Received 11 April 2024; accepted 5 May 2024. Date of current version 3 June 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3399605

Multi-RIS-Assisted 3D Localization and Synchronization via Deep Learning

ALIREZA FADAKAR [1], MARYAM SABBAGHIAN [1] (Member, IEEE), AND HENK WYMEERSCH [2] (Fellow, IEEE)

[1]School of Electrical and Computer Engineering, University of Tehran, Tehran 1439957131, Iran
[2]Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden

CORRESPONDING AUTHOR: M. SABBAGHIAN (e-mail: msabbaghian@ut.ac.ir)

The work of Henk Wymeersch was supported in part by the SNS JU Project 6G-DISAC under the EU's Horizon Europe Research and Innovation Program under Grant 101139130, and in part by the Swedish Research Council under Project 2022-03007.

ABSTRACT Reconfigurable intelligent surfaces (RISs) have received considerable attention in applications related to localization. However, operation in multi-path scenarios is challenging from both complexity and performance perspectives. This study presents a two-stage low complexity method for joint three-dimensional (3D) localization and synchronization using multiple RISs. Firstly, the received signals are preprocessed, and an efficient deep learning architecture is proposed to initially estimate the angles of departure (AODs) of the virtual line of sight paths from the RISs to the user. Then, a hybrid asynchronous AOD time-of-arrival-based approach is proposed in the first stage to estimate an initial guess of the position of the user equipment (UE). Finally, in the second stage, an optimization problem is formulated to refine the position of the UE by effectively utilizing the estimated delays and the clock offset. Our comparative study reveals that the proposed method outperforms the existing methods in terms of accuracy and complexity. Notably, the proposed method showcases enhanced robustness against multipath effects when compared to the state-of-the-art approaches.

INDEX TERMS 3D localization, deep learning, mmWave, reconfigurable intelligent surface, synchronization.

I. INTRODUCTION

THE APPLICATION of localization technologies has witnessed extensive adoption in various domains, including wireless networks, disaster rescue, augmented reality, and the Internet of Things (IoT) [1], [2]. Consequently, there has been a growing interest in advancing localization systems by leveraging innovative technologies over the past few decades [3], [4], [5]. In the presence of noise-corrupted sensor measurements, various approaches are utilized to estimate the position of the source, encompassing time-of-arrival (TOA) [6], time-difference-of-arrival (TDOA) [7], angle-of-arrival (AOA) [3], [4], angle-of-departure (AOD), received-signal-strength (RSS) [8], [9], and their hybrid counterparts [10]. Time-based measurements rely on synchronization [4], [6], while RSS-based measurements are subject to propagation effects [8]. Finally, AOA- or AOD-based localization has been a focal point in recent research [4], [11], [12], as it eliminates the need for synchronization, thus expanding its applicability to a broader range of scenarios [3], [4].

In 5G, accurate radio localization is achievable through the utilization of a multitude of antennas and a large radio bandwidth, combining AOA, AOD, and TDOA. Looking ahead to 6G, radio localization is anticipated to become even more ubiquitous, with the integration of reconfigurable intelligent surfaces (RISs) [13]. RISs have attracted enormous interest, mainly for their ability to overcome line-of-sight (LoS) blockages in millimeter wave (mmWave) communications. They are composed of multiple nearly passive metamaterials and are poised to revolutionize the performance of wireless applications. This includes enhancing capabilities in localization and sensing through dynamic manipulation of the signal propagation environment [12], [14]. RISs have emerged as a pivotal tool in propelling cost-effective and energy-efficient

approaches for optimizing wireless communication. By modifying the characteristics of electromagnetic waves in the environment, RISs enable significant improvements in wireless communication performance [11], [12], [15], [16].

RIS-aided localization approaches leverage RISs to actively manipulate multipath propagation, which enhances the precision and robustness of localization systems, extending the concept of multipath-aided localization to RIS [17], [18], [19]. In [12], a closed-form least squares (LS) based localization method is proposed for three-dimensional (3D) localization. This method utilizes partially connected receiving RISs and employs the popular subspace-based root multiple signal classification (R-MUSIC) method for angle estimation. However, it is worth noting that the R-MUSIC method requires a high computation. In [19], a single RIS is utilized for localization, with the base station (BS) receiving signals reflected by the RIS for AOA and TOA estimation. The proposed method involves an algorithm conducting two 1D searches over the TOAs and one two-dimensional (2D) search over the AODs, resulting in high computational complexity. Additionally, due to the propagation path of the probing signal traversing through the BS-RIS-user equipment (UE)-RIS-BS, significant signal attenuation is incurred. Consequently, the method demonstrates limited accuracy, particularly over longer ranges. The study in [20] achieves environment mapping and user localization through array signal processing and atomic norm denoising, utilizing two RISs for channel parameter estimation. More recently, [21] demonstrates that the absolute positions of UEs can be estimated through the assistance of at least two RISs and sidelink communication between the UEs, even in the absence of BSs. In recent years, various maximum likelihood (ML) based approaches have been proposed for localization using RISs [11], [17], [22]. ML-based methods, while asymptotically optimal, depend on a known signal model at the receiver to achieve optimal performance. Thus in the presence of multipath, due to the unknown number of non-line-of-sight (NLoS) paths, ML methods have an inconsistent performance. Moreover, the nonconvex nature of ML estimators necessitates the use of iterative solvers, leading to significant computational demands at each iteration. In [22], a semi-passive RIS is employed which utilizes the MUSIC algorithm for AOA estimation. Subsequently, the received sensor data is transformed into frequency-domain sequences using discrete Fourier transform (DFT), enabling the ML estimator to jointly estimate the TOA and the UE related coefficient. Among ML methods, [17] also addresses the synchronization using a single BS and a single RIS in 2D scenarios. In particular, first, it proposes a relaxed ML-based method (RML), which searches the uncertainty environment to obtain an initial guess for both the location of the UE and the clock offset for synchronization. Thus, it depends on the dimensions of the environment, which requires significant computation, especially in large 3D environments. Next, it proposes a joint ML based method (JML) that aims to optimize the ML estimator accurately. This is achieved by utilizing the derivative-free Nelder-Mead algorithm. However, when generalizing this method to 3D muti-RIS scenarios, the algorithm is computationally demanding and requires a high number of iterations to converge.

A critical sub-problem in RIS-aided localization is AOD estimation, which is necessarily performed in beamspace. Although deep learning (DL) has been applied for channel estimation or sensing in RIS-assisted systems [23], [24], [25], to the best of our knowledge, our work is among the first to utilize deep neural networks (DNNs) for 2D-AOD estimation for RIS-assisted localization and synchronization. Recent advancements in AOA estimation have introduced DL methods, in order to deal with various imperfections, low-SNR conditions, or multipath, or to provide lower-complexity solutions than model-based counterparts. In [26] a hierarchical DNN framework for 1D-AOA estimation is proposed, utilizing a multitask autoencoder and parallel multilayer perceptron (MLP) classifiers for denoising and AOA classification across spatial subregions. However, limitations in low-SNR conditions and a lack of generalization to 2D scenarios are observed in previous studies [27], [28]. A subsequent study in [27] presents a deeper DNN architecture with three residual blocks, relying heavily on computation for training and output spectrum calculation. Both methods employ the upper triangular elements of the correlation matrix as input vectors, overlooking the 2D matrix structure and resulting in suboptimal AOA performance due to lost features. Later in [29], a two-stage method is introduced for 2D-AOA estimation. Firstly, using a similar technique to [26], a deep convolutional network (DCN) classifies 2D angles into discrete subregions of the arrival plane. Subsequently, the orthogonal matching pursuit algorithm estimates 2D angles within each subregion. In [30], the authors introduce a similar approach using MLP classifiers for 1D-AOA estimation. Initially, they employ a full-row Toeplitz matrices reconstruction (FTMR) algorithm to utilize all rows of the sample covariance matrix (SCM), followed by computing the sum of squares of these matrices. Utilizing the forward/backward spatial smoothing (FBSS) technique, similar to the MUSIC algorithm, they then perform eigenvalue decomposition (EVD) to identify the noise subspace. The resulting polynomial coefficients are used as features for the MLP classifiers. Although this method reduces input size compared to previous DL-based approaches, the integration of FTMR, FBSS, and EVD significantly increases computational demands. Moreover, [28] notes its performance limitations in low-SNR scenarios. In [31], a convolutional neural network (CNN) is proposed for 1D-AOA estimation, utilizing the 2D covariance matrix as input. The CNN comprises four convolutional layers and three fully-connected layers. However, it exhibits high computational complexity due to its complex architecture. Recently, [32] proposed a deep residual network (ResNet) utilizing raw in-phase

