as shown in Fig. 2, the UE has an unknown state $\boldsymbol{s}$, which should be inferred from observations of the form (2). The state comprises the position $\boldsymbol{x} \in \mathbb{R}^3$, the clock bias $B \in \mathbb{R}$, and possibly the orientation $\boldsymbol{o} \in \mathbb{R}^3$, which is often described with over-parameterized representation, such as a rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ subject to $\boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{I}$ and $\det(\boldsymbol{R}) = 1$, or a quaternion $\boldsymbol{q} \in \mathbb{R}^4$ subject to $\|\boldsymbol{q}\|^2 = 1$ [16]. The infrastructure nodes (BSs $i \in \{1, \ldots, N_{\text{B}}\}$) have known states, i.e., position $\boldsymbol{x}^{(i)}$ and orientation $\boldsymbol{o}^{(i)}$, and are time synchronized. Localization can be user-centric[2] in downlink (DL) or network-centric in uplink (UL). In DL, each BS i transmits signals over orthogonal subcarriers, leading to observations $\boldsymbol{y}_{k,n}^{(i)}$ at the UE Rx over channels $\boldsymbol{H}_{n,k}^{(i)}$, where i has been added to make the BS index explicit. In UL, the UE transmits a signal, which reaches the BS Rx i. Note that under time division duplexing, UL and DL channels are each other's transpose.[3]

1) Channel Decomposition: While the model (2) is generic and also widely used in communications, the localization aspect is revealed when we consider the channel $\boldsymbol{H}_{n,k}^{(i)}$ and break it up into the line-of-sight (LoS) path and the non-line-of-sight (NLoS) paths: $\boldsymbol{H}_{n,k}^{(i)} = \boldsymbol{H}_{n,k,\text{LoS}}^{(i)} + \boldsymbol{H}_{n,k,\text{NLoS}}^{(i)}$. The LoS path $\boldsymbol{H}_{n,k,\text{LoS}}^{(i)} = \alpha_1^{(i)} \boldsymbol{a}_{\text{rx}}(\boldsymbol{\theta}_1^{(i)}) \boldsymbol{a}_{\text{tx}}^\top(\boldsymbol{\phi}_1^{(i)}) e^{-\jmath 2\pi n \Delta_f \tau_1^{(i)}} e^{\jmath 2\pi k T_s \nu_1^{(i)}}$ (if it is visible and resolvable) contains geometric information related to the UE state $\boldsymbol{s}$, via the parameters $\boldsymbol{\eta}_1^{(i)} = [(\boldsymbol{\theta}_1^{(i)})^\top, (\boldsymbol{\phi}_1^{(i)})^\top, \tau_1^{(i)}, \nu_1^{(i)}]^\top$. The impact of non-resolvable LoS is discussed in [17] and of non-visible LoS in [18]. The geometric information brought by each component of $\boldsymbol{\eta}_1^{(i)}$ and by $\alpha_1^{(i)}$ is now detailed.

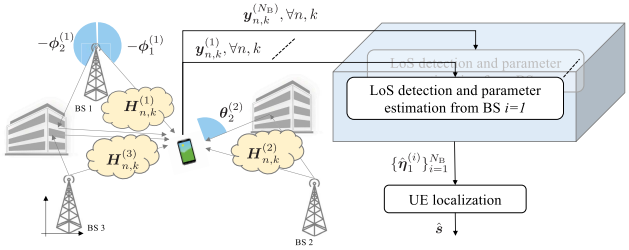


Fig. 2. Localization (here shown in DL) involves estimating the LoS channel parameters, denoted by $\boldsymbol{\eta}_1^{(i)}$, from each BS i.

