IEEE Open Journal of Vehicular Technology

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version 1 April, 2025.

Digital Object Identifier 10.1109/OJVT.2025.XXXXXXX

Indoor Localization and Tracking in Reconfigurable Intelligent Surface Aided mmWave Systems

Kunlun Li, *Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*, Chao Xu, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

ABSTRACT Millimeter wave (mmWave) carriers have a high available bandwidth, which can be beneficial for high-resolution localization in both the angular and temporal domains. However, the limited coverage due to severe path loss and line-of-sight (LoS) blockage are considered to be major challenges in mmWave. A promising solution is to employ reconfigurable intelligent surfaces (RIS) to circumvent the lack of line-of-sight paths, which can assist in localization. Furthermore, radio localization and tracking are capable of accurate real-time monitoring of the UE's locations and trajectories. In this paper, we propose a three-stage indoor tracking scheme. In the first stage, channel sounding is harnessed in support of the transmitter beamforming and receiver combining design. Based on the estimation in the first stage, a simplified received signal model is obtained, while using a discrete Fourier transform (DFT) matrix for the configuration of the RIS phase shifter for each time block. Based on the simplified received signal model, tracking initialization is carried out. Finally, in the third stage, Kalman filtering is employed for tracking. Our results demonstrate that the proposed scheme is capable of improving both the accuracy and robustness of tracking compared to single-shot successive localization. Additionally, we derive the position error bounds (PEB) of single-shot localization.

INDEX TERMS mmWave, localization/positioning, tracking, channel estimation, reconfigurable intelligent surfaces, sparse Bayesian learning.

Nomenclature

β	Complex Rician fading	T	Number of time blocks
$\mathbf{b}$	Location of BS	T_S	Sampling period
$\mathbf{r}$	Location of RIS	$\mathbf{A}_{\mathrm{U}}$,$\mathbf{A}_{\mathrm{R}}$,$\mathbf{A}_{\mathrm{B}}$	Steering matrices for UE, RIS and BS
$\boldsymbol{\omega}$	Phase shift vector	$\mathbf{C}$	Correlation matrix
$\boldsymbol{\Psi}$	Sensing matrix	$\mathbf{F}$	Random beamformer at BS
λ_c	Carrier frequency	$\mathbf{H}$	Channel model
ρ	Path-loss	$\mathbf{H}_{\mathrm{eff}}$	Effective channel
τ	Delay	$\mathbf{H}_{\mathrm{v}}$	Beamspace channel model
$\widehat{\mathbf{H}}$	Estimated channel matrix	$\mathbf{I}$	Fisher information matrix
$\widehat{\mathbf{h}}$	Vectorized estimated channel	$\mathbf{K}$	Kalman gain
$\widehat{\mathbf{P}}$	Updated estimation covariance matrix	$\mathbf{M}$	Measurement matrix
$\widehat{\mathbf{u}}_k$	Estimated location of UE for the k-th state	$\mathbf{P}$	Estimated covariance matrix for Kalman filtering
$\widehat{\mathbf{y}}$	Vectorized received signal	$\mathbf{Q}$	Process noise variance
$\widehat{\boldsymbol{\Gamma}}$	Updated hyperparameter	$\mathbf{R}$	Noise covariance matrix
B	Bandwidth	$\mathbf{T}$	State transition matrix
G	Number of OFDM symbols for channel sounding	$\mathbf{u}_k$	Location of UE for the k-th state
G_r	Beamspace resolution	$\mathbf{U}_{\mathrm{B}}$,$\mathbf{U}_{\mathrm{U}}$ and $\mathbf{U}_{\mathrm{R}}$	DFT matrix for beamspace samples
K	Number of motion states	$\mathbf{V}$	Measurement noise variance
L_{BR}	Number of paths between BS and RIS	$\mathbf{W}$	Random combiner at UE
L_{RU}	Number of paths between UE and RIS	$\mathbf{Y}$	Received signal
M	Number of OFDM symbols for each time block	$\boldsymbol{\Omega}$	Phase shift matrix at RIS
N	Number of subcarriers	ϕ_{BR}	AoA of RIS
N_{B}	Number of BS antennas	ϕ_{RM}	AoA of UE
N_{R}	Number of RIS elements	$\boldsymbol{\theta}_{\mathrm{BR}}$	AoD of BS
N_{U}	Number of UE antennas	$\boldsymbol{\theta}_{\mathrm{RM}}$	AoR of RIS
		$\boldsymbol{\Gamma}$	Hyperparamters for iterations

ζ	Maximum number of iteration
AMP	Approximate Message Passing
AoA	Angle of Arrival
AoD	Angle of Departure
AoR	Angle of Reflection
BS	Base Station
CDF	Cumulative Density Function
CIS	Continuous Intelligent Surface
CRLB	Cramer-Rao Lower Bound
CS	Compressed Sensing
DFT	Discrete Fourier Transform
DL	Downlink
GPS	Global Positioning System
IMM	Interacting Multiple Model
ISAC	Integrated Sensing and Communication
LoS	Line-of-Sight
MIMO	Multiple-Input-Multiple-Output
MMV	Multiple Measurement Vector
mmWave	Millimeter Wave
mSBL	Modified Sparse Bayesian Learning
NLN	Network Localization and Navigation
OFDM	Orthogonal Frequency Division Multiplexing
OTFS	Orthogonal Time-Frequency Space
PDF	Probability Density Function
PEB	Position Error Bound
PRSs	Position Reference Signals
RC	Receiver Combining
RIS	Reconfigurable Intelligent Surfaces
RMSE	Root Mean Squared Error
RSS	Received Signal Strength
SC	Scatterer
SNR	Signal-to-Noise Ratio
SOMP	Simultaneous Orthogonal Matching Pursuit
TBF	Transmitter Beamforming
ToA	Time of Arrival
UE	User Equipment
ULA	Uniform Linear Array
V2I	Vehicle to Infrastructure

I. Introduction

A. Motivation and Background

THERE has been a growing interest in localization and tracking due to its practical applications across various domains, including intelligent transportation systems and unmanned aerial vehicles [1]. Conventional localization methods, such as the global positioning system (GPS), often suffer from limited accuracy and high latency, especially in indoor environments [2]. To address these limitations, millimeter wave (mmWave)-based localization techniques have garnered significant attention [3]. These methods typically rely on a high number of antenna elements, facilitating high angular resolution [4]. Nevertheless, mmWave localization faces several practical challenges, such as high path loss, which can be mitigated through beamforming relying on massive antenna arrays [4]–[7]. The potential line-of-sight (LoS) blockages may be circumvented by reconfigurable intelligent surfaces (RIS) [8], which are capable of attaining potential performance improvements [8]–[11]. A RIS is composed of numerous reflectors typically positioned either in the vicinity of the transmitter or a receiver, where each individual element of the RIS can alter the phase and/or amplitude of the impinging signal [12]. This enables potential improvements in the energy efficiency, spectrum efficiency, positioning accuracy, communication security, etc. [13], [14]. While estimating the cascaded two-hop channels in the RIS-supported systems is a challenging task, the principles of passive beamforming or the combination of active and passive beamforming [15]–[17] can be beneficially exploited for localization.

Numerous studies have been conducted on RIS-assisted localization [16], [18]–[21]. Specifically in [18], the concept of continuous intelligent surfaces (CIS) was introduced, and the limits of RIS-aided localization and communication systems were discussed. More specifically, in [18], a general signal model was presented for RIS-aided localization and communication systems, when considering both far and near-field scenarios. In [19], holographic network localization and navigation (NLN) was proposed, where RISs relying on specific antenna patterns were used to improve the robustness of holographic localization against obstructions. In [20], a RIS-assisted localization scheme supported by adaptive beamforming using a hierarchical codebook based algorithm was proposed for joint localization and communication, when assuming the absence of LoS paths. In [21], a received signal strength (RSS)-based positioning scheme was investigated in a RIS-aided mmWave system. Furthermore, in [16], a joint active and passive beamforming codebook based localization scheme was conceived for RIS-aided mmWave systems. It was shown that the scheme is capable of striking an attractive performance vs. training overhead tradeoff.

However, these works only considered static scenarios. When the user equipment (UE) is moving, the localization problem becomes a tracking problem, and the complexity of training for the successive single-shot localization may become excessive [16], [22].

There are many solutions [23]–[29] for tackling the tracking problem. In [23], the authors propose integrated sensing and communication (ISAC) techniques to enhance vehicle-to-infrastructure (V2I) networks by dynamically adjusting the beamwidth and tracking the vehicle positions using predictive beamforming strategies employing extended Kalman filtering (EKF). In [24], a novel channel tracking algorithm was developed for mmWave systems operating in temporally correlated channels, advocating a hybrid analog/digital precoding approach for better adaptability to dynamic channel conditions. The authors of [25] introduced an extended Kalman filter-based beam tracking algorithm for improving the angle-of-departure (AoD) and angle-of-arrival (AoA) estimates, while offering enhanced beam alignment in mobile mmWave multiple-input-multiple-output (MIMO) systems. Furthermore, in [26], a sparsity based approach was conceived for multi-target tracking using orthogonal frequency division multiplexing (OFDM) radar, leveraging compressive sensing techniques to efficiently track targets in the delay-Doppler domain at a reduced computational overhead. In [27], the authors investigated orthogonal time-

frequency space (OTFS) systems, using hybrid digital-analog architectures and nested array techniques for improving the angle, delay, and Doppler shift estimation in high-mobility scenarios. In [28], a radar sensing and tracking scheme was proposed for dynamic targets, using a monostatic MIMO system integrated with ISAC functionalities for estimating parameters such as distance, angles, and velocities. In [29], the authors developed an algorithm for localizing and tracking vehicular users equipped with RISs. While the above studies provide valuable insights into tracking using EKF [23], sparsity-aware estimation [26], or hybrid beamforming under ISAC and radar frameworks [27], they primarily focus on either angle-delay-Doppler estimation or beam-level mobility support in MIMO systems. In contrast, this work proposes a unified three-stage RIS-aided localization and tracking scheme that explicitly models time-varying scatterers, incorporates phase configuration diversity through multiple RIS states, and employs an MMV-based sparse recovery framework for enhanced multipath estimation. Furthermore, we introduce a bias-compensated Kalman filter to track the UE trajectories using noisy position estimates extracted from the compressed sensing stage.

In this work, we propose a three-stage localization and tracking method indoor for RIS-aided localization scenarios, which are characterized by multi-path reflections, angular coupling, and signal degradation due to beamforming misalignment. In addition, we simulate low-SNR scenarios down to –10 dB to reflect realistic indoor deployment conditions, where mmWave signals may experience severe attenuation and blockage. The 2D geometry assumed in our simulations further reflects practical indoor use cases such as smart rooms or factory floors. We note that our previous work in [22], where we proposed a two-stage localization scheme, assumes a stationary UE. In contrast, firstly, we have upgraded our system model for the mobile UE scenario. Secondly, we have adapted the two-stage localization scheme considered in [22] for the mobile UE scenario, where a DFT-based RIS phase configuration is conceived to improve the performance of the random phase shift scenario in [22]. Thirdly, in this paper we propose to add a third stage constituted by a Kalman filter, so that the proposed three-stage tracking scheme becomes capable of achieving centimeter level localization accuracy. Table 1 contrasts the novelty of this paper to previous work, which is detailed as follows.

- We first propose a RIS-aided localization scheme for single-shot localization, using the multiple measurement vector based modified sparse Bayesian learning (MMV-mSBL) algorithm relying on the time, frequency and angular domains.
- To find the UE's location, we transfer the localization problem to an angle of reflection (AoR)/time of arrival (ToA) estimation problem. Specifically, to reduce the complexity of the MMV model, we exploit the frequency domain diversity for identifying the specific path having the highest power, where the cardinality of the angular domain in the MMV model is reduced to 1.
- Successive single-shot localization requires channel estimation for each motion state, which imposes both high complexity and training overhead. To reduce the overhead, we propose a three-stage tracking scheme, in which the single-shot localization is trained offline for inferring the measurement noise distribution. For online training, channel estimation is only required once in the first two stages. Briefly, in the first stage, random beamforming is employed for channel sounding in support of the transmitter beamforming (TBF) and receiver combining (RC) design. Based on the result of channel sounding in the first stage, a simplified received signal model is constructed. Then, a DFT matrix is harnessed for the configuration of the RIS phase shifter in each time block for initializing the Kalman filter in the second stage. Finally, in the third stage, the Kalman filter is employed for tracking exploiting the prior information gleaned during the offline training, which improves the accuracy and robustness compared to single-shot localization.
- Finally, a position error bound (PEB) is derived, and the localization error distribution is characterized.

B. Organization of the Paper and Notations

The rest of the paper is organized as follows. Section II introduces the model of our localization and tracking system. Section III introduces the framework of single-shot localization, while Section IV presents the three-stage tracking technique conceived. The simulation setup and our simulation results are provided in Section V. Finally, we conclude in Section VI.

Notations: a, $\mathbf{a}$, $\mathbf{A}$ stand for scalar, vector and matrix, respectively. $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, $\mathbf{A}^{\dagger}$, $\|\mathbf{a}\|_2$ and $\|\mathbf{A}\|_F$ represent the transpose, Hermitian transpose, pseudoinverse, Euclidean norm and Frobenius norm of matrix $\mathbf{A}$, respectively. The (i,j)-th entry of $\mathbf{A}$ is $[\mathbf{A}]_{i,j}$, and $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix formed by the diagonal elements of $\mathbf{a}$. $\mathrm{Trace}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$, $\mathbb{E}(\mathbf{A})$ is the expectation of $\mathbf{A}$, $\mathrm{vec}(\mathbf{A})$ is the vectorization operation of $\mathbf{A}$, $\mathrm{mod}(i,j)$ denotes the modulo operation, and $j=\sqrt{-1}$; $(\mathbf{A})^*$ represents the conjugate of matrix $\mathbf{A}$.

II. System Model

We consider an indoor downlink (DL) MIMO localization and tracking system, where the BS and the moving UE are equipped with N_{B} and N_{U} antennas, respectively. Furthermore, a RIS having N_{R} antennas is deployed between the BS and the UE to overcome the line-of-sight (LoS) blockage problem, as shown in Fig. 1. More explicitly, the OFDM modulated position reference signals (PRSs) are transmitted in the DL by the BS, which are then reflected by the RIS to the UE, to circumvent the LoS blockage. Based on the reflected signals received by the UE, the UE estimates the channel parameters and its location for the k-th motion state, $k=1,...,K$, where each motion state represents the real-time location of the UE.