(I) and quadrature (Q) components of the received signal as input. However, its input is dependent on the number of snapshots, limiting its generalizability. Additionally, the increased depth of ResNets leads to higher complexity during testing. Furthermore, [28] improves 2D-AOA estimation by integrating attention mechanisms into CNNs, while [33] investigates multi-source 2D-AOA estimation using uniform circular arrays (UCAs) under receiver mobility. In all of the aforementioned DL-based methods for AOA estimation, scenarios involving multipath and RISs are not considered. This paper addresses this issue by employing two techniques: effective dataset generation taking into account the multipath and the utilization of a less common loss function to mitigate the impact of multipath on AOA (or AOD) estimation performance.

The presence of multipath imposes requirements on resolution. Resolution can be provided in the delay domain (by the use of larger bandwidths) or in the angle domain (by the use of larger arrays). Various array structures can be used to provide 2D angle resolution, including uniform planar arrays (UPA), circular arrays, conformal arrays, and L-shaped arrays. While UPA structures have been more widely considered in RIS-aided localization [11], [12], [14], [19], [21], the advantages of other array topologies are given less attention in previous studies. Linear arrays for RISs and BSs have been recently considered in [34] for UE localization. Recently, [35] introduced a novel circular RIS architecture for precise environmental information acquisition and unique decoupling of channel parameters for user localization, not feasible with traditional UPA RIS topologies. Furthermore, in [36], the authors suggest employing conformal metasurfaces on vehicles' bodies to alleviate blockage effects by generating artificial reflections. This compensates for the non-flat shape of the vehicle's body using appropriate phase patterns.

In this paper, we introduce RISs with L-shaped topologies which have several advantages over other topologies including UPAs. First of all, for a given aperture (and thus angular resolution) L-shaped arrays exhibit a significantly reduced number of elements compared to UPAs [37], [38], [39], [40]. Secondly, by employing L-shaped arrays, we can decompose RIS elements into two symmetric ULAs. Thus, the 2D-AOD estimation problem can be easily decomposed into two simpler 1D-AOD estimation problems, which results in a considerable reduction in computational complexity [38], [39], [40], [41], compared to UPAs. For a given number of elements, L-shaped arrays, thanks to their higher subarray widths, have better AOD estimation performance compared to UPAs, so there is no loss in terms of SNR.

In this paper, we consider the multi-RIS-aided localization problem in a 3D complex multipath environment using L-shaped RIS. We present a novel two-stage hybrid AOD/TOA localization method aided by deep learning for joint user localization and synchronization. The main contributions of this work are as follows:

- *A novel architecture for deep learning-based AOD estimation:* An efficient low-complexity deep learning-based architecture is proposed to initially estimate the AODs of the virtual line of sight (VLoS) paths between the RISs and the UE. In contrast to existing DL-based methods, our approach effectively addresses the multipath effect, ensuring its robustness against this phenomenon. For simplicity, in the rest of the paper, this network is referred to as VNet.
- *A low-complexity 3D localization routine:* A novel low-complexity two-stage method is proposed to estimate the location and the clock offset of the UE in the presence of a single BS and multiple RISs. In the first stage (SG1), a hybrid asynchronous AOD/TOA-based method is proposed to estimate the delays of the received signals and an initial guess for the UE's position, using the proposed VNet and beamforming techniques. Moreover, the clock offset is also estimated in this stage. In the second stage (SG2), a hybrid AOD/TOA-based optimization problem is formulated to refine the estimated position in SG1. An efficient iterative approach is proposed for solving this problem using the estimated position from SG1 as the starting point of the algorithm.
- *Comprehensive complexity and performance analysis:* Our comparative study validates that the proposed method shows more robustness against the multipath effect compared to the existing methods. Moreover, our complexity analysis demonstrates the superiority and efficiency of the proposed method.

The rest of this paper is organized as follows. Section II presents the system model and problem formulation. Section III simplifies the signal model and defines some key notations and metrics. Moreover, it details the design of beamforming at the BS and RIS phase profiles. Section IV introduces and investigates the details of the signal pre-processing and the proposed VNet. Section V, presents the details of the proposed two-stage method. Section VI carries out extensive simulations to demonstrate the effectiveness and efficiency of the proposed method. Finally, Section VII concludes the paper.

Notations: Matrices are denoted by capital bold letters, such as $\boldsymbol{X}$, while vectors are represented by bold lowercase letters, such as $\boldsymbol{x}$. The submatrix $\boldsymbol{X}_{\boldsymbol{u},\boldsymbol{v}}$ is defined as the selection of rows indexed by $\boldsymbol{u}$ and columns indexed by $\boldsymbol{v}$ from the matrix $\boldsymbol{X}$. If : is used instead of $\boldsymbol{u}$ or $\boldsymbol{v}$, it indicates the selection of all rows or columns, respectively, from $\boldsymbol{X}$. The superscripts $(.)^T$, $(.)^H$ and $(.)^{-1}$ denote vector or matrix transpose, hermitian, and inverse, respectively. $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ shows the horizontal concatenation of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. The notation $\text{diag}(\boldsymbol{x})$ represents a diagonal matrix constructed using the vector $\boldsymbol{x}$ as its diagonal. Similarly, $\text{diag}(\boldsymbol{X})$ refers to a vector composed of the diagonal elements of the matrix $\boldsymbol{X}$. For a matrix $\boldsymbol{X}$ whose elements are in dB, $\boldsymbol{X}_{\text{lin}}$ represents

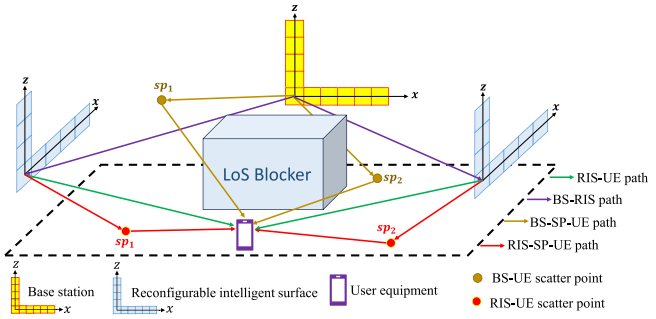


FIGURE 1. System model where the LoS path between the UE and BS is blocked. The G RISs provide VLoS paths, useful for localization of the UE.

the linear values of $\boldsymbol{X}$, where $[\boldsymbol{X}_{\text{lin}}]_{i,j} = 10^{[\boldsymbol{X}]_{i,j}/10}$. The $n \times n$ identity matrix is shown by $\boldsymbol{I}_n$, and $\mathbf{1}_M$ denotes an $M \times 1$ all ones vector. $\|\boldsymbol{X}\|_F$ and $\|\boldsymbol{x}\|$ show the Frobenius and l_2 norms, respectively, and finally, $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{A} \odot \boldsymbol{B}$ are the Kronecker and Hadamard productions, respectively.

II. SYSTEM MODEL

As illustrated in Fig. 1, we consider a BS, G RISs, and a UE in an indoor or outdoor environment. We aim to estimate the UE location and clock offset in the presence of NLoS paths using the downlink signal received from the BS via established VLoS paths with RISs.