- *LoS complex gain* $\alpha_1^{(i)}$*:* Since the phase of $\alpha_1^{(i)}$ varies 2π for every movement over one wavelength, it is challenging to account for this (more on this in Part II). The power of $\alpha_1^{(i)}$ can be determined by the path loss equation[4]

$$\left|\alpha_1^{(i)}\right|^2 = \frac{\lambda^2}{(4\pi)^2} \frac{G_{\text{rx}}(\boldsymbol{\theta}_1^{(i)}) G_{\text{tx}}(\boldsymbol{\phi}_1^{(i)})}{\|\boldsymbol{x} - \boldsymbol{x}^{(i)}\|^2}, \tag{3}$$

  where λ is the wavelength at the carrier and $G_{\text{tx}}(\cdot)$ and $G_{\text{rx}}(\cdot)$ denote the antenna element response at the Tx and Rx, respectively. Since these element responses are often only partially known and are affected by environmental variations, the dependence of $|\alpha_1^{(i)}|^2$ on the distance $\|\boldsymbol{x} - \boldsymbol{x}^{(i)}\|$ is usually not utilized in the development of algorithms, except for fingerprinting [19].
- *LoS AoA* $\boldsymbol{\theta}_1^{(i)}$ *and AoD* $\boldsymbol{\phi}_1^{(i)}$*:* In DL, $\boldsymbol{\theta}_1^{(i)}(\boldsymbol{x}, \boldsymbol{o})$ is a function of the UE position and orientation, while $\boldsymbol{\phi}_1^{(i)}(\boldsymbol{x})$ only depends on the UE position. This means that DL-AoA can only be used when the UE orientation is either known or also estimated as part of the state. In UL, these dependencies are reversed. The specific expressions depend on how the coordinate systems and angles are defined. For examples, see [20] or [21, Appendix A].
- *LoS delay* $\tau_1^{(i)}$*:* It is given by $\tau_1^{(i)}(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}^{(i)}\|/c + B$, where c is the speed of light. The clock bias B is due to the lack of synchronization between the BSs and UE and may drift over time. While setting B to a known value is occasionally an assumption in academic letters on localization, it is overly optimistic and leads to misleading designs and results. Dealing with the clock bias can be avoided by using round-trip-time (RTT) measurements, but should otherwise be estimated as part of the UE state.
- *LoS Doppler* $\nu_1^{(i)}$*:* Due to short transmission times in localization within a coherence interval, $|KT_s\nu_1^{(i)}| \ll 1$, so that Doppler is generally not considered. Nevertheless, Doppler can improve positioning [22], provided it can be disambiguated from the carrier frequency offset (CFO).

[1]Due to complex propagation effects, such as material propagation, diffraction, Rayleigh scattering limited shadowing, and multi-bounce scattering.

[2]In principle, user-centric monostatic localization is possible without any BS, using a priori environment information. See Part II.

[3]Note also the transpose in (1) rather than the Hermitian, typically used in communication.

[4]The path loss exponent relates to the LoS path only, not any average behavior, as in usual communication channel models.

The NLoS channel $\boldsymbol{H}^{(i)}_{n,k,\text{NLoS}}$ contains all the multipath, both specular and diffuse, which is traditionally considered as a disturbance. More on the geometric nature of the NLoS channel, the ability to modify the NLoS channel via reconfigurable intelligent surface (RIS) [23], and how to harness multipath is deferred to Part II.

C. The Sensing Problem

We consider several point objects (for extended objects, see, e.g., [24]) with state $\boldsymbol{s}_l$, including position $\boldsymbol{x}_l$ and velocity $\boldsymbol{v}_l$ for object l. An important difference in sensing compared to localization is that the number of objects is a priori unknown in sensing. Moreover, objects may appear and disappear from the sensor's field of view and/or may be occluded, leading to missed detections. In addition, clutter may lead to non-existing objects being detected, leading to false alarms. Hence, sensing combines both detection and estimation, while localization is essentially an estimation problem. The signal and channel model are again of the form (1)–(2), but the interpretation of the channel parameters is different, depending on whether the transmitter and receiver are co-located (monostatic sensing, as in automotive radar) or not (bistatic or multistatic sensing), as shown in Fig. 1. Localization can be seen as a special case of bistatic sensing, where only the LoS path is of interest.

1) Channel Decomposition: In both monostatic and bistatic sensing, the channel $\boldsymbol{H}_{n,k}$ is broken up as $\boldsymbol{H}_{n,k} = \boldsymbol{H}^{\text{object}}_{n,k} + \boldsymbol{H}^{\text{clutter}}_{n,k}$, where $\boldsymbol{H}^{\text{object}}_{n,k}$ captures the part of the channel related to the objects, while $\boldsymbol{H}^{\text{clutter}}_{n,k}$ describes the part of the channel related to clutter, e.g., ground reflections, and is modeled statistically. The different components in the channel bring the following geometric information per resolvable path.

