A Projective Geometric View for 6D Pose Estimation in mmWave MIMO Systems

Citation for the original published paper (version of record):
Shen, S., Wymeersch, H. (2024). A Projective Geometric View for 6D Pose Estimation in mmWave MIMO Systems. IEEE Transactions on Wireless Communications, 23(8): 9144-9159.
http://dx.doi.org/10.1109/TWC.2024.3359253

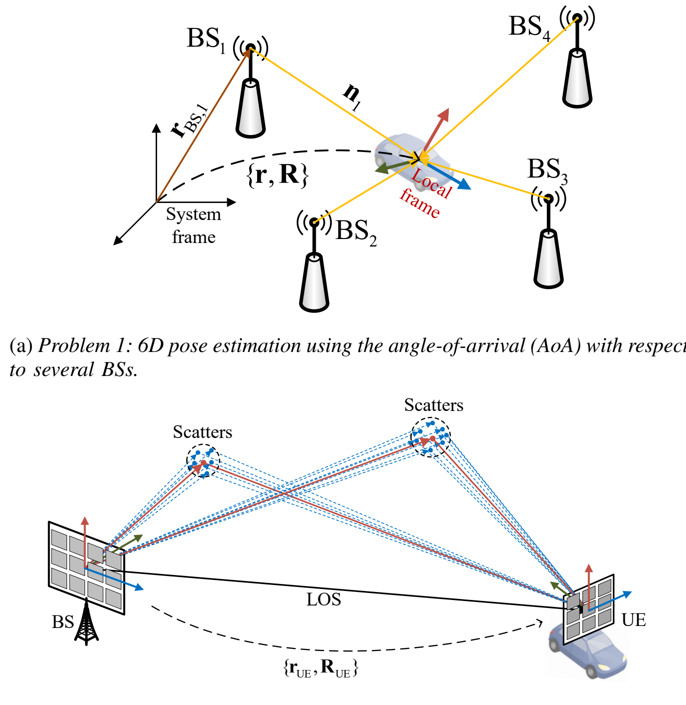
(article starts on next page)

A Projective Geometric View for 6D Pose Estimation in mmWave MIMO Systems

Shengqiang Shen, *Member, IEEE*, Henk Wymeersch, *Fellow, IEEE*

***Abstract*—Millimeter-wave (mmWave) systems in the 30–300 GHz bands are among the fundamental enabling technologies of 5G and beyond 5G, providing large bandwidths, not only for high data rate communication but also for precise positioning services, in support of high accuracy demanding applications such as for robotics, extended reality, or remote surgery. With the possibility to introduce relatively large arrays on user devices with a small footprint, the ability to determine the user orientation becomes unlocked. The estimation of the full user pose (joint 3D position and 3D orientation) is referred to as 6D localization. Conventionally, the problem of 6D localization using antenna arrays has been considered difficult and was solved through a combination of heuristics and optimization. In this paper, we reveal a close connection between the angle-of-arrivals (AoAs) and angle-of-departures (AoDs) and the well-studied perspective projection model from computer vision. This connection allows us to solve the 6D localization problem, by adapting state-of-the-art methods from computer vision. More specifically, two problems, namely 6D pose estimation from AoAs from multiple single-antenna base stations and 6D simultaneous localization and mapping (SLAM) based on single- base station (BS) mmWave communication, are first modeled with the perspective projection model, and then solved. Numerical simulations show that the proposed estimators operate close to the theoretical performance bounds. Moreover, the proposed SLAM method is effective even in the absence of the line-of-sight (LoS) path, or knowledge of the LoS/non-line-of-sight (NLoS) condition.**

***Index Terms*—AoD, AoA, pose estimation, SLAM, antenna arrays, mmWave communication.**



(a) *Problem 1: 6D pose estimation using the angle-of-arrival (AoA) with respect to several BSs.*

(b) *Problem 2: 6D pose estimation from a single-mmWave BS in an environment with several scatterers with unknown locations.*

Fig. 1: Two common configurations for 6D pose estimation in mmWave MIMO systems.

I. Introduction

Continuous development of the fifth-generation (5G) network intends to broaden its uses beyond traditional mobile broadband and to enable a key capability of precise positioning, which is expected to be required in a variety of new applications [1], in particular, inter-robot coordination and extended reality applications [2]. Among the developments towards 5G Advances are the support of sidelinks, improved integrity, carrier phase positioning, and the support of reduced capacity (RedCap) devices [3], [4]. As one of the key enabling components of 5G and beyond 5G, millimeter-wave (mmWave) provides massive bandwidths for high data rates and empowers precise positioning services using cellular technology rather than a separate infrastructure. The possibility to introduce a relatively large array on user devices in mmWave multiple-input multiple-output (MIMO) systems brings the ability to estimate user orientation, in addition to the user position [5]. Information of the full user pose (i.e., joint 3D position and 3D orientation) not only benefits the performance of communication systems [6], but also can be used for various important applications, such as in intelligent transport systems for driving assistance [7]; in assisted living facilities for tracking health status [8]; in search-and-rescue operations with unmanned aerial vehicles (UAVs) for control, self-localization, and victim recovery [9]; in augmented reality applications [10], and several 6G use cases [2]. As the external device that is able to provide the pose information is not always available due to cost and/or size limits, there has been an increasing interest in the use of antenna arrays for joint position and orientation estimation, i.e., pose estimation, referred to here as 6D localization [11].

We consider two particular problems of 6D pose estimation (see Fig. 1). The motivation for these two cases stems from

Shengqiang Shen is with the School of Information and Control Engineering, China University of Mining and Technology, 221000 Xuzhou, China (e-mail: sshen@cumt.edu.cn). Henk Wymeersch is with the Department of Electrical Engineering, Chalmers University of Technology, 41258 Gothenburg, Sweden (e-mail: henkw@chalmers.se).

This work was supported, in part, by the European Commission through the H2020 project Hexa-X (Grant Agreement no. 101015956) and by the Swedish Research Council under grant No. 2018-03701. The work of Shengqiang Shen is supported in part by the Fundamental Research Funds for the Central Universities under Grant 2023QN107. (*Corresponding author: Shengqiang Shen.*)

their challenging nature, meaning that solutions for these problems can be readily simplified to easier cases (e.g., with more measurements or more BSs). Pose estimation in mmWave MIMO necessarily relies on angle measurements at the UE, so that there is a natural breakdown depending on the available bandwidth: when bandwidth is limited, the challenging angle-only pose estimation is obtained (Problem 1 in Fig. 1), while when bandwidth is plentiful, angle and delay information can be used. The most challenging problem in that case is the single-BS pose estimation problem (Problem 2 in Fig. 1).In the multi-BS case (Problem 1), the user attempts to achieve 6D localization using pose-related information with respect to multiple bases. For this case, most works [12]–[14] focus on the position estimation, while the orientation estimation is relatively less investigated. In [15], the authors investigated the orientation estimation using AoAs, assuming the position has been estimated. However, to the best of our knowledge, the joint position and orientation estimation is still missing. At the same time, for the single-BS case (denoted by Problem 2 in this paper), with the increasing depth of the research, the problem of joint position and orientation estimation has been investigated by many works. Enabled by the large bandwidth and antenna arrays with mmWave, in the single-BS case, the estimator exploits the information provided by the multipath to achieve the joint position and orientation estimation. Such multipath exploitation naturally leads to simultaneous localization and mapping (SLAM) problems, in either snapshot [16] or tracking contexts [17]. Initial studies on this topic usually place restrictions on the degrees of freedom or focus on planar scenarios. For example, [16], [18]–[20] studied 2D position and 1D orientation for single-BS case. In these works, the analysis on different aspects, such as estimation method, clock biases, and theoretical bound are more and more profound. In [21]–[23], the authors investigated 3D position and 2D orientation under the synchronized condition. Until recently, 3D position and 3D orientation is investigated in [24], [25]. In [24], the authors studied 3D position and 3D orientation, i.e., 6D localization, in the presence of line-of-sight (LoS) path for the unsynchronized case, where both the estimation method and theoretical bound are analyzed. In [25], the authors investigated the impact of hardware impairment on the joint estimation of 3D orientation and 3D position under the synchronized condition. However, it can be seen that a unified framework that deals with 6D localization in the absence and presence of LoS path is still needed.

The 6D localization problem is not unique to mmWave MIMO systems, but can also be found in robotics and visual SLAM [26], [27], to track the 6D camera poses over time. The perspective projection model from computer vision [28] provides a natural approach to solve the 6D localization problem in this context, with a rich and efficient set of tools [29]. In the visual SLAM setting, there is no notion of a LoS path, since a camera cannot see itself in the past, while the measurements are different compared to the mmWave MIMO setting, since distance measurements are affected by clock biases. Hence, computer vision methods have seen limited application in the context of mmWave MIMO positioning.

In this paper, we reveal a connection between the AoAs and angle-of-departures (AoDs) and the perspective projection model. This connection allows us to solve the 6D localization problem, by adapting state-of-the-art methods from computer vision. In contrast to [12]–[14], the unknown orientation is naturally accounted for, while in contrast to [16], [18]–[23] a general 3D unknown orientation can be estimated. The proposed projective geometric view also solves a broader class of problems than [24], [25], since the LoS is not required. Moreover, by relying on state-of-the-art methods from computer vision, the 6D localization problem can be solved efficiently, without the need for grid searches. The two previously mentioned configurations (pose estimation from AoAs from multiple base stations (BSs) and 6D SLAM based on a single mmWave BS), selected for their generality and close relation to existing mmWave MIMO positioning problems, are modeled with the perspective projection model and then solved, building on computer vision tools. The main contributions of our work are as follows.

- **A projective geometric view for the AoA/AoD:** We reveal a connection between projective geometry and the perspective projection model from computer vision and the AoA/AoD of antenna arrays. With the help of this model, we derive an explicit expression relating the AOD/AOA to the position of the base station (BS) and the pose of the user equipment (UE). These expressions from the basis for new methods in the mmWave MIMO domain, requiring tailored modifications to account for the properties of radio signals.
- **Novel methods for solving Problem 1:** Through the perspective projection modeling, we provide two novel methods for 6D pose estimation using AoAs, a closed-form one and an iterative one based on the least squares (LS) principle. The required number of paths, i.e., the number of BSs, is also determined.
- **Novel methods for solving Problem 2:** On the basis of perspective projection modeling, the geometric relation between AoD, AoA, and scatter points for single BS and single UE scenarios is further modeled with the epipolar model.[1] Based on this modeling, we propose two novel algorithms for SLAM based on mmWave communication, a closed-form one and an iterative one as well. The required number of paths (scatter points) is also determined.
- **Detailed performance evaluation:** The performance of all the proposed algorithms is assessed using Monte Carlo simulations, and the tightness of the results with the Cramér-Rao bound (CRB) is evaluated to show that the method is efficient. For Problem 2, the proposed method is thoroughly evaluated in a 3D propagation environment, proving its performance at various degrees of surface roughness.

The rest of the paper is organized as follows. The projective geometric modeling is presented in Section III. The two use cases are investigated in Section III, where the proposed methods as well as the derived theoretical lower bound are

[1]For the reader unfamiliar with epipolar models, we refer to Fig. 3 and Fig. 5 for a quick view of such models.

given. The Monte Carlo numerical comparison is given in Section IV. Finally, some concluding remarks are given in Section V.

Notations

We introduce the unit vectors $\mathbf{e}_1 = [1\,0\,0]^\mathrm{T}$, $\mathbf{e}_2 = [0\,1\,0]^\mathrm{T}$, and $\mathbf{e}_3 = [0\,0\,1]^\mathrm{T}$. The operator $\bar{\cdot}$ converts a vector from Cartesian coordinates $\mathbf{x}$ into the homogeneous coordinates, i.e., $\bar{\mathbf{x}} = [\mathbf{x}^\mathrm{T}, 1]^\mathrm{T}$. A line between two points $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\overrightarrow{\mathbf{xy}}$. The operator $\mathbf{x}_\times$ generates a skew-symmetric matrix

$$\mathbf{x}_\times = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}. \tag{1}$$

The group of all rotation matrices, i.e., the special orthogonal group, is denoted by $SO(3)$, and the associated Lie algebra is denoted by $\mathfrak{so}(3)$. The group of all pose matrices, i.e., the special Euclidean group, is denoted by $SE(3)$, and the associated Lie algebra is denoted by $\mathfrak{se}(3)$. The operator $\cdot^\wedge$ converts a 6×1 vector into a member of $\mathfrak{se}\,(3)$ by

$$\left([\mathbf{y}^\mathrm{T}, \mathbf{x}^\mathrm{T}]^\mathrm{T}\right)^\wedge = \begin{bmatrix} \mathbf{x}_\times & \mathbf{y} \\ \mathbf{0}^\mathrm{T} & 0 \end{bmatrix} \in \mathbb{R}^{4\times4}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^{3\times1}. \tag{2}$$

We also introduce the shorthand $[x_{\ell\in\{1,2,3\}}] = [x_1, x_2, x_3]$.

II. System Model and Problem Formulation

A. General System Model

We consider a scenario with a single mmWave multi-antenna BS with known position $\mathbf{r}_\mathrm{BS}$ and a single multi-antenna UE with unknown position $\mathbf{r}_\mathrm{UE}$ and unknown orientation $\mathbf{R}_\mathrm{UE} \in SO(3)$. Under downlink transmission, assuming $L \geq 1$ paths between the BS and UE, the channel for subcarrier f_n, represented by $\mathsf{H}[n] \in \mathbb{C}^{N_r \times N_t}$, is given by[2] [30]

$$\mathsf{H}[n] = \sum_{\ell\in\mathcal{S}_L} \alpha_\ell e^{-j2\pi\tau_\ell f_n} \mathbf{a}_{N_r}(\boldsymbol{\psi}_{\mathrm{R},\ell})\, \mathbf{a}_{N_t}^\mathrm{T}(\boldsymbol{\psi}_{\mathrm{T},\ell}), \tag{3}$$

where α_ℓ is the ℓ^th channel gain, τ_ℓ the ℓ^th propagation delay and

$$\mathcal{S}_L = \begin{cases} \{0, \ldots, L\} \text{ for LoS case,} \\ \{1, \ldots, L\} \text{ for NLoS case.} \end{cases} \tag{4}$$

The propagation delays are related to the UE position by

$$\tau_\ell = \begin{cases} B + \|\mathbf{r}_\mathrm{UE} - \mathbf{r}_\mathrm{BS}\|/c & \ell = 0 \\ B + (\|\mathbf{p}_\ell - \mathbf{r}_\mathrm{BS}\| + \|\mathbf{p}_\ell - \mathbf{r}_\mathrm{UE}\|)/c & \ell > 0, \end{cases} \tag{5}$$

where B is the UE's unknown clock bias and $\mathbf{p}_\ell$ is a location of a scatterer, assuming at most single-bounce reflections.[3] In addition, $N_r = N_{r,x}N_{r,y}$ is the number antenna of elements in the UE array, comprising $N_{r,x}$ elements along the local x-axis and $N_{r,y}$ along the local y-axis. Similarly, $N_t = N_{t,x}N_{t,y}$ is the number of antenna elements in the BS array. We focus on planar antenna arrays, and, without loss of generality, we consider the uniform rectangular array. The angular information of the wavefront, e.g., AoD/AoA, is conventionally represented by the azimuth and elevation angles $\boldsymbol{\psi} = [\phi, \theta]^\mathrm{T}$, which specify the normal vector

$$\mathbf{n} = \begin{bmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{bmatrix} \tag{6}$$

of the wavefront direction. Then, the steering vector $\mathbf{a}_{N_r}(\boldsymbol{\psi}_{\mathrm{R},\ell}) \in \mathbb{C}^{N_r\times1}$ is defined as

$$\mathbf{a}_{N_r}(\boldsymbol{\psi}_{\mathrm{R},\ell}) = \mathfrak{a}_1(\boldsymbol{\psi}_{\mathrm{R},\ell}) \otimes \mathfrak{a}_2(\boldsymbol{\psi}_{\mathrm{R},\ell}), \tag{7}$$

with

$$[\mathfrak{a}_1(\boldsymbol{\psi})]_m = e^{j\pi m\sin(\theta)\sin(\phi)},\ m \in \{0, \ldots, N_x\} \tag{8}$$

$$[\mathfrak{a}_2(\boldsymbol{\psi})]_m = e^{j\pi m\sin(\theta)\cos(\phi)},\ m \in \{0, \ldots, N_y\}, \tag{9}$$

assuming the antenna spacing is of half-wavelength. The AoD $\boldsymbol{\psi}_{\mathrm{T},\ell} \in \mathbb{R}^{2\times1}$ and steering vector $\mathbf{a}_{N_t}(\boldsymbol{\psi}_{\mathrm{T},\ell})$ are defined similarly.