A. GEOMETRY MODEL

Both the BS and each of the G RISs are equipped with an L-shaped array, comprising two orthogonal subarrays. Specifically, one subarray is oriented in the elevation direction (perpendicular to the ground), while the other is in the azimuth direction (parallel to the ground). Let M_b^θ and M_g^θ denote the number of elements in the elevation subarray of the BS and the g-th RIS, respectively. Similarly, M_b^ψ and M_g^ψ present the number of elements in azimuth subarrays. Thus, the total of $M_b = M_b^\theta + M_b^\psi$ and $M_g = M_g^\theta + M_g^\psi$ elements are utilized at the BS and the g-th RIS, respectively. Moreover, $\boldsymbol{p}_b \in \mathbb{R}^{3\times 1}$, $\boldsymbol{p}_{r_g} \in \mathbb{R}^{3\times 1}$, and $\boldsymbol{p}_u \in \mathbb{R}^{3\times 1}$ present the position of the centers of the BS, the g-th RIS, and the UE, respectively, where $g = 1, \ldots, G$. We assume that the LoS path between the BS and UE is blocked. In addition, scatter points (SPs) are utilized to model the NLoS paths for the multipath effect as shown in Fig. 1.

In this work, it is assumed that the UE resides within the far-field (FF) region relative to each RIS. According to [42], the FF condition is met when the distance d between the UE and any RIS surpasses

$$D_F = 2D^2/\lambda, \tag{1}$$

with λ indicating the carrier wavelength and D specifying the RIS's aperture size, defined as the greatest distance between any pair of elements on the RIS. This assumption will be verified later when presenting the results.

B. SIGNAL MODEL

The BS communicates with the UE via G RISs by transmitting orthogonal frequency division multiplexing (OFDM) pilots over N_s subcarriers across N_t transmissions. The subcarrier spacing, denoted by Δf, can be determined by $\Delta f = B/N_s$, where B represents the bandwidth. In particular, the t-th transmission over the n-th subcarrier is given by

$$y_t[n] = \sqrt{P}\boldsymbol{h}_t^T[n]\boldsymbol{f}_t s_t[n] + v_t[n], \tag{2}$$

where $s_t[n] \in \mathbb{C}$ is the n-th component of the OFDM symbol $\mathcal{S}_t \in \mathbb{C}^{N_s\times 1}$ in the t-th transmission, $\boldsymbol{f}_t \in \mathbb{C}^{M_b\times 1}$ represents the beamforming vector at the BS in the t-th transmission. Moreover, P denotes the total transmitted power, $v_t[n]$ is the zero-mean complex Gaussian noise with variance σ_v^2, and $\boldsymbol{h}_t[n] \in \mathbb{C}^{M_b\times 1}$ signifies the channel between the BS and UE for the n-th subcarrier and t-th transmission. This channel can be expressed as follows, assuming the LoS path between the UE and BS is blocked

$$\boldsymbol{h}_t^T[n] = \boldsymbol{h}_{b,u}^T[n] + \sum_{g=1}^{G} \boldsymbol{h}_{r_g,u}^T[n]\boldsymbol{\Omega}_{g,t}\boldsymbol{H}_{b,r_g}[n], \tag{3}$$

where $\boldsymbol{h}_{b,u}[n] \in \mathbb{C}^{M_b\times 1}$ denotes the channel from the BS to the UE, $\boldsymbol{h}_{r_g,u}[n] \in \mathbb{C}^{M_g\times 1}$ is the channel from the g-th RIS to the UE, and $\boldsymbol{\Omega}_{g,t} \in \mathbb{C}^{M_g\times M_g}$ represents the phase control coefficients of the g-th RIS during the t-th transmission for which we adopt a commonly used model, as seen in prior literature [17], [21], [34],

$$\boldsymbol{\Omega}_{g,t} = \text{diag}\left(e^{j\omega_{g,t}^1}, \ldots, e^{j\omega_{g,t}^{M_g}}\right), \tag{4}$$

where $\omega_{g,t}^m \in [0, 2\pi)$ is the amount of phase change by the m-th element of the g-th RIS at the t-th transmission. In addition, $\boldsymbol{H}_{b,r_g} \in \mathbb{C}^{M_g\times M_b}$ denotes the channel from the BS to the g-th RIS.

C. CHANNEL MODEL

The tandem channel $\boldsymbol{h}_t[n]$ comprises three parts: the channel from BS to UE $\boldsymbol{h}_{b,u}[n]$, the channel from BS to RIS $\boldsymbol{H}_{b,r_g}[n]$ and the channel from RIS to UE $\boldsymbol{h}_{r_g,u}[n]$. These are detailed next.

1) BS-UE CHANNEL

Based on the geometric channel model [43], since the LoS channel between the BS and UE is blocked, the first tandem channel $\boldsymbol{h}_{b,u}[n]$ is defined as follows assuming the presence of $I_{b,u}$ SPs located at $\{\boldsymbol{p}_{b,u}^{(i)}\}_{i=1}^{I_{b,u}}$ affecting this channel:

$$\boldsymbol{h}_{b,u}[n] = \sum_{i=1}^{I_{b,u}} \alpha_{b,u}^{(i)} e^{-j2\pi\tau_{b,u}^{(i)} n\Delta f} \boldsymbol{a}_b\left(\boldsymbol{\Phi}_{b,u}^{d^{(i)}}\right), \tag{5}$$

where $\alpha_{b,u}^{(i)} = \rho_{b,u}^{(i)} e^{j\varphi_{b,u}^{(i)}}$ is the complex gain between BS and UE through the i-th SP, with $\rho_{b,u}^{(i)}$ and $\varphi_{b,u}^{(i)}$ denoting its modulus and phase components. $\tau_{b,u}^{(i)}$ represents the delay

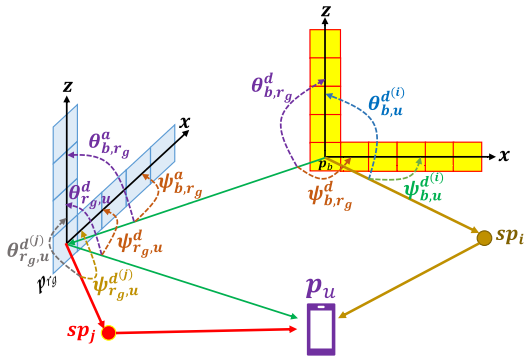


FIGURE 2. Definitions of AODs from RISs and from the BS and AOAs at the RISs.

from BS to UE via the i-th SP up to a clock offset δ, which can be obtained as follows

$$\tau^{(i)}_{b,u} = \frac{\|\boldsymbol{p}_b - \boldsymbol{p}^{(i)}_{b,u}\| + \|\boldsymbol{p}^{(i)}_{b,u} - \boldsymbol{p}_u\|}{c} + \delta, \tag{6}$$

where c is the speed of light. In (5), $\boldsymbol{a}_b(.) \in \mathbb{C}^{M_b \times 1}$ denotes the BS array steering vector. With the assumption that the element located at the origin of the array belongs to the elevation subarray, this vector is defined as follows for the elevation (θ) and spatial angle (ψ) pair $\boldsymbol{\Phi} = [\theta, \psi]^T$:

$$\boldsymbol{a}_b(\boldsymbol{\Phi}) = \Big[1, e^{jkd\cos(\theta)}, \ldots, e^{jk(M^\theta_b - 1)d\cos(\theta)}, e^{jkd\cos(\psi)}, \ldots, e^{jkM^\psi_b d\cos(\psi)}\Big]^T, \tag{7}$$

where $k = 2\pi/\lambda$ denotes the wave number with $\lambda = c/f_c$ being the wavelength and f_c the carrier frequency, and d is the distance between any two adjacent elements. The spatial angle is split into $\boldsymbol{\Phi}^{d(i)}_{b,u} = [\theta^{d(i)}_{b,u}, \psi^{d(i)}_{b,u}]^T$, where $\theta^{d(i)}_{b,u}$ is the elevation AOD, and $\psi^{d(i)}_{b,u}$ denotes the spatial AOD with respect to the x-axis of the array, which are shown in Fig. 2. It is important to highlight that the latter differs from the actual azimuth angle. Specifically, the relationship between the spatial angle ψ and the elevation and azimuth pair (θ, ϕ) is given by $\cos(\psi) = \sin(\theta)\cos(\phi)$. The geometric relationships of the angles $\theta^{d(i)}_{b,u}$ and $\psi^{d(i)}_{b,u}$ with the positions of the BS, UE, and SPs are presented as follows:

$$\theta^{d(i)}_{b,u} = \arccos\left(\frac{\left[\boldsymbol{p}^{(i)}_{b,u;b}\right]_3}{\|\boldsymbol{p}^{(i)}_{b,u;b}\|}\right), \quad \psi^{d(i)}_{b,u} = \arccos\left(\frac{\left[\boldsymbol{p}^{(i)}_{b,u;b}\right]_1}{\|\boldsymbol{p}^{(i)}_{b,u;b}\|}\right), \tag{8}$$

where $\boldsymbol{p}^{(i)}_{b,u;b} = \boldsymbol{R}^T_b(\boldsymbol{p}^{(i)}_{b,u} - \boldsymbol{p}_b)$ denotes the coordinates of the i-th SP with respect to the local coordinate system of the BS. Here, $\boldsymbol{R}_b \in \mathbb{R}^{3\times 3}$ depicts the rotation matrix defining the orientation of the BS.