- *Channel gain α_l:* For monostatic sensing, due to the two-way propagation, the gain is often much smaller than in (3), and is given by [9, Vol. I, Ch. 2]
$$|\alpha_l|^2 = \frac{\lambda^2 \sigma_{\text{RCS},l} G_{\text{rx}}(\boldsymbol{\theta}_l) G_{\text{tx}}(\boldsymbol{\phi}_l)}{(4\pi)^3 \|\boldsymbol{x}_l\|^4}, \tag{4}$$
where $\|\boldsymbol{x}_l\|$ denotes the distance from the sensor (which is seen as the center of the coordinate system) to the object and $\sigma_{\text{RCS},l}$ is the radar cross section (RCS) of the l-th object, which depends on the object type. The RCS is expressed in m^2 and can range from 1 m^2 for a person to 100 m^2 for a car. The very small values of $|\alpha_l|^2$ are compensated by longer integration times, enabled by including the Doppler shift in the observation model. For bistatic sensing, the model is more involved, see, e.g., [21, Section 2.3].
- *AoA $\boldsymbol{\theta}_l$ and AoD $\boldsymbol{\phi}_l$:* In monostatic sensing, the AoA and AoD are identical and depend on the object position $\boldsymbol{x}_l$. In bistatic sensing, the AoA and AoD provide independent information about the object [21, Appendix A].
- *ToA τ_l:* For monostatic sensing $\tau_l = 2\|\boldsymbol{x}_l\|/c$, measured with respect to the sensor, while for bistatic sensing $\tau_l = (\|\boldsymbol{x}_l - \boldsymbol{x}_{\text{tx}}\| + \|\boldsymbol{x}_l - \boldsymbol{x}_{\text{rx}}\|)/c + B$, where $B = 0$ when Tx and Rx are time-synchronized.
- *Doppler ν_l:* In monostatic sensing, the Doppler is measured in the reference frame of the sensor and given by $\nu_l = 2(\boldsymbol{v}_l^\top \boldsymbol{u}_l)/\lambda$, where $\boldsymbol{u}_l$ is a unit vector pointing from the object to the sensor and $\boldsymbol{v}_l$ is the relative velocity. For bistatic sensing, the Doppler depends on both the relative velocity and the unit vectors to the target from Tx and Rx, as well the CFO.

III. Performance Metrics

While there are many metrics that are of importance, such as latency (the time between the positioning request and the position being available), availability (the fraction of space or time that the localization and sensing service is available with sufficient accuracy), and scalability (density of UEs that can be simultaneously supported), our focus will be on accuracy and resolution.

A. Accuracy

The main performance metric in localization and sensing is accuracy. Let $\boldsymbol{e}$ denote the random estimation error, e.g., for localization $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, and $\|\boldsymbol{e}\|^2$ be the ℓ_2 error norm. Based on percentile or mean values of the error norm, the accuracy is determined, e.g., the root mean squared error (RMSE) or the 90% percentile. For unbiased estimators with $\mathbb{E}\{\boldsymbol{e}\} = \boldsymbol{0}$ the RMSE can, under certain conditions, be lower bounded by the Cramér-Rao bound (CRB) [25]. The CRB is a powerful tool not only for benchmarking algorithms and predicting performance, but also for deployment and waveform optimization [14], [20] and for including prior knowledge [26]. Combining all the observations (2) yields a long vector $\boldsymbol{y} = [\boldsymbol{y}_{1,1}^\top, \ldots, \boldsymbol{y}_{N,K}^\top]^\top$, which depends on parameters of interest $\boldsymbol{\eta}$ (e.g., the UE location) of length $d_{\boldsymbol{\eta}}$, and nuisance parameters, say, $\boldsymbol{\xi}$ (e.g., channel gains and clock bias). Then, the Fisher information matrix (FIM) $\boldsymbol{J}(\boldsymbol{\kappa})$ of $\boldsymbol{\kappa} = [\boldsymbol{\eta}^\top, \boldsymbol{\xi}^\top]^\top$ has as elements[5]