B. Problem 1 – AoA-only Pose Estimation

In this problem (as visualized in Fig. 1a), the goal is to estimate the receiver's pose based on the AoA with respect to multiple single-antenna BSs at known positions, from narrowband downlink signals under *pure LoS propagation*. Here, narrowband refers to the fact that there are no distinct subcarriers and hence no ability to estimate propagation delays. In that case, the channel from BS i to the UE simplifies to $\boldsymbol{h}_i = \alpha_i \mathbf{a}_{N_r}(\boldsymbol{\psi}_{\mathrm{R},i})$. The focus is on processing after the AoAs have been estimated (there are many algorithms to estimate AoA, such as the MUSIC algorithm [32, Sect. 9.3.2]), since the AoA-based pose estimation was only discussed partially in the literature, in that either positioning or orientation estimation, e.g., [15], [33], was considered. Therefore, the joint estimation of position and orientation, that is, the pose estimation has not been discussed previously.

C. Problem 2 – mmWave MIMO Snapshot SLAM

As shown in Fig. 1b, the objective is to simultaneously estimate the UE pose $\{\mathbf{r}_\mathrm{UE}, \mathbf{R}_\mathrm{UE}\}$ as well as the position of scatters $\mathbf{p}_\ell$, $\ell \in \{1\ldots, L\}$, based on an estimate of the channels $\mathsf{H}[n]$. This objective of mmWave communication-based SLAM is usually achieved through a sequence of multi-step processes [16], [17]:

1) Estimation of the channel matrices $\mathsf{H}[n]$ for $n \in \{1, \ldots, N_f\}$ based on the observed pilot signal (this is not the core of this work and will be discussed in Section V);
2) Estimation of the (effective) parameter vector $\mathbf{z}_\ell = [\boldsymbol{\psi}_{\mathrm{R},\ell}, \boldsymbol{\psi}_{\mathrm{T},\ell}, \tau_\ell, \alpha_\ell]^\mathrm{T}$ from $\mathsf{H}[n]$ for all N_f subcarriers,
3) Estimation of the UE pose and the position of scatters, i.e., SLAM, based on the parameter vector set $\mathcal{Z} = \{\mathbf{z}_0, \ldots, \mathbf{z}_{\hat{L}}\}$, where $\hat{L}$ is the detected number of paths.

As in Problem 1, the current paper focuses mainly on step 3, assuming the channel parameter vectors are available. Since the first two steps give the estimate of $\mathbf{z}_\ell$, the ambiguities caused by steps 1-2 are parameterized by the estimate's

[2] Doppler is not considered in this work and its effect is absorbed in the complex channel gain, which is a correct model under the assumption of short transmission periods and/or medium mobility.

[3] Due to the mmWave propagation characteristics, the received power contributed by multiple-bounce reflections is negligible [16], [31]. Multiple-bounce reflections should be removed prior to the application of the method. On the other hand, effects such as scattering or diffraction can still be accounted for, as each resolved path would correspond to a unique incidence point.

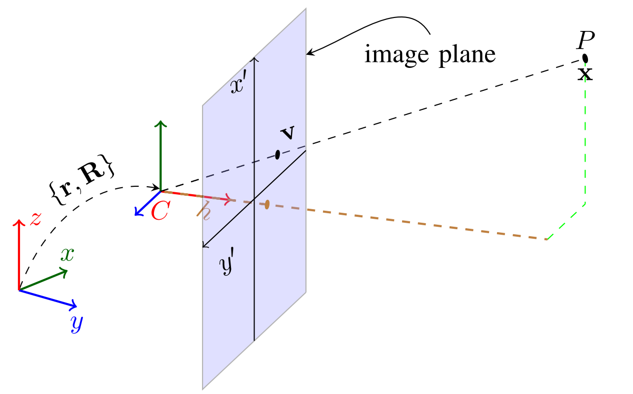


Fig. 2: The perspective projection model of a camera with position $\mathbf{r}$ and orientation $\mathbf{R}$, showing the image coordinate $\mathbf{v}$ of a point P.

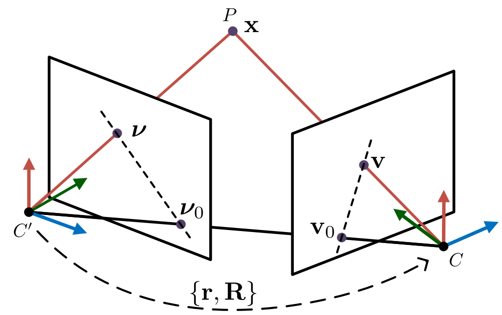


Fig. 3: The epipolar modeling, involving the image coordinates with respect to two cameras (C, C') of a common point P with system coordinates $\mathbf{x}$. The epipoles are $\mathbf{v}_0$ and $\boldsymbol{\nu}_0$. Two epipolar lines are shown in dashed.

PDF. Such ambiguities can be accounted for by solving the SLAM problem for each possible case, followed by ruling out physically inconsistent cases. While mmWave MIMO channel parameter estimation is challenging, there exist several in the literature, such as tensor-ESPRIT [21], [22], orthogonal matching pursuit [34], or received power per beam [35]. Unless stated otherwise, we will assume that *we do not know whether the LoS path is present.*

III. Background on Computer Vision and Relation to AoA and AoD

In this section, a short primer on basic results from computer vision is given, followed by the application of these results to expression AoAs and AoDs in mmWave MIMO communication systems.

A. Perspective Projection Model

Consider an ideal perspective camera and a point P with homogeneous coordinates $\bar{\mathbf{x}}_{\text{cam}} \in \mathbb{R}^4$ in the camera frame and homogeneous coordinates $\bar{\mathbf{x}} \in \mathbb{R}^4$ in the system frame. These coordinates are linked by the coordinate transformation

$$\bar{\mathbf{x}}_{\text{cam}} = \mathbf{T}\bar{\mathbf{x}}, \tag{10}$$

where the matrix $\mathbf{T} \in SE\,(3)$ belongs to the Special Euclidean group $SE\,(3)$ and is defined through its inverse

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{R} & \mathbf{r} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}, \tag{11}$$

in which $\{\mathbf{r}, \mathbf{R}\}$ represent the pose of the camera, comprising a displacement vector $\mathbf{r} \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R} \in SO(3)$, depicted in Fig. 2.

The image coordinates $\mathbf{v} \in \mathbb{R}^2$ of P (see Fig. 2) are given by the so-called perspective projection model as [28, Sect. 6.1]

$$\lambda\bar{\mathbf{v}} = \mathbf{P}\bar{\mathbf{x}}_{\text{cam}}, \tag{12}$$

where $\bar{\mathbf{v}}$ are the homogeneous coordinate of the image point $\mathbf{v}$, λ is a scale factor ($\lambda = 1/\mathbf{e}_3^{\mathrm{T}}\mathbf{x}_{\text{cam}}$), and $\mathbf{P} \in \mathbb{R}^{3\times 4}$ is a projection matrix, given by

$$\mathbf{P} = \begin{bmatrix} h, 0, 0 \\ 0, h, 0 \\ 0, 0, 1 \end{bmatrix} [\mathbf{I}_{3\times 3}\ \mathbf{0}_{3\times 1}], \tag{13}$$

in which h the camera focal length (i.e., the distance of the image plane).

Taking into account (10)–(12) and keeping only the first two elements of $\bar{\mathbf{v}}$, we have a concise relation between the 2D image coordinates $\mathbf{v}$ of a 3D point P and its 3D coordinates $\mathbf{x}$ in the system frame:

$$\mathbf{v} = \frac{\mathbf{K}\mathbf{T}\bar{\mathbf{x}}}{\mathbf{e}_3^{\mathrm{T}}\mathbf{T}\bar{\mathbf{x}}} \in \mathbb{R}^{2\times 1}, \tag{14}$$

where

$$\mathbf{K} = \begin{bmatrix} h, 0, 0, 0 \\ 0, h, 0, 0 \end{bmatrix} \in \mathbb{R}^{2\times 4}. \tag{15}$$

The image coordinates will later be connected to the angle measurements so that the perspective projection model provides a way to relate angle measurements to the pose.

B. Epipolar Model

Consider now two cameras, denoted by C and C' and a point P. We now denote by $\{\mathbf{r}, \mathbf{R}\}$ the relative pose of camera C to camera C' (see Fig. 3). The point P leads to two image coordinates, say $\boldsymbol{\nu}$ and $\mathbf{v}$, which are related to one another, via the relative pose $\{\mathbf{r}, \mathbf{R}\}$. In this section, we will describe this relation.

The two camera centers can be connected by a line. This line is called the baseline in computer vision, and it intersects the image planes at two points, i.e., $\mathbf{v}_0$ and $\boldsymbol{\nu}_0$, which are called the *epipoles*. The line joining an image point and the epipole in each image plane is called an *epipolar line*, e.g., the line $\overrightarrow{\mathbf{v}_0\mathbf{v}}$. It follows that for each point $\mathbf{v}$ in one image, there exists a corresponding epipolar line $\overrightarrow{\boldsymbol{\nu}_0\boldsymbol{\nu}}$ in the other image. Any point $\boldsymbol{\nu}$ in the second image matching the point $\mathbf{v}$ must lie on the epipolar line $\overrightarrow{\boldsymbol{\nu}_0\boldsymbol{\nu}}$. The above relation between these image points is characterized by the epipolar model in computer vision, which is [28, Sect. 9.6]

$$\bar{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{E}\bar{\mathbf{v}} = 0, \tag{16}$$

where

$$\mathbf{E} = \mathbf{r}_{\times}\mathbf{R} \in \mathbb{R}^{3\times 3} \tag{17}$$

is the so-called *essential matrix*. The essential matrix is used in computer vision to determine relative poses between two cameras, based on matched image points. The epipolar model can thus be used to relate the scatterers seen by the BS with those seen by the UE.

C. A Projective Geometric View of the AoD/AoA

We are now ready to relate the perspective projection model and the epipolar model to AoAs and AoDs in mmWave MIMO systems. Due to the antenna reciprocity, we focus on the AoA

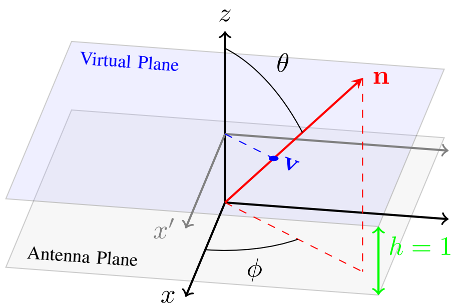


Fig. 4: Illustration of the virtual plane of a uniform rectangular array. A direction $\mathbf{n}$ defines a line that can be intersected with a plane (called the virtual plane) parallel to the array at height $h = 1$. The intersection point $\mathbf{n}$ is related to the pose.

at the receiver end in the following, and a similar conclusion can be inferred for the AoD at the transmitter end.

We recall that the angular information of the wavefront, e.g., AoD/AoA, is conventionally represented by the azimuth and elevation angles (ϕ, θ), which specify the normal vector

$$\mathbf{n} = \begin{bmatrix} \cos\phi \sin\theta \\ \sin\phi \sin\theta \\ \cos\theta \end{bmatrix} \tag{18}$$

of the wavefront direction. Consider now the physical antenna plane (say, the XY plane in the array frame) and a virtual plane parallel to the XY plane at $z = 1$, as shown in Fig. 4. It can be readily seen that each wavefront direction $\mathbf{n} \in \mathbb{R}^3$ can be bijectively mapped to a 2D point on the plane, i.e., the intersection $\mathbf{v} \in \mathbb{R}^{2\times 1}$ of the virtual plane and the line of the wavefront direction through the origin, with (in homogeneous coordinates)

$$\bar{\mathbf{v}} = \frac{\mathbf{n}}{(\mathbf{e}_3^{\mathrm{T}}\mathbf{n})} = \begin{bmatrix} \cos\phi \tan\theta \\ \sin\phi \tan\theta \\ 1 \end{bmatrix}. \tag{19}$$

The line $\overrightarrow{\mathbf{r}_{\mathrm{BS}}\mathbf{r}_{\mathrm{UE}}}$ specifies the wavefront direction at the UE receiver end of the direct path. By analogy with a camera, if the BS is seen as the object point, the UE antenna center as the camera center, after introducing the virtual plane, $\mathbf{v}$ can be seen as the "projected" point on the virtual plane at the UE of the point $\mathbf{r}_{\mathrm{BS}}$ along the line $\overrightarrow{\mathbf{r}_{\mathrm{BS}}\mathbf{r}_{\mathrm{UE}}}$, complying with the perspective projection model from Fig. 2. Consequently, if we further assume that the orientation of the receiver in the system frame is described by $\mathbf{R}_{\mathrm{UE}} \in SO(3)$, then, following (14), the coordinates $\mathbf{v}$ are given by

$$\mathbf{v} = \frac{\mathbf{K}\mathbf{T}_{\mathrm{UE}}\bar{\mathbf{r}}_{\mathrm{BS}}}{\mathbf{e}_3^{\mathrm{T}}\mathbf{T}_{\mathrm{UE}}\bar{\mathbf{r}}_{\mathrm{BS}}} \in \mathbb{R}^{2\times 1}, \tag{20}$$

where the matrix $\mathbf{T}_{\mathrm{UE}} \in SE(3)$ is defined through its inverse

$$\mathbf{T}_{\mathrm{UE}}^{-1} = \begin{bmatrix} \mathbf{R}_{\mathrm{UE}} & \mathbf{r}_{\mathrm{UE}} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}, \tag{21}$$

and $\mathbf{K}$ was defined in (15), with $h = 1$. Hence, based on (18)–(20), we have related the angular information (represented by either $\mathbf{v}$, $\mathbf{n}$ or (θ, ϕ)) to the BS position $\mathbf{r}_{\mathrm{BS}}$, and UE pose $\mathbf{T}_{\mathrm{UE}}$ with the perspective projection model. For the AoD, we have a similar relation, i.e.,

$$\boldsymbol{\nu} = \frac{\mathbf{K}\mathbf{T}_{\mathrm{BS}}\bar{\mathbf{r}}_{\mathrm{UE}}}{\mathbf{e}_3^{\mathrm{T}}\mathbf{T}_{\mathrm{BS}}\bar{\mathbf{r}}_{\mathrm{UE}}} \in \mathbb{R}^{2\times 1}, \tag{22}$$

where $\mathbf{T}_{\mathrm{BS}}$ specifies the pose of the BS.