2) BS-RIS CHANNEL

Assuming LoS path between BS and RISs, the second tandem channel $\boldsymbol{H}_{b,r_g}[n]$ is defined as

$$\boldsymbol{H}_{b,r_g}[n] = \alpha_{b,r_g} e^{-j2\pi\tau_{b,r_g}n\Delta f} \boldsymbol{a}_{r_g}\left(\boldsymbol{\Phi}^a_{b,r_g}\right)\boldsymbol{a}^T_b\left(\boldsymbol{\Phi}^d_{b,r_g}\right), \tag{9}$$

where $\alpha_{b,r_g} = \rho_{b,r_g} e^{j\varphi_{b,r_g}}$ is the complex gain between BS and the g-th RIS, with ρ_{b,r_g} and φ_{b,r_g} being its modulus and phase components, respectively. Here, τ_{b,r_g} expresses the delay between the g-th RIS and the BS as $\tau_{b,r_g} = \|\boldsymbol{p}_b - \boldsymbol{p}_{r_g}\|/c$. The g-th RIS steering vector $\boldsymbol{a}_{r_g}(.) \in \mathbb{C}^{M_g \times 1}$ is defined similarly to (7) as

$$\boldsymbol{a}_{r_g}(\boldsymbol{\Phi}) = \Big[1, e^{jkd\cos(\theta)}, \ldots, e^{jk\left(M^\theta_g - 1\right)d\cos(\theta)}, e^{jkd\cos(\psi)}, \ldots, e^{jkM^\psi_g d\cos(\psi)}\Big]^T. \tag{10}$$

Here, $\boldsymbol{\Phi}^a_{b,r_g} = [\theta^a_{b,r_g}, \psi^a_{b,r_g}]^T$ and $\boldsymbol{\Phi}^d_{b,r_g} = [\theta^d_{b,r_g}, \psi^d_{b,r_g}]^T$, as illustrated in Fig. 2, denote the AOAs and AODs from the BS to the g-th RIS, respectively. θ^a_{b,r_g}, θ^d_{b,r_g} are the elevation AODs, and ψ^a_{b,r_g}, ψ^d_{b,r_g} correspond to the spatial angles with respect to the x-axis of the array, which are defined similar to (8) as

$$\theta^a_{b,r_g} = \arccos\left(\frac{\left[\boldsymbol{p}_{b;r_g}\right]_3}{\|\boldsymbol{p}_{b;r_g}\|}\right), \quad \theta^d_{b,r_g} = \arccos\left(\frac{\left[\boldsymbol{p}_{r_g;b}\right]_3}{\|\boldsymbol{p}_{r_g;b}\|}\right), \tag{11}$$

$$\psi^a_{b,r_g} = \arccos\left(\frac{\left[\boldsymbol{p}_{b;r_g}\right]_1}{\|\boldsymbol{p}_{b;r_g}\|}\right), \quad \psi^d_{b,r_g} = \arccos\left(\frac{\left[\boldsymbol{p}_{r_g;b}\right]_1}{\|\boldsymbol{p}_{r_g;b}\|}\right), \tag{12}$$

where $\boldsymbol{p}_{r_g;b} = \boldsymbol{R}^T_b(\boldsymbol{p}_{r_g} - \boldsymbol{p}_b)$ and $\boldsymbol{p}_{b;r_g} = \boldsymbol{R}^T_g(\boldsymbol{p}_b - \boldsymbol{p}_{r_g})$, with $\boldsymbol{R}_g \in \mathbb{R}^{3\times 3}$ being the rotation matrix corresponding to the orientation of the g-th RIS.

3) RIS-UE CHANNEL

Assuming the presence of $I_{r_g,u}$ SPs affecting the channel from the g-th RIS to UE, the third tandem channel $\boldsymbol{h}_{r_g,u}[n]$ in (3) is defined as

$$\boldsymbol{h}_{r_g,u}[n] = \alpha_{r_g,u} e^{-j2\pi\tau_{r_g,u}n\Delta f} \boldsymbol{a}_{r_g}\left(\boldsymbol{\Phi}^d_{r_g,u}\right) + \sum_{i=1}^{I_{r_g,u}} \alpha^{(i)}_{r_g,u} e^{-j2\pi\tau^{(i)}_{r_g,u}n\Delta f} \boldsymbol{a}_{r_g}\left(\boldsymbol{\Phi}^{d(i)}_{r_g,u}\right), \tag{13}$$

in which $\boldsymbol{\Phi}^d_{r_g,u} = [\theta^d_{r_g,u}, \psi^d_{r_g,u}]^T$ denotes the elevation and spatial (with respect to x-axis) AODs of the LoS path from the g-th RIS to the UE, defined similarly to (11). In addition, $\alpha_{r_g,u} = \rho_{r_g,u} e^{j\varphi_{r_g,u}}$ is the complex channel gain between the g-th RIS and the UE. The notation $\alpha^{(i)}_{r_g,u} = \rho^{(i)}_{r_g,u} e^{j\varphi^{(i)}_{r_g,u}}$ is defined similarly, which corresponds to the i-th SP with position $\boldsymbol{p}^{(i)}_{r_g,u} \in \mathbb{R}^{3\times 1}$. $\boldsymbol{\Phi}^{d(i)}_{r_g,u} = [\theta^{d(i)}_{r_g,u}, \psi^{d(i)}_{r_g,u}]^T$ is defined as the AODs from the g-th RIS to $\boldsymbol{p}^{(i)}_{r_g,u}$, obtained in a similar manner to (11). In (13), $\tau_{r_g,u}$ is the delay between the g-th RIS and the UE up to a clock offset δ,

$$\tau_{r_g,u} = \frac{\|\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\|}{c} + \delta. \tag{14}$$

Similarly, $\tau^{(i)}_{r_g,u}$ denotes the NLoS delay between the g-th RIS and the UE through the i-th SP,

$$\tau^{(i)}_{r_g,u} = \frac{\|\boldsymbol{p}^{(i)}_{r_g,u} - \boldsymbol{p}_{r_g}\| + \|\boldsymbol{p}_u - \boldsymbol{p}^{(i)}_{r_g,u}\|}{c} + \delta. \tag{15}$$

It is noteworthy that in (14) and (15), the inclusion of the clock offset δ, addresses synchronization discrepancies specifically between the BS and the UE. Importantly, since the RISs operate passively, they inherently do not introduce any clock offset.

Moreover, it is important to note that while the detection and estimation of NLoS paths (i.e., SP locations) have the potential to enhance system performance, the realized improvement is limited by the considerable path loss associated with these paths and the fact that the number of such paths is a priori unknown [21]. In this study, similar to [17], [21], [34], we interpret these NLoS paths as interference. Unlike conventional methodologies, our proposed method neither ignores nor attempts to estimate or detect these paths and thus, we do not estimate the numbers $I_{b,u}$ or $\{I_{r_g,u}\}_{g=1}^{G}$. Instead, as shown in Section IV-B, it leverages machine learning techniques to intelligently mitigate the impact of NLoS paths.