$$[\boldsymbol{J}(\boldsymbol{\kappa})]_{\iota,\iota'} = -\mathbb{E}\Big\{ \frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\kappa})}{\partial [\boldsymbol{\kappa}]_\iota} \frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\kappa})}{\partial [\boldsymbol{\kappa}]_{\iota'}} \Big\}, \tag{5}$$

where $\log p(\boldsymbol{y}|\boldsymbol{\kappa})$ is the log-likelihood and $\mathbb{E}\{\cdot\}$ indicates the expectation over the noise. The following inequality holds:

$$\sqrt{\mathbb{E}\left\{\|\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}\|^2\right\}} \geq \sqrt{\text{trace}[\boldsymbol{J}^{-1}(\boldsymbol{\kappa})]_{1:d_{\boldsymbol{\eta}},1:d_{\boldsymbol{\eta}}}}, \tag{6}$$

where the square root is used for easier interpretation of the numerical values. When $\boldsymbol{\eta}$ is the position, orientation, or velocity, the right-hand side of (6) is known as the position error bound (PEB) (expressed in meters), orientation error bound (OEB), or velocity error bound (VEB), respectively. We denote $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = [\boldsymbol{J}^{-1}(\boldsymbol{\kappa})]_{1:d_{\boldsymbol{\eta}},1:d_{\boldsymbol{\eta}}}$ as the covariance of the estimation of $\boldsymbol{\eta}$. In certain cases, the FIM lends itself to analytical manipulation, providing deep insights into the nature of the performance due to various factors, such as sensor deployment or bandwidth [14]. When the FIM is not invertible, the problem is non-identifiable, i.e., there are infinitely many solutions based on the measurements.

B. Resolution

Resolution is the ability to separate correlated signals. For example, if two objects, say, l and l', in (1) have similar AoA, AoD, ToA, and Doppler, they would appear to the receiver as a single object with complex gain $\alpha_l + \alpha_{l'}$. Resolution in one

[5] In most practical cases, the expression is simplified by using the Slepian-Bangs formula $\boldsymbol{J}(\boldsymbol{\kappa}) = 2/N_0 \sum_{n,k} \Re\{ (\frac{\partial \boldsymbol{\mu}_{n,k}}{\partial \boldsymbol{\kappa}})^{\text{H}} \frac{\partial \boldsymbol{\mu}_{n,k}}{\partial \boldsymbol{\kappa}} \}$, where $\boldsymbol{\mu}_{n,k}$ is the noise-free observation in $\boldsymbol{y}_{n,k}$.

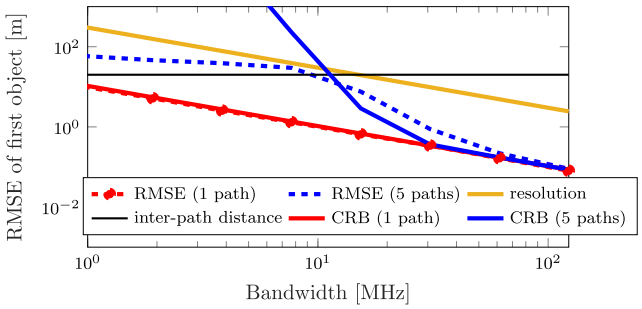


Fig. 3. An evaluation of the CRB (discussed in Section III-C) and estimation RMSE (discussed in Section IV-A) of the closest among 5 objects as a function of the signal bandwidth. Here, '1 path' refers to the case with only one object.

domain is sufficient for objects to be separable. Resolution is also applicable to the localization problem (2), to separate the LoS path from the NLoS channel. We have several domains of resolution [10].

- *Delay resolution:* Two objects can be resolved if their delay difference $|\tau_l - \tau_{l'}|$ is greater than $1/W$. Hence, a larger bandwidth leads to better delay resolution. For example, 400 MHz of bandwidth leads to a distance resolution of 75 cm.
- *Doppler resolution:* Two objects can be resolved if their Doppler difference $|\nu_l - \nu_{l'}|$ is greater than $1/(KT_s)$. Hence, a larger coherent integration time leads to better Doppler resolution. For example, to obtain a radial velocity resolution of 1 m/s with a signal at 30 GHz carrier, the integration time should be at least 10 ms.
- *Angular resolution:* Two objects can be resolved if the difference in azimuth (a similar argument holds for elevation) angle $|\theta_l^{\text{az}} - \theta_{l'}^{\text{az}}|$ is greater than (approximately, since the exact expression depends on the angle itself) $2/N^{\text{az}}$, where N^{az} is the number of $\lambda/2$-spaced elements along the axis with respect to which the azimuth angle is computed. Hence, larger arrays lead to better angle resolution. For example, to obtain 5 degree angular resolution at boresight, around 23 antennas are required.