IV. Projective Geometry Solutions to mmWave MIMO Pose Estimation

Based on the relations (20) and (22), we can now reformulate the problems from Section II-B and Section II-C in terms of a perspective projection model, in order to apply methods from computer vision. We start with Problem 1, as it is less complex.

A. Solution to Problem 1 – AoA-only Pose Estimation

The perspective projection model provides us with a new mathematical description to simplify the problem from Fig. 1a. After obtaining virtual points from AoAs, the problem is converted into estimating the UE pose from 3D to 2D point correspondences $(\mathbf{r}_{\mathrm{BS},i}, \mathbf{v}_i)$, $i = 1, \ldots, I$ in analogy to the problem of estimating the camera's pose from 3D landmark to 2D image correspondences. We denote by $\tilde{\mathbf{V}} = [\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_I]$ the observation matrix containing the observed virtual points $\tilde{\mathbf{v}}_i$ converted from AoA (ϕ_i, θ_i) according to (19), for $i \in \{1, \ldots, I\}$, We also introduce $\mathbf{V}(\mathbf{T}_{\mathrm{UE}}) = [\mathbf{v}_1, \ldots, \mathbf{v}_I]$ as a function of $\mathbf{T}_{\mathrm{UE}}$ with $\mathbf{v}_i$ modeled by (20). With these formulations, standard computer vision methods can be applied, including closed-form solutions such as the Perspective-n-Point (PnP) algorithm [36, 12.2], as well as the iterative method given by

$$\text{minimize} \quad \left\| \tilde{\mathbf{V}} - \mathbf{V}(\mathbf{T}_{\mathrm{UE}}) \right\|_F^2 \tag{23a}$$

$$\text{s.t.} \quad \mathbf{T}_{\mathrm{UE}} \in SE(3) \tag{23b}$$

in order to solve for $\mathbf{T}_{\mathrm{UE}}$.

The LS algorithm for iterative solving the problem (23) is specified by each iteration, which is [37, Eq. 7.196]

$$\mathbf{T}_{\mathrm{UE}}^{t+1} = \exp\left(\left(\kappa^t \boldsymbol{\Delta}^t \right)^{\wedge} \right) \mathbf{T}_{\mathrm{UE}}^t, \tag{24}$$

where $\kappa^t > 0$ control the incremental step size for $\mathbf{T}_{\mathrm{UE}}$. In (24), the update direction is calculated by

$$\boldsymbol{\Delta} = \left(\nabla_{\mathbf{T}_{\mathrm{UE}}} \mathrm{vec}\left(\mathbf{V}\right) \right)^{\dagger} \left(\mathrm{vec}\left(\tilde{\mathbf{V}} - \mathbf{V} \right) \right) \tag{25}$$

where $(\cdot)^{\dagger}$ is the pseudoinverse and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_I]$ For the Jacobian matrix, refer to [38, Appendix A].. In addition, the problem (23) can be solved by off-the-shelf algorithm toolboxes for cameras, and closed-form algorithms, such as the PnP algorithm, can be used for initialization.

Finally, we note that since the pose has six degrees of freedom and that each correspondence generates two constraints, at least $I = 3$ bases are needed to estimate the pose [28, Sect. 7.3].

B. Solution to Problem 2 – mmWave MIMO Snapshot SLAM

Moving on to Problem 2, we focus on step 3 from Section II-C, considering that estimates of the AoAs and AoDs are given. To simplify the analysis, we will set the BS pose so that $\mathbf{R}_{\mathrm{BS}} = \mathbf{I}_3$ and $\mathbf{r}_{\mathrm{BS}} = \mathbf{0}_{3\times 1}$, i.e., set the system frame to align with the BS antenna, then the UE pose is given by the relative pose to the BS. We apply the perspective projection model to both the UE and BS antenna arrays, so that each scatter point makes a virtual point on each virtual plane, as shown in Fig. 5. It can be seen that (i) the LoS path is represented by the baseline, which intersects the virtual planes at $\mathbf{v}_0$ (at

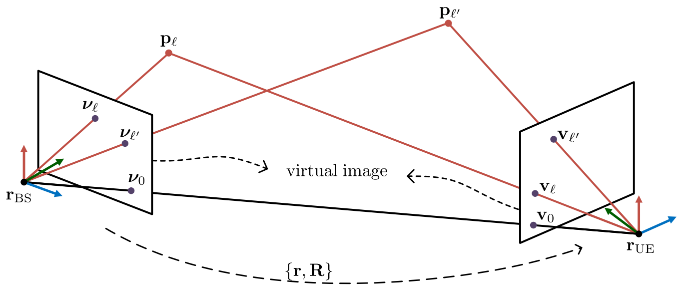


Fig. 5: Illustration of 3D MIMO channel in the epipolar modeling. The BS and UE each act similarly to a camera, which sees the same landmarks (scatter points) from different perspectives. The local virtual images, comprising the virtual points determined by the scattering points in the local frame of reference, as well as the baseline, can be related to the UE pose via the essential matrix.

UE) and $\boldsymbol{\nu}_0$ (at BS); (ii) each scatter point $\mathbf{p}_\ell$ is "projected" in two virtual points, at $\mathbf{v}_\ell$ (at the UE), and $\boldsymbol{\nu}_\ell$ (at the BS). As a result, the geometric relation between the transmitter, receiver, and scatter points can be described by the epipolar model from Section III-B:

$$\bar{\boldsymbol{\nu}}_\ell^{\mathrm{T}}\mathbf{E}\bar{\mathbf{v}}_\ell = 0, \quad \text{for } \ell \in \mathcal{S}_L, \tag{26}$$

where the essential matrix $\mathbf{E} = (\mathbf{r}_{\mathrm{UE}})_\times \mathbf{R}_{\mathrm{UE}}$ fully describes the UE pose.

We now proceed to solve the SLAM problem, relying on existing computer vision methods and applying modifications where needed. Note that the methods ignore the effect of the estimation errors in the angles and channel delays, in order to obtain a closed-form SLAM method, summarized in Alg. 1. These effects can be later accounted for in suitable maximum-likelihood methods, initialized by our approach. We also ignore the effect that the paths in (3) may not all be resolvable. Since the gross effect of these factors is mathematically intractable, all these effects will be considered in the numerical results in Section V, in which the robustness of the proposed closed-form and iterative solutions are analyzed numerically.

We start from estimates of the AoA $\tilde{\boldsymbol{\psi}}_{\mathrm{R},\ell}$, AoD $\tilde{\boldsymbol{\psi}}_{\mathrm{T},\ell}$, and delays $\tilde{\tau}_\ell$, for $\ell \in \mathcal{S}_L$. Here $\tilde{\cdot}$ is used to denote observations (inputs), while $\hat{\cdot}$ is used to denote estimates (outputs).

Algorithm 1 Summary of the Closed-Form Algorithm

Require: $\tilde{\boldsymbol{\psi}}_{\mathrm{R},\ell}$ $\tilde{\boldsymbol{\psi}}_{\mathrm{T},\ell}$, $\tilde{\tau}_\ell$, for $\ell \in \mathcal{S}_L$
- Obtain $\tilde{\boldsymbol{\nu}}_\ell$, and $\tilde{\mathbf{v}}_\ell$ based on (19)
- Estimate essential matrix $\hat{\mathbf{E}}$ based on $\tilde{\boldsymbol{\nu}}_\ell$, and $\tilde{\mathbf{v}}_\ell$
- Estimate $\hat{\mathbf{R}}_{\mathrm{UE}}$, and $\hat{\mathbf{n}}_r$ based on $\hat{\mathbf{E}}$
- Triangulation for $\breve{\mathbf{p}}_\ell$ based on $\hat{\mathbf{R}}_{\mathrm{UE}}$, $\hat{\mathbf{n}}_{\mathbf{r}}$, $\tilde{\mathbf{v}}_\ell$ and $\tilde{\boldsymbol{\nu}}_\ell$
- Metric reconstruction of $\hat{s}$ (29), and recover $\hat{\mathbf{p}}_\ell$ and $\hat{\mathbf{r}}_{\mathrm{UE}}$,

1) Phase 1: Estimation of the virtual points and the essential matrix from the AoDs and AoAs: We first convert AoDs and AoAs estimates into the virtual points, i.e., from $\tilde{\boldsymbol{\psi}}_{\mathrm{R},\ell}$ and $\tilde{\boldsymbol{\psi}}_{\mathrm{T},\ell}$ to $\tilde{\mathbf{v}}_\ell$ and $\tilde{\boldsymbol{\nu}}_\ell$ for $\ell \in \mathcal{S}_L$, based on (19). Then, based on multiple pairs of $\tilde{\mathbf{v}}_\ell$ and $\tilde{\boldsymbol{\nu}}_\ell$, the essential matrix $\mathbf{E}$ can be estimated by using, for example, the algorithm given in [39], [40]. Estimation of the essential matrix is very involved, but in principle it is a technology, meaning that it is a standard function in available computer vision toolboxes with well-defined interfaces, e.g., [41], [42].

2) Phase 2: Computation of the relative UE pose from the estimate $\hat{\mathbf{E}}$: This step can be implemented by applying the SVD-based algorithm from [28, Sects. 9.6.2 and 9.6.3] to $\hat{\mathbf{E}}$ to estimate the orientation $\mathbf{R}_{\mathrm{UE}}$ and the normal vector of the position $\mathbf{n}_{\mathbf{r}} = \mathbf{r}_{\mathrm{UE}}/\|\mathbf{r}_{\mathrm{UE}}\|$. Note that in this step, the UE position can only be estimated up to scale, therefore, the position is only estimated up to the unit vector $\mathbf{n}_{\mathbf{r}}$. The SVD-based algorithm [28, Sect. 9.6.2] for this step gives four candidate solutions, i.e.,

$$\begin{bmatrix}\hat{\mathbf{R}}_{\mathrm{UE}} & \hat{\mathbf{n}}_{\mathbf{r}}\end{bmatrix} = \begin{bmatrix}\mathbf{U}_E\mathbf{W}_E\mathbf{V}_E^{\mathrm{T}} & +\mathbf{u}_{\mathrm{E},3}\end{bmatrix} \tag{27a}$$
$$\text{or } \begin{bmatrix}\mathbf{U}_{\mathrm{E}}\mathbf{W}_{\mathrm{E}}\mathbf{V}_{\mathrm{E}}^{\mathrm{T}} & -\mathbf{u}_{\mathrm{E},3}\end{bmatrix} \tag{27b}$$
$$\text{or } \begin{bmatrix}\mathbf{U}_{\mathrm{E}}\mathbf{W}_{\mathrm{E}}^{\mathrm{T}}\mathbf{V}_{\mathrm{E}}^{\mathrm{T}} & +\mathbf{u}_{\mathrm{E},3}\end{bmatrix} \tag{27c}$$
$$\text{or } \begin{bmatrix}\mathbf{U}_{\mathrm{E}}\mathbf{W}_{\mathrm{E}}^{\mathrm{T}}\mathbf{V}_{\mathrm{E}}^{\mathrm{T}} & -\mathbf{u}_{\mathrm{E},3}\end{bmatrix} \tag{27d}$$

where $\mathbf{W}_{\mathrm{E}} = \begin{bmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$, $\mathbf{u}_{\mathrm{E},3} = \mathbf{U}_{\mathrm{E}}\mathbf{e}_3$, and $\mathbf{U}_{\mathrm{E}}\boldsymbol{\Sigma}_{\mathrm{E}}\mathbf{V}_{\mathrm{E}}^{\mathrm{T}}$ composes the SVD of $\hat{\mathbf{E}}$. The final solution is determined from the 4 candidate solutions by physical feasibility. A physically realizable solution is the one that puts reconstructed 3D points in front of both virtual planes.

3) Phase 3: Triangulation for computing the 3D scatter points: The scatter positions are estimated (again up to a scale as with $\mathbf{n}_{\mathbf{r}}$ [43, Sect. 5.2.2]) and denoted by $\breve{\mathbf{p}}_\ell, \ell \in \{1,\ldots,L\}$, based on $\hat{\mathbf{R}}_{\mathrm{UE}}$, $\hat{\mathbf{n}}_{\mathbf{r}}$, $\tilde{\mathbf{v}}_\ell$ and $\tilde{\boldsymbol{\nu}}_\ell$ by using, for example, the homogeneous method [28, Sect. 12.2]. In particular, the homogeneous method from [28, Sect. 12.2] (for this step) operates as follows: the triangulation solution for the scatter $\breve{\mathbf{p}}_\ell, \ell \in \{1,\ldots,L\}$ is given by the smallest singular value of

$$\mathbf{A}_\ell = \begin{bmatrix}[\tilde{\mathbf{v}}_\ell]_1\,\mathbf{t}_{(3)}^{\mathrm{T}} - \mathbf{t}_{(1)}^{\mathrm{T}}\\ [\tilde{\mathbf{v}}_\ell]_2\,\mathbf{t}_{(3)}^{\mathrm{T}} - \mathbf{t}_{(2)}^{\mathrm{T}}\\ [\tilde{\boldsymbol{\nu}}_\ell]_1\,\mathbf{t}_{(3)}'^{\mathrm{T}} - \mathbf{t}_{(1)}'^{\mathrm{T}}\\ [\tilde{\boldsymbol{\nu}}_\ell]_2\,\mathbf{t}_{(3)}'^{\mathrm{T}} - \mathbf{t}_{(2)}'^{\mathrm{T}}\end{bmatrix}, \tag{28}$$

where $\mathbf{t}_{(i)}^{\mathrm{T}}$ and $\mathbf{t}_{(i)}'^{\mathrm{T}}$ are the i-th row of $\mathbf{T}_{\mathrm{BS}} = \mathbf{I}$ and $\hat{\mathbf{T}}_{\mathrm{UE}} = \begin{bmatrix}\hat{\mathbf{R}}_{\mathrm{UE}} & \hat{\mathbf{n}}_{\mathbf{r}}\\ \mathbf{0}_{1\times 3} & 1\end{bmatrix}^{-1}$, respectively, and $[\tilde{\mathbf{v}}_\ell]_i$ and $[\tilde{\boldsymbol{\nu}}_\ell]_i$ are the i-th element of $\tilde{\mathbf{v}}_\ell$ and $\tilde{\boldsymbol{\nu}}_\ell$, respectively.

4) Phase 4: Scale recovery: In order to recover the scale factor so that the relative UE pose and scatter positions can be fully recovered, computer vision methods require knowledge of the overall scale of the scene. As this is not possible in mmWave MIMO pose estimation, we instead rely on the estimated propagation delays. We introduce the vector of estimated delays, $\tilde{\boldsymbol{\tau}}$ with $[\tilde{\boldsymbol{\tau}}]_\ell = \tilde{\tau}_\ell$, $\ell \in \mathcal{S}_L$, and a corresponding vector of scaled path lengths $[\breve{\mathbf{d}}]_\ell = \breve{d}_\ell$, $\ell \in \mathcal{S}_L$, and $\breve{d}_\ell = \|\breve{\mathbf{p}}_\ell\| + \|\breve{\mathbf{p}}_\ell - \hat{\mathbf{n}}_{\mathbf{r}}\|$. Under the correct scaling s, we recall that $\tilde{\tau}_\ell = s\breve{d}_\ell/c + B$. To get rid of the clock bias, we introduce differential measurements $\mathbf{D}_s\tilde{\boldsymbol{\tau}}$ and $\mathbf{D}_s\breve{\mathbf{d}}$, where $\mathbf{D}_s$ is a $|\mathcal{S}_L| \times |\mathcal{S}_L|$ binary matrix. This matrix is constructed as $\mathbf{D}_s = \mathbf{I}_{|\mathcal{S}_L|} - \mathbf{D}_s'$, where $\mathbf{D}_s'$ has exactly one '1' on each row, off the main diagonal.[4] Then, $\mathbf{D}_s\tilde{\boldsymbol{\tau}} = s\mathbf{D}_s\breve{\mathbf{d}}/c$, so that the

[4]In the special case where UE and BS are synchronized, B is known to be 0, so that $\mathbf{D}_s = \mathbf{I}_{|\mathcal{S}_L|}$ can be used.

scaling is found as

$$\hat{s} = \frac{c}{|\mathcal{S}_L|} \sum_{\ell \in \mathcal{S}_L} \frac{[\mathbf{D}_s \tilde{\boldsymbol{\tau}}]_\ell}{[\mathbf{D}_s \breve{\mathbf{d}}]_\ell}. \tag{29}$$

Finally, the positions of the receiver and scatter point are recovered by $\hat{\mathbf{r}}_{\mathrm{UE}} = \hat{s}\hat{\mathbf{n}}_{\mathbf{r}}$, and $\hat{\mathbf{p}}_\ell = \hat{s}\breve{\mathbf{p}}_\ell$, respectively.

C. Problem 2 - Discussion, Variations, and Refinement

1) Identifiability without LoS knowledge: The epipolar model also gives the minimum number of distinctive scatter points required for SLAM using the estimates of AoD, AoA, and propagation delay: at least five distinctive pairs of virtual point correspondences are required to estimate the essential matrix [36, Sect. 13.3.2] by, for example, using the five-point algorithm (see Appendix A for a brief introduction and refer to [39], [40] for more details). Therefore, in order to achieve the SLAM problem of the 6D UE pose and 3D scatter positions estimation using the estimates of AoD, AoA, and propagation delay, the estimator requires four and five distinctive scatter points for the LoS and NLoS scenarios, respectively.

2) Identifiability with LoS knowledge: If we know that the LoS path is present, the methods can be further improved (refer to Appendix B) and the number of scatter points can be reduced to one, in line with the literature [24]. Note that in this case, the element within $\breve{\mathbf{d}}$ in (29) corresponding to the LoS path is a constant 1, since the scaled path lengths are normalized so that the scaled LoS distance is norm-1.

3) Iterative Refinement: In addition to the above closed-form solution consisting of sequential steps, we propose a direct estimation based on the weighted least-squares (LS) principle, which is given by

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{y}, \boldsymbol{\Theta}), \quad \text{s.t. } \mathbf{T}_{\mathrm{UE}} \in SE(3) \tag{30}$$

where the parameter set is $\boldsymbol{\Theta} = \{\mathbf{T}_{\mathrm{UE}}, \mathbf{P}\}$, $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_L]$, and the observation vector $\mathbf{y} = \mathrm{vec}(\mathbf{y}_{\ell \in \mathcal{S}_L})$ with $\mathbf{y}_\ell = \left[\tilde{\boldsymbol{\nu}}_\ell^{\mathrm{T}}, \tilde{\mathbf{v}}_\ell^{\mathrm{T}}, \tilde{\tau}_\ell\right]^{\mathrm{T}}$. In (30), the objective function $\mathcal{L}(\mathbf{y}, \boldsymbol{\Theta})$ is given by

$$\mathcal{L}(\mathbf{y}, \boldsymbol{\Theta}) = \tag{31}$$
$$\sum_{\ell \in \mathcal{S}_L} w_{1,\ell}\|\tilde{\boldsymbol{\nu}}_\ell - \boldsymbol{\nu}_\ell\|^2 + w_{2,\ell}\|\tilde{\mathbf{v}}_\ell - \mathbf{v}_\ell\|^2 + w_{3,\ell}\left[\mathbf{D}_s(\tilde{\boldsymbol{\tau}} - \boldsymbol{\tau})\right]_\ell^2,$$

where w_i for $i \in \{1, 2, 3\}$ is the weight factor and should properly reflect the precision of estimates $\tilde{\boldsymbol{\psi}}_{\mathrm{R},\ell}$, $\tilde{\boldsymbol{\psi}}_{\mathrm{T},\ell}$ and $\tilde{\tau}_\ell$. The constraint in (30) implies $\mathbf{R}_{\mathrm{UE}}^{\mathrm{T}}\mathbf{R}_{\mathrm{UE}} = \mathbf{I}, \det(\mathbf{R}_{\mathrm{UE}}) = +1$, and considering that $SE(3)$ is a manifold, the optimization of (30) can be solved by the Gauss-Newton method on the corresponding manifold [44] to convert the above optimization problem into an unconstrained optimization problem on the manifold, which is obtained with an iterative procedure. At each iteration, the update step is [37, Eq. 7.196]

$$\mathbf{T}_{\mathrm{UE}}^{t+1} = \exp\left(\left(\kappa_{\mathbf{T}}^t \boldsymbol{\Delta}_{\mathbf{T}}^t\right)^\wedge\right)\mathbf{T}_{\mathrm{UE}}^t, \tag{32}$$

$$\mathbf{p}_\ell^{t+1} = \mathbf{p}_\ell^t + \kappa_{\mathbf{p}_\ell}^t \boldsymbol{\Delta}_{\mathbf{p}_\ell}^t \quad \ell \in \{1, \dots, L\} \tag{33}$$

where $\kappa_{\mathbf{T}}^t > 0$ and $\kappa_{\mathbf{p}_\ell}^t > 0$ control the incremental step size for $\mathbf{T}_{\mathrm{UE}}$ and $\mathbf{p}_\ell$, respectively. In (32), the update direction is calculated by

$$\left[\boldsymbol{\Delta}_{\mathbf{T}}, \boldsymbol{\Delta}_{\mathbf{p}_1}, \dots, \boldsymbol{\Delta}_{\mathbf{p}_L}\right]^{\mathrm{T}} = \left(\nabla_{\boldsymbol{\Theta}}\boldsymbol{\mu}\right)^\dagger(\tilde{\mathbf{y}} - \boldsymbol{\mu}) \tag{34}$$

where $(\cdot)^\dagger$ is the weighted pseudoinverse defined by [45, Eq. 7.32]

$$(\mathbf{J})^\dagger = (\mathbf{J}^{\mathrm{T}}\mathbf{W}\mathbf{J})^{-1}\mathbf{J}^{\mathrm{T}}\mathbf{W}, \tag{35}$$

$\mathbf{W} = \mathrm{diag}(\mathrm{vec}(\mathbf{w}_{\ell \in \mathcal{S}_L}))$, $\mathbf{w}_\ell = [w_{1,\ell}, w_{1,\ell}, w_{2,\ell}, w_{2,\ell}, w_{3,\ell}]^{\mathrm{T}}$, $\tilde{\mathbf{y}} = \mathrm{vec}(\tilde{\mathbf{y}}_{\ell \in \mathcal{S}_L}) \in \mathbb{R}^{5|\mathcal{S}_L| \times 1}$ with $\tilde{\mathbf{y}}_\ell = \left[\tilde{\boldsymbol{\nu}}_\ell^{\mathrm{T}}, \tilde{\mathbf{v}}_\ell^{\mathrm{T}}, [\mathbf{D}_s\tilde{\boldsymbol{\tau}}]_\ell\right]^{\mathrm{T}}$, $\boldsymbol{\mu} = \mathrm{vec}(\boldsymbol{\mu}_{\ell \in \mathcal{S}_L}) \in \mathbb{R}^{5|\mathcal{S}_L| \times 1}$ with $\boldsymbol{\mu}_\ell = \left[\boldsymbol{\nu}_\ell^{\mathrm{T}}, \mathbf{v}_\ell^{\mathrm{T}}, [\mathbf{D}_s\boldsymbol{\tau}]_\ell\right]^{\mathrm{T}}$, $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\mu} = [\nabla_{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{\mu}_{\ell \in \mathcal{S}_L}]^{\mathrm{T}} \in \mathbb{R}^{5|\mathcal{S}_L| \times (6+3L)}$ with $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\mu}_\ell = \left[\nabla_{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{\nu}_\ell, \nabla_{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{v}_\ell, \left[\mathbf{D}_s\nabla_{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{\tau}\right]_\ell\right]^{\mathrm{T}}$ the gradient of $\boldsymbol{\mu}_\ell$ with respect to $\boldsymbol{\Theta}$. The involving derivatives are given in the Appendix C. Note that the differential measurements $\mathbf{D}_s\tilde{\boldsymbol{\tau}}$ are used as effective observations, while the differential delays $\mathbf{D}_s\boldsymbol{\tau}$ are updated in the iteration. These differential values cancel out the clock bias, such that the bias has no direct effect during the iteration. Further, the initialization of the Gauss-Newton method can be achieved with the closed-form solution given previously.

4) Computation Complexity: For the closed-form algorithm with a set of 5 paths, in phase 1, the calculation of virtual points has a complexity of $\mathcal{O}(2 \times 7)$. In addition, the computation of the essential matrix is dominated by the SVD for (40) and by the matrix inversion and SVD for calculating the solution to the 10 third-order polynomial equations. These two SVDs and the matrix inversion have a complexity of $\mathcal{O}(5 \times 9^2)$, $\mathcal{O}(10^3)$, and $\mathcal{O}(10^3)$, respectively. In phase 2, the SVD-based algorithm is dominated by the computation of an SVD and matrix multiplication, which have a complexity of $\mathcal{O}(3^3)$, and $\mathcal{O}(3^3)$, respectively. In phase 3, the computation of the homogeneous method for each path is dominated by the SVD of a 4×4 matrix, which leads to a complexity of $\mathcal{O}(5 \times 4^3)$. In phase 4, the calculation of the scale factor has a complexity of $\mathcal{O}(21)$. Hence, for 5 paths, the complexity is dominated by the estimation of the essential matrix in Phase 1.

To account for more than 5 paths, we rely on random sample consensus (RANSAC). In this strategy [39], a number of random samples containing five correspondences each are first taken. Then the five-point algorithm is applied to each sample and the estimated essential matrix is scored with the Sampson distance [28, Sect. 11.4.3]. The estimate with the best score is chosen as the final estimate.

Finally, for the iterative refinement, the complexity is dominated by the computation of the pseudoinverse, which has a complexity of $\mathcal{O}((6 + 3|\mathcal{S}_L|)(5|\mathcal{S}_L|)^2)$ per iteration.

V. Numerical Results

In this section, we numerically analyze the performance of the proposed estimators in the two problems from Section II.

A. Error metrics

The error vector measuring the residual error of the estimate is defined as $\boldsymbol{\epsilon}(\mathring{\boldsymbol{\Theta}}, \boldsymbol{\Theta}) = \left[\boldsymbol{\epsilon}_{\mathbf{r}}^{\mathrm{T}}, \boldsymbol{\epsilon}_{\mathbf{u}}^{\mathrm{T}}, \boldsymbol{\epsilon}_{\mathbf{p}}^{\mathrm{T}}\right]^{\mathrm{T}}$ with $\boldsymbol{\epsilon}_{\mathbf{r}} = \mathring{\mathbf{r}}_{\mathrm{UE}} - \mathbf{r}_{\mathrm{UE}}$, $\boldsymbol{\epsilon}_{\mathbf{u}} = \log(\mathbf{R}_{\mathrm{UE}}\mathring{\mathbf{R}}_{\mathrm{UE}}^{-1})^\vee$, and $\boldsymbol{\epsilon}_{\mathbf{p}} = \mathrm{vec}(\mathring{\mathbf{P}} - \mathbf{P})$, where $\boldsymbol{\epsilon}_{\mathbf{u}}$ measures of orientation error vector between $\mathbf{R}_{\mathrm{UE}}$ and $\mathring{\mathbf{R}}_{\mathrm{UE}}$ [46]. The performance of the estimator is measured in two

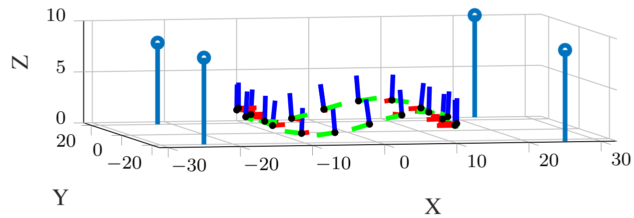


Fig. 6: Simulation setup for the receiver. The three orthonormal vectors in three different colors (the red, green, and blue vectors represent the x-axis, y-axis, and z-axis, respectively) at each sample on the path represent the frame of the receiver. The stems represent the BSs.

ways. First, by the root mean squared errors (RMSEs), which are defined as

$$\text{RMSE}_{\mathbf{r}} = \sqrt{\mathbb{E}\left\{\boldsymbol{\epsilon}_{\mathbf{r}}^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathbf{r}}\right\}} \tag{36}$$

$$\text{RMSE}_{\mathbf{u}} = \sqrt{\mathbb{E}\left\{\boldsymbol{\epsilon}_{\mathbf{u}}^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathbf{u}}\right\}} \tag{37}$$

$$\text{RMSE}_{\mathbf{p}} = \sqrt{\frac{1}{|\mathcal{S}_L|}\mathbb{E}\left\{\boldsymbol{\epsilon}_{\mathbf{p}}^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathbf{p}}\right\}}, \tag{38}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator. The proposed methods are compared to the corresponding CRB to demonstrate their efficiency. Second, once the efficiency is established, a more detailed analysis of the error statistics is conducted via the cumulative distribution function (CDF) of the errors.

As noted in Section I, neither Problem 1, nor Problem 2 have standard solutions (only simplified cases have been solved previously), so we are not able to report comparisons against other methods.

B. Problem 1 – AoA-only Pose Estimation: Establishing the Tightness of the Estimator

In this section, we evaluate the asymptotic tightness of the proposed algorithm given by (23) by comparing it with the theoretical bound (See [46] for their derivation) under the assumption that the observation matrix $\tilde{\mathbf{V}}$ is given.

1) Scenario: In evaluating the problem of using the AoA from multiple BSs for pose estimation, we consider a scenario with 4 single-antenna BSs at $\mathbf{r}_{\text{BS},1} = [-24, -20, 8.5]^{\mathrm{T}}$ m, $\mathbf{r}_{\text{BS},2} = [25, -25, 9]^{\mathrm{T}}$ m, $\mathbf{r}_{\text{BS},3} = [-22, 20, 8]^{\mathrm{T}}$ m, and $\mathbf{r}_{\text{BS},4} = [23, 25, 10]^{\mathrm{T}}$ m. The UE is equipped with a uniform rectangular array of 10×10 elements at half-wavelength spacing. To evaluate the performance of the proposed estimators, we consider a path with a circular pattern in the xy plane and a sinusoidal pattern in the z direction, as shown in Fig. 6. The radius of the circle in the xy plane is 15 m, and the amplitude of the sinusoid is 1 m. The circle is centered at $[0, 0, 1.5]^{\mathrm{T}}$ m. Starting at the coordinates $[0, 15, 1.5]^{\mathrm{T}}$ m, the path oscillates sinusoidally in the z direction and completes the path in three periods. For each sample point on the circle, the receiver is always in LoS condition. The observed virtual points are assumed to be $\tilde{\mathbf{V}} = \mathbf{V} + \mathbf{W}_{\mathbf{V}}$ with $\text{vec}\left(\mathbf{W}_{\mathbf{V}}\right) \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{V}}^2\mathbf{I}_4)$ of units 1 for the virtual points.