We define the location of the BS as $\mathbf{b}=[b_x,b_y]^{\mathrm{T}}\in\mathbb{R}^2$, and the location of the UE for the k-th state as $\mathbf{u}_k=$

TABLE 1: Contrasting our novelty to the literature.

	[16]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	This work
System Model										
Time-vary random scatterers between the RIS and the UE in successive localization										✓
Radio localization										
User self-sensing	✓	✓								✓
MMV channel estimation based localization		✓								✓
RIS-assisted localization	✓	✓							✓	✓
Orthogonal phase shift configuration for MMV model										✓
PEB analysis for MMV model										✓
Tracking process										
Arbitrary tracking			✓	✓						✓
Kalman filter refines localization accuracy							✓			✓

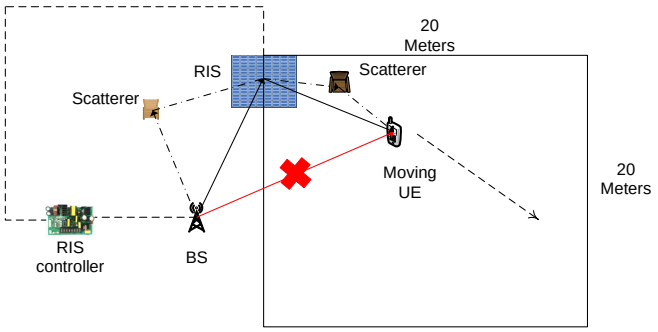


FIGURE 1: Illustration of the RIS aided localization and tracking model.

$[u_x^k, u_y^k]^{\mathrm{T}} \in \mathbb{R}^2$, while the location of the RIS is given by $\mathbf{r} = [r_x, r_y]^{\mathrm{T}} \in \mathbb{R}^2$. Furthermore, a scatterer (SC) is assumed to exist between the BS and the RIS in a fixed location, but other SCs exist between the RIS and the UE, which have different positions for different states k. The locations of both the BS and of the RIS are fixed and assumed to be known as a reference point for localization, while the location of the UE for the k-th state can be estimated based on the estimated channel parameters.

Again, a mmWave channel is considered with the LoS path blocked, where the channel parameters, such as the AoA, AoD, AoR and ToA, are determined by the geometry of the BS, UE and RIS. Therefore, downlink transmission has to rely on the path reflected by the RIS, as shown in Fig. 1. The AoD of the BS and the AoA of the RIS are denoted by θ_{BR} and ϕ_{BR}, while the AoR of the RIS and the AoA of the UE are expressed as $\theta_{\mathrm{RU},k}$ and $\phi_{\mathrm{RU},k}$ for the k-th state, respectively. Moreover, the ToA between the BS and the RIS, as well as between the RIS and the UE are τ_{BR} and $\tau_{\mathrm{RU},k}$, respectively. The study assumes that the BS, RIS and UE are equipped with a uniform linear array (ULA) for simplicity[1]. Hence, the steering vectors $\mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{BR}})$ and $\mathbf{a}_{\mathrm{R},n}(\phi_{\mathrm{BR}})$ at the BS and RIS in the context of subcarrier n are [30]

$$\mathbf{a}_{\mathrm{B},n}(\theta_{\mathrm{BR}}) = \frac{1}{\sqrt{N_{\mathrm{B}}}}\Big[1, e^{-j2\pi\frac{d}{\lambda_n}\sin(\theta_{\mathrm{BR}})}, ..., e^{-j2\pi\frac{d}{\lambda_n}\sin(\theta_{\mathrm{BR}})(N_{\mathrm{B}}-1)}\Big]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{B}}\times 1}, \quad (1)$$

$$\mathbf{a}_{\mathrm{R},n}(\phi_{\mathrm{BR}}) = \frac{1}{\sqrt{N_{\mathrm{R}}}}\Big[1, e^{-j2\pi\frac{d}{\lambda_n}\sin(\phi_{\mathrm{BR}})}, ..., e^{-j2\pi\frac{d}{\lambda_n}\sin(\phi_{\mathrm{BR}})(N_{\mathrm{R}}-1)}\Big]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}, \quad (2)$$

where d represents the element spacing, and λ_n is the wavelength of the n-th subcarrier. For simplicity, we assume that the signal bandwidth obeys $B \ll f_c$, yielding $\lambda_n \approx \lambda_c$, where λ_c represents the wavelength of the main carrier [31]. Then, the mmWave channel model introduced in [32], [33] can be applied to obtain the $(N_{\mathrm{R}} \times N_{\mathrm{B}})$-dimensional frequency domain channel matrix of the line between the BS and the RIS, which can be represented as [34]:

$$\mathbf{H}_{\mathrm{BR}}[n] = \mathbf{A}_{\mathrm{R}}(\boldsymbol{\phi}_{\mathrm{BR}})\boldsymbol{\Sigma}_{\mathrm{BR}}[n]\mathbf{A}_{\mathrm{B}}^{\mathrm{H}}(\boldsymbol{\theta}_{\mathrm{BR}}), \quad (3)$$

where we have

$$\mathbf{A}_{\mathrm{B}}(\boldsymbol{\theta}_{\mathrm{BR}}) = [\mathbf{a}_{\mathrm{B}}(\theta_{\mathrm{BR},0}), \mathbf{a}_{\mathrm{B}}(\theta_{\mathrm{BR},1})..., \mathbf{a}_{\mathrm{B}}(\theta_{\mathrm{BR},L_{\mathrm{BR}}-1})], \quad (4)$$

$$\mathbf{A}_{\mathrm{R}}(\boldsymbol{\phi}_{\mathrm{BR}}) = [\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{BR},0}), \mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{BR},1})..., \mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{BR},L_{\mathrm{BR}}-1})], \quad (5)$$

and the diagonal matrix $\boldsymbol{\Sigma}_{\mathrm{BR}}[n]$ is given by [16]

$$\boldsymbol{\Sigma}_{\mathrm{BR}}[n] = \sqrt{N_{\mathrm{B}}N_{\mathrm{R}}} \times \mathrm{diag}\{\beta_{\mathrm{BR},0}\rho_{\mathrm{BR},0}e^{\frac{-j2\pi n\tau_{\mathrm{BR},0}}{NT_s}}, ..., \beta_{\mathrm{BR},L_{\mathrm{BR}}-1}\rho_{\mathrm{BR},L_{\mathrm{BR}}-1}e^{\frac{-j2\pi n\tau_{\mathrm{BR},L_{\mathrm{BR}}-1}}{NT_s}}\}. \quad (6)$$

In (4)-(6), L_{BR} is the number of paths between the BS and the RIS, $\beta_{\mathrm{BR},l}$ and $\rho_{\mathrm{BR},l}$ are respectively the complex Rician fading gain and path-loss of the l-th path between the BS and the RIS, and $T_S = 1/B$ is the sampling period. In (6), the time delay $\tau_{\mathrm{BR},l}$, i.e. ToA, is given by $\tau_{\mathrm{BR},l} = d_{\mathrm{BR},l}/c$, where c denotes the speed of light, while $d_{\mathrm{BR},l}$ is the propagation distance of the l-th path. Specifically, for the LoS path ($l = 0$), the distance between the BS and the RIS is evaluated as $d_{\mathrm{BR},0} = \|\mathbf{r} - \mathbf{b}\|_2$, while for the NLoS path ($l > 0$), the distance is $d_{\mathrm{BR},l} = \|\mathbf{s}_l - \mathbf{b}\|_2 + \|\mathbf{r} - \mathbf{s}_l\|_2$, where $\mathbf{s}_l$ is the

[1]This can be generalized to any antenna structures.

location of the SC. Again, we consider an indoor localization scenario in this paper, where the Doppler effect is ignored for simplicity[2], but would have to be considered and estimated in vehicular systems [29].

We note that the RIS-UE link is time-varying, where the delays of all reflected links are different due to their different AoAs. Therefore, for each motion state k, its position $\mathbf{s}_{2,k} \in \mathbb{R}^2$ is modeled as a uniformly distributed random point along the line segment that connects the RIS and the UE, defined as $\mathbf{s}_{2,k} = \mathbf{r} + \lambda_k (\mathbf{u}_k - \mathbf{r}), \quad \lambda_k \sim \mathcal{U}(0,1)$. This construction reflects the physical intuition that local SCs on the RIS-UE path often lie in proximity of the geometric propagation trajectory.

Similarly, for the k-th state of the UE's motion model, the channel between the RIS and the UE can be represented as

$$\mathbf{H}_{\text{RU}}[k,n] = \mathbf{A}_{\text{U}}(\boldsymbol{\phi}_{\text{RU},k})\boldsymbol{\Sigma}_{\text{RU},k}[n]\mathbf{A}_{\text{R}}^{\text{H}}(\boldsymbol{\theta}_{\text{RU},k}), \tag{7}$$

where $\mathbf{A}_{\text{U}}(\boldsymbol{\phi}_{\text{RM},k})[n]$ and $\mathbf{A}_{\text{R}}(\boldsymbol{\theta}_{\text{RM},k})[n]$ are defined similarly as in (3), by adding the index of the k-th state to the channel parameters $\boldsymbol{\phi}_{\text{RU},k}$, $\boldsymbol{\theta}_{\text{RU},k}$ and $\boldsymbol{\tau}_{\text{RU},k}$, determined by the k-th real-time location of the UE.

By combining (3) and (7), the frequency domain cascaded channel spanning from the BS to the UE for the k-th state can be represented as

$$\mathbf{H}[k,n] = \mathbf{H}_{\text{RU}}[k,n]\boldsymbol{\Omega}_k\mathbf{H}_{\text{BR}}[n], \tag{8}$$

where $\boldsymbol{\Omega}_k \in \mathbb{C}^{N_{\text{R}} \times N_{\text{R}}}$ is the RIS's phase shift matrix associated with the motion state $k = 0, \ldots, K$. Furthermore, $\boldsymbol{\Omega}_k$ is a diagonal matrix that has unit-modulus on the diagonal elements [35], [36]. Specifically, the diagonal element is $[\boldsymbol{\Omega}_k]_{i,i} = e^{j\omega_i}$, where $\omega_i \in [0, 2\pi]$. According to (8), we define the effective channel as

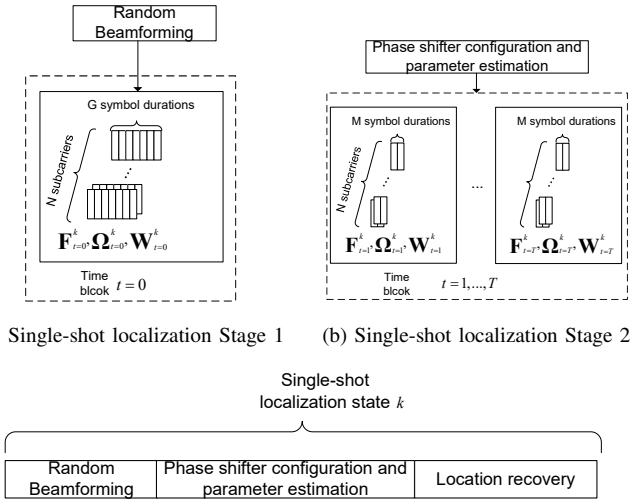
$$\begin{aligned}\mathbf{H}_{\text{eff}}[k,n] =& \text{diag}(\widehat{\boldsymbol{\rho}}_{\text{RU},k}[n])\mathbf{A}_{\text{R}}^{\text{H}}(\boldsymbol{\theta}_{\text{RU},k})\boldsymbol{\Omega}_k\mathbf{A}_{\text{R}}(\boldsymbol{\phi}_{\text{BR}}) \\ &\times \text{diag}(\widehat{\boldsymbol{\rho}}_{\text{BR}}[n]),\end{aligned} \tag{9}$$

where $\widehat{\boldsymbol{\rho}}_{\text{RU},k}[n] = [\beta_{\text{RU},0}^k \rho_{\text{RU},0}^k e^{\frac{-j2\pi n \tau_{\text{RU},0}^k}{NT_s}}, \ldots, \beta_{\text{RU},L_{\text{RU}}-1}^k \rho_{\text{RU},L_{\text{RU}}-1}^k e^{\frac{-j2\pi n \tau_{\text{RU},L_{\text{RU}}-1}^k}{NT_s}}]^{\text{T}}$, and $\widehat{\boldsymbol{\rho}}_{\text{BR}}[n]$ is similarly defined according to (6). Thus, the frequency domain channel can then be represented as

$$\mathbf{H}[k,n] = \mathbf{A}_{\text{U}}(\boldsymbol{\phi}_{\text{RU},k})\mathbf{H}_{\text{eff}}[k,n]\mathbf{A}_{\text{B}}^{\text{H}}(\boldsymbol{\theta}_{\text{BR}}), \tag{10}$$

where $\mathbf{H}_{\text{eff}}[k,n]$ defined in (9) is a function of the phase shifter matrix $\boldsymbol{\Omega}_k$ for the k-th state, which shows the importance of the angular parameters $\boldsymbol{\theta}_{\text{RU}}$ and $\boldsymbol{\phi}_{\text{BR}}$, as well as of the corresponding phase shifter design for both channel estimation and localization. Observe based on (10) that the localization is also related to the estimate of the channel parameters, hence it may be viewed as a channel estimation problem. Hence, in the next section, we will formulate localization as MMV based channel estimation.

[2]For instance, when the carrier frequency f_c is 28 GHz and the velocity of UE is 1 m/s, the Doppler shift is $f_D = 93.33$ Hz. When we consider the total bandwidth of $B = 100$ MHz and the number of subcarriers N is relatively small, e.g. $N = 20$, the subcarrier spacing is $\Delta f = 5$ MHz, which means that the Doppler shift has very little effect in the frequency domain. However, we note that the proposed architecture can be extended to dynamic outdoor environments, such as vehicular or UAV-based systems. In such cases, Doppler-induced frequency shifts become significant and hence should be jointly estimated with delay parameters, which is our future research.