C. Case Study

To show that resolution is needed to achieve high accuracy, it is instructive to consider a simple example. In Fig. 3, a single-antenna bistatic sensing scenario with 5 objects is evaluated, with inter-object spacing of 20 meters, each with same channel magnitude, set to achieve a 10 dB SNR per object. The figure shows the RMSE of the first object as a function of bandwidth. First, consider the inter-path spacing (black) and resolution (orange). When the resolution curve is below the inter-path spacing (this happens at about 20 MHz), we expect the objects to become resolved. This is confirmed by the CRB: the CRB (blue curve) is high for low bandwidths and around 20 MHz starts to reach the CRB of the single-object case (red curve). Hence, performance far better than the waveform resolution is possible, e.g., by increasing the integrated SNR, provided the signal paths can be resolved. Without sufficient resolution, accuracy is limited.

IV. Localization and Sensing Methods

From the observations $\boldsymbol{y}_{n,k}$, it would be tempting to solve the localization or sensing problem by direct optimization. Letting $\boldsymbol{\kappa}$ contain the UE state (for localization) or object states (for sensing) as well as any nuisance parameters (channel gains), then

$$\hat{\boldsymbol{\kappa}} = \arg\max_{\boldsymbol{\kappa}} \log p(\boldsymbol{y}|\boldsymbol{\kappa}), \tag{7}$$

where the relevant quantities were already defined in Section III-A. This is known as *direct positioning* and has as benefit that all possible information is used, though at a high computational cost [27]. Most practical methods apply a two-stage approach, already hinted at in Fig. 2, whereby first the geometric channel parameters (angles, delays, Dopplers) are estimated, and then the UE / object state is recovered.

A. Channel Parameter Estimation

As the channel estimation problem is also present in wireless communication, there exists a variety of estimators, including FFT/Periodograms [28], ESPRIT [29], generalized approximate message passing [30], orthogonal matching pursuit [18], and RIMAX/SAGE [31], which exploit either underlying sparsity or principles of harmonic retrieval (or both). A common approach is to first obtain an unstructured estimate $\hat{\boldsymbol{H}}_{n,k}$ of the channel (1) from a least-squares (LS) estimator. Introducing $\boldsymbol{a}_{\text{d}}(\tau) = [1, \ldots, e^{-\jmath 2\pi(N-1)\Delta_f \tau_l}]^\top$ and $\boldsymbol{a}_{\text{D}}(\nu) = [1, \ldots, e^{\jmath 2\pi(K-1)T_s \nu_l}]^\top$, after vectorizing and stacking of $\hat{\boldsymbol{H}}_{n,k}$, we can write[6]

$$\hat{\boldsymbol{h}} = \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_{\text{d}}(\tau_l) \otimes \boldsymbol{a}_{\text{D}}(\nu_l) \otimes \boldsymbol{a}_{\text{rx}}(\boldsymbol{\theta}_l) \otimes \boldsymbol{a}_{\text{tx}}(\boldsymbol{\phi}_l) + \boldsymbol{n}, \tag{8}$$

which is in an appropriate form for compressive sensing methods [33]. This was applied in Fig. 3, where orthogonal matching pursuit was used to detect the number of paths and estimate their delays. When the resolution is above the inter-path distance, the number of detected paths is too few, leading to biased estimates, which explains the RMSE (dashed) below the CRB (in blue). When the resolution is below the inter-path distance, the RMSE follows the correspond CRB quite well, and then attains the RMSE for the single-path case.