2) Results: In Fig. 7, we show the RMSE of the iterative algorithm solution for the position and orientation as a function of the $\sigma_{\mathbf{V}}^{-1}$. The results show that the proposed estimator is indeed asymptotically tight for large SNR. There, the RMSE for all types of errors reduces in inverse proportion to the SNR,

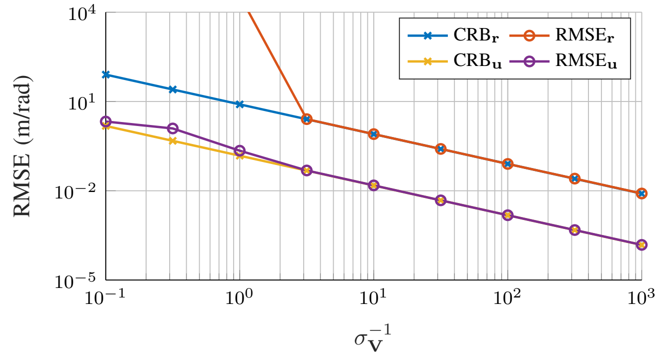


Fig. 7: RMSE as a function of SNR $\sigma_{\mathbf{V}}^{-1}$. It can be seen that the proposed estimator is asymptotically tight for large SNR, implying that the proposed algorithm is able to estimate the UE pose with high accuracy from the observed virtual points.

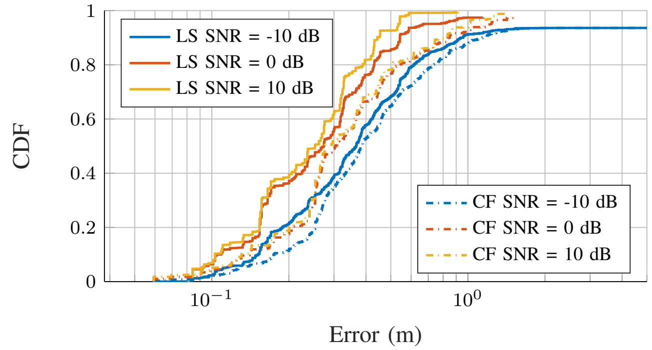
i.e., the proposed algorithm is able to estimate the UE pose with high accuracy from the observed virtual points. We note that Problem 1 was solved in a challenging narrowband setting using one or very few subcarriers. In an easier wideband setting (e.g., in 5G mmWave using many subcarriers or in ultra-wideband (UWB) systems) pose estimation performance will improve, due to the ability to also estimate delays.

C. Problem 1 – AoA-only Pose Estimation: In-depth Performance Evaluation

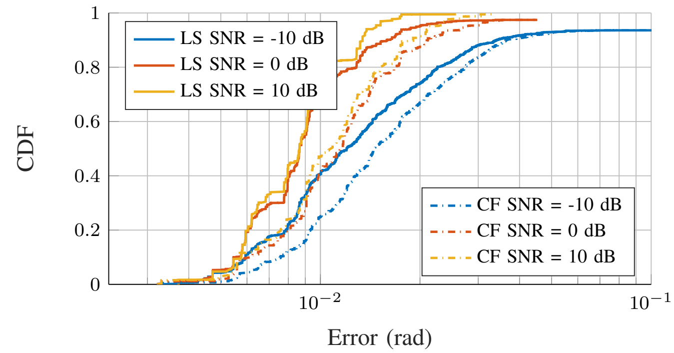
In this section, we evaluate the method proposed in Section IV-A, for determining the pose of a UE based on signal from several single-antenna BSs. In this part, the simulation includes estimating channel parameters from the reference signal, based on which the proposed iterative method initiated with the closed-form solution is then employed, which is expected to illustrate the validity of the proposed method in the overall system for Problem 1.

1) Scenario: The same scenario setting of the previous subsection is used in this part. However, the BSs are assumed to operate at the carrier frequency of 28 GHz, and i^{th} BS is assigned with a baseband transmit signal of $s_i(t) = e^{j2\pi f_{s,i}t}$ with $f_{s,i} = 100 + 50i$ Hz. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = |s_i|^2/\sigma_s^2$, where σ_s^2 is the variance of the additive Gaussian noise. At the receiver side, the MUSIC algorithm [32, Sect. 9.3.2] is used to estimate AoA, and then the association of the AOA estimate and corresponding BS is achieved by MVDR beamforming [32, Sect. 6.2.1] to identify the transmit signal frequency. Finally, the iterative method initialized with the solution of the P3P algorithm [47] is used to estimate the pose, using the virtual points converted from the AOA estimates.

2) Results: We evaluate the proposed estimator with respect to the SNR, and the pose estimate over the path is evaluated. Fig. 8 shows the CDF of the pose error. The performance of both closed-form estimation and LS estimation is included and denoted as CF and LS, respectively, for comparison. Here, we see an outperforming of the LS method against the CF method. The figure also reveals that our algorithm converges approximately for 95% of the sample positions when SNR = 0 dB, while this reduces to 90% when SNR = -10 dB. Hence, as expected, the coverage degrades if the receiver has a lower



(a) Position estimation error



(b) Orientation estimation error

Fig. 8: CDF of pose estimation errors for different SNRs for the AoA-only pose estimation problem. It can be seen that higher SNR leads to improved coverage of both methods, but the gains diminish with higher SNR. In addition, an outperforming of the LS method against the CF method can be seen.

SNR. Furthermore, it can be seen that the performance of the estimator improves as the SNR increases, however, there are diminishing performance gains with increasing SNR.

D. Problem 2 – mmWave MIMO Snapshot SLAM: Establishing the Tightness of the Estimator

In this section, we evaluate the asymptotic tightness of the proposed algorithm given by (30) by comparing it with the theoretical bound (See [46] for their derivation) under the assumption that the parameter vector estimate $\hat{\mathbf{z}}_\ell$ is given.

1) Scenario: We first consider a channel model of 3D geometry-based stochastic model [48], as shown in Fig. 9, where N_s scatter points are randomly generated on the surface of an ellipsoid. The principal axes of the ellipsoid are set to $[16, 12, 8]^{\mathrm{T}}$ m. The UE and BS are placed at the foci and point toward each other. The field of view of the receiver is set to $4\pi/9$ rad. The observed parameter vector is assumed to be $\mathbf{y} = \boldsymbol{\mu} + \mathbf{w}_{\boldsymbol{\mu}}$ with $\mathbf{w}_{\boldsymbol{\mu}} \sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \boldsymbol{\Sigma} \otimes \mathbf{I}_{N_s})$ and $\boldsymbol{\Sigma} = 10^{-3}\mathrm{diag}(\left[1, 1, 1, 1, 10^{-12}\right]^{\mathrm{T}})$ with units 1 for the virtual points and s^2 for the ToA, which is chosen to characterize the magnitude order relation between different components based on the MSE of estimates given by the Tensor-ESPRIT algorithm.

2) Results: First, in Fig. 10a, we show the RMSE of the iterative algorithm solution (denoted as weighted least squares (WLS)) for the position, orientation, and scatter positions as a function of the σ_μ^{-1} for $N_s = 10$ scatter points. For the

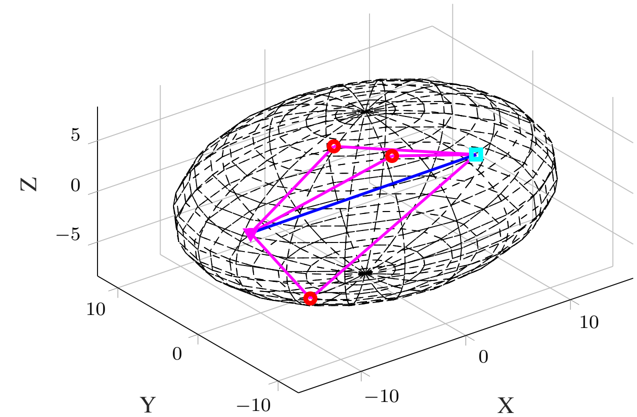
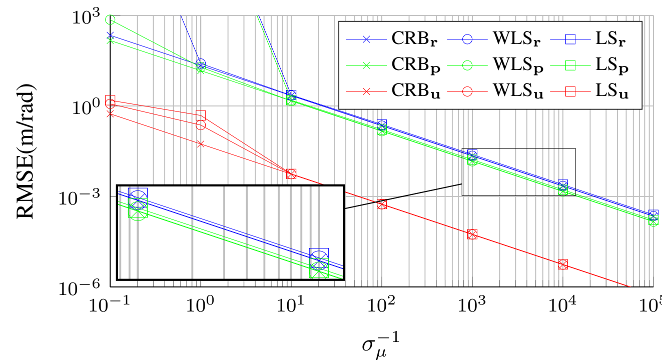
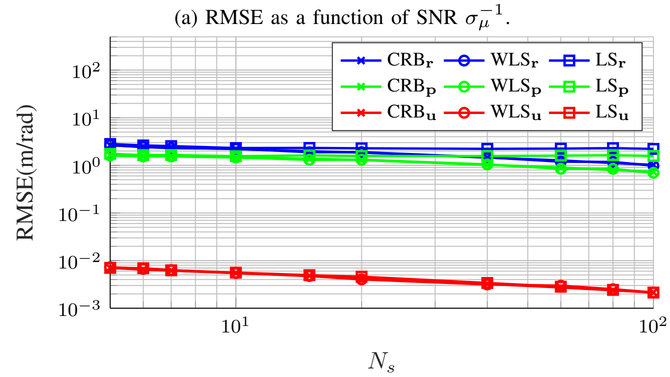


Fig. 9: Simulation setup of 3D ellipsoid channel model, used to establish the tightness of the method from Section IV-B for the mmWave MIMO snapshot SLAM problem. The UE and BS (shown with a triangle and square) are the foci of the ellipsoid.



(a) RMSE as a function of SNR σ_μ^{-1}.



(b) RMSE as a function of number of scatter points N_s.

Fig. 10: RMSE for position, orientation, and the position of scatters as a function of (a) σ_μ^{-1} and (b) N_s. The results show that the proposed estimator is asymptotically tight for large SNR and is consistently tight for all N_s, suggesting that the proposed algorithm can achieve a high level of accuracy in estimating the UE pose and the positions of scatterers based on the observed parameter vector.

weighted least squares method, the weight[5] of the ℓ^{th} path is set to $\mathbf{w}_\ell = |\alpha_\ell|^2[1, 1, 1, 1, 10^{12}]^{\mathrm{T}}$. We also added results for equal weight, i.e., the standard LS, (denoted as LS) for comparison. The results show that the proposed estimator is indeed asymptotically tight for large SNR. There, the RMSE

[5] In practice, the weights could be tabulated after an offline calibration stage (where measurements are compared to the ground truth in order to determine the error statistics) [49] or determined online from the measurements themselves, e.g., via the Fisher information, evaluated at the measurement [50].

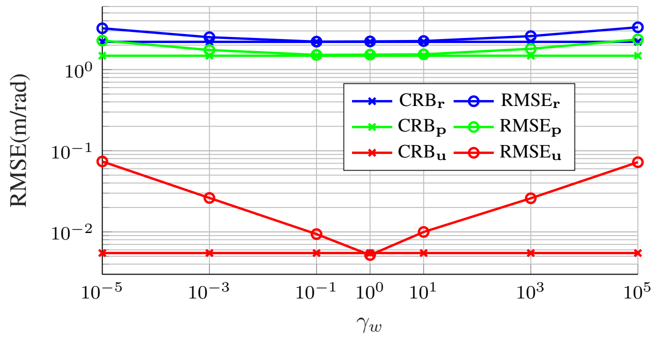


Fig. 11: RMSE as a function of the weights provided to the estimator. Note that $\gamma_w = 1$ corresponds to the correct generative weight.

for all types of errors reduces in inverse proportion to the SNR, i.e., the proposed algorithm is able to estimate the UE pose and the position of scatters with high accuracy from the observed parameter vector. Next, we show in Fig. 10b the RMSE for the position, orientation, and scatter positions as a function of the number N_s of scatters, for $\sigma_\mu^{-1} = 10$. Also here, we see a tightness between the RMSEs and their respective lower bounds, while it can be seen that the RMSEs decrease slowly with respect to the logarithm of N_s. However, the dependency is insignificant, so that in a smaller range of N_s, the RMSE appears almost constant. In addition, it can be seen that WLS achieves noticeable improvement at low SNR and a large number of scatter points. However, even in other cases, there is only a marginal improvement.

To understand the impact of mismatches in the weights, we performed additional simulations to investigate the performance when some weights deviate from the correct generative value. The generative model is set as used previously, and the weight of the ℓ^{th} path that is provided to the estimator is set to $\mathbf{w}_\ell = |\alpha_\ell|^2[\gamma_w, \gamma_w, \gamma_w^{-1}, \gamma_w^{-1}, 10^{12}]^\mathrm{T}$, where γ_w denotes the factor controlling the degree of mismatch. The result is given in Fig. 11, where the RMSE for the position, orientation, and scatter positions as a function of γ_w is given. As expected, the RMSE performance degrades with the degree of mismatch. However, it can be seen that the position estimate is more sensitive to the mismatch than the orientation estimate. Most importantly, the degradation due to a wrong selection of the weight is rather limited, indicating that the method is relatively robust to weight mismatch.

E. Problem 2 – mmWave MIMO Snapshot SLAM: In-depth Performance Evaluation

In this section, we evaluate the estimator in the simulation setting given by [21]. In this part, the simulation comprises an end-to-end approach that includes estimating channel parameters from the pilot signal, based on which the proposed iterative method initiated with the closed-form solution is then employed, which is expected to illustrate the validity of the proposed method in the overall system for Problem 2.

1) Scenario: We consider a scene consisting of two surfaces, representing the building and ground surfaces respectively, as shown in Fig. 12. The building facade's center is at $[10, 10, 5]^\mathrm{T}$ m with a facade length of 20 m, facade height of 10 m, and direction $[0, 1, 0]^\mathrm{T}$. The ground surface

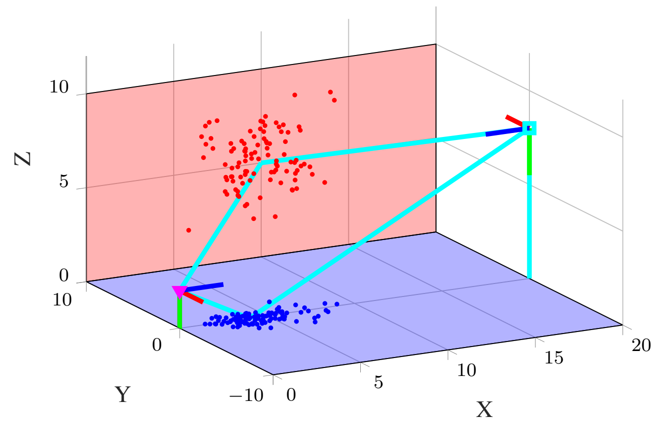


Fig. 12: Simulation setup for channel estimation and SLAM with 2 clusters. The BS is on the right and has a known pose, while the UE is on the left and has an unknown pose. The two walls have unknown locations and each give rise to several scatter points.

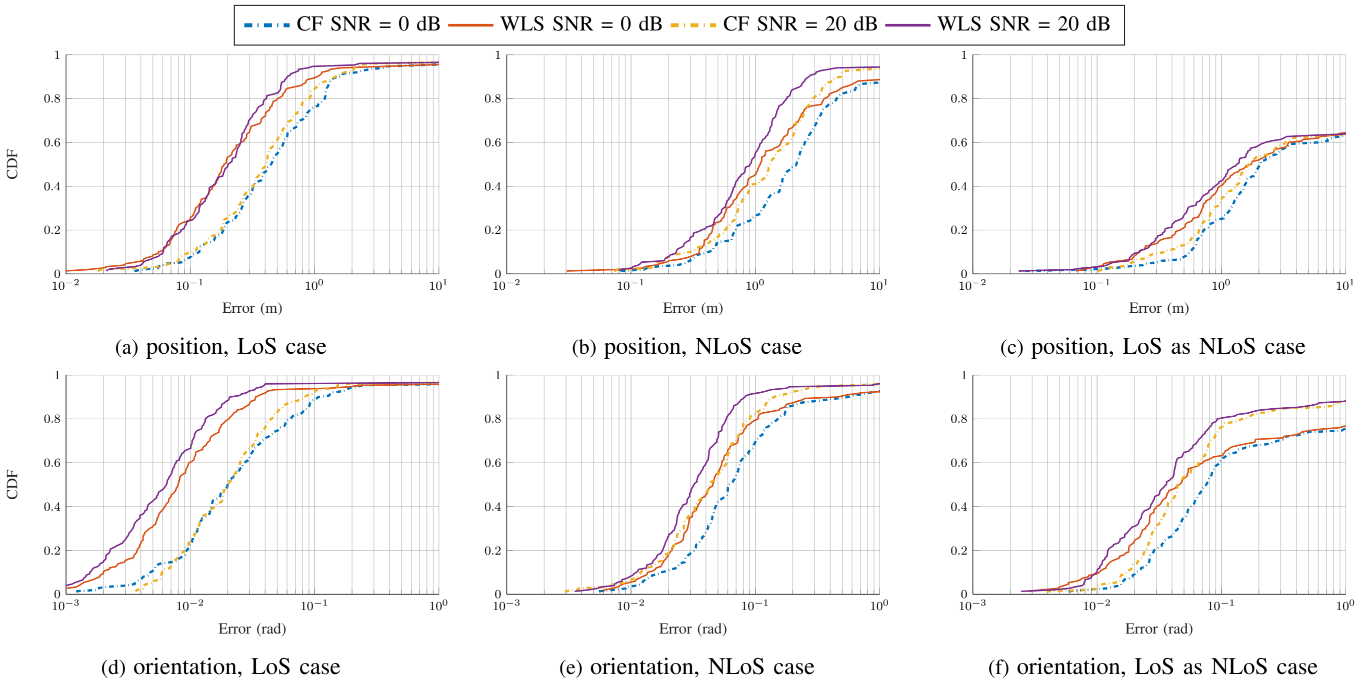
is at $[10, 0, 0]^\mathrm{T}$ m with direction $[0, 0, 1]^\mathrm{T}$, surface dimension 20×20 m^2. The positions of BS and UE are set to $\mathbf{r}_{\mathrm{BS}} = [20, 0, 8]^\mathrm{T}$ m and $\mathbf{r}_{\mathrm{UE}} = [0, 0, 2]^\mathrm{T}$ m, respectively, while their orientations are

$$\mathbf{R}_{\mathrm{BS}} = \begin{bmatrix} 0, & 0, & -1 \\ 1, & 0, & 0 \\ 0, & -1, & 0 \end{bmatrix}, \quad \mathbf{R}_{\mathrm{UE}} = \begin{bmatrix} 0, & 0, & 1 \\ -1, & 0, & 0 \\ 0, & -1, & 0 \end{bmatrix}, \tag{39}$$

pointing at each other. The BS and UE are equipped with a uniform rectangular array with 10×10 half-wavelength spaced elements and a carrier frequency of 28 GHz. For the k^{th} facade, $L_k = 100$ scatters are generated ($L = 200$ in total), whose positions $\mathbf{p}_\ell$ are randomly generated according to the model from [17], [21], [22], [51], described in Appendix D. This model is parameterized by $\beta \in \mathbb{N}$, which describes the directivity of the scattering (i.e., larger β means more directive scattering). The clock bias is assumed to comply with a uniform distribution, i.e., $B \sim \mathcal{U}\left(-\frac{T_{\mathrm{sym}}}{2}, \frac{T_{\mathrm{sym}}}{2}\right)$ with symbol duration T_{sym}. The channel parameter estimation method is detailed in Appendix E, providing $\mathcal{Z} = \{\mathbf{z}_0, \ldots, \mathbf{z}_{\hat{L}}\}$ (where we recall that $\mathbf{z}_\ell = [\boldsymbol{\psi}_{\mathrm{R},\ell}, \boldsymbol{\psi}_{\mathrm{T},\ell}, \tau_\ell, \alpha_\ell]^\mathrm{T}$). The weight factors for the proposed WLS algorithm are specified with the squared magnitude of the estimated channel gain $|\hat{\alpha}_\ell|^2$. While our focus is on a coarse estimator in this case, further performance improvements can be obtained by considering the actual error statistics of the channel parameter estimator from Appendix E. Due the finite resolution in angle and delay domain, the estimated number of paths $\hat{L}$ is usually much less than L.[6] For the methods to work, $\hat{L}$ needs to be sufficiently large (i.e., $\hat{L} \geq 4$ in LoS and $\hat{L} \geq 5$ in NLoS). If fewer paths are detected, the proposed methods cannot be applied. If more paths are detected, either all paths can be used, or a subset can be selected with maximal diversity among the angles and delays.

2) Results: First, the averaged estimated number of paths for each setting is given in Tab. I for reference. In Fig. 13, we show the CDF of the estimation errors, for different SNRs for $\beta = 10$. The performance of both closed-form estimation

[6]This observation implies that the model order L is random and unknown, since each noise realization will give rise to a different value of $\hat{L}$ with corresponding different path parameters. Hence, neither the maximum likelihood estimator nor the CRB are well-defined in this case.



(a) position, LoS case

(b) position, NLoS case

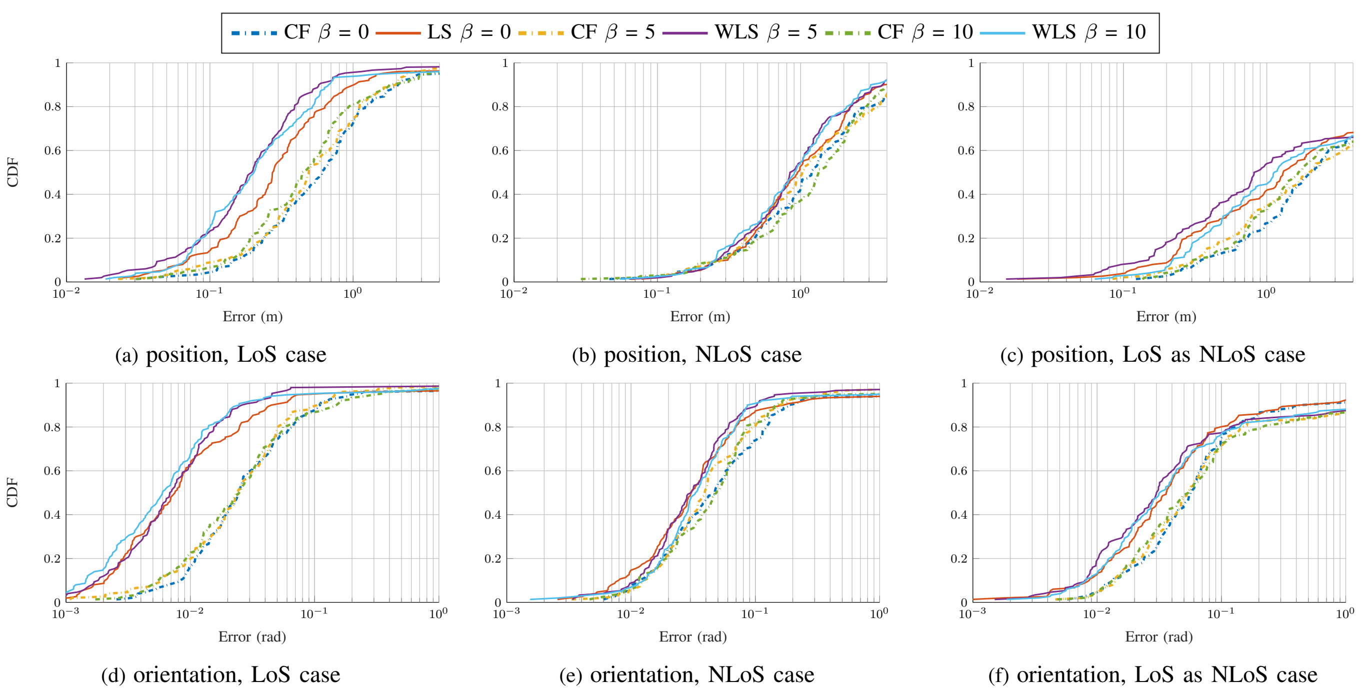
(c) position, LoS as NLoS case

(d) orientation, LoS case

(e) orientation, NLoS case

(f) orientation, LoS as NLoS case

Fig. 13: CDF of pose errors for different SNRs for the closed form (CF) and weighted least squares (WLS) solution. The position and orientation errors are given in top and bottom rows, respectively. The first column contains the performance in the LoS case and the knowledge of the presence of LoS path is utilized, the second column contains the performance in the NLoS case, and the last column contains the performance in the LoS case but the LoS path is treated as the NLoS paths.



(a) position, LoS case

(b) position, NLoS case

(c) position, LoS as NLoS case

(d) orientation, LoS case

(e) orientation, NLoS case

(f) orientation, LoS as NLoS case

Fig. 14: CDF of pose errors for different values of the directivity β for the closed form (CF) and weighted least squares (WLS) solution. The position and orientation errors are given in the top and bottom rows, respectively. The first column contains the performance in the LoS case and the knowledge of the presence of LoS path is utilized, the second column contains the performance in the NLoS case, and the last column contains the performance in the LoS case but the LoS path is treated as the NLoS paths.

TABLE I: The averaged estimated number of paths for each setting.

Cases	SNR \ β	0	5	10
LoS	0 (dB)	8.03	7.73	7.33
	20 (dB)	9	9	9
NLoS	0 (dB)	8.58	8.44	8.12
	20 (dB)	9	9	9

and WLS estimation is included and denoted as CF and WLS, respectively, for comparison. Under the condition of asynchronization, we evaluated the three study cases. More specifically, the first column in Fig. 13 contains the performance in the LoS case and the knowledge of the presence of LoS path is utilized, the second column contains the performance in the NLoS case, and the last column contains the performance in the LoS case but the LoS path is treated as the NLoS paths. It can be seen that the WLS algorithm achieves an accuracy of 1 m and 0.04 rad or better in approximately 70% of the cases when the LoS channel is present and utilized properly, while this accuracy becomes 3 m and 0.1 rad in approximately 70% of the cases when the LoS is absent. This indicates that even the derivation of the closed-form solutions ignores some impairments, i.e., the effect of the estimation errors in the angles and channel delay, the CF method demonstrates a remarkable level of robustness, standing out as a reliable solution that maintains its performance integrity even when subjected to these impairments. In addition, we observe that the WLS algorithm outperforms the CF method, in all cases. The figure further reveals that the WLS algorithm achieves an accuracy of 1 m and 0.03 rad or better in approximately 90% of the cases when the LoS channel is present and utilized properly, while this accuracy becomes 3 m and 0.2 rad in approximately 80% of the cases when the LoS is absent. However, if the LoS path is utilized improperly, the accuracy degrades significantly. This is because the LoS path does not have an associated scattering point and the algorithm dealing with the NLoS will always associate one with the estimated path, such that a fake scattering point is expected to be produced for the LoS path, biasing the final estimate. It can also be seen that the performance does not improve significantly with increasing SNR, which means that SNR is not the dominant factor in determining performance. This is due to the fact that both the WLS and CF algorithms utilize the estimated parameter set $\tilde{\mathcal{Z}}$, which is provided in the form of effective scatter points and has limited resolution, so the performance is mainly limited by the accuracy of $\tilde{\mathcal{Z}}$. Next, we show in Fig. 14 the CDF of the estimation errors for different values of the directivity β, for SNR $= 40$ dB. Also here, we see an outperforming of the WLS method against the CF method. The figure further reveals that the WLS algorithm achieves an accuracy of 1 m and 0.03 rad or better in approximately 90% of the cases when the LoS channel is present and utilized properly, while this accuracy becomes 3 m and 0.1 rad in approximately 90% of the cases when the LoS is absent. However, if the LoS path is utilized improperly, the accuracy degrades significantly. The results shown in Fig. 14 reflect that β has a smaller impact on performance when the LoS channel is present and utilized properly.

VI. Conclusion

In this paper, we investigate the estimation of the full 6D user pose (joint 3D position and 3D orientation) using antenna arrays by providing a projective geometric view of AoDs and AoAs in the context of 5G and beyond 5G positioning. To this end, the directional angular information is first modeled in terms of the receiver's pose using the perspective projection model from computer vision. Then, two pose estimation problems, namely 6D pose estimation using AoAs from multiple base stations and 6D SLAM based on single-BS mmWave communication, are investigated with the perspective projection model. Particularly, we show that the 6D SLAM problem, when modeled with the perspective projection model, can be further modeled with the epipolar model. For each problem, we propose two estimation algorithms, a closed-form one and an iterative one based on the principle of LS. The simulation results confirm the effectiveness of the proposed algorithms and demonstrate that reliable 6D localization of a user is achievable even in the absence of the LoS path, while the performance is significantly improved when the LoS path is present and utilized properly. However, improper utilization of the LoS path degrades the performance significantly, prompting the necessity of an effective LoS detection stage.

Appendix A
Brief introduction to the five-point algorithm for estimating the essential matrix

Base on (16), for a set of five correspondences $\{(\bar{\mathbf{v}}_i, \bar{\boldsymbol{\nu}}_i)\}$, we have

$$\underbrace{\begin{bmatrix} \bar{\mathbf{v}}_1^{\mathrm{T}} \otimes \bar{\boldsymbol{\nu}}_1^{\mathrm{T}} \\ \vdots \\ \bar{\mathbf{v}}_5^{\mathrm{T}} \otimes \bar{\boldsymbol{\nu}}_5^{\mathrm{T}} \end{bmatrix}}_{\mathbf{A} \in \mathbf{R}^{5\times 9}} \mathrm{vec}(\mathbf{E}) = \mathbf{0} \tag{40}$$

where $\mathrm{vec}(\mathbf{E}) = \left[\mathfrak{e}_1^{\mathrm{T}}, \mathfrak{e}_2^{\mathrm{T}}, \mathfrak{e}_3^{\mathrm{T}}\right]^{\mathrm{T}}$. (40) implies that $\mathrm{vec}(\mathbf{E})$ is the right null space of $\mathbf{A}$, whose base vectors $\mathrm{vec}(\mathbf{E}_i)$, $i \in \{1, \ldots 4\}$ can be obtained with SVD. Since $\mathbf{E}$ is defined up to a scale, we have $\mathbf{E} = x\mathbf{E}_1 + y\mathbf{E}_2 + z\mathbf{E}_3 + \mathbf{E}_4$, for certain coefficients $\mathcal{C} = (x, y, z)$. Further, the following constraints,

$$\det(\mathbf{E}) = 0 \tag{41}$$

and

$$\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{E} - \frac{1}{2}\mathrm{tr}(\mathbf{E}\mathbf{E}^{\mathrm{T}})\mathbf{E} = \mathbf{0}_3, \tag{42}$$

are used to build 10 third-order polynomial equations in $\mathcal{C}$. In general, $K \leq 10$ sets of real solutions to $\mathcal{C}$ are obtained from the polynomial equations. Since the five-point algorithm is used together with the RANSAC strategy, in order to keep only the optimal solution, these K solutions are assessed with the remaining correspondences other than the chosen five correspondences.