(a) Single-shot localization Stage 1 (b) Single-shot localization Stage 2

(c) Single-shot localization Stage 3

FIGURE 2: Illustration of the single-shot localization scheme.

III. Compressed Sensing Based Single-Shot Localization

In mmWave localization systems, one of the conventional methods is to exploit the channel-sparsity using CS algorithms in both the angular and temporal domains, in order to glean the location-related information, such as the AoD and ToA [34], [37], [38]. For conventional ToA-based localization, either predefined waveform structures [39], or alternative schemes such as time difference of arrival (TDoA) relying on multiple synchronized anchors [40] are required. In our framework, both the delay and angle information are estimated from multicarrier channel observations. Specifically, the delay is inferred from the phase shifts across subcarriers, while angular features are extracted by exploiting spatial sparsity. This motivates our unified channel-estimation-based localization scheme, which operates within a multicarrier communication system and avoids the need for either multiple synchronized receivers or sensing-specific waveform designs. However, predefined precoding, combining and time-domain DFT matrices cause quantization error, in both of these domains which degrades localization performance [16]. On the other hand, in RIS-aided localization, it is challenging to exploit the AoR for localization, when the channel state information is unknown. In this section, as shown in Fig. 2a, for the single-shot localization carried out in the first stage, we employ random beam training for estimating the angle parameters such as the AoA and AoD at the BS and the UE, as detailed in Subsection A. In Subsection B, we design the phase shifter

matrix to prevent the encountering path crossing[3] problem first identified in [17], and derive a simplified received signal model. The sparse formulation and MMV-mSBL algorithm designed for determining the channel parameters are introduced in Subsection C and Subsection D, respectively. Furthermore, the PEB is derived in Subsection E.

A. Random Beamforming for Channel Sounding

For each motion state k, the first time block $t = 0$, as shown in Fig. 2a, a random precoding matrix $\mathbf{F}^0[k,n] \in \mathbb{C}^{N_\mathrm{B}\times G}$ associated with uniform Gaussian distribution and a random combining matrix $\mathbf{W}^0[k,n] \in \mathbb{C}^{N_\mathrm{U}\times G}$ with uniform Gaussian distribution are employed for eliminating directional bias and achieving isotropic coverage, where G denotes the number of symbol durations, as shown in Fig. 2b. The parameter G has to be larger than the number of propagation paths for assumption of angular sparsity to hold. Furthermore, at this stage, the phase shift matrix $\mathbf{\Omega}_k^0$ at the RIS is set randomly and it is fixed during the next G symbol durations. Thus, for the k-th state and $t = 0$-th time block, the observations at the UE can be further written as

$$\begin{aligned}\mathbf{Y}^0[k,n] =&\sqrt{P}(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{H}^0[k,n]\mathbf{F}^0[k,n]\\ &+(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{N}^0[k,n].\end{aligned} \tag{11}$$

Let us introduce the beamspace channel representation [31], which is obtained via uniformly sampling the spatial angles in the beamspace, yielding:

$$\begin{aligned}\mathbf{U}_\mathrm{B} =&\left[\mathbf{u}_\mathrm{B}\left(q_0\right),\ldots,\mathbf{u}_\mathrm{B}\left(q_{G_\mathrm{r}-1}\right)\right],\\ \mathbf{u}_\mathrm{B}\left(q_r\right) =&\left[1,\ldots,e^{j2\pi(N_\mathrm{B}-1)q_r}\right]^{\mathrm{T}}.\end{aligned} \tag{12}$$

In (12), $\mathbf{U}_\mathrm{B}$ is a unitary DFT matrix determined by the beamspace grid indices of $q_r = \frac{(r-1)+(-\frac{G_\mathrm{r}-1}{2})}{G_\mathrm{r}}, \forall r \in [0, G_\mathrm{r}-1]$, where G_r is the beamspace resolution of the spatial angles. Then, the beamspace channel representation of $\mathbf{H}^0[k,n]$ in (11) for the k-th state and for the $t = 0$ block can be written as

$$\mathbf{H}_\mathrm{v}^0[k,n] = \mathbf{U}_\mathrm{U}^{\mathrm{H}}\mathbf{H}^0[k,n]\mathbf{U}_\mathrm{B} \in \mathbb{C}^{N_\mathrm{U}\times N_\mathrm{B}}, \tag{13}$$

where the DFT dictionary matrix $\mathbf{U}_\mathrm{u}$ for the receiver array is obtained similarly as $\mathbf{U}_\mathrm{B}$. Upon substituting (13) into (11), we obtain

$$\begin{aligned}\mathbf{Y}^0[k,n] =&\sqrt{P}(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{U}_\mathrm{U}\mathbf{H}_\mathrm{v}^0[k,n]\mathbf{U}_\mathrm{B}^{\mathrm{H}}\mathbf{F}^0[k,n]\\ &+(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{N}^0[k,n].\end{aligned} \tag{14}$$

Furthermore, to represent the channel vector in a standard CS form, let us vectorize the observations $\mathbf{Y}^0[n]$, yielding [41]:

$$\begin{aligned}\mathrm{vec}(\mathbf{Y}^0[k,n]) =&\mathrm{vec}(\sqrt{P}(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{U}_\mathrm{U}\mathbf{H}_\mathrm{v}^0[k,n]\mathbf{U}_\mathrm{B}^{\mathrm{H}}\mathbf{F}^0[k,n]\\ &+(\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{N}^0[k,n])\\ =&[(\mathbf{F}^0[k,n])^{\mathrm{T}}\otimes(\mathbf{W}^0[k,n])^{\mathrm{H}}][(\mathbf{U}_\mathrm{B}^0)^*\otimes\mathbf{U}_\mathrm{U}^0]\\ &\times\mathrm{vec}(\mathbf{H}_\mathrm{v}^0[k,n])+\mathrm{vec}((\mathbf{W}^0[k,n])^{\mathrm{H}}\mathbf{N}^0[k,n])\\ =&\mathbf{\Phi}^0[k,n]\boldsymbol{\chi}^0\mathbf{h}_\mathrm{v}^0[k,n]+\mathbf{n}^0[k,n],\end{aligned} \tag{15}$$

where $\mathbf{\Phi}^0[k,n] = (\mathbf{F}^0[k,n])^{\mathrm{T}}\otimes(\mathbf{W}^0[k,n])^{\mathrm{H}}$ is the sensing matrix[4], and $\boldsymbol{\chi}^0 = (\mathbf{U}_\mathrm{B}^0)^*\otimes\mathbf{U}_\mathrm{U}^0$ is the overcomplete dictionary or beamspace transformation matrix [31], [41]. Furthermore, $\mathbf{h}_\mathrm{v}^0[n]$ is the vectorized beamspace channel vector to be estimated, while $\mathbf{n}^0[k,n]$ is the noise vector. The AoD $\widehat{\boldsymbol{\theta}}_{\mathrm{BR},k}$ and AoA $\widehat{\boldsymbol{\phi}}_{\mathrm{RU},k}$ can either be estimated according to (11), using an *off-grid* atomic norm minimization algorithm, or estimated based on (15) as an *on-grid* CS problem [22].

B. Phase Shifter Configuration and Received Signal Simplification

As shown in Fig. 2b, for the k-th state, when the precoding matrices $\mathbf{F}^1[k,n] = \mathbf{F}^2[k,n] = ... = \mathbf{F}^T[k,n]$ at the BS and the combining matrix $\mathbf{W}^1[k,n] = \mathbf{W}^2[k,n] = ... = \mathbf{W}^T[k,n]$ at the UE are obtained for the blocks $t > 0$, as shown in the previous subsection, the AoR at the RIS can then be estimated using the time block of $t > 0$, as shown in Fig. 2c, which can be used in the single-shot localization of the UE.

More specifically, as mentioned in Section II, after the estimation of $\boldsymbol{\theta}_{\mathrm{BR},k}$ and $\boldsymbol{\phi}_{\mathrm{RU},k}$ in the first stage, the corresponding beam training matrices at the BS and the UE for the RIS can be designed as follows:

$$\begin{aligned}\mathbf{W}^t[k,n] =&\mathbf{A}_\mathrm{U}(\widehat{\boldsymbol{\phi}}_{\mathrm{RU},k}),\\ \mathbf{F}^t[k,n] =&\mathbf{A}_\mathrm{B}(\widehat{\boldsymbol{\theta}}_{\mathrm{BR},k}),\end{aligned} \tag{16}$$

where we have $t = 1, ..., T$, $k = 0, ..., K$ and $n = 0, ..., N-1$. Recalling the time block $t = 0$ for each state k, we note that the number G of symbol durations required in the first stage, as shown in Fig. 2a is much higher than the number of paths L_{BR} and L_{RU}, where L_{RU} denotes that between the RIS and the UE, while L_{BR} of those between the BS and RIS, respectively. In contrast, for $t > 0$, as shown in Fig. 2b, the number of symbol durations M only has to be larger than the number of paths, e.g. $M \geq L_{\mathrm{RU}}$ and $M \geq L_{\mathrm{BR}}$, resulting in $\mathbf{W}^t[k,n] \in \mathbb{C}^{N_\mathrm{B}\times L_{\mathrm{RU}}}$ and $\mathbf{F}^t[k,n] \in \mathbb{C}^{N_\mathrm{B}\times L_{\mathrm{BR}}}$. In this regard, the training overhead can be significantly reduced for the $t > 0$ blocks.

Recalling (11) and (16), the received signal at the UE corresponding to $t > 0$ for the k-th motion state can be

[3] In [17], the phenomenon of "path crossing", referring to the effective multiplication of multipath components from the BS–RIS and RIS–UE links after applying a random phase shift matrix, which implicitly indicates that when both the BS–RIS and RIS–UE links contain multipath (e.g., two paths each), their combination results in four distinct effective propagation paths (2×2 combinations).

[4] In the literature, the random beamformers conceived for constructing sensing matrices can be predefined or selected from a shared codebook that is known to both sides [37], [42], [43]. This enables robust sparse recovery without requiring exhaustive beam training and significantly reduces the implementation complexity.

formulated as

$$\begin{aligned}\mathbf{Y}^t[k,n] =&\sqrt{P}(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{H}^t[k,n]\mathbf{F}^t[k,n]\\ &+(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{N}^t[k,n]\\ =&\sqrt{P}(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{A}_{\mathrm{U}}(\boldsymbol{\phi}_{\mathrm{RU}})\mathbf{H}^t_{\mathrm{eff}}[k,n]\\ &\times\mathbf{A}_{\mathrm{B}}(\boldsymbol{\theta}_{\mathrm{BR}})\mathbf{F}^t[k,n]+(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{N}^t[n]\\ \approx&\sqrt{P}\mathbf{H}^t_{\mathrm{eff}}[k,n]+(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{N}^t[k,n],\end{aligned}\tag{17}$$

where $\mathbf{W}^t[k,n]^{\mathrm{H}}\mathbf{A}_{\mathrm{U}}(\phi_{\mathrm{RU}})\approx\mathbf{I}$ and $\mathbf{A}_{\mathrm{B}}(\boldsymbol{\theta}_{\mathrm{BR}})\mathbf{F}^t[k,n]\approx\mathbf{I}$, owing to the precoder and combiner designed for the BS and the UE, as shown in (16). However, in practical settings, the estimated directions $\hat{\boldsymbol{\theta}}_{\mathrm{BR},k}$, $\hat{\boldsymbol{\phi}}_{\mathrm{RU},k}$ may deviate from the true angles due to resolution limitations or noise. This causes a mismatch in the beamforming directions, resulting in a partial misalignment of $\mathbf{F}^t[k,n]$ and $\mathbf{W}^t[k,n]$. To model this, we introduce a signal fidelity coefficient $\eta_{\mathrm{error}}\in(0,1]$, representing the cosine of the alignment between the true and estimated beam directions. Assuming that both the TBF and RC suffer from this alignment-induced loss, the received signal can be written as

$$\mathbf{Y}^t[k,n]\approx\eta^2_{\mathrm{error}}\mathbf{H}^t_{\mathrm{eff}}[k,n]+(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{N}^t[k,n].\tag{18}$$

We emphasize that the noise term $(\mathbf{W}^t[k,n])^{\mathrm{H}}\mathbf{N}^t[k,n]$ is kept unchanged to preserve analytical simplicity, even though the combiner mismatch could slightly affect the noise projection. Then, the effective channel $\mathbf{H}^t_{\mathrm{eff}}[k,n]$ is defined in (9). It can be shown that the (a,b)-th entry of the effective channel $\mathbf{H}^t_{\mathrm{eff}}[k,n]$ can be expressed as [12]

$$\begin{aligned}[\mathbf{H}^t_{\mathrm{eff}}[k,n]]_{(a,b)}=&(\widehat{\boldsymbol{\rho}}_{\mathrm{RU}}[k,n])_a(\boldsymbol{\omega}^t_k)^{\mathrm{T}}\\ &\times\mathbf{a}([\boldsymbol{\theta}_{\mathrm{spatial},k}]_{(a,b)})(\widehat{\boldsymbol{\rho}}_{\mathrm{BR}}[k,n])_b,\end{aligned}\tag{19}$$

where $\boldsymbol{\omega}^t_k$ is a vector denoting the diagonal elements of the phase shift matrix $\boldsymbol{\Omega}^t_k$, and $\boldsymbol{\theta}_{\mathrm{spatial},k}=\mathrm{asin}(\sin([\boldsymbol{\phi}_{\mathrm{BR},k}]_b)-\sin([\boldsymbol{\theta}_{\mathrm{RU},k}]_a))$ is the spatial frequency [12]. By vectorizing (9), the effective channel vector can be represented as $\mathbf{h}^t_{\mathrm{eff}}[k,n]=\mathrm{vec}(\mathbf{H}^t_{\mathrm{eff}}[k,n])\in\mathbb{C}^{L_{\mathrm{RU}}L_{\mathrm{BR}}\times1}$, and the elements of $\mathbf{h}^t_{\mathrm{eff}}[k,n]$ can be written as

$$[\mathbf{h}^t_{\mathrm{eff}}[k,n]]_l=\widehat{\boldsymbol{\rho}}_l[k,n]\boldsymbol{\omega}^t_k\mathbf{a}(\widehat{\boldsymbol{\theta}}_{\mathrm{spatial},k,l}),l=1,...,L_{\mathrm{RU}}L_{\mathrm{BR}},\tag{20}$$

where we have $\widehat{\boldsymbol{\rho}}_l[k,n]=(\widehat{\boldsymbol{\rho}}_{\mathrm{RU}}[k,n])_a(\widehat{\boldsymbol{\rho}}_{\mathrm{BR}}[k,n])_b$, $\widehat{\boldsymbol{\theta}}_{\mathrm{spatial},k,l}=\arcsin(\sin([\boldsymbol{\phi}_{\mathrm{BR},k}]_b)-\sin([\boldsymbol{\theta}_{\mathrm{RU},k}]_a))$, while $a=\mod(l-1,L_{\mathrm{RU}})+1$ and $b=\lceil l/L_{\mathrm{RU}}\rceil$. Consequently, when considering all the observations over T blocks for the k-th state, according to (17), we have

$$\begin{aligned}\mathbf{Y}[k,n]=&[\mathrm{vec}(\mathbf{Y}^1)[k,n],\mathrm{vec}(\mathbf{Y}^2)[k,n],...,\mathrm{vec}(\mathbf{Y}^T)[k,n]],\\ &n=0,...,N-1.\end{aligned}\tag{21}$$