Alternatively, we can express the LS estimates in a tensor form

$$\hat{h}_{i_1,i_2,\ldots,i_D} = \sum_{l=1}^{L} \alpha_l \prod_{d=1}^{D} e^{\jmath i_d \omega_{d,l}} + n_{i_1,i_2,\ldots,i_D}, \tag{9}$$

where $\omega_{d,l}$ is a so-called spatial frequency. For instance, if index d refers to the subcarrier dimension, then $\omega_{d,l} = -2\pi\Delta_f\tau_l$. Now, (9) is a classic harmonic retrieval problem in D dimensions ($D = 4$ in (8) but can be expanded to $D = 6$ if the Tx and Rx array admit a Kronecker structure). Once the number of objects $\hat{L}$, their gain $\hat{\alpha}_l$ and their geometric parameters $\hat{\boldsymbol{\eta}}_l = [\hat{\boldsymbol{\theta}}_l^\top, \hat{\boldsymbol{\phi}}_l^\top, \hat{\tau}_l, \hat{\nu}_l]^\top$ have been estimated, they can be further refined by optimization of the log-likelihood function $\log p(\boldsymbol{y}|\boldsymbol{\kappa})$ around this initial estimate. If the initial estimate is good enough, this will lead to an efficient estimate, close to the CRB with inverse FIM, say $\boldsymbol{\Sigma}(\hat{\boldsymbol{\eta}}_l)$, which can be used as an uncertainty estimate.

B. Position Estimation

Estimating the state of a UE or an object now relies on the relationship between the estimated channel parameters,

[6]The vectors $\boldsymbol{a}_{\text{tx}}(\boldsymbol{\phi})$ and $\boldsymbol{a}_{\text{D}}(\nu)$ are coupled, which can be resolved by imposing additional assumptions [32].

expressed as $\hat{\boldsymbol{\eta}}$ and the corresponding uncertainty $\boldsymbol{\Sigma}(\hat{\boldsymbol{\eta}})$, to the state of interest. We focus on the localization problem for concreteness.[7] Starting from $\hat{\boldsymbol{\eta}}_1^{(i)} = [(\hat{\boldsymbol{\theta}}_1^{(i)})^\top, (\hat{\boldsymbol{\phi}}_1^{(i)})^\top, \hat{\tau}_1^{(i)}, \hat{\nu}_1^{(i)}]^\top$ and associated uncertainty $\boldsymbol{\Sigma}(\hat{\boldsymbol{\eta}}_1^{(i)})$ of the LoS path from each BS i (see Fig. 2), the UE state is related to these channel parameters measurements through $\hat{\boldsymbol{\eta}}_1^{(i)} = \boldsymbol{h}^{(i)}(\boldsymbol{s}) + \boldsymbol{n}^{(i)}$, where $\boldsymbol{n}^{(i)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\eta}}_1^{(i)}))$ and $\boldsymbol{h}^{(i)}(\boldsymbol{s})$ is the known mapping from UE state to the geometric channel parameters, as described in Section II-B.1. This enables us to express the problem[8]

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \sum_{i=1}^{N_\mathrm{B}} \big(\hat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{h}^{(i)}(\boldsymbol{s})\big)^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\eta}}_1^{(i)}) \big(\hat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{h}^{(i)}(\boldsymbol{s})\big)$$

which is non-convex and can be solved by first obtaining a coarse estimate (e.g., using geometric reasoning, linearization, or relaxation). It is then refined by local optimization of the likelihood function. The final estimate $\hat{\boldsymbol{s}}$ is then used to compute a covariance $\boldsymbol{\Sigma}(\hat{\boldsymbol{s}})$ from

$$\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{s}}) = \sum_{i=1}^{N_\mathrm{B}} \left(\frac{\partial \boldsymbol{\eta}_1^{(i)}}{\partial \boldsymbol{s}}\right)^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\eta}}_1^{(i)}) \frac{\partial \boldsymbol{\eta}_1^{(i)}}{\partial \boldsymbol{s}} \bigg|_{\boldsymbol{s}=\hat{\boldsymbol{s}}}. \quad (10)$$

The couple $(\hat{\boldsymbol{s}}, \boldsymbol{\Sigma}(\hat{\boldsymbol{s}}))$ can then be further processed, e.g., in a tracking filter or sensor fusion engine.

V. Conclusion

In this letter, we have provided an overview of the radio localization and sensing problems, described the basic models, performance metrics and methods. An important focus was on modeling of channels and signals, which is needed to develop practical methods with high accuracy and reasonable uncertainty information. We also emphasized the need for high resolution as a prerequisite for high accuracy. This overview provides the background for the more advanced principles in Part II.

References

[1] S. Thrun *et al.*, *Probabilistic Robotics. Intelligent Robotics and Autonomous Agents*. Cambridge, MA, USA: MIT Press, 2005.

[2] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.

[3] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven defining features of terahertz (THz) wireless systems: A fellowship of communication and sensing," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 967–993, 2nd Quart., 2022.

[4] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, Mar. 2017.

[5] A. Shastri *et al.*, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1708–1749, 3rd Quart., 2022.

[6] E. D. Kaplan *et al.*, *Understanding GPS/GNSS: Principles and Applications*. Norwood, MA, USA: Artech House, 2017.

[7] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1124–1148, 2nd Quart., 2017.

[8] S. Dwivedi *et al.*, "Positioning in 5G networks," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 38–44, Dec. 2021.

[9] M. L. William and S. A. James, *Principles of Modern Radar: Radar Applications*, vol. 3. Trenton, NJ, USA: SciTech, 2014.

[7]For sensing, a similar process is performed for each detected target.

[8]The Gaussian model is justified through the extended invariance principle [34, Thm. 1].

[10] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 20–31, Sep. 2019.

[11] D. Ma *et al.*, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 85–97, Jul. 2020.

[12] C. B. Barneto, S. D. Liyanaarachchi, M. Heino, T. Riihonen, and M. Valkama, "Full duplex radio/radar technology: The enabler for advanced joint communication and sensing," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 82–88, Feb. 2021.

[13] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[14] K. Witrisal *et al.*, "High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 59–70, Mar. 2016.

[15] W. Wang, T. Jost, C. Gentner, S. Zhang, and A. Dammann, "A semiblind tracking algorithm for joint communication and ranging with OFDM signals," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5237–5250, Jul. 2016.

[16] T. D. Barfoot, *State Estimation for Robotics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[17] S. Aditya, A. F. Molisch, and H. M. Behairy, "A survey on the impact of multipath on wideband time-of-arrival based localization," *Proc. IEEE*, vol. 106, no. 7, pp. 1183–1203, Jul. 2018.

[18] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[19] Q. D. Vo and P. De, "A survey of fingerprint-based outdoor localization," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 491–506, 1st Quart., 2015.

[20] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Error bounds for uplink and downlink 3D localization in 5G millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 4939–4954, Aug. 2018.

[21] Y. Ge *et al.*, "5G SLAM using the clustering and assignment approach with diffuse multipath," *Sensors*, vol. 20, no. 16, p. 4656, Aug. 2020.

[22] Y. Han, Y. Shen, X.-P. Zhang, M. Z. Win, and H. Meng, "Performance limits and geometric properties of array localization," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 1054–1075, Feb. 2016.

[23] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[24] K. Granstrom, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview and applications," 2016, *arXiv:1604.00970*.

[25] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2004.

[26] R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative sensor network localization: Algorithms and practical issues," *Proc. IEEE*, vol. 106, no. 6, pp. 1089–1114, Jun. 2018.

[27] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2475–2487, May 2017.

[28] K. M. Braun, "OFDM radar algorithms in mobile communication networks," Ph.D. dissertation, Karlsruher Inst. für Technol. (KIT), Karlsruhe, Germany, 2014.

[29] F. Roemer, M. Haardt, and G. Del Galdo, "Analytical performance assessment of multi-dimensional Matrix- and tensor-based ESPRIT-type algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2611–2625, 2014.

[30] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3205–3219, May 2019.

[31] R. Thomä *et al.*, "RIMAX—A maximum likelihood framework for parameter estimation in multidimensional channel sounding," in *Proc. Int. Symp. Antennas Propag.*, vol. 2004, pp. 53–56.

[32] Y. Cheng, H. Gu, and W. Su, "Joint 4-D angle and Doppler shift estimation via tensor decomposition for MIMO array," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 917–920, Jun. 2012.

[33] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[34] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, no. 2, pp. 421–441, Feb. 1998.