III. PRELIMINARIES: METRICS AND PHASE PROFILE DESIGN

In this section, the signal model and beamforming models are further elaborated and specialized, in order to simplify the processing.

A. SIGNAL SIMPLIFICATION

In this section, we simplify the signal model and derive equations that will be used in the following sections. By substituting (5), (9) and (13) into (3), the expression is simplified to the following compact form:

$$\boldsymbol{h}_t^T[n] = \sum_{i=1}^{I_{b,u}} \alpha^{(i)}_{b,u}\Big[\boldsymbol{d}\Big(\tau^{(i)}_{b,u}\Big)\Big]_n \boldsymbol{a}_b^T\Big(\boldsymbol{\Phi}^{d^{(i)}}_{b,u}\Big) \tag{16}$$

$$+ \sum_{g=1}^{G}\Big[\alpha_{r_g}[\boldsymbol{d}(\tau_{r_g})]_n \boldsymbol{a}_{r_g}^T\Big(\boldsymbol{\Phi}^{d}_{r_g,u}\Big)\boldsymbol{\Omega}_{g,t}\boldsymbol{a}_{r_g}\Big(\boldsymbol{\Phi}^{a}_{b,r_g}\Big)\boldsymbol{a}_b^T\Big(\boldsymbol{\Phi}^{d}_{b,r_g}\Big) \tag{17}$$

$$+ \sum_{i=1}^{I_{r_g,u}} \alpha^{(i)}_{r_g}\Big[\boldsymbol{d}\Big(\tau^{(i)}_{r_g}\Big)\Big]_n \boldsymbol{a}_{r_g}^T\Big(\boldsymbol{\Phi}^{d^{(i)}}_{r_g,u}\Big)\boldsymbol{\Omega}_{g,t}\boldsymbol{a}_{r_g}\Big(\boldsymbol{\Phi}^{a}_{b,r_g}\Big)\boldsymbol{a}_b^T\Big(\boldsymbol{\Phi}^{d}_{b,r_g}\Big)\Big],$$

where $\alpha_{r_g} = \alpha_{r_g,u}\alpha_{b,r_g}$ and $\tau_{r_g} = \tau_{r_g,u} + \tau_{b,r_g}$ denote the total complex gain and the total delay between the BS and UE through the g-th RIS, respectively. The delay steering vector $\boldsymbol{d}(.) \in \mathbb{C}^{N_s\times 1}$ is defined as

$$\boldsymbol{d}(\tau) = \Big[1, e^{-j\xi_1\tau}, \ldots, e^{-j\xi_{N_s-1}\tau}\Big]^T, \tag{18}$$

where $\xi_n = 2\pi n\Delta f$. Similarly, the notations $\alpha^{(i)}_{r_g} = \alpha^{(i)}_{r_g,u}\alpha^{(i)}_{b,r_g}$ and $\tau^{(i)}_{r_g} = \tau^{(i)}_{r_g,u} + \tau^{(i)}_{b,r_g}$ are defined for the i-th SP between the g-th RIS and UE. We can express

$$\boldsymbol{a}_{r_g}^T\Big(\boldsymbol{\Phi}^{d}_{r_g,u}\Big)\boldsymbol{\Omega}_{g,t}\boldsymbol{a}_{r_g}\Big(\boldsymbol{\Phi}^{a}_{b,r_g}\Big) = \boldsymbol{b}_{r_g}^T\Big(\boldsymbol{\Phi}^{d}_{r_g,u}, \boldsymbol{\Phi}^{a}_{b,r_g}\Big)\boldsymbol{\omega}_{g,t}, \tag{19}$$

where $\boldsymbol{\omega}_{g,t} = \mathrm{diag}(\boldsymbol{\Omega}_{g,t})$ and the notation $\boldsymbol{b}_{r_g}(.,.) \in \mathbb{C}^{M_g\times 1}$ is defined as

$$\boldsymbol{b}_{r_g}(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = \boldsymbol{a}_{r_g}(\boldsymbol{\Phi}_1) \odot \boldsymbol{a}_{r_g}(\boldsymbol{\Phi}_2), \tag{20}$$

for $\boldsymbol{\Phi}_1 = [\theta_1, \psi_1]^T$ and $\boldsymbol{\Phi}_2 = [\theta_2, \psi_2]^T$. For simplicity, in the rest of the paper, we assume that $s_t[n] = 1$ for all $t = 1, \ldots, N_t$ and $n = 1, \ldots, N_s$. After simplifying (16) using (19) and substituting the simplified equation into (2), and stacking all N_sN_t received signals over all N_t transmissions and N_s subcarriers, we obtain the matrix $\boldsymbol{Y} \in \mathbb{C}^{N_s\times N_t}$ as

$$\boldsymbol{Y} = \underbrace{\sum_{g=1}^{G} \boldsymbol{X}^{(g)}_{\mathsf{VLoS}}}_{\boldsymbol{X}_{\mathsf{VLoS}}} + \underbrace{\sum_{i=1}^{I_{b,u}} \boldsymbol{X}^{(b,i)}_{\mathsf{NLoS}} + \sum_{g=1}^{G}\sum_{i=1}^{I_{r_g,u}} \boldsymbol{X}^{(g,i)}_{\mathsf{NLoS}}}_{\boldsymbol{X}_{\mathsf{NLoS}}} + \boldsymbol{V}, \tag{21}$$

where $\boldsymbol{X}^{(g)}_{\mathsf{VLoS}}$ and $\boldsymbol{X}^{(g,i)}_{\mathsf{NLoS}}$ are the noise-free received VLoS and i-th SP NLoS signals through the g-th RIS, respectively, and $\boldsymbol{X}^{(b,i)}_{\mathsf{NLoS}}$ denotes the i-th NLoS signal from the BS to UE through the i-th SP. Moreover, $\boldsymbol{V} \in \mathbb{C}^{N_s\times N_t}$ denotes the noise matrix with elements $[\boldsymbol{V}]_{n,t} = v_t[n]$. $\boldsymbol{X}^{(g)}_{\mathsf{VLoS}}$ can be obtained as

$$\boldsymbol{X}^{(g)}_{\mathsf{VLoS}} = \sqrt{P}\alpha_{r_g}\boldsymbol{d}(\tau_{r_g})\boldsymbol{b}_{r_g}^T\Big(\boldsymbol{\Phi}^{d}_{r_g,u}, \boldsymbol{\Phi}^{a}_{b,r_g}\Big)\boldsymbol{W}_g \times \mathrm{diag}\Big(\boldsymbol{a}_b^T\Big(\boldsymbol{\Phi}^{d}_{b,r_g}\Big)\boldsymbol{F}\Big), \tag{22}$$

with $\boldsymbol{W}_g = [\boldsymbol{\omega}_{g,1}, \ldots, \boldsymbol{\omega}_{g,N_t}]$, $\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{N_t}]$. In (21) $\boldsymbol{X}^{(g,i)}_{\mathsf{NLoS}}$ is defined similarly to $\boldsymbol{X}^{(g)}_{\mathsf{VLoS}}$ in (22), corresponding to the i-th SP from the g-th RIS to UE. Finally, $\boldsymbol{X}^{(b,i)}_{\mathsf{NLoS}}$ can be represented as

$$\boldsymbol{X}^{(b,i)}_{\mathsf{NLoS}} = \sqrt{P}\alpha^{(i)}_{b,u}\boldsymbol{d}\Big(\tau^{(i)}_{b,u}\Big)\boldsymbol{a}_b^T\Big(\boldsymbol{\Phi}^{d^{(i)}}_{b,u}\Big)\boldsymbol{F}. \tag{23}$$

The signal-to-noise ratio (SNR) and the line-of-sight to multipath ratio (LMR) metrics are defined as follows:

$$\mathrm{SNR} = \frac{\|\boldsymbol{X}_{\mathsf{VLoS}}\|_F^2}{\|\boldsymbol{V}\|_F^2}, \quad \mathrm{LMR} = \frac{\|\boldsymbol{X}_{\mathsf{VLoS}}\|_F^2}{\|\boldsymbol{X}_{\mathsf{NLoS}}\|_F^2}. \tag{24}$$

These terms will be used in Section VI to evaluate the performance of the proposed method.

B. BEAMFORMING AND RIS PHASE PROFILE DESIGN

In order to cover the environment around the RISs, based on [44], [45], [46], an optimized choice of codebooks for phase profiles $\{\boldsymbol{W}_g\}_{g=1}^{G}$ in terms of improving the received signal power at the UE, is the DFT matrix $\mathcal{W}_g$, whose (i,j)-th elements is given by $[\mathcal{W}_g]_{i,j} = \varphi_g^{(j-1)(i-1)}$, $1 \le \forall i,j \le M_g$,and $\varphi_g = e^{-j2\pi/M_g}$ for $g = 1, \ldots, G$.

Regarding BS beamforming design, we recall that the LoS path between the BS and UE is assumed to be blocked (see Fig. 1). So in this paper we use VLoS paths between BS and the UE established by RISs for joint localization and synchronization. Since the locations of the BS and RISs $\boldsymbol{p}_b$ and $\{\boldsymbol{p}_{r_g}\}_{g=1}^{G}$ are known beforehand, the angles $\{\boldsymbol{\Phi}^{a}_{b,r_g}, \boldsymbol{\Phi}^{d}_{b,r_g}\}_{g=1}^{G}$ are also known. Thus, we design the BS

beamforming matrix $\boldsymbol{F}$ using the optimal directional beamforming to concentrate the power P towards individual RISs, which maximizes the SNR at the RISs [17], [34]. Hence, the codebook $\mathcal{F} \in \mathbb{C}^{M_b \times G}$ for designing the beamforming matrix $\boldsymbol{F}$ is chosen as

$$\mathcal{F} = \left[\boldsymbol{a}_b^*\left(\boldsymbol{\Phi}_{b,r_1}^d\right), \ldots, \boldsymbol{a}_b^*\left(\boldsymbol{\Phi}_{b,r_G}^d\right)\right]. \tag{25}$$

Hence, using the codebooks $\mathcal{W}_g$ and $\mathcal{F}$, the BS beamforming matrix $\boldsymbol{F}$ and RIS phase profiles $\{\boldsymbol{W}_g\}_{g=1}^G$ are designed as follows:

$$\boldsymbol{F} = \frac{1}{\sqrt{M_b N_t}}\left[[\mathcal{F}]_{:,1} \otimes \mathbf{1}^{1\times M_1}, \ldots, [\mathcal{F}]_{:,G} \otimes \mathbf{1}^{1\times M_g}\right], \tag{26}$$

$$[\boldsymbol{W}_g]_{:,t} = \begin{cases} [\mathcal{W}_g]_{:,(t-L_{g-1})}, & L_{g-1}+1 \le t \le L_g \\ \mathbf{0} \text{ (absorption mode)}, & \text{otherwise,} \end{cases}$$

where $L_g = \sum_{g'=1}^{g} M_g$ for $g = 1, \ldots, G$ with $L_0 = 0$. Hence for each subcarrier, the total number of transmissions is equal to $N_t = \sum_{g=1}^{G} M_g$. In (26), each beamforming vector is divided by a factor of $\sqrt{M_b}$ to normalize the vectors and ensure that the total transmitted power remains constant throughout the entire period. Moreover, we divide each vector by $\sqrt{N_t}$ to satisfy the condition $\text{tr}(\boldsymbol{F}\boldsymbol{F}^H) = 1$. Only M_g beams with indices $L_{g-1}+1 \le t \le L_g$ are concentrated towards the g-th RIS. Thus, almost no power is received from other samples at the UE from the BS through the g-th RIS. Thus, for $t \le L_{g-1}$ or $L_g + 1 \le t$, the g-th RIS is configured in absorption mode to prevent signal scattering.

In conclusion to this section, we emphasize that the focus of this paper is on a downlink scenario. In this framework, all methodologies proposed in the subsequent sections are specifically designed for autonomous operation at the UE level, for instance within a mobile device. The beamforming designs in (26) ensures compatibility with scenarios involving multiple UEs. As a result, our proposed approach supports independent implementation at each UE, thereby enabling personalized 3D localization and synchronization based on the signals each device receives.

IV. PROPOSED VNET AOD ESTIMATION

For clarity and systematic presentation, the proposed method is divided into sequential stages. In the first stage, the received signal $\boldsymbol{Y}$, as defined in (21), undergoes preprocessing to yield G matrices, which are provided to VNet, to yield AOD estimates. The subsequent positioning and refinement stages are deferred to Section V.

A. GENERATING THE VNET INPUT

First, $\boldsymbol{Y}$ from (21) is partitioned into G submatrices $\boldsymbol{Y}_g \in \mathbb{C}^{N_s \times M^{(g)}}$ as

$$\boldsymbol{Y}_g = [\boldsymbol{Y}]_{:,(L_{g-1}+1):L_g}, \quad g = 1, \ldots, G. \tag{27}$$

According to the designed RIS phase profiles in (26), the g-th RIS only reflects in M_g transmissions with indices $t = L_{g-1}+1, \ldots, L_g$. Thus, $\boldsymbol{Y}_g$ can be obtained as:

$$\begin{aligned}\boldsymbol{Y}_g = {} & \sqrt{P}\alpha_{r_g}\boldsymbol{d}(\tau_{r_g})\boldsymbol{b}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^d, \boldsymbol{\Phi}_{b,r_g}^a\right)\mathcal{W}_g \,\text{diag}(\boldsymbol{u}_g) \\ & + \sum_{i=1}^{I_{r_g,u}} \sqrt{P}\alpha_{r_g}^{(i)}\boldsymbol{d}\left(\tau_{r_g}^{(i)}\right)\boldsymbol{b}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^{d^{(i)}}, \boldsymbol{\Phi}_{b,r_g}^a\right)\mathcal{W}_g \,\text{diag}(\boldsymbol{u}_g) \\ & + \sum_{i=1}^{I_{b,u}} \sqrt{P}\alpha_{b,u}^{(i)}\boldsymbol{d}\left(\tau_{b,u}^{(i)}\right)\boldsymbol{u}_g^{(i)} + \boldsymbol{V}_g, \end{aligned} \tag{28}$$

where $\boldsymbol{V}_g = [\boldsymbol{V}]_{:,(L_{g-1}+1):L_g}$, and

$$\boldsymbol{u}_g = \boldsymbol{a}_b^T\left(\boldsymbol{\Phi}_{b,r_g}^d\right)\boldsymbol{F}_g, \;\; \boldsymbol{u}_g^{(i)} = \boldsymbol{a}_b^T\left(\boldsymbol{\Phi}_{b,u}^{d^{(i)}}\right)\boldsymbol{F}_g \tag{29}$$

where $\boldsymbol{F}_g = [\boldsymbol{F}]_{:,(L_{g-1}+1):L_g} = [\mathcal{F}]_{:,g} \otimes \mathbf{1}^{1\times M_g}$. Hence, according to BS precoder design in (25) and (26), we deduce that $\text{diag}(\boldsymbol{u}_g) = \sqrt{M_b/N_t}\boldsymbol{I}_{M_b}$ and $\boldsymbol{u}_g^{(i)} = \beta_i \mathbf{1}_{M_g}^T/\sqrt{M_b N_t}$, where $\beta_i = \boldsymbol{a}_b^T(\boldsymbol{\Phi}_{b,u}^{d^{(i)}})\boldsymbol{a}_b^*(\boldsymbol{\Phi}_{b,r_g}^d)$ denotes the gain of the BS array at the direction of the i-th SP. Observe that β_i depends on the number of BS array elements M_b and the position of the i-th SP. We assume that the SPs are not located on the LoS paths between the BS and RISs, so that $|\beta_i|$ becomes negligible, especially for larger M_b. Hence, we conclude that the SPs between the BS and UE approximately do not interfere with the VLoS paths (though they will be included in the simulations). On the contrary, since the phase profiles of RISs are designed to illuminate the space in front of them uniformly, the SPs between RISs and UE are more likely to interfere with the VLoS paths at the receiver. Our proposed deep learning-based AOD estimator explained in Section IV-B2, meticulously considers these SPs to mitigate their impact.