The effective channel is further expressed as $\widehat{\mathbf{H}}_{\mathrm{eff}}[k,n]=[\mathbf{h}^1_{\mathrm{eff}}[k,n],\mathbf{h}^2_{\mathrm{eff}}[k,n],...,\mathbf{h}^T_{\mathrm{eff}}[k,n]]$. Furthermore, based on (17), (19) and (20), it can be shown that the l-th row of $\mathbf{Y}[k,n]$ is

$$\begin{aligned}\mathbf{Y}_{l,:}[k,n]\approx&\widehat{\mathbf{H}}_{\mathrm{eff},l,:}[k,n]^{\mathrm{T}}+\widehat{\mathbf{N}}_{l,:}[k,n]\\ =&\widehat{\boldsymbol{\Omega}}_k\widehat{\boldsymbol{\rho}}_l[k,n]\mathbf{a}(\widehat{\boldsymbol{\theta}}_{\mathrm{spatial},k,l})+\widehat{\mathbf{N}}_{l,:}[k,n],\\ l=&1,...,L_{\mathrm{RU}}L_{\mathrm{BR}}.\end{aligned}\tag{22}$$

Phase shift configuration: In (22), $\widehat{\boldsymbol{\Omega}}_k=[\boldsymbol{\omega}^1_k,\boldsymbol{\omega}^2_k,...,\boldsymbol{\omega}^T_k]^{\mathrm{T}}$ is from the phase shift vectors used for the transmissions over T blocks. Explicitly, we define a $T\times T$ DFT matrix in the form:

$$\mathbf{F}_{\mathrm{R}}=\begin{bmatrix}1&1&1&\cdots&1\\1&\omega&\omega^2&\cdots&\omega^{T-1}\\1&\omega^2&\omega^4&\cdots&\omega^{2(T-1)}\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&\omega^{T-1}&\omega^{2(T-1)}&\cdots&\omega^{(T-1)(T-1)}\end{bmatrix},\tag{23}$$

where $\widehat{\boldsymbol{\Omega}}_k$ contains the first N_{R} rows of $\mathbf{F}_{\mathrm{R}}$ ($N_{\mathrm{R}}\le T$), e.g. $\widehat{\boldsymbol{\Omega}}_k=[\mathbf{F}_{\mathrm{R}}]_{1:N_{\mathrm{R}},:}$ and each column of $\widehat{\boldsymbol{\Omega}}_k$ denotes $\boldsymbol{\omega}^t_k$, which is the phase shift vector for the k-th state and t-th time block. Moreover, the stacked noise matrix is expressed as $\widehat{\mathbf{N}}[k,n]=[\mathrm{vec}(\mathbf{W}^1[k,n]^{\mathrm{H}}\mathbf{N}^1[k,n]),\mathrm{vec}(\mathbf{W}^2[k,n]^{\mathrm{H}}\mathbf{N}^2[k,n]),...,\mathrm{vec}(\mathbf{W}^T[k,n]^{\mathrm{H}}\mathbf{N}^T[k,n])]$, where the covariance matrix of the AWGN vector with respect to the t-th block is $\mathbf{R}^t_k=(\sigma^t_k)^2\mathbf{W}^t[k,n]^{\mathrm{H}}\mathbf{W}^t[k,n]$. Moreover, the covariance matrix $\mathbf{R}_k\in\mathbb{C}^{TL_{\mathrm{RU}}L_{\mathrm{BR}}\times TL_{\mathrm{RU}}L_{\mathrm{BR}}}$ of the vectorized noise matrix $\mathrm{vec}(\widehat{\mathbf{N}}[k,n])$ is a block diagonal matrix, with the matrices $\mathbf{R}^t_k$ on its diagonal.

C. Sparse Formulation

To estimate the AoR $\boldsymbol{\theta}_{\mathrm{RU},k}$ and the ToA $\boldsymbol{\tau}_{\mathrm{RU},k}$ between the RIS and the UE for the k-th state, the problem can be formulated as the estimation of $\widehat{\boldsymbol{\rho}}_l[k,n]$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{spatial},k,l}$, since the locations of both the BS and RIS are usually fixed in most applications in practice. Hence the related angles such as $\boldsymbol{\theta}_{\mathrm{BR}}$, $\boldsymbol{\phi}_{\mathrm{BR}}$, and the ToA $\boldsymbol{\tau}_{\mathrm{BR}}$ can be determined in advance [20]. Let us introduce the DFT matrix $\mathbf{U}_{\mathrm{R}}$ for the sparse formulation of $\mathbf{a}(\widehat{\boldsymbol{\theta}}_{\mathrm{spatial},k,l})$ in (22). Hence, $\mathbf{U}_{\mathrm{R}}$ is defined in the same way as (12). Then, the received signal of (22) can be alternatively expressed as

$$\mathbf{Y}_{l,:}[k,n]=\boldsymbol{\Psi}_k\mathbf{h}_l[k,n]+\widehat{\mathbf{n}}[k,n],\tag{24}$$

where $\boldsymbol{\Psi}_k=\widehat{\boldsymbol{\Omega}}^{\mathrm{T}}_k\mathbf{U}_{\mathrm{R}}$ can be viewed as a sensing matrix, and $\mathbf{h}_l[k,n]$ is the desired sparse vector, which embeds the location information AoR and ToA for the k-th state. Based on (24), for each state k, the received signals have three distinguishing dimensions, namely time blocks, subcarriers, and possible propagation paths. For localization purposes, the path having the highest power represents the LoS path, which can be extracted from (22) via the power measurement, represented as

$$\widehat{l}=\arg\max\|\mathbf{Y}_{l,:}[k,n]\|^2_2,\tag{25}$$

which reduces the number of paths from $L_{\mathrm{RU}}L_{\mathrm{BR}}$ to 1. Consequently, when considering T blocks, N subcarriers, and $\widehat{l}$-th path components, corresponding to (24), the received signals can be expressed as

$$\widehat{\mathbf{Y}}_k=\boldsymbol{\Psi}_k\widehat{\mathbf{H}}_k+\mathbf{Z}_k,\tag{26}$$

where $\widehat{\mathbf{Y}}_k\in\mathbb{C}^{T\times N}$, $\boldsymbol{\Psi}_k\in\mathbb{C}^{T\times N_{\mathrm{R}}}$ and $\widehat{\mathbf{H}}_k\in\mathbb{C}^{N_{\mathrm{R}}\times N}$, while each column of $\mathbf{Z}_k$ reprensents the noise matrix. Finally, after the vectorization of (26), the received signal can be represented in a group sparse [44] format as

$$\widehat{\mathbf{y}}_k=\widehat{\boldsymbol{\Psi}}_k\widehat{\mathbf{h}}_k+\mathbf{z}_k.\tag{27}$$

In (27), we have $\widehat{\mathbf{y}}_k = \mathrm{vec}(\widehat{\mathbf{Y}}_k^{\mathrm{T}}) \in \mathbb{C}^{NT\times 1}$,$\widehat{\mathbf{\Psi}}_k = (\mathbf{\Psi}_k \otimes \mathbf{I}_N) \in \mathbb{C}^{NT\times NN_{\mathrm{R}}}$, and $\widehat{\mathbf{h}}_k = \mathrm{vec}(\widehat{\mathbf{H}}_k^{\mathrm{T}}) \in \mathbb{C}^{NN_{\mathrm{R}}\times 1}$. Furthermore, according to (22), the noise variance for the k-th state, t-th time block and n-th subcarrier can be expressed as $(\widehat{\sigma}_k^t)^2 = \sigma_k^2 \mathbf{W}_{:,\widehat{l}}^t[k,n]^{\mathrm{H}} \mathbf{W}_{:,\widehat{l}}^t[k,n]$. Hence, the covariance of the elements in $\mathbf{z}_k$ can be written as $\widehat{\mathbf{R}}_k^{t,t} = (\widehat{\sigma}_k^t)^2$, where $\widehat{\mathbf{R}}_k^{t,t}$ denotes the t-th diagonal element of $\widehat{\mathbf{R}}_k$. Therefore, in (27), the covariance matrix of the noise vector $\mathbf{z}_k$ can be expressed as $\widetilde{\mathbf{R}}_k = \mathbb{E}\left\{\mathbf{z}_k \mathbf{z}_k^{\mathrm{H}}\right\} = (\widehat{\mathbf{R}}_k \otimes \mathbf{I}_N)$.

Hence, (27) is formulated as an MMV channel estimation problem to estimate the AoR and the ToA for the k-th state. In the next section, we will detail the single-shot localization using the modified MMV-SBL algorithm to estimate the channel parameters.

D. Single-Shot Localization Using Modified MMV-SBL Algorithm

Unlike conventional methods such as orthogonal matching pursuit (OMP) and AMP, the proposed algorithm effectively captures the shared and structured sparsity inherent in the MMV model [22]. In addition, they provide higher robustness against noise, avoid reliance on strict assumptions such as near-Gaussian sensing matrices, and provide superior estimation accuracy, especially in complex propagation environments [45]. As mentioned in the previous subsection, the desired group-sparse vector $\widehat{\mathbf{h}}$ is sparse in both the time and spatial domain, which contains the location information represented by the AoR and ToA. The *a priori* information of the vector $\widehat{\mathbf{h}}_k$ for the k-th state can be mathematically expressed as

$$p\left(\widehat{\mathbf{h}}_k; \mathbf{\Gamma}_k, \mathbf{C}_k\right) = \prod_{j=1}^{N_R} p\left(\mathbf{h}_k^j; \gamma_k^j, \mathbf{C}_k\right), \tag{28}$$

where $\mathbf{C}_k$ denotes the correlation matrix of hyperparamters for the k-th motion state, γ_k^j is the hyperparameter controlling the variance of each group, while $\mathbf{h}_k^j$ denotes the j-th group of $\widehat{\mathbf{h}}_k$. Specifically, for a given γ_k^j and $\mathbf{C}_k$, the probability density function (PDF) of $\mathbf{h}_k^j$ is

$$p\left(\mathbf{h}_k^j; \gamma_k^j, \mathbf{C}_k\right) = \frac{1}{\left(\pi\gamma_k^j\right)^N \det\left(\mathbf{C}_k\right)} e^{\left(-\frac{(\mathbf{h}_k^j)^{\mathrm{H}} \mathbf{C}_k^{-1} \mathbf{h}_k^j}{\gamma_k^j}\right)}. \tag{29}$$

Explicitly, $\mathbf{h}_j$ can be generated on the basis of the hyperparameter γ_j. When considering all the N_{R} groups, we have a vector $\boldsymbol{\gamma}$ that controls the prior variance of the group of elements in $\mathbf{h}_k$. Let $\mathbf{\Gamma}_k = \mathrm{diag}(\boldsymbol{\gamma}_k) \in \mathbb{R}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ be a diagonal matrix with the hyperparameter vector $\boldsymbol{\gamma}_k$ on its diagonal.