Appendix B
Estimation with the knowledge of the LoS path

Prior knowledge of the existence of the LoS path can improve the performance of SLAM, since the AoD and AoA associated with the LoS channel are specifically represented by epipoles in epipolar geometry. For the sake of simplification, we ignore the subscript of $\mathbf{R}_{\mathrm{UE}}$ in this section. As pointed out in [52, Sect. 2.2], the epipoles $\boldsymbol{\nu}_0$ and $\mathbf{v}_0$ have the following two properties

$$\bar{\boldsymbol{\nu}}_0 = \mathbf{r}_{\mathrm{UE}} \tag{43}$$

$$\bar{\mathbf{v}}_0 = -\mathbf{R}^{\mathrm{T}}\mathbf{r}_{\mathrm{UE}}, \tag{44}$$

where these two equalities are defined up to a scale. These two properties can be exploited to recover the pose partially. On the one hand, the direction of $\mathbf{r}_{\mathrm{UE}}$ can be recovered according to (43). On the other hand, (44) implies that the rotated vector $\mathbf{R}\bar{\mathbf{v}}_0$ specifies the direction of $-\bar{\boldsymbol{\nu}}_0$, with which we can recover $\mathbf{R}$ partially. To illustrate this, substituting (43) into (44) and left multiplying $\mathbf{R}$, we have, up to a scale,

$$-\bar{\boldsymbol{\nu}}_0 = \mathbf{R}\bar{\mathbf{v}}_0. \tag{45}$$

Then following the swing-twist parametrization [53, Sect. 5], the rotation $\mathbf{R}$ is decomposed into $\mathbf{R} = \mathbf{R}_{\perp}\mathbf{R}_{\parallel}$, where $\mathbf{R}_{\parallel} = \exp\left(\theta(\mathbf{n}_{\parallel})_{\times}\right)$ is a rotation matrix over an arbitrary unit vector $\mathbf{n}_{\parallel}$ and an arbitrary angle θ, and $\mathbf{R}_{\perp} = \exp\left((\mathbf{u}_{\perp})_{\times}\right)$ is the rotation matrix of a rotation vector $\mathbf{u}_{\perp}$ residing in the plane perpendicular to $\mathbf{n}_{\parallel}$, i.e., $\mathbf{u}_{\perp}^{\mathrm{T}}\mathbf{n}_{\parallel} = 0$. Note that for arbitrary θ we have $\mathbf{n}_{\parallel} = \mathbf{R}_{\parallel}\mathbf{n}_{\parallel}$. Then if we choose $\mathbf{n}_{\parallel} = \frac{\bar{\mathbf{v}}_0}{\|\bar{\mathbf{v}}_0\|}$, for arbitrary θ, (45) can be rewritten as

$$-\bar{\boldsymbol{\nu}}_0 = \mathbf{R}_{\perp}\mathbf{R}_{\parallel}\bar{\mathbf{v}}_0 \tag{46}$$

$$= \mathbf{R}_{\perp}\bar{\mathbf{v}}_0. \tag{47}$$

Since $\mathbf{R}_{\perp}$ brings $\bar{\mathbf{v}}_0$ to the direction of $-\bar{\boldsymbol{\nu}}_0$, the axis of rotation $\mathbf{R}_{\perp}$ is therefore $\frac{\bar{\boldsymbol{\nu}}_0\times\bar{\mathbf{v}}_0}{\|\bar{\boldsymbol{\nu}}_0\times\bar{\mathbf{v}}_0\|}$ and the rotation angle is the angle between $\bar{\mathbf{v}}_0$ and $-\bar{\boldsymbol{\nu}}_0$, then we have [54, Eq. 3]

$$\mathbf{u}_{\perp} = \frac{\bar{\boldsymbol{\nu}}_0 \times \bar{\mathbf{v}}_0}{\|\bar{\boldsymbol{\nu}}_0 \times \bar{\mathbf{v}}_0\|} \arccos \frac{-\bar{\mathbf{v}}_0^{\mathrm{T}}\bar{\boldsymbol{\nu}}_0}{\|\bar{\mathbf{v}}_0\|\|\bar{\boldsymbol{\nu}}_0\|}, \tag{48}$$

which has two degrees of freedom [53, Sect. 5]. It can be seen that based merely on the AoA and AoD of the LoS path, two degrees of freedom specified by $\mathbf{u}_{\perp}$ can be recovered, but not the remaining one, θ (i.e., $\mathbf{R} = \exp\left((\mathbf{u}_{\perp})_{\times}\right)\exp\left(\theta(\mathbf{n}_{\parallel})_{\times}\right)$ where only θ remains unknown). To this end, we use the information provided by the scattering paths. Based on the facts that $\bar{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{E}\bar{\mathbf{v}} = 0$ from (16), $\mathbf{E} = (\mathbf{r}_{\mathrm{UE}})_{\times}\mathbf{R}$ from (17) and the relation (43), an additional pair of correspondences $(\boldsymbol{\nu}_l, \mathbf{v}_l)$ associated with the l^{th} scatter point leads to

$$\begin{aligned}
&\bar{\boldsymbol{\nu}}_l^{\mathrm{T}}(\bar{\boldsymbol{\nu}}_0)_{\times}\mathbf{R}\bar{\mathbf{v}}_l = \bar{\boldsymbol{\nu}}_l^{\mathrm{T}}(\bar{\boldsymbol{\nu}}_0)_{\times}\mathbf{R}_{\perp}\mathbf{R}_{\parallel}\bar{\mathbf{v}}_l = \mathfrak{v}_l^{\mathrm{T}}\exp(\theta(\mathbf{n}_{\parallel})_{\times})\bar{\mathbf{v}}_l \\
&= \mathfrak{v}_l^{\mathrm{T}}\bar{\mathbf{v}}_l + \mathfrak{v}_l^{\mathrm{T}}(\mathbf{n}_{\parallel})_{\times}\bar{\mathbf{v}}_l\sin\theta + \mathfrak{v}_l^{\mathrm{T}}(\mathbf{n}_{\parallel})_{\times}^2(1-\cos\theta)\bar{\mathbf{v}}_l \qquad (49) \\
&= \mathfrak{v}_l^{\mathrm{T}}(\mathbf{n}_{\parallel})_{\times}\bar{\mathbf{v}}_l\sin\theta + \mathfrak{v}_l^{\mathrm{T}}\bar{\mathbf{v}}_l\cos\theta \qquad (50) \\
&= \mathfrak{n}_l^{\mathrm{T}}\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\\ 1\end{bmatrix} = 0. \qquad (51)
\end{aligned}$$

where $\mathfrak{v}_l^{\mathrm{T}} = \bar{\boldsymbol{\nu}}_l^{\mathrm{T}}(\bar{\boldsymbol{\nu}}_0)_{\times}\mathbf{R}_{\perp}$, and

$$\mathfrak{n}_l = \begin{bmatrix}\mathfrak{v}_l^{\mathrm{T}}\bar{\mathbf{v}}_l\\ \mathfrak{v}_l^{\mathrm{T}}(\mathbf{n}_{\parallel}\times\mathbf{v}_l)\\ 0\end{bmatrix}. \tag{52}$$

Eq. (49) holds due to the Rodrigues' rotation formula, given by,

$$\exp(\theta\mathbf{n}_{\times}) = \mathbf{I} + \mathbf{n}_{\times}\sin\theta + \mathbf{n}_{\times}^2(1-\cos\theta), \tag{53}$$

while (50) holds due to the cross product properties of $\mathbf{a}\times\mathbf{a} = \mathbf{0}$ and $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}^{\mathrm{T}}\mathbf{c})\mathbf{b} - (\mathbf{a}^{\mathrm{T}}\mathbf{b})\mathbf{c}$ and to the equality $\mathfrak{v}_l^{\mathrm{T}}\mathbf{n}_{\parallel} = 0$ resulted from (47). It can be seen that θ corresponds to the intersections of the unit circle centered at the origin with the line specified by the normal vector $\mathfrak{n}_l$ and the origin. There are two intersections, thus two values θ_i, $i \in \{1, 2\}$, which lead to two estimates of $\mathbf{R}$. By first noticing that the rotated vector $\mathbf{R}\bar{\mathbf{v}}_l$ resides in the epipolar plane and that

$$(\mathbf{R}|_{\theta_1}\bar{\mathbf{v}}_l)\times(\mathbf{R}|_{\theta_1}\mathbf{n}_{\parallel}) = (\mathbf{R}|_{\theta_2}\mathbf{n}_{\parallel})\times(\mathbf{R}|_{\theta_2}\bar{\mathbf{v}}_l), \tag{54}$$

$$\mathbf{R}|_{\theta_1}\mathbf{n}_{\parallel} = \mathbf{R}|_{\theta_2}\mathbf{n}_{\parallel}, \tag{55}$$

where $\mathbf{R}|_{\theta} \doteq \mathbf{R}_{\perp}\exp(\theta(\mathbf{n}_{\parallel})_{\times})$, then the two estimates can be illustrated by the vectors $\mathbf{R}|_{\theta_1}\bar{\mathbf{v}}_l$ and $\mathbf{R}|_{\theta_2}\bar{\mathbf{v}}_l$ forming reflection with respect to $\mathbf{R}\mathbf{n}_{\parallel}$ within the epipolar plane. As a result, we can choose the appropriate value for θ so that $\bar{\boldsymbol{\nu}}_0\times\bar{\boldsymbol{\nu}}_l$ and $\bar{\boldsymbol{\nu}}_0\times(\mathbf{R}\bar{\mathbf{v}}_l)$ have the identical sign. Thus, one additional scattering point suffices to uniquely recover the rotation matrix.

The sequential application of (43) and (49) can be used to estimate the pose up to a scale factor. However, in this approach, the direction estimation of $\mathbf{r}_{\mathrm{UE}}$ is based exclusively on the observation of the LoS channel, while the observation of scattering paths does not contribute to this estimation. To deal with this problem and to make the estimation compatible with existing algorithms, we can convert the constraints (43) and (44) into a set of extra correspondences $\mathcal{S} = \{(\mathbf{e}_1,\bar{\mathbf{v}}_0),(\mathbf{e}_2,\bar{\mathbf{v}}_0),(\mathbf{e}_3,\bar{\mathbf{v}}_0),(\bar{\boldsymbol{\nu}}_0,\mathbf{e}_1),(\bar{\boldsymbol{\nu}}_0,\mathbf{e}_2),(\bar{\boldsymbol{\nu}}_0,\mathbf{e}_3)\}$, which can then be fed into existing algorithms along with the virtual points of scattering paths.

Appendix C
Gradients in the Gauss-Newton Method

The gradient of τ_ℓ with respect to $\boldsymbol{\Theta}$ is given by

$$\nabla_{\boldsymbol{\Theta}}\tau_\ell = \frac{1}{c}\Big(\frac{\mathbf{p}_\ell^{\mathrm{T}}}{\|\mathbf{p}_\ell\|}\nabla_{\boldsymbol{\Theta}}\mathbf{p}_\ell + \frac{\mathbf{p}_{\ell,u}^{\mathrm{T}}}{\|\mathbf{p}_{\ell,u}\|}\nabla_{\boldsymbol{\Theta}}\mathbf{p}_{\ell,u}\Big), \tag{56}$$

where $\nabla_{\boldsymbol{\Theta}}\mathbf{p}_\ell = \left[\nabla_{\mathbf{T}_{\mathrm{UE}}}^{\mathrm{T}}\mathbf{p}_\ell, \nabla_{\mathbf{p}_1}^{\mathrm{T}}\mathbf{p}_\ell, \ldots, \nabla_{\mathbf{p}_L}^{\mathrm{T}}\mathbf{p}_\ell\right]$, $\nabla_{\mathbf{T}_{\mathrm{UE}}}\mathbf{p}_\ell = \mathbf{0}_{3\times 6}$,

$$\nabla_{\mathbf{p}_i}\mathbf{p}_\ell = \begin{cases}\mathbf{I}_3 & \text{if } \ell = i\\ \mathbf{0}_{3\times3} & \text{if } \ell \neq i\end{cases}, \tag{57}$$

$\bar{\mathbf{p}}_{\ell,u} = \mathbf{T}_{\mathrm{UE}}\bar{\mathbf{p}}_\ell$, and

$$\begin{aligned}
\nabla_{\mathbf{T}_{\mathrm{UE}}}\mathbf{p}_{\ell,u} &= \nabla_{\mathbf{T}_{\mathrm{UE}}}\left([\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]^{\mathrm{T}}\,\mathbf{T}_{\mathrm{UE}}\bar{\mathbf{p}}_\ell\right)\\
&= [\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]^{\mathrm{T}}\,\bar{\mathbf{p}}_\ell^{\odot}
\end{aligned} \tag{58}$$

$$\begin{aligned}
\nabla_{\mathbf{p}_i}\mathbf{p}_{\ell,u} &= \nabla_{\mathbf{p}_i}\left([\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]^{\mathrm{T}}\,\mathbf{T}_{\mathrm{UE}}\bar{\mathbf{p}}_\ell\right)\\
&= \begin{cases}\mathbf{R}_{\mathrm{UE}}^{\mathrm{T}} & \text{if } \ell = i\\ \mathbf{0}_{3\times3} & \text{if } \ell \neq i\end{cases}.
\end{aligned} \tag{59}$$

where the gradient with respect to $\mathbf{T}_{\mathrm{UE}}$ is calculated with the infinitesimal perturbation [37, Sect. 7.1], and the operator $(\cdot)^{\odot}$ is

$$\left(\left[\boldsymbol{\xi}^{\mathrm{T}},\eta\right]^{\mathrm{T}}\right)^{\odot} = \begin{bmatrix}\eta\mathbf{I}_3 & -\boldsymbol{\xi}_{\times}\\ \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}}\end{bmatrix}, \tag{60}$$

The gradients of $\boldsymbol{\nu}_\ell$ and $\mathbf{v}_\ell$ with respect to $\boldsymbol{\Theta}$ are given by

$$\nabla_{\boldsymbol{\Theta}}\boldsymbol{\nu}_\ell = \begin{bmatrix}\mathbf{e}_1^{\mathrm{T}}\\ \mathbf{e}_2^{\mathrm{T}}\end{bmatrix}\left(\frac{\mathbf{I}_3}{\mathbf{e}_3^{\mathrm{T}}\mathbf{p}_\ell} - \frac{\mathbf{p}_\ell\mathbf{e}_3^{\mathrm{T}}}{(\mathbf{e}_3^{\mathrm{T}}\mathbf{p}_\ell)^2}\right)\nabla_{\boldsymbol{\Theta}}\mathbf{p}_\ell, \tag{61}$$

and

$$\nabla_{\boldsymbol{\Theta}}\mathbf{v}_\ell = \begin{bmatrix}\mathbf{e}_1^{\mathrm{T}}\\ \mathbf{e}_2^{\mathrm{T}}\end{bmatrix}\left(\frac{\mathbf{I}_3}{\mathbf{e}_3^{\mathrm{T}}\mathbf{p}_{\ell,u}} - \frac{\mathbf{p}_{\ell,u}\mathbf{e}_3^{\mathrm{T}}}{(\mathbf{e}_3^{\mathrm{T}}\mathbf{p}_{\ell,u})^2}\right)\nabla_{\boldsymbol{\Theta}}\mathbf{p}_{\ell,u}, \tag{62}$$

respectively.

Appendix D
Generative Model for the Scatter Locations

The scatter point locations are synthesized by applying Markov Chain Monte Carlo (MCMC) sampling to the PDF of $\mathbf{p}_\ell$ given by [17], [21], [22], [51]

$$p_k(\mathbf{p}_\ell|\mathbf{r}_{\mathrm{UE}},\mathbf{r}_{\mathrm{BS}}) \propto \begin{cases} R_k(\mathbf{p}_\ell,\mathbf{r}_{\mathrm{UE}},\mathbf{r}_{\mathrm{BS}},\beta) & \text{if } \mathbf{p}_\ell \in \mathcal{S}_{F,k} \\ 0 & \text{otherwise,} \end{cases} \tag{63}$$

where $\mathcal{S}_{F,k}$ for $i \in \{1,2\}$ denotes the space of the k^{th} facade, and the pattern function $R_k(\mathbf{p}_\ell,\mathbf{r}_{\mathrm{UE}},\mathbf{r}_{\mathrm{BS}},\beta)$ is [55]

$$R_k(\mathbf{p}_\ell,\mathbf{r}_{\mathrm{UE}},\mathbf{r}_{\mathrm{BS}},\beta) \propto \begin{cases} \frac{\cos\theta_i\cos\theta_s}{d_i^2 d_s^2} & \text{if } \beta=0 \\ \frac{\cos\theta_i(1+\cos\psi_b)^\beta}{F_\beta d_i^2 d_s^2} & \text{otherwise.} \end{cases} \tag{64}$$

In (64), $d_i = \|\mathbf{r}_{\mathrm{BS}} - \mathbf{p}_\ell\|$, $d_s = \|\mathbf{r}_{\mathrm{UE}} - \mathbf{p}_\ell\|$, the angles θ_i and θ_s are, respectively, the incidence and scattering directions with respect to $\mathbf{p}_\ell$, ψ_b denotes the angle between the reflection and scattering directions at $\mathbf{p}_\ell$, $\beta \in \mathbb{N}$ describes the directivity of the scattering at $\mathbf{p}_\ell$, and the normalization factor F_β is given by $F_\beta = \frac{1}{2^\beta}\sum_{j=0}^{\beta}\binom{\beta}{j}\cdot I_j$ and

$$I_j = \frac{2\pi}{j+1}\left[\cos\theta_i \sum_{w=0}^{\frac{j-1}{2}}\binom{2w}{w}\cdot\frac{\sin^{2w}\theta_i}{2^{2w}}\right]^{\left(\frac{1-(-1)^j}{2}\right)}. \tag{65}$$

In this paper, the channel gain α_ℓ of the ℓ^{th} path is modeled with the exponential decay model [56]. More specifically, the k^{th} cluster is assigned with a power of $P_k = e^{-\frac{\tau_k}{D_c}}10^{\frac{Z_k}{10}}$, where $Z_k \sim \mathcal{N}(0,\sigma_Z^2)$, and τ_k is the speculator delay contributed by the k^{th} facade. Further, within the k^{th} cluster, the n^{th} scatter is associated with a channel gain of

$$\alpha_{k,n} = \frac{P'_{k,n}}{\sum_i^N P'_{k,i}}P_k, \tag{66}$$

where $P'_{k,n} = e^{-\frac{\tau_{k,n}}{D_s}}10^{\frac{U_n}{10}}$, $U_n \sim \mathcal{N}(0,\sigma_U^2)$, and $\tau_{k,n}$ is the delay associated with n^{th} scatter within the k^{th} facade. Following the parameters given in [56], it is chosen such that $D_c = 25.9$ ns, $\sigma_Z = 1$ dB, $D_s = 16.9$ ns, and $\sigma_U = 6$ dB.

Appendix E
Channel Parameter Estimation

Following [21], [22], we assume that the received pilot signal at n^{th} subcarrier is

$$\mathbf{Y}[n] = \mathbf{H}[n]\mathbf{X}[n] + \mathbf{W}[n], \tag{67}$$

where $\mathbf{W}[n]$ models the additive Gaussian noise, and $\mathbf{X}[n]$ is chosen such that $(\mathbf{X}[n])(\mathbf{X}[n])^{\mathrm{H}} = \mathbf{I}$. Then, the channel estimate is given by $\hat{\mathbf{H}}[n] = \mathbf{Y}[n](\mathbf{X}[n])^{\mathrm{H}}$ [21], and after stacking and rearranging the channel estimates of all N_f subcarriers, we have a tensor representation of the channel estimates, given by

$$\hat{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{W}} \in \mathbb{C}^{N_{r,x}\times N_{r,y}\times N_{t,x}\times N_{t,y}\times N_f}, \tag{68}$$

where the stacking order is implicit from the dimensions of the tensor. In (68), the tensor $\boldsymbol{\mathcal{H}}$ represents the true value of the channel, and the tensor $\boldsymbol{\mathcal{W}}$ represents the estimation error contained in $\hat{\boldsymbol{\mathcal{H}}}$, which is Gaussian. The SNR is defined as $\mathrm{SNR} = \|\boldsymbol{\mathcal{H}}\|_F^2/\|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}\|_F$ with $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}$ the covariance of $\boldsymbol{\mathcal{W}}$. The MATLAB package Tensorlab [57] is used to perform tensor-ESPRIT algorithm [21], [22] to estimate the parameter vector set $\mathcal{Z}$ from $\hat{\boldsymbol{\mathcal{H}}}$. The number of paths is estimated by the minimum description length (MDL) [58], this estimated number $\hat{L}$ is usually much less than L, representing the number of effective scattering paths.

References

[1] M. Säily, O. N. Yilmaz, D. S. Michalopoulos, E. Pérez, R. Keating, and J. Schaepperle, "Positioning technology trends and solutions toward 6G," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021.

[2] A. Behravan, V. Yajnanarayana, M. F. Keskin, H. Chen, D. Shrestha, T. E. Abrudan, T. Svensson, K. Schindhelm, A. Wolfgang, S. Lindberg *et al.*, "Positioning and sensing in 6G: Gaps, challenges, and opportunities," *IEEE Vehicular Technology Magazine*, 2022.

[3] J. Nikonowicz, A. Mahmood, M. I. Ashraf, E. Björnson, and M. Gidlund, "Indoor positioning trends in 5G-Advanced: Challenges and solution towards centimeter-level accuracy," *arXiv preprint arXiv:2209.01183*, 2022.

[4] W. Chen, J. Montojo, J. Lee, M. Shafi, and Y. Kim, "The standardization of 5G-Advanced in 3GPP," *IEEE Communications Magazine*, 2022.

[5] H. Chen, H. Sarieddeen, T. Ballal, H. Wymeersch, M.-S. Alouini, and T. Y. Al-Naffouri, "A tutorial on terahertz-band localization for 6G communication systems," *IEEE Communications Surveys & Tutorials*, 2022.

[6] J. Zhao, F. Gao, Q. Wu, S. Jin, Y. Wu, and W. Jia, "Beam tracking for UAV mounted SatCom on-the-move with massive antenna array," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 2, pp. 363–375, 2018.

[7] S. Bartoletti, H. Wymeersch, T. Mach, O. Brunnegård, D. Giustiniano, P. Hammarberg, M. F. Keskin, J. O. Lacruz, S. M. Razavi, J. Rönnblom, F. Tufvesson, J. Widmer, and N. B. Melazzi, "Positioning and sensing for vehicular safety applications in 5G and beyond," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 15–21, 2021.

[8] K. Witrisal, P. Meissner, E. Leitinger, Y. Shen, C. Gustafson, F. Tufvesson, K. Haneda, D. Dardari, A. F. Molisch, A. Conti *et al.*, "High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 59–70, 2016.

[9] A. Albanese, V. Sciancalepore, and X. Costa-Pérez, "First responders got wings: UAVs to the rescue of localization operations in beyond 5G systems," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 28–34, 2021.

[10] J. Struye, F. Lemic, and J. Famaey, "Millimeter-wave beamforming with continuous coverage for mobile interactive virtual reality," *arXiv preprint arXiv:2105.11793*, 2021.

[11] A. Shastri, N. Valecha, E. Bashirov, H. Tataria, M. Lentmaier, F. Tufvesson, M. Rossi, and P. Casari, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Communications Surveys & Tutorials*, 2022.

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

[13] O. Kanhere and T. S. Rappaport, "Position location for futuristic cellular communications: 5G and beyond," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 70–75, 2021.

[14] G. Kwon, A. Conti, H. Park, and M. Z. Win, "Joint communication and localization in millimeter wave networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1439–1454, 2021.

[15] M. A. Nazari, G. Seco-Granados, P. Johannisson, and H. Wymeersch, "3D orientation estimation with multiple 5G mmWave base stations," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

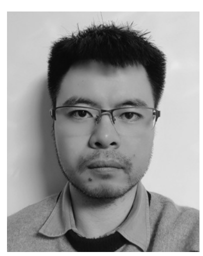
[16] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1822–1835, 2018.

[17] Y. Ge, F. Wen, H. Kim, M. Zhu, F. Jiang, S. Kim, L. Svensson, and H. Wymeersch, "5G SLAM using the clustering and assignment approach with diffuse multipath," *Sensors*, vol. 20, no. 16, 2020.

[18] J. Li, M. F. Da Costa, and U. Mitra, "Joint localization and orientation estimation in millimeter-wave mimo ofdm systems via atomic norm minimization," *arXiv preprint arXiv:2203.00892*, 2022.

[19] R. Mendrzik, H. Wymeersch, G. Bauch, and Z. Abu-Shaban, "Harnessing NLOS components for position and orientation estimation in 5G millimeter wave MIMO," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 93–107, 2019.

[20] A. Kakkavas, M. H. Castañeda García, R. A. Stirling-Gallacher, and J. A. Nossek, "Performance limits of single-anchor millimeter-wave positioning," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5196–5210, 2019.
[21] F. Wen, J. Kulmer, K. Witrisal, and H. Wymeersch, "5G positioning and mapping with diffuse multipath," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1164–1174, 2021.
[22] F. Wen and H. Wymeersch, "5G synchronization, positioning, and mapping from diffuse multipath," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 43–47, 2021.
[23] A. Guerra, F. Guidi, and D. Dardari, "Single-anchor localization and orientation performance limits using massive arrays: MIMO vs. beamforming," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5241–5255, 2018.
[24] M. A. Nazari, G. Seco-Granados, P. Johannisson, and H. Wymeersch, "Mmwave 6D radio localization with a snapshot observation from a single BS," *arXiv preprint arXiv:2204.05189*, 2022.
[25] H. Chen, M. F. Keskin, S. R. Aghdam, H. Kim, S. Lindberg, A. Wolfgang, T. E. Abrudan, T. Eriksson, and H. Wymeersch, "Modeling and analysis of 6G joint localization and communication under hardware impairments," *arXiv preprint arXiv:2301.01042*, 2023.
[26] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
[27] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual SLAM algorithms," *Robotics*, vol. 11, no. 1, p. 24, 2022.
[28] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
[29] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EP-n-P: An accurate O(n) solution to the P-n-P problem," *International journal of computer vision*, vol. 81, pp. 155–166, 2009.
[30] 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 Transactions on Wireless Communications*, vol. 17, no. 8, pp. 4939–4954, 2018.
[31] H. Deng and A. Sayeed, "Mm-wave MIMO channel modeling and user localization using sparse beamspace signatures," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2014, pp. 130–134.
[32] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2002.
[33] B. Jachimczyk, D. Dziak, and W. J. Kulesza, "Customization of UWB 3D-RTLS based on the new uncertainty model of the AoA ranging technique," *Sensors*, vol. 17, no. 2, 2017.
[34] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.
[35] M. Koivisto, J. Talvitie, E. Rastorgueva-Foi, Y. Lu, and M. Valkama, "Channel parameter estimation and TX positioning with multi-beam fusion in 5g mmWave networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3192–3207, 2021.
[36] W. Förstner and B. P. Wrobel, *Photogrammetric computer vision*. Springer, 2016.
[37] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.
[38] S. Shen, J. M. Menéndez Sánchez, S. Li, and H. Steendam, "Pose estimation for visible light systems using a quadrature angular diversity aperture receiver," *Sensors*, vol. 22, no. 14, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/14/5073
[39] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, pp. 756–770, 2004.
[40] Z. Kukelova, M. Bujnak, and T. Pajdla, "Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems." in *BMVC*, vol. 2, no. 5, 2008, p. 2008.
[41] P. I. Corke, W. Jachimczyk, and R. Pillat, *Robotics, vision and control: fundamental algorithms in MATLAB*. Springer, 2011, vol. 73.
[42] P. Corke, *Robotics, Vision and Control: Fundamental Algorithms in Python*. Springer Nature, 2023, vol. 146.
[43] Y. Ma, S. Soatto, J. Košecká, and S. Sastry, *An invitation to 3-D vision: from images to geometric models*. Springer, 2004, vol. 26.
[44] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
[45] B. Siciliano, O. Khatib, and T. Kröger, *Springer handbook of robotics*. Springer, 2008, vol. 200.
[46] S. Shen, S. Li, and H. Steendam, "Simultaneous position and orientation estimation for visible light systems with multiple LEDs and multiple PDs," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1866–1879, 2020.
[47] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930–943, 2003.
[48] H. Jiang, M. Mukherjee, J. Zhou, and J. Lloret, "Channel modeling and characteristics for 6G wireless communications," *IEEE Network*, vol. 35, no. 1, pp. 296–303, 2021.
[49] G. De Angelis, A. Moschitta, and P. Carbone, "Positioning techniques in indoor environments based on stochastic modeling of UWB round-trip-time measurements," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2272–2281, 2016.
[50] X. Li, E. Leitinger, M. Oskarsson, K. Åström, and F. Tufvesson, "Massive MIMO-based localization and mapping exploiting phase information of multipath components," *IEEE transactions on wireless communications*, vol. 18, no. 9, pp. 4254–4267, 2019.
[51] J. Kulmer, "High-accuracy positioning exploiting multipath for reducing the infrastructure," *Ph. D. dissertation, Graz University of Technology*, vol. 3, 2019.
[52] O. D. Faugeras and S. Maybank, "Motion from point matches: multiplicity of solutions," *International Journal of Computer Vision*, vol. 4, no. 3, pp. 225–246, 1990.
[53] F. S. Grassia, "Practical parameterization of rotations using the exponential map," *Journal of graphics tools*, vol. 3, no. 3, pp. 29–48, 1998.
[54] M. Kallmann, "Analytical inverse kinematics with body posture control," *Computer animation and virtual worlds*, vol. 19, no. 2, pp. 79–91, 2008.
[55] V. Degli-Esposti, F. Fuschini, E. M. Vitucci, and G. Falciasecca, "Measurement and modelling of scattering from buildings," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 1, pp. 143–153, 2007.
[56] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2225, 2016.
[57] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, 2016. [Online]. Available: http://www. tensorlab. net
[58] T. Yokota, N. Lee, and A. Cichocki, "Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1196–1206, 2017.



Shengqiang Shen (Member, IEEE) received the B.S. degree in information engineering from the China University of Mining and Technology, Xuzhou, China, in 2013, and the Ph.D. degree in electrical engineering from Ghent University, Ghent, Belgium, in 2021. He is currently a Lecturer with the School of Information and Control Engineering, China University of Mining and Technology. His research interests include wireless positioning and channel sounding.

Henk Wymeersch (Fellow, IEEE) received the Ph.D. degree in electrical engineering/applied sciences from Ghent University, Ghent, Belgium, in 2005. He is currently a Professor of communication systems with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. His current research focuses on the convergence of communication and sensing, in a 5G and beyond 5G context. Dr. Wymeersch is currently a Senior Member of the IEEE Signal Processing Magazine Editorial Board. During 2019–2021, he was an IEEE Distinguished Lecturer with the Vehicular Technology Society.