Next, the LS estimates of the channels are obtained by multiplying $\sqrt{M_g}\mathcal{W}_g^{-1}$[1] by the right side of (28) as

$$\begin{aligned}\hat{\boldsymbol{X}}_g = {} & \sqrt{PM_gM_b/N_t}\alpha_{r_g}\boldsymbol{d}(\tau_{r_g})\boldsymbol{b}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^d, \boldsymbol{\Phi}_{b,r_g}^a\right) \\ & + \sum_{i=1}^{I_{r_g,u}} \sqrt{PM_gM_b/N_t}\alpha_{r_g}^{(i)}\boldsymbol{d}\left(\tau_{r_g}^{(i)}\right)\boldsymbol{b}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^{d^{(i)}}, \boldsymbol{\Phi}_{b,r_g}^a\right) \\ & + \boldsymbol{I}_{b,u,g} + \boldsymbol{V}_g', \end{aligned} \tag{30}$$

where $\boldsymbol{I}_{b,u,g}$ is the (small) interference term stemming from $\boldsymbol{h}_{b,u}[n]$. Moreover, $\boldsymbol{V}_g' = \sqrt{M_g}\boldsymbol{V}_g\mathcal{W}_g^{-1}$ denotes the noise matrix after LS estimation with zero-mean and the same covariance matrix $\sigma_v^2\boldsymbol{I}_{M_g}$ as $\boldsymbol{V}_g$. This LS estimation has a complexity of $O(N_sM_g^2)$ for the g-th RIS. Subsequently, since the AOAs $\boldsymbol{\Phi}_{b,r_g}^a$ are known by the UE, we multiply (30) by the matrix $\text{diag}(\boldsymbol{a}_{r_g}^*(\boldsymbol{\Phi}_{b,r_g}^a))$ to eliminate the dependency of $\hat{\boldsymbol{X}}_g$ on $\boldsymbol{\Phi}_{b,r_g}^a$, which results the following calibrated signal:

$$\hat{\boldsymbol{X}}_g^c = \sqrt{PM_gM_b/N_t}\alpha_{r_g}\boldsymbol{d}(\tau_{r_g})\boldsymbol{a}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^d\right)$$

[1]Since $\mathcal{W}_g$ is a DFT matrix, $\mathcal{W}_g^{-1} = \frac{1}{M_g}\mathcal{W}_g^H$. For simplicity, the scalar $\sqrt{M_g}$ is also multiplied to keep the variance of the noise, unchanged after the LS estimation.

$$+\sum_{i=1}^{I_{r_g,u}}\sqrt{PM_gM_b/N_t}\alpha_{r_g}^{(i)}\boldsymbol{d}\left(\tau_{r_g}^{(i)}\right)\boldsymbol{a}_{r_g}^T\left(\boldsymbol{\Phi}_{r_g,u}^{d^{(i)}}\right)+\boldsymbol{I}_{b,u,g}^c+\boldsymbol{V}_g'', \tag{31}$$

which is done with complexity $O(N_sM_g)$ for the g-th RIS. In (31), $\boldsymbol{V}_g'' = \boldsymbol{V}_g' \operatorname{diag}(\boldsymbol{a}_{r_g}^*(\boldsymbol{\Phi}_{b,r_g}^a))$ denotes the calibrated noise matrix with zero-mean and the covariance matrix of $\sigma_v^2\boldsymbol{I}_{M_g}$.

B. THE PROPOSED VNET

Although (31) represents a conventional model for angle/delay estimation, this paper diverges from traditional methods by introducing a deep learning-based approach designed to mitigate interference impact with lower complexity. Specifically designed to excel in challenging scenarios characterized by low-SNR, low-LMR conditions, and proximity of SPs to the UE, the proposed approach addresses these complexities effectively, as will be verified later in Section VI in comparison to existing methods.

This section proposes a low-complexity deep learning structure to obtain an initial estimate of the VLoS AODs $\{\boldsymbol{\Phi}_{r_g,u}^d\}_{g=1}^G$ which will be refined in the subsequent sections.

1) VNET INPUT

Recall from (10), that the first M_g^θ elements of $\boldsymbol{a}_{r_g}(\boldsymbol{\Phi}_{r_g,u}^d)$ solely depend on the elevation AOD $\theta_{r_g,u}^d$ and the rest M_g^ψ elements only depend on the spatial AOD $\psi_{r_g,u}^d$. Hence $\hat{\boldsymbol{X}}_g^c$ can be partitioned into two submatrices $\hat{\boldsymbol{X}}_g^\theta = [\hat{\boldsymbol{X}}_g^c]_{1:M_g^\theta,:}$ $\hat{\boldsymbol{X}}_g^\psi = [\hat{\boldsymbol{X}}_g^c]_{(M_g^\theta+1):(M_g^\theta+M_g^\psi),:}$. To eliminate the dependency of these two submatrices on unknown variables $\{\alpha_{r_g}, \tau_{r_g}, \alpha_{r_g}^{(i)}, \tau_{r_g}^{(i)}\}$, we compute the sample covariance matrix (SCM) as follows:

$$\begin{aligned}\widehat{\boldsymbol{R}}_g^\theta &= \frac{1}{N_s}\hat{\boldsymbol{X}}_g^{\theta^T}\hat{\boldsymbol{X}}_g^{\theta^*}\\ &= P_g\boldsymbol{a}_{r_g,\theta}\left(\theta_{r_g,u}^d\right)\boldsymbol{a}_{r_g,\theta}^H\left(\theta_{r_g,u}^d\right)+\boldsymbol{C}_{g,I}^\theta+\widehat{\boldsymbol{R}}_{g,v}^\theta,\end{aligned} \tag{32}$$

where $\boldsymbol{a}_{r_g,\theta}(.) = [\boldsymbol{a}_{r_g}(.)]_{1:M_g^\theta,:}$, and the first term, dominated by power, depends on the elevation AOD $\theta_{r_g,u}^d$. Additionally, $\widehat{\boldsymbol{R}}_{g,v}^\theta$ is the SCM of the noise covariance matrix $\boldsymbol{R}_{g,v}^\theta = \sigma_v^2\boldsymbol{I}_{M_g^\theta}$, and $\boldsymbol{C}_{g,I}^\theta \in \mathbb{C}^{M_g^\theta\times M_g^\theta}$ denotes all remaining interference terms. In addition, $P_g = PM_gM_b\rho_{r_g}^2/N_t$ represents the overall gain of the g-th VLoS path through the g-th RIS. Similar notations, such as $\widehat{\boldsymbol{R}}_g^\psi$, $\boldsymbol{a}_{r_g,\psi}(.) = [\boldsymbol{a}_{r_g}(.)]_{((M_g^\theta+1):M_g),:}$, $\boldsymbol{C}_{g,I}^\psi$, $\boldsymbol{R}_{g,v}^\psi$ and $\widehat{\boldsymbol{R}}_{g,v}^\psi$ are defined accordingly. Since the neurons in neural networks only take real values, we decompose the complex-valued SCMs $\widehat{\boldsymbol{R}}_g^\theta$ and $\widehat{\boldsymbol{R}}_g^\psi$ into their real and imaginary parts, yielding tensors $\mathcal{R}_g^\theta \in \mathbb{R}^{M_g^\theta\times M_g^\theta\times 2}$ and $\mathcal{R}_g^\psi \in \mathbb{R}^{M_g^\psi\times M_g^\psi\times 2}$ that can be represented as two-channel images. Moreover, $\mathcal{R}_g^\theta$ and $\mathcal{R}_g^\psi$ are normalized to reduce the input variability and sensitivity of the network as follows:

$$\mathcal{R}_g^\theta = \frac{\left[\Re\{\widehat{\boldsymbol{R}}_g^\theta\};\ \Im\left\{\widehat{\boldsymbol{R}}_g^\theta\right\}\right]}{\|\widehat{\boldsymbol{R}}_g^\theta\|_F},\quad \mathcal{R}_g^\psi = \frac{\left[\Re\left\{\widehat{\boldsymbol{R}}_g^\psi\right\};\ \Im\left\{\widehat{\boldsymbol{R}}_g^\psi\right\}\right]}{\|\widehat{\boldsymbol{R}}_g^\psi\|_F}. \tag{33}$$

The resulting normalized tensors are then fed into the proposed network. The overall complexity order of the derivations for equations (32) and (33) can be expressed as $O(((M_g^\theta)^2+(M_g^\psi)^2)N_s)$.

2) VNET ARCHITECTURE

As presented in (33), the input of the proposed VNet depends on the number of elements in each subarray of RISs. Therefore, a separate neural network must be trained for subarrays with a different number of elements. To simplify the exposition, in the rest of the paper, we assume that all subarrays in the RISs, whether horizontal or vertical, have the same number of elements, i.e., $M = M_g^\theta = M_g^\psi$ for $g = 1,\ldots,G$. This assumption is made without loss of generality and can be relaxed in practice by defining separate networks for different subarray configurations.

The proposed architecture of VNet is illustrated in Fig. 3. It consists of Q distinct subnetworks with identical structures, each responsible for predicting the AODs for a different region of the output scope (i.e., non-overlapping ranges of angles). The primary reason for considering Q subnetworks instead of a single network is to enable generalization and scalability with low complexity. By employing this technique, which is also used in [30], [47], we can seamlessly expand the angle scope by incorporating a new subnetwork, without the necessity of retraining existing networks. Conversely, for a more targeted angle scope, we can efficiently focus on the corresponding subnetworks.

- *Input:* The input is shared across all Q subnetworks. The input is denoted by $\mathcal{R}$ and is either $\mathcal{R}_\theta^{(g)}$ or $\mathcal{R}_\psi^{(g)}$.
- *Layers:* Given the input in (33), the output of the q-th subnetwork is given by

$$\boldsymbol{g}^{(q)} = o^{(q)}\left(d^{(q)}\left(l^{(q)}\left(f_2^{(q)}\left(f_1^{(q)}(\mathcal{R})\right)\right)\right)\right). \tag{34}$$

  The functions $f_1^{(q)}(.)$ and $f_2^{(q)}(.)$ denote the outputs of convolutional layers followed by a batch normalization layer and a rectified linear unit (ReLU) layer used as an activation function. The function $l^{(q)}(.)$ is a flatten layer, and the functions $d^{(q)}(.)$ and $o^{(q)}(.)$ present the fully connected layers (or dense layer), which construct the classification part of the network.
- *Output:* The output $\boldsymbol{g}^{(q)}$ is of length K and relates to the desired output angle scope (denoted by $[\varphi_{\min}^{(q)}, \varphi_{\max}^{(q)}]$, where φ presents either elevation (θ) or spatial (ψ) AODs) and resolution r, so that $K = (\varphi_{\max}^{(q)} - \varphi_{\min}^{(q)})/(Q\cdot r)$.

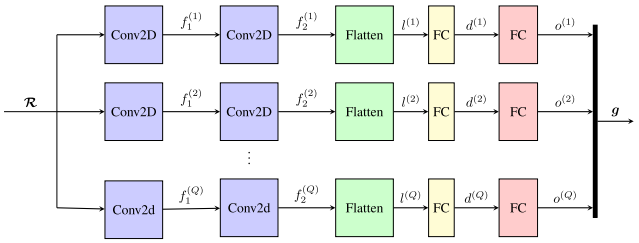


FIGURE 3. Proposed architecture for the VNet, including Q parallel branches, each covering a specific range of AODs.

To obtain the overall output of the VNet, the outputs of all Q subnetworks are concatenated as follows: $\boldsymbol{g} = \left[\boldsymbol{g}^{(1)^T}, \ldots, \boldsymbol{g}^{(Q)^T}\right]^T \in \mathbb{R}_+^{QK\times 1}$.

3) VNET LOSS FUNCTION

We use one-hot binary vectors for the labels of the training dataset. These vectors have exactly one non-zero element, corresponding to the true AOD of the VLoS path. To reduce the impact of NLoS paths and thus to down-weight the contribution of the zero classes in the loss function, we employ an asymmetric loss function (ASL) [48] which generalizes the commonly used binary cross-entropy (BCE). For the output $\boldsymbol{g} \in \mathbb{R}^{QK}$ with the ground truth label $\boldsymbol{y} \in \mathbb{R}^{QK}$, the total classification loss is defined as

$$\mathcal{L}_{\text{ASL}}(\boldsymbol{g}, \boldsymbol{y}) = \sum_{l=1}^{QK} \mathcal{L}(g_l, y_l), \quad \text{where} \tag{35}$$

$$\mathcal{L}(g, y) = -y(1-g)^{\gamma_+}\log(g) - (1-y)g_\epsilon^{\gamma_-}\log(1-g_\epsilon)$$

where $g_\epsilon = \max(0, g-\epsilon)$, for a small $\epsilon > 0$, and the parameters γ_+ and γ_- serve as positive and negative focusing parameters, respectively. The dynamic adjustment in the loss function via the parameters $\epsilon, \gamma_+, \gamma_-$ allows for reduced emphasis on simple classes.[2] Note that when $\gamma_+ = \gamma_- = \epsilon = 0$, ASL reverst to BCE.

4) VNET DATASET GENERATION STRATEGY

The proposed VNet outputs the desired angle scope $[\varphi_{\min}, \varphi_{\max}]$ for either the elevation AOD (θ) or spatial AOD (ψ), where φ denotes the AOD. This interval is evenly divided into $Z = QK$ grid points to obtain the vector $\boldsymbol{\varphi} = [\varphi_1, \ldots, \varphi_Z]^T$, as explained in the previous section. For simplicity, a single network is trained to estimate both the elevation (θ) and spatial (ψ) AODs due to their symmetry. Hence, we assume that $\varphi_{\min}$ and $\varphi_{\max}$ are determined based on the minimum and maximum values of these AODs.

Each of the signals $\hat{\boldsymbol{X}}_g^\theta$ and $\hat{\boldsymbol{X}}_g^\psi$, defined in Section IV-B1 (obtained from (31)), represents a single-input multiple-output (SIMO) system with multiple subcarriers and a linear array comprising M elements at the receiver, operating in the presence of multipath and noise. Hence, to generate training data, as detailed in Algorithm 1, we propose this equivalent interpretation to generate a LoS SIMO signal with a random delay and an AOD φ_z (line 7), in the presence of multipath (line 13) and noise (line 15) with the correct statistical properties. We outline our proposed approach as follows:

- $[\Gamma_{\min}, \Gamma_{\max}]$ is defined to denote the desired SNR range (in dB). This range is evenly divided into N_Γ values to obtain the vector $\boldsymbol{\Gamma} = [\Gamma_1, \ldots, \Gamma_{N_\Gamma}]^T$, with $\Gamma_1 = \Gamma_{\min}$ and $\Gamma_{N_\Gamma} = \Gamma_{\max}$.
- $[\chi_{\min}, \chi_{\max}]$ presents the desired LMR range, and it is divided evenly into N_χ values to obtain the vector $\boldsymbol{\chi} = [\chi_1, \ldots, \chi_{N_\chi}]^T$, where $\chi_1 = \chi_{\min}$ and $\chi_{N_\chi} = \chi_{\max}$.
- Let $\boldsymbol{I} = [I_1, \ldots, I_{N_I}]^T$ be a set of possible values for the number of SPs between one of RISs and UE denoted by $I \in \{I_{r_1,u}, \ldots, I_{r_G,u}\}$. It is important to note that for a fixed LMR, higher values of I correspond to smaller power levels for each NLoS path. As a result, fewer NLoS paths (I) leads to more intricate scenarios.
- Let $\Gamma \in \boldsymbol{\Gamma}$ and $\chi \in \boldsymbol{\chi}$ be the selected SNR and LMR for a specific data sample during dataset generation,

Algorithm 1 Proposed Dataset Generation for Training VNet

Inputs: $\boldsymbol