Specifically, to solve our problem, the expected value for the $(i-1)$-st iteration of the log-likelihood function $\ell(\mathbf{\Gamma}_k, \mathbf{C}_k | \widehat{\mathbf{\Gamma}}_k^{i-1}, \widehat{\mathbf{C}}_k^{i-1})$, corresponding to the complete data $\{\widehat{\mathbf{y}}_k, \widehat{\mathbf{h}}_k\}$, is determined by the E-step formulated as

$$\begin{aligned}&\ell(\mathbf{\Gamma}_k, \mathbf{C}_k | \widehat{\mathbf{\Gamma}}_k^{i-1}, \widehat{\mathbf{C}}_k^{i-1}) \\ &\quad = \mathbb{E}_{\widehat{\mathbf{h}}_k | \widehat{\mathbf{y}}_k; \widehat{\mathbf{\Gamma}}_k^{i-1}, \widehat{\mathbf{C}}_k^{i-1}} \{\ln p(\widehat{\mathbf{y}}_k, \widehat{\mathbf{h}}_k; \mathbf{\Gamma}_k, \mathbf{C}_k)\}.\end{aligned} \tag{30}$$

Then, by employing Bayes' rule to (30), the M-step designed for maximizing the log-likelihood function $\ell(\mathbf{\Gamma}_k, \mathbf{C}_k | \widehat{\mathbf{\Gamma}}_k^{i-1}, \widehat{\mathbf{C}}_k^{i-1})$ to update $\mathbf{\Gamma}_k$ and $\mathbf{C}_k$ can be expressed as

$$(\widehat{\mathbf{\Gamma}}_k^i, \widehat{\mathbf{C}}_k^i) = \underset{\mathbf{\Gamma}_k, \mathbf{C}_k}{\arg\max}\, \mathbb{E}\{\ln p(\widehat{\mathbf{h}}_k; \mathbf{\Gamma}_k, \mathbf{C}_k)\}. \tag{31}$$

Then, based on (28) and (29), each γ_k^j is decoupled by the M-step, and according to [45], the *a posteriori* PDF for the i-th iteration is given by $p\left(\widehat{\mathbf{h}}_k | \mathbf{y}_k; \widehat{\mathbf{\Gamma}}_k^{i-1}, \widehat{\mathbf{C}}_k^{i-1}\right) \sim \mathcal{CN}\left(\boldsymbol{\mu}_k^i, \boldsymbol{\kappa}_k^i\right)$ with

$$\begin{aligned}\boldsymbol{\mu}_k^i =& \boldsymbol{\kappa}_k^i \widehat{\mathbf{\Psi}}_k^{\mathrm{H}} \widetilde{\mathbf{R}}_k^{-1} \widehat{\mathbf{y}}_k, \\ \boldsymbol{\kappa}_k^i =& \left[(\widehat{\mathbf{\Gamma}}_k^{i-1} \otimes \widehat{\mathbf{C}}_k^{i-1})^{-1} + \widehat{\mathbf{\Psi}}_k^{\mathrm{H}} \widetilde{\mathbf{R}}_k^{-1} \widehat{\mathbf{\Psi}}_k\right]^{-1}.\end{aligned} \tag{32}$$

To reduce complexity, we adopt the simplified covariance approximation of [22], [45], which yields

$$\boldsymbol{\kappa}_k^i \approx ((\widehat{\mathbf{\Gamma}}_k^{i-1})^{-1} + \mathbf{\Psi}_k^{\mathrm{H}} \widehat{\mathbf{R}}_k^{-1} \mathbf{\Psi}_k)^{-1} \otimes \widehat{\mathbf{C}}_k^{i-1}, \tag{33}$$

and the posterior mean efficiently is updated as

$$\begin{aligned}\boldsymbol{\mu}_k^i \approx& [(\widehat{\boldsymbol{\kappa}}_k^i \mathbf{\Psi}_k^{\mathrm{H}} \widehat{\mathbf{R}}_k^{-1}) \otimes \mathbf{I}_N] \mathrm{vec}(\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{T}}) \\ =& \mathrm{vec}([\widetilde{\widehat{\mathbf{H}}}_{k,i}]^{\mathrm{T}}),\end{aligned} \tag{34}$$

where $\widetilde{\widehat{\mathbf{H}}}_{k,i} = \widehat{\boldsymbol{\kappa}}_k^i \mathbf{\Psi}_k^{\mathrm{H}} \widehat{\mathbf{R}}_k^{-1} \widehat{\mathbf{Y}}_k$. By substituting the approximations of (33) and (34) into (32), the hyperparameter $\gamma_k^{j,i}$ can be expressed as

$$\gamma_k^{j,i} = [\widehat{\boldsymbol{\kappa}}_k^i]_{j,j} + \frac{1}{N} ([\widetilde{\widehat{\mathbf{H}}}_k^i]_{j,:})^{\mathrm{H}} (\widehat{\mathbf{C}}_k^{i-1})^{-1} [\widetilde{\widehat{\mathbf{H}}}_k^i]_{j,:}, \tag{35}$$

where the correlation matrix $\widehat{\mathbf{C}}_k^i$ is represented as

$$\begin{aligned}\widetilde{\mathbf{C}}_k^i =& \sum_{j=1}^{N_{\mathrm{R}}} \frac{1}{\gamma_k^{j,i}} [\widetilde{\widehat{\mathbf{H}}}_k^i]_{j,:} ([\widetilde{\widehat{\mathbf{H}}}_k^i]_{j,:})^{\mathrm{H}} + \eta \mathbf{I}_N, \\ \widehat{\mathbf{C}}_{iter} =& \frac{\widetilde{\mathbf{C}}_k^i}{\left\|\widetilde{\mathbf{C}}_k^i\right\|_F},\end{aligned} \tag{36}$$

with η having a constant. The estimation of the desired vector $\widetilde{\widehat{\mathbf{h}}}_k$ is given by the converged *a posteriori* mean $\boldsymbol{\mu}_k^i$. To recover the estimated channel vector into matrix for the k-th state, we have $\widehat{\mathbf{H}}_k = \mathrm{reshape}(\widehat{\mathbf{h}}_k, N) \in \mathbb{C}^{N_{\mathrm{R}}\times N}$. After obtaining the channel parameters, the location of UE can be recovered as shown in Fig. 2c.

Delay estimation: the estimated channel gain $\widetilde{\widehat{\boldsymbol{\rho}}}_{k,\widehat{l}}$ and the ToA $\widehat{\tau}_{\mathrm{RU},k,\widehat{l}}$ for the $\widehat{l}$-th path is recovered as

$$\begin{aligned}\widetilde{\widehat{\boldsymbol{\rho}}}_{k,\widehat{l}} =& \max(\widetilde{\widehat{\mathbf{H}}}_k, 2) \in \mathbb{C}^{1\times N}, \\ \widehat{\tau}_{\mathrm{BRU},k,\widehat{l}} =& \underset{\tau_{\mathrm{BRU},k,l}}{\arg\max} \left|\mathbf{g}(\tau_{\mathrm{BRU},k,\widehat{l}}) \widetilde{\widehat{\boldsymbol{\rho}}}_{k,\widehat{l}}\right|^2 \\ \widehat{\tau}_{\mathrm{RU},k,\widehat{l}} =& \widehat{\tau}_{\mathrm{BRU},k,\widehat{l}} - \tau_{\mathrm{BR},\widehat{l}},\end{aligned} \tag{37}$$

where $\mathbf{g}(\tau_{\mathrm{BRU},k,\widehat{l}}) = \left[1, ..., e^{-j2\pi(N-1)\tau_{\mathrm{BRU},k,\widehat{l}}/(NT_s)}\right]^{\mathrm{T}}$ represents the on-grid candidates of the delay between the BS and the UE, and $\widehat{\tau}_{\mathrm{BRU},k,\widehat{l}}$ denotes the total delay between the BS and the UE, while $\tau_{\mathrm{BR},\widehat{l}}$ is assumed to be known,

because again, the locations of the BS and RIS are normally assumed to be fixed in DL transmission [20], [46].

AoR estimation: The spatial frequency $\widehat{\widetilde{\boldsymbol{\theta}}}_{\text{spatial},k,\widehat{l}}$ for the $\widehat{l}$-th path can be expressed as

$$\begin{aligned}\widehat{s} &= \arg\max_{s=1,\ldots,N_R} \sum_{n=1}^{N} \widehat{\widetilde{\mathbf{H}}}_{:,n},\\ \widehat{\theta}_{\text{spatial},k,\widehat{l}} &= \arcsin\left(\frac{\lambda_c}{d}\frac{\widehat{s}-(N_R-1)/2-1}{N_R}\right),\\ \widehat{\theta}_{\text{RU},k,\widehat{l}} &= \arcsin[\sin(\phi_{\text{BR},\widehat{l}}) - \sin(\widehat{\theta}_{\text{spatial},k,\widehat{l}})]. \end{aligned} \tag{38}$$

Finally, the location of the UE for the k-th state can be recovered as

$$\widehat{\mathbf{u}}_k = \mathbf{r} + c\widehat{\tau}_{\text{RU},k,1}[\cos(\widehat{\theta}_{\text{RU},k,1}), \sin(\widehat{\theta}_{\text{RU},k,1})]^{\text{T}}, \tag{39}$$

where $\mathbf{r}$ is the location of the RIS, as the reference node of localization. In successive motion states, directly applying single-shot localization at each time step can lead to quantization errors in both the angular and time domains. Therefore, we incorporate a Kalman filter to perform temporal smoothing of the position estimates by exploiting the motion dynamics of the user. Note that we do not aim for improving the Kalman filter itself, but rather for the nontrivial construction of the measurement model for the Kalman filter in the RIS-aided mmWave localization scenario. When all the estimates of the K states of $\widehat{\mathbf{u}}_k$ are collected, the Kalman filtering can be used for improving trajectory recovery, which will be detailed in the next section.

E. Cramer-Rao Lower Bounds

In this subsection, we obtain the CRLBs of θ_{RU} and τ_{RU} for the k-th state, and also the PEB of the UE. Based on (22), the log-likelihood function is

$$\ln p(\mathbf{Y} \mid \mathbf{u}_k) \propto -\sum_{l,n}^{L,N} \left\| \mathbf{Y}_{l:}[k,n] - \hat{\boldsymbol{\Omega}}_k \hat{\rho}_l[k,n] \mathbf{a}(\theta_{\text{spatial},k,l}) \right\|^2_{\mathbf{R}_k^{-1}}, \tag{40}$$

where we have $L = L_{\text{BR}}L_{\text{RU}}$, $\|\mathbf{x}\|_{\mathbf{R}_k^{-1}} = \mathbf{x}^{\text{H}}\mathbf{R}_k^{-1}\mathbf{x}$. Furthermore, the Fisher information matrix (FIM), given $L_{\text{RU}}L_{\text{BR}}$ and the number of subcarriers N, can be defined as

$$\mathbf{I} = \sum_{l,n}^{L,N} \frac{2}{\sigma^2} \Re\left\{\mathbf{J}_{l,n}^{\text{H}} \mathbf{J}_{l,n}\right\}, \tag{41}$$

where we have $\mathbf{J}_{l,n} = \left[\frac{\partial \boldsymbol{\mu}_{l,n}}{\partial u_x^k}, \frac{\partial \boldsymbol{\mu}_{l,n}}{\partial u_y^k}\right]$, and $\boldsymbol{\mu}_{l,n}$ is the noiseless version of (22). The derivation of $\mathbf{J}_{l,n}$ is detailed in Appendix A. The PEB is defined as

$$\text{PEB} = \sqrt{\text{Trace}(\mathbf{I})^{-1}}. \tag{42}$$

IV. Kalman Filter with Bias-Compensated Localization Measurements

In this section, initialization at $k = 0$ is based on the method of Section III. Then, Kalman filtering is employed to improve the performance of the single-shot localization for trajectory estimation, based on the statistical characterization of the noise distribution in the measurement model of Section III and the process noise distribution in the state evolution model.

The procedure of the proposed three-stage tracking scheme is shown in Fig. 3. The Kalman filter-based approach

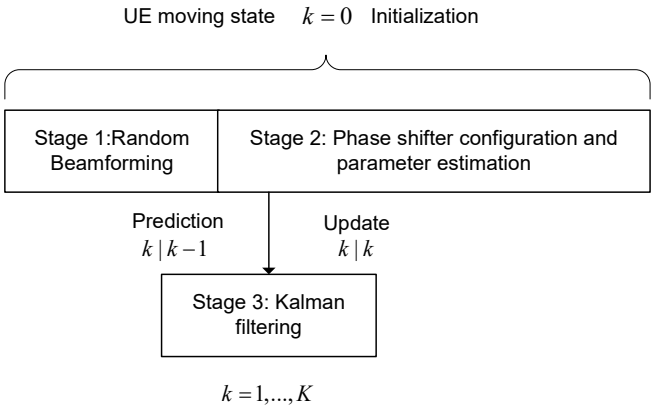


FIGURE 3: Illustration of the proposed three-stage tracking scheme.

used for tracking the user for $k > 0$, where we consider the state transition model with fixed velocity, may be expressed as

$$\mathbf{x}_k = \mathbf{T}\mathbf{x}_{k-1} + \mathbf{q}_k, \tag{43}$$

where the state variables $\mathbf{x}_k = [u_x^k, v_x^k, u_y^k, v_y^k]^{\text{T}}$, and the process noise lead to variations in both direction and speed, defined as $\mathbf{q}_k \sim \mathcal{CN}(0, \sigma_{q_i}^2)$. As introduced in Section II, the Cartesian coordinate of the UE is represented as $\mathbf{u}_k = [u_x^k, u_y^k]^{\text{T}} \in \mathbb{R}^2$. Furthermore, v_x^k and v_y^k denote the velocity of the UE along the x-axis and y-axis, respectively. The state transition matrix $\mathbf{T}$ is given by[5]

$$\mathbf{T} = \begin{bmatrix} 1 & \Delta T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta T \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{44}$$

In (44), ΔT is the interval of input samples. On the other hand, the measurement variable is given by $\widehat{\mathbf{x}}_k = [\widehat{u}_x^k, \widehat{v}_x^k, \widehat{u}_y^k, \widehat{v}_y^k]^{\text{T}}$. At each time step, the position is estimated via single-shot localization based on (39), which is denoted as $\hat{\mathbf{u}}_k = [\hat{u}_x^k, \hat{u}_y^k]^{\text{T}}$. Due to the use of discrete delay-angle grids and practical estimation errors, this estimate contains both a systematic bias $\mathbf{b}_k$ and measurement noise $\mathbf{r}_k$, which can be modeled as

$$\widetilde{\mathbf{x}}_k = \mathbf{x}_k + \mathbf{b}_k + \mathbf{r}_k. \tag{45}$$

To enable standard Kalman filtering, which assumes zero-mean Gaussian noise, we apply an offline bias compensation strategy. The average estimation bias $\mathbf{b}_k$ is statistically estimated from Section III. This bias is subtracted from $\hat{\mathbf{u}}_k$.

[5]In this paper, we assume the low-speed pedestrian velocity is considered in indoor scenario, where the abrupt speed change (acceleration or turn) is not taken into consideration, which implies less complicated trajectory is considered and requires no sophisticated filter, such as interacting multiple model (IMM) [47].

Therefore, the Kalman filter assumes a linear measurement model [6] formulated as follows

$$\widetilde{\mathbf{x}}_k = \mathbf{M}_k \mathbf{x}_k + \mathbf{r}_k, \tag{46}$$

where the measurement matrix is $\mathbf{M}_k = \mathbf{I}_4$ and the measurement noise is expressed as $\mathbf{r}_k \sim \mathcal{CN}(0, \sigma_r^2)$. Alternatively, when only the measurements of locations are considered in (46), the measurement matrix can be further written as

$$\mathbf{M}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{47}$$

In the tracking process, at the first state of $k = 0$, we estimate $\mathbf{x}_{[0|0]}$ as the initial location of the UE via (39). Then the *a priori* prediction can be formulated as

$$\widehat{\mathbf{x}}_{k|k-1} = \mathbf{T}\widehat{\mathbf{x}}_{k-1|k-1}, \tag{48}$$

where $\widehat{\mathbf{x}}_{k-1|k-1}$ is the *a posteriori* prediction gleaned from the previous state. Thus, the estimated covariance matrix is updated as

$$\mathbf{P}_{k|k-1} = \mathbf{T}\mathbf{P}_{k-1|k-1}\mathbf{T}^{\mathrm{T}} + \mathbf{Q}, \tag{49}$$

where $\mathbf{Q} = \mathrm{diag}(\sigma_{q_1}^2, \sigma_{q_2}^2, \sigma_{q_3}^2, \sigma_{q_4}^2)$. Based on the prediction stage in (48), the Kalman gain is updated as

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}(\mathbf{M}_k)^{\mathrm{T}} \left[\mathbf{M}_k \mathbf{P}_{k|k-1}(\mathbf{M}_k)^{\mathrm{T}} + \mathbf{V}_k\right]^{-1}, \tag{50}$$

where the measurement noise covariance matrix is given by[7] $\mathbf{V}_k = \mathrm{diag}(\sigma_{r_{k,1}}^2, \sigma_{r_{k,2}}^2)$. Accordingly, the prediction is updated as

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left(\widetilde{\mathbf{x}}_k - \mathbf{M}_k \widehat{\mathbf{x}}_{k|k-1}\right), \tag{51}$$

and the updated covariance matrix is expressed as

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k \mathbf{M}_k\right) \mathbf{P}_{k|k-1}. \tag{52}$$

Finally, we define the smoothed trajectory $\{\hat{\mathbf{u}}_{k|k}\}_{k-1}^{K}$

$$\mathbf{U}_{\mathrm{KF}} = [\hat{\mathbf{u}}_{1|1}, \hat{\mathbf{u}}_{2|2}, \ldots, \hat{\mathbf{u}}_{K|K}]. \tag{50}$$

The proposed three-stage tracking is summarized in Algorithm 1.

Algorithm 1 Three-stage tracking scheme

Inputs: Observations $\widehat{\mathbf{Y}}_k$; Sensing matrix $\mathbf{\Psi}_k$; Maximum iteration ζ. Process noise covariance $\mathbf{Q}$ Measurement noise covariance $\mathbf{V}_k$

Objectives: To estimate $\widehat{\mathbf{u}}_k$.

for $k = 0, ..., K$ **do**
  if $k < 1$ **then**
    Initialize hyperparameters: $\widehat{\mathbf{\Gamma}}_k^0 = \mathbf{I}_{N_\mathrm{R}}$, $\widehat{\mathbf{\Gamma}}_k^{-1} = \mathbf{0}_{N_\mathrm{R}}$
    for $i = 0, ..., \zeta$ **do**
      Expectation: Evaluate the *a postriori* mean $\boldsymbol{\mu}_k^i$ and covariance matrix $\boldsymbol{\kappa}_k^i$ according to (34) and (33), respectively.
      Maximization: Update the hyperparameters $\gamma_k^{j,i}$ based on (35), and the correlation matrix $\widehat{\mathbf{C}}_k^i$ based on (36), with $\widehat{\mathbf{\Gamma}}_{iter} = \mathrm{diag}(\gamma_{1,iter}, \gamma_{2,iter}, ..., \gamma_{N_\mathrm{R},iter})$.
    end for
    ToA and AoR estimation based on (37) and (38), respectively.
    Kalman filter initialization $\widehat{\mathbf{u}}_{k=0}$ is obtained by (39).
  else
    Prediction procedures based on (48) and (49).
    Update procedures based on (50) to (52).
  end if
end for
return $\widehat{\mathbf{u}}_k$ from $\widehat{\mathbf{x}}_{k|k}$.

V. Performance Results

In this section, we consider the system of Fig. 1, where both the BS and the UE employing 16 antennas and the RIS has 16 elements. The number of time blocks is set to $T = 32$, while the number of pilot subcarriers is $N = 20$. The locations of the BS and the RIS are at $\mathbf{b} = [0, 0]^{\mathrm{T}}$ and $\mathbf{r} = [2, 10]^{\mathrm{T}}$ in order to investigate either an unreasonable RIS deployment or $\mathbf{r} = [2, 20]^{\mathrm{T}}$ for the other cases, respectively. Furthermore, the location of the first scatterer is fixed at $\mathbf{s}_1 = [1, 3]$, while the second scatterer $\mathbf{s}_2$ is randomly located between the RIS and the UE. The carrier frequency f_c is set to 28 GHz, while the bandwidth is B = 100 MHz, unless specified otherwise. The SNR is defined based on the reconstructed signal energy using the sensing matrix and sparse coefficients. Specifically, SNR (dB) $= 10\log_{10}(\frac{\left\|\hat{\mathbf{\Psi}}_k \hat{\mathbf{h}}_k\right\|^2}{\sigma_k^2})$. Finally, the estimation performance is characterized by the root mean squared error (RMSE), defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\mathbf{q}} - \mathbf{q}\|_2^2}, \tag{53}$$

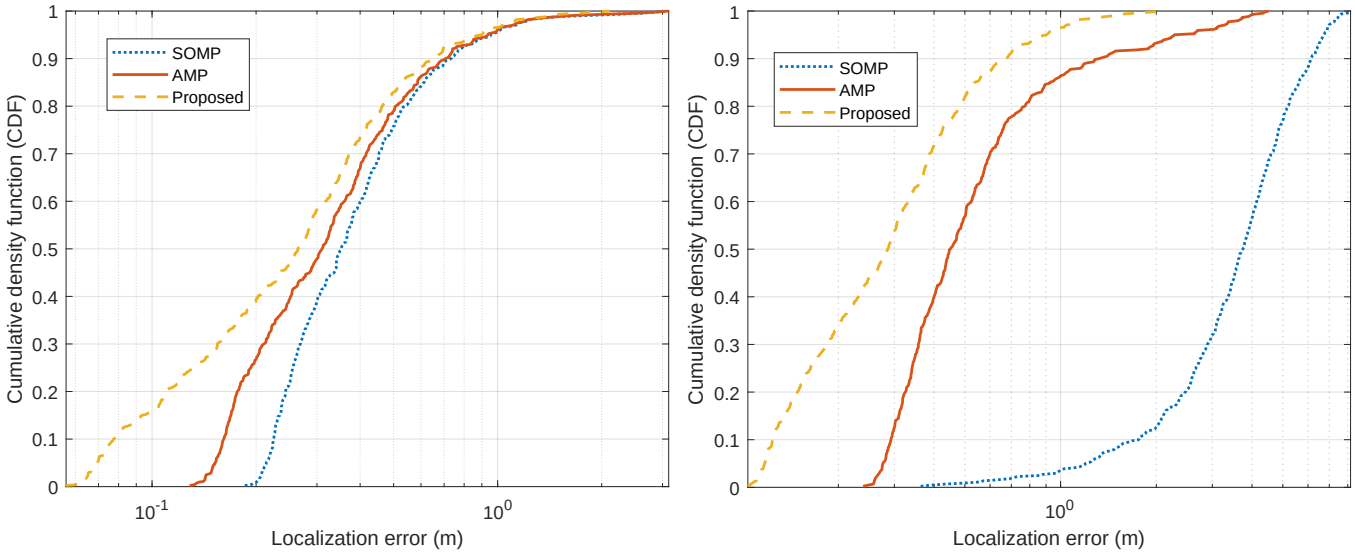
where K denotes the number of Monte Carlo trials, while $\mathbf{q}$ and $\widehat{\mathbf{q}}$ are the true and estimated UE locations, respectively.

A. Single-Shot Localization Performance

In the simulation results of single-shot localization, the location of the RIS is set to $\mathbf{r} = [2, 20]^{\mathrm{T}}$, and the UE is moving in a 20×20-meter indoor scenario, as shown in Fig. 1. In Fig. 4, we compare the CDF of the localization error at different SNR values using different compressed sensing algorithms. From the simulation results, we can conclude that the proposed method is more accurate and robust than the SMV based approximate message passing algorithm (AMP) and the MMV based simultaneous orthogonal matching pursuit (SOMP) algorithm. Moreover, it can be observed that the proposed algorithm outperforms both the SOMP and AMP algorithms in mmWave channel estimation due to its ability to exploit the temporal correlation, effectively handled by the MMV structure, maintain robustness in low SNR conditions, and achieve stable convergence at a moderate computational complexity. Explicitly, the maximum/minimum localization

[6] Note that the observation model can be assumed to be non-linear, which can be solved by the extended Kalman filter that has been studied in [29], [48].

[7] The noise distribution of $\mathbf{V}_k$ for the k-th state can be obtained based on offline statistics using the method detailed in Section III.



(a) SNR = 0 dB

(b) SNR = -5 dB

FIGURE 4: CDF of the UE's estimation error using different channel estimation algorithms at SNR = 0 and -5 dB.

error of the proposed algorithm is significantly lower than that of the rest.

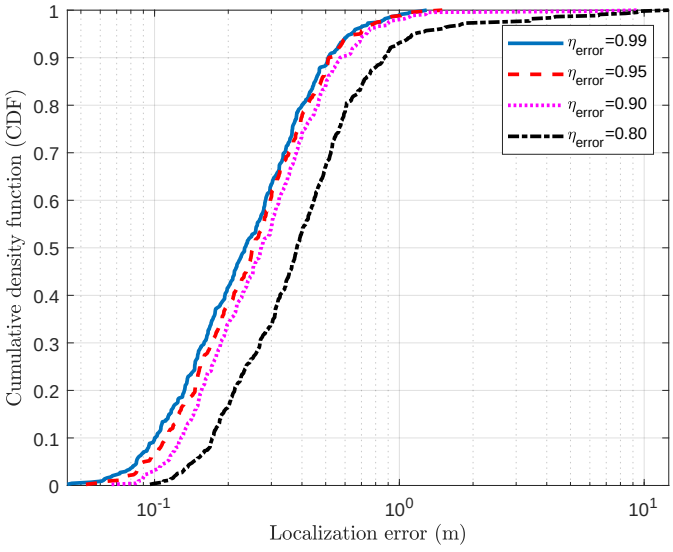


FIGURE 5: CDF of localization error under different beam misalignment levels η_{error} at SNR=5 dB and $T = 16$.

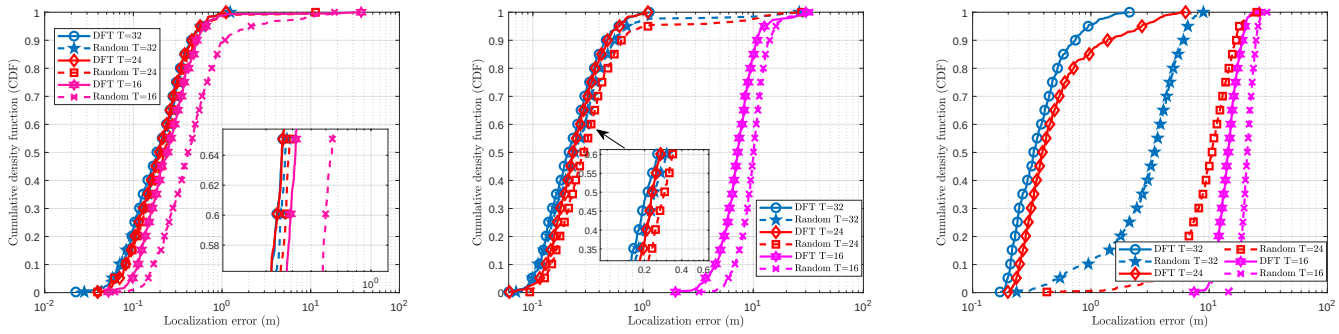
To assess this impact, we characterized various levels of misalignment values $\eta_{\text{error}} = \{0.99, 0.95, 0.90, 0.80\}$, and evaluated their effect on the localization performance. Fig. 5 shows that the proposed method retains similar accuracy even when the beamforming misalignment factor is $\eta_{\text{error}} = 0.90$, which confirms that the proposed method maintains robust performance under moderate beamforming inaccuracies.

In Fig. 6, we compare the performance of DFT-based and random RIS phase shift matrices for different SNR levels and different number of time blocks. Fig. 6 presents the CDF of localization error for SNR values of -10, -5 and 0 dB, and $T = \{16, 24, 32\}$. From the simulation results, we can find that the random configuration causes severe degradation due to the high column-correlation in the sensing matrix. When the number of time blocks increases, the performance gap between the DFT-based and random-based configurations reduces, since more measurements help average the randomness. On the other hand, Fig. 6 shows that the DFT configuration consistently yields lower localization error and steeper CDF transitions, reflecting more reliable support recovery.

Fig. 7 compares the CDF of localization errors for the proposed scheme and a benchmark IFFT-based pulse estimator [49]. It can be seen in Fig. 7 that the proposed scheme substantially outperforms the benchmark, which is due to the fact that the resolution of the IFFT-based method is limited to $1/B$, where B refers to the system bandwidth. By contrast, the proposed scheme relies on the super-resolution delay steering vector $\mathbf{g}(\tau)$ of (37), which is constructed from an overcomplete delay dictionary with arbitrarily fine spacing.

In Fig. 8, we can observe that as the distance between the RIS and the UE increases, the PEB also increases, as expected. Furthermore, it can be observed that the maximum error occurs when the reflected angle is 0 degrees, which leads to a blind spot. Moreover, the minimum PEB occurs when the UE is near the RIS.

In Fig. 9, the angular resolution G_{R} is set to 16, 32, 64, 96, 128, corresponding to the angles of 22.5, 11.25, 5.625, 3.75 and 2.8125 degrees. Fig. 9 presents the cumu-



(a) SNR = 0 dB (b) SNR = -5 dB (c) SNR = -10 dB

FIGURE 6: CDF of the UE's estimation error using different RIS phase shift configurations, including the DFT-based and random-based configurations, when the number of time blocks $T = 16, 24, 32$ for different SNR values.

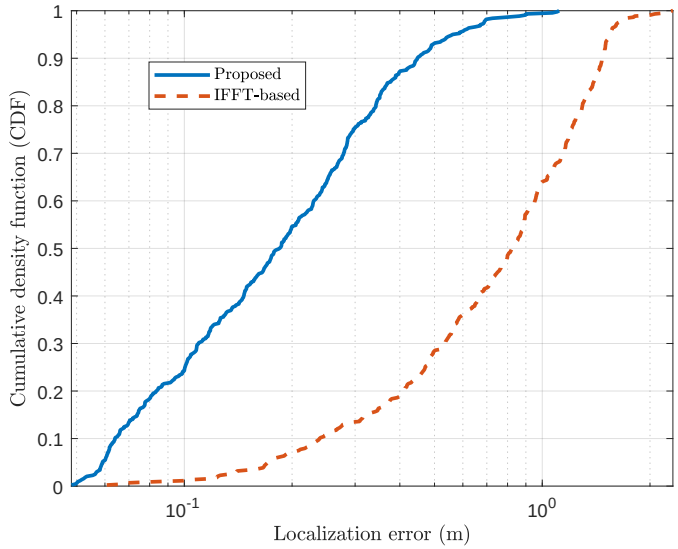


FIGURE 7: CDF of localization error under different ToA estimation methods, e.g., channel estimation based and IFFT based pulse estimation, when SNR=5 dB and $T = 32$.

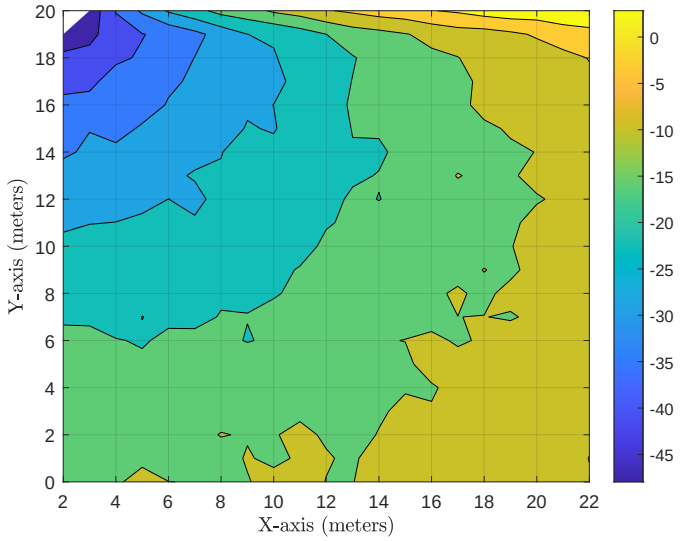


FIGURE 8: PEB (dB) distribution of the UE's location, at SNR = 0 dB.

lative distribution function (CDF) of single-shot localization accuracy under varying angular resolutions. As the angular resolution increases, the localization error decreases, resulting in higher accuracy. The steeper CDF curves highlight the significant advantage of high angular resolution in reducing the localization errors. This suggests that optimizing the antenna arrays or signal processing can effectively enhance localization performance.

B. Kalman Filtered Tracking Performance

1) Scenario 1 - Tracking for ideal straight movement

In Fig. 10, the tracking trajectories of the ground truth, single-shot localization, and the proposed tracking method are compared, when the UE follows a straight movement pattern. This analysis evaluates the effectiveness of the Kalman filter in refining the trajectory estimation. Observe in Fig. 10 that the single-shot localization method exhibits noticeable deviations from the ground truth caused by the angular and temporal quantization errors, making the estimated trajectory less reliable for continuous tracking. By contrast, the Kalman filter-based tracking method provides a smoother and more accurate trajectory. It significantly reduces sudden deviations, ensuring that the estimated path remains close to the true movement.

Fig. 11 investigates the tracking performance, when the RIS is disadvantageously placed. In this case, the UE moves to a position behind the RIS, which can only cover half the plane. In Fig. 11, we compare the estimated trajectories of single-shot localization and Kalman filtering against the true path. It can be observed that the single-shot localization results show significantly higher errors, which implies that the lack of strong reflected paths leads to poor localization accuracy. Kalman filtering succeeds in mitigating the errors, but still fails to fully compensate for the poor RIS placement.

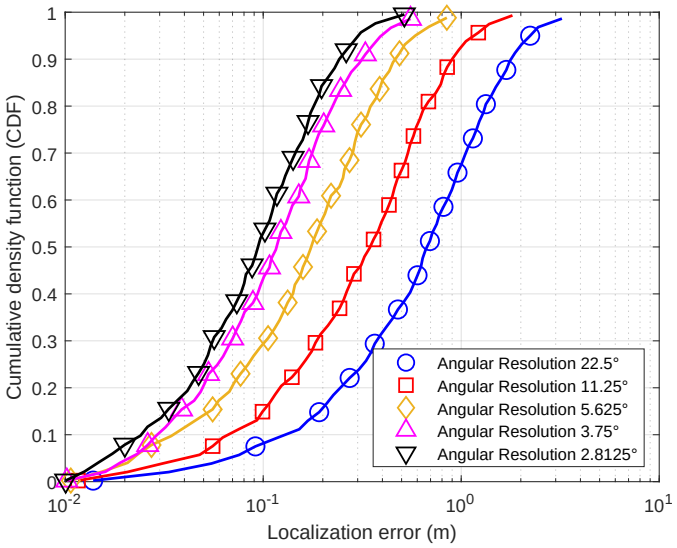


FIGURE 9: CDF of single-shot localization accuracy for $N_{\mathrm{B}} = N_{\mathrm{U}} = N_{\mathrm{R}} = 16$ and $T = 32$, at different angular resolutions.

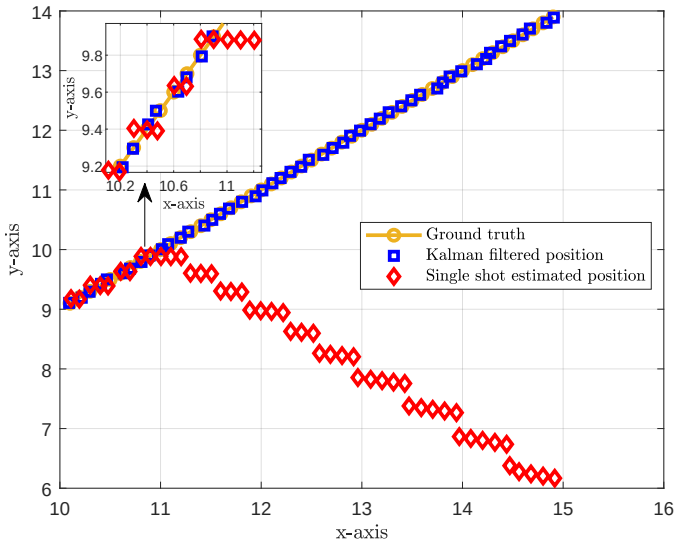


FIGURE 11: The trajectory of the ground truth versus the single-shot localization and the proposed method, when the UE's trajectory is behind the RIS.

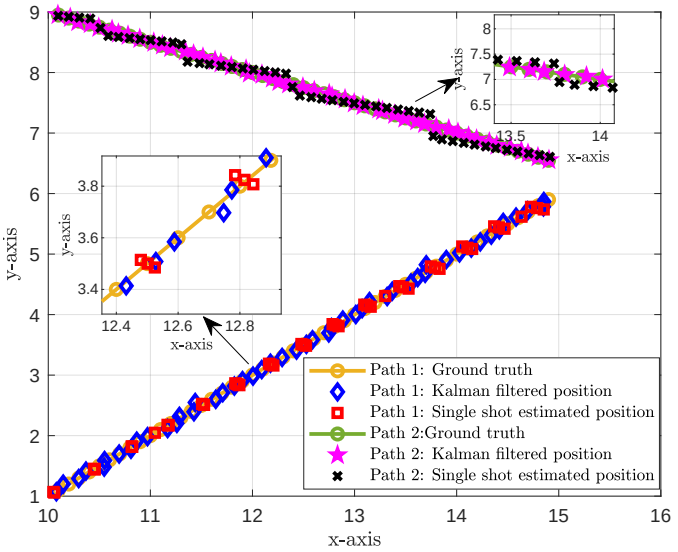


FIGURE 10: The trajectory of the ground truth versus the single-shot localization and the proposed method, when $N_{\mathrm{B}} = N_{\mathrm{U}} = N_{\mathrm{R}} = 16$, $K = 50$ and $T = 32$ for straight movement.

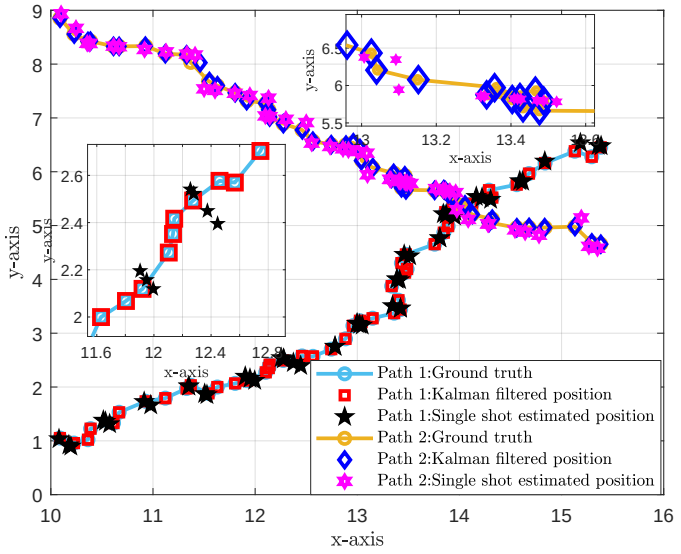


FIGURE 12: The trajectory of the ground truth versus the single-shot localization and the proposed method in arbitrary tracking, when $N_{\mathrm{B}} = N_{\mathrm{U}} = N_{\mathrm{R}} = 16$, $K = 50$ and $T = 32$.

The estimated trajectory still deviates from the ground truth, indicating that RIS positioning plays a crucial role in tracking performance. To further enhance performance, simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RISs) [50]–[52] and more robust tracking models can be employed, which will be part of our future research.

2) Scenario 2 - Tracking of arbitrary tracks

Fig. 12 compares the ground truth trajectories, single-shot localization, and the proposed tracking method for two different trajectories. The single-shot localization shows noticeable deviations, especially in dynamically fluctuating scenarios. By contrast, the Kalman filter-based tracking method is closely aligned with the ground truth. This demonstrates the higher robustness of the proposed method in dynamic environments, which is able to improve the tracking accuracy compared to single-shot localization.

Fig. 13 compares the ground truth trajectories, single-shot localization, and Kalman filter-based tracking under different

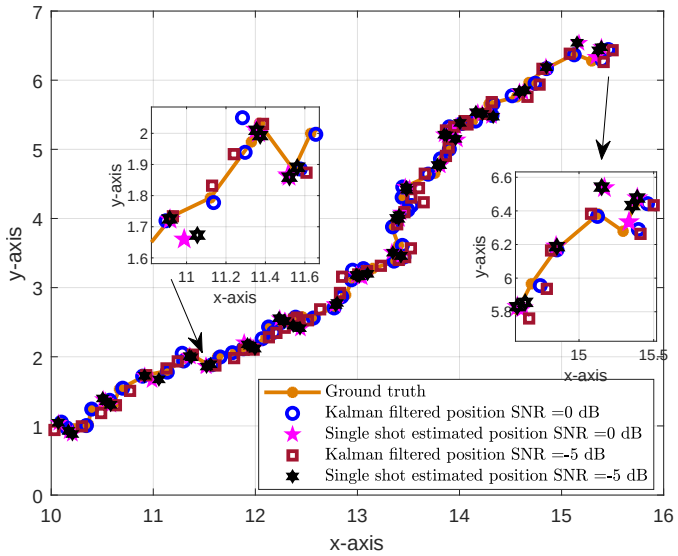


FIGURE 13: Investigation of the impact of different SNRs: the trajectory of the ground truth versus the single-shot localization and the proposed method in arbitrary tracking, when $N_{\mathrm{B}} = N_{\mathrm{U}} = N_{\mathrm{R}} = 16$, $K = 50$ and $T = 32$.

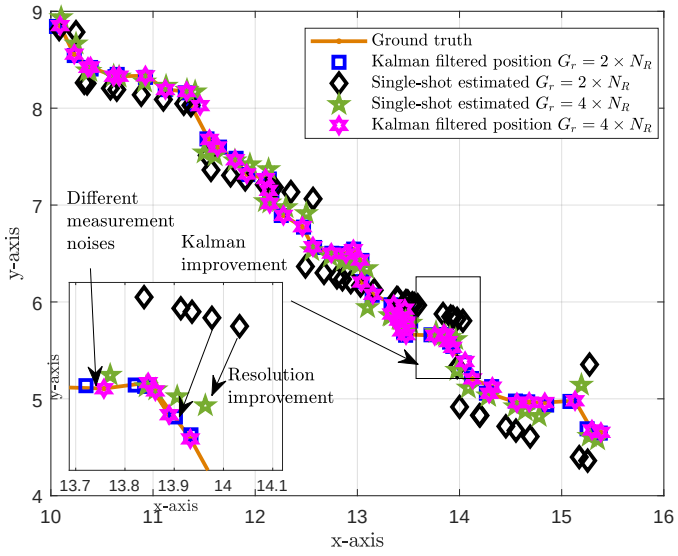


FIGURE 14: Investigation of the impact of different beamspace resolutions: the trajectory of the ground truth versus the single-shot localization and the proposed method in arbitrary tracking, when $N_{\mathrm{B}} = N_{\mathrm{U}} = N_{\mathrm{R}} = 16$, $K = 50$ and $T = 32$.

SNR conditions. The study examines how different noise levels affect the localization accuracy. At higher SNR levels, both single-shot localization and Kalman tracking closely follow the ground truth. The tracking error remains minimal, showing that as expected, a stronger signal facilitates more accurate positioning. At lower SNR levels, single-shot localization experiences larger deviations from the true trajectory. The Kalman filter mitigates some of the noise but still struggles to fully compensate for the degraded measurements. Nonetheless, the Kalman filter outperforms single-shot localization in all cases, demonstrating its effectiveness in smoothing noisy measurements.

Fig. 14 investigates the influence of different beamspace resolutions on the tracking performance. It compares Kalman filtering and single-shot localization using two different resolutions. Observe from Fig. 14 that the higher beamspace resolution improves tracking accuracy. The trajectories associated with increased resolution are aligned more closely with the ground truth. On the other hand, a lower resolution results in higher localization errors, especially for single-shot localization. The reduced spatial resolution makes it harder to precisely estimate user positions. Moreover, we may conclude that Kalman filtered tracking outperforms single-shot localization in both cases, but the performance gap narrows at higher resolutions.

VI. Conclusions

In this paper, we addressed the problem of UE trajectory tracking in an indoor environment having multiple random scatterers. We proposed a three-stage localization and tracking method using RIS-aided localization and a enchanced MMV-mSBL algorithm. A structured RIS phase shift design relying on a DFT matrix was introduced to resolve the AoR path crossing issue. To reduce complexity, we exploited the frequency diversity for path selection, simplifying the AoR and ToA estimation. A three-stage tracking scheme was designed for mitigating the training overhead. It relied on offline training for measurement noise distribution, random beamforming for channel sounding, and Kalman filtering for trajectory tracking on simulations showed that the method proposed outperforms SMV-AMP and MMV-SOMP, especially at low SNRs. The PEB derived validated its accuracy, and different movement scenarios were analyzed. Our framework improved thelocalization robustness and efficiency, offering a promising solution for indoor UE tracking.

REFERENCES

[1] X. Cheng, F. Shu, Y. Li, Z. Zhuang, D. Wu, and J. Wang, "Optimal measurement of drone swarm in RSS-based passive localization with region constraints," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 1–11, 2022.

[2] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[3] J. Lota, S. Ju, O. Kanhere, T. S. Rappaport, and A. Demosthenous, "MmWave V2V localization in MU-MIMO hybrid beamforming," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 210–220, 2022.

[4] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 870–913, 2018.

[5] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[6] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE*

Transactions on Vehicular Technology, vol. 69, no. 3, pp. 2943–2956, 2020.

[7] A. A. Nasir, H. D. Tuan, E. Dutkiewicz, and L. Hanzo, "Finite-resolution digital beamforming for multi-user millimeter-wave networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9647–9662, 2022.

[8] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[9] R. Kaur, B. Bansal, S. Majhi, S. Jain, C. Huang, and C. Yuen, "A survey on reconfigurable intelligent surface for physical layer security of next-generation wireless communications," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 172–199, 2024.

[10] Z. Chen, X. Ma, C. Han, and Q. Wen, "Towards intelligent reflecting surface empowered 6G Terahertz communications: A survey," *China Communications*, vol. 18, no. 5, pp. 93–119, 2021.

[11] Y. Dong, Q. Li, S. X. Ng, and M. El-Hajjar, "Reconfigurable intelligent surface relying on low-complexity joint sector non-diagonal structure," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 1106–1123, 2024.

[12] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 883–917, 2022.

[13] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[14] H. Wymeersch, J. He, B. Denis, A. Clemente, and M. Juntti, "Radio localization and mapping with reconfigurable intelligent surfaces: Chall enges, opportunities, and research directions," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 52–61, 2020.

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[16] K. Li, M. El-Hajjar, and L.-L. Yang, "Reconfigurable intelligent surface aided position and orientation estimation based on joint beamforming with limited feedback," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 748–767, 2023.

[17] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7830–7833.

[18] Z. Wang, Z. Liu, Y. Shen, A. Conti, and M. Z. Win, "Location awareness in beyond 5G networks via reconfigurable intelligent surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2011–2025, 2022.

[19] M. Z. Win, Z. Wang, Z. Liu, Y. Shen, and A. Conti, "Location awareness via intelligent surfaces: A path toward holographic NLN," *IEEE Vehicular Technology Magazine*, vol. 17, no. 2, pp. 37–45, 2022.

[20] J. He, H. Wymeersch, T. Sanguanpuak, O. Silvén, and M. Juntti, "Adaptive beamforming design for mmWave RIS-aided joint localization and communication," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2020, pp. 1–6.

[21] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Towards ubiquitous positioning by leveraging reconfigurable intelligent surface," *IEEE Communications Letters*, vol. 25, no. 1, pp. 284–288, 2020.

[22] K. Li, J. He, M. El-Hajjar, and L.-L. Yang, "Localization in reconfigurable intelligent surface aided mmWave systems: A multiple measurement vector based channel estimation method," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 9, pp. 13 326–13 339, 2024.

[23] Z. Du, F. Liu, W. Yuan, C. Masouros, Z. Zhang, S. Xia, and G. Caire, "Integrated sensing and communications for V2I networks: Dynamic predictive beamforming for extended vehicle targets," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 3612–3627, 2022.

[24] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *2014 IEEE Globecom Workshops (GC Wkshps)*, 2014, pp. 416–421.

[25] F. Liu, P. Zhao, and Z. Wang, "EKF-based beam tracking for mmWave MIMO systems," *IEEE Communications Letters*, vol. 23, no. 12, pp. 2390–2393, 2019.

[26] S. Sen and A. Nehorai, "Sparsity-based multi-target tracking using OFDM radar," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1902–1906, 2010.

[27] Y. Shi and Y. Huang, "Integrated sensing and communication-assisted user state refinement for OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 2, pp. 922–936, 2023.

[28] H. Luo, F. Gao, F. Liu, and S. Jin, "6D radar technology in monostatic integrated sensing and communications system," in *2024 IEEE/CIC International Conference on Communications in China (ICCC)*, 2024, pp. 1104–1109.

[29] S. Aghashahi, Z. Zeinalpour-Yazdi, A. Tadaion, M. B. Mashhadi, and A. Elzanaty, "Single antenna tracking and localization of RIS-enabled vehicular users," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2024.

[30] H. Liu, S. Lu, M. El-Hajjar, and L.-L. Yang, "Machine learning assisted ddaptive index modulation for mmWave communications," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1425–1441, 2020.

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

[32] Y. Chen, D. Chen, and T. Jiang, "Beam-squint mitigating in reconfigurable intelligent surface aided wideband mmwave communications," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–6.

[33] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1847–1860, 2020.

[34] K. Li, M. El-Hajjar, and L.-l. Yang, "Millimeter-wave based localization using a two-stage channel estimation relying on few-bit ADCs," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1736–1752, 2021.

[35] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. M. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6367–6383, 2022.

[36] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. M. Mourad, and L. Hanzo, "Reconfigurable intelligent surface aided amplitude- and phase-modulated downlink transmission," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8146–8151, 2023.

[37] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[38] 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.

[39] J. Veen and P. Van Der Wiellen, "The application of matched filters to PD detection and localization," *IEEE Electrical Insulation Magazine*, vol. 19, no. 5, pp. 20–26, 2003.

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

[41] Y. Zhang, M. El-Hajjar, and L.-L. Yang, "Adaptive codebook-based channel estimation in OFDM-aided hybrid beamforming mmWave systems," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1553–1562, 2022.

[42] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, 2016.

[43] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE transactions on wireless communications*, vol. 14, no. 11, pp. 6481–6494, 2015.

[44] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 635–648, 2022.

[45] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[46] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5786–5797, 2021.

[47] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Transactions on aerospace and electronic systems*, vol. 34, no. 1, pp. 103–123, 1998.
[48] Y. Zhu, M. Li, F. E. Bouanani, and Y. Zhao, "IRS-assisted collaborative tracking for ISAC in vehicular networks," in *2024 7th International Conference on Advanced Communication Technologies and Networking (CommNet)*, 2024, pp. 1–6.
[49] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE journal on selected areas in communications*, vol. 40, no. 6, pp. 1728–1767, 2022.
[50] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on STAR-RIS: Use cases, recent advances, and future research challenges," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 689–14 711, 2023.
[51] X. Zhai, G. Han, Y. Cai, Y. Liu, and L. Hanzo, "Simultaneously transmitting and reflecting (STAR) RIS assisted over-the-air computation Systems," *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1309–1322, 2023.
[52] R. Zhong, Y. Liu, X. Mu, Y. Chen, X. Wang, and L. Hanzo, "Hybrid reinforcement learning for STAR-RISs: A coupled phase-shift model based beamformer," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2556–2569, 2022.

Appendix A
Elements in (41)

To obtain the Jacobian matrix $\mathbf{J}_{l,n}$ of l-th path and n-th subcarrier for the k-th state, each element of $\mathbf{J}_{l,n} = \left[\frac{\partial \boldsymbol{\mu}_{l,n}}{\partial u_x^k}, \frac{\partial \boldsymbol{\mu}_{l,n}}{\partial u_y^k}\right]$, and $\boldsymbol{\mu}_{l,n}$ is given by

$$\frac{\partial \boldsymbol{\mu}_{l,n}}{\partial u_x^k} = \widehat{\boldsymbol{\Omega}}_k \left(\frac{\partial \widehat{\rho}_l[k,n]}{\partial \tau_{\mathrm{RU},l,k}} \frac{\partial \tau_{\mathrm{RU},l,k}}{\partial u_x^k} \mathbf{a} + \widehat{\rho}_l[k,n] \frac{\partial \mathbf{a}}{\partial \theta_{\mathrm{RU},l,k}} \frac{\partial \theta_{\mathrm{RU},l,k}}{\partial u_x^k} \right), \tag{54}$$

where the derivatives in (54) is indexed by a in (20) defined as

$$\frac{\partial \widehat{\rho}_l[k,n]}{\partial \tau_{\mathrm{RU},l,k}} = \widehat{\rho}_l[k,n](-j2\pi nB/N), a \geq 1, \tag{55}$$

$$\frac{\partial \tau_{\mathrm{RU},l,k}}{\partial u_x^k} = \frac{u_x^k - r_x}{c\left\|\mathbf{u}_k - \mathbf{r}\right\|_2}, a = 1, \tag{56}$$

$$\frac{\partial \tau_{\mathrm{RU},l,k}}{\partial u_y^k} = \frac{u_y^k - r_y}{c\left\|\mathbf{u}_k - \mathbf{r}\right\|_2}, a = 1, \tag{57}$$

$$\frac{\partial \theta_{\mathrm{RU},l,k}}{\partial u_x^k} = \frac{-u_y^k + r_y}{\left\|\mathbf{u}_k - \mathbf{r}\right\|_2^2}, a = 1, \tag{58}$$

$$\frac{\partial \theta_{\mathrm{RU},l,k}}{\partial u_y^k} = \frac{-u_x^k + r_x}{\left\|\mathbf{u}_k - \mathbf{r}\right\|_2^2}, a = 1, \tag{59}$$

$$\frac{\partial \tau_{\mathrm{RU},l,k}}{\partial u_x^k} = \frac{u_x^k - s_{2,x}}{c\left\|\mathbf{u}_k - \mathbf{s}_2\right\|_2}, a \neq 1, \tag{60}$$

$$\frac{\partial \tau_{\mathrm{RU},l,k}}{\partial u_y^k} = \frac{u_y^k - s_{2,y}}{c\left\|\mathbf{u}_k - \mathbf{s}_2\right\|_2}, a \neq 1, \tag{61}$$

$$\frac{\partial \theta_{\mathrm{RU},l,k}}{\partial u_x^k} = 0, a \neq 1, \tag{62}$$

$$\frac{\partial \theta_{\mathrm{RU},l,k}}{\partial u_y^k} = 0, a \neq 1. \tag{63}$$

$$\frac{\partial \mathbf{a}(\theta_{\mathrm{spatial},k,l})}{\partial \theta_{\mathrm{RU},l,k}} = j\pi \frac{d}{\lambda_c} \cos\left(\theta_{\mathrm{spatial},k,l}\right) \mathbf{D}\mathbf{a}(\theta_{\mathrm{spatial},k,l}), \tag{64}$$

where $\mathbf{s}_2$ denotes the location of the second SC. Furthermore, $\mathbf{D}$ is defined as $\mathbf{D} = \mathrm{diag}\,(0, \ldots, N_{\mathrm{R}} - 1)$, and $\mathbf{a} \triangleq \mathbf{a}(\theta_{\mathrm{spatial},k,l})$.

Kunlun Li (Member, IEEE) received his M.Sc and PhD degrees in wireless communications from the University of Southampton, U.K. in 2018 and 2024, respectively, where he is a research fellow with the Next Generation Wireless Research Group. His research interests include millimeter-wave, channel estimation, integrated sensing and communications (ISAC) and reconfigurable intelligent surface for wireless positioning and communications.

Mohammed El-Hajjar (Senior Member, IEEE) is a Professor of Signal Processing for Wireless Communications in the School of Electronics and Computer Science in the University of Southampton. He is the recipient of several academic awards and has published a Wiley-IEEE book and more than 150 IEEE journal and conference papers and in excess of 10 patents. Mohammed's research interests include the design of intelligent and energy efficient transceivers, millimeter wave communications and localisation, and machine learning for wireless communications. Mohammed's research is funded by the Engineering and Physical Sciences Research Council, the Royal Academy of Engineering and many industrial partners.

CHAO XU (Senior Member, IEEE) received the B.Eng. degree in telecommunications from Beijing University of Posts and Telecommunications, Beijing, China, the B.Sc. (Eng.) degree (with First Class Hons.) in telecommunications from the Queen Mary, University of London, London, U.K., through a Sino-U.K. joint degree Program in 2008, and the M.Sc. degree (with Distinction) in radio frequency communication systems and the Ph.D. degree in wireless communications from the University of Southampton, Southampton, U.K., in 2009 and 2015, respectively. He is currently a Senior Lecturer with Next Generation Wireless Research Group, University of Southampton. His research interests include index modulation, reconfigurable intelligent surfaces, noncoherent detection, and turbo detection. He was the recipient of the Best M.Sc. Student in Broadband and Mobile Communication Networks by the IEEE Communications Society United Kingdom and Republic of Ireland Chapter in 2009, 2012 Chinese Government Award for Outstanding Self-Financed Student Abroad, 2017 Dean's Award, Faculty of Physical Sciences and Engineering, University of Southampton, 2023 Marie Sklodowska-Curie Actions Global Postdoctoral Fellowships with the highest evaluation score of 100/100.

Lajos Hanzo (Life Fellow, IEEE) received Honorary Doctorates from the Technical University of Budapest (2009) and Edinburgh University (2015). He is a Foreign Member of the Hungarian Science-Academy, Fellow of the Royal Academy of Engineering (FREng), of the IET, of EURASIP. He holds the IEEE Edwin Howard Armstrong Award and the IEEE Eric Sumner Technical Field Award. For further details please see http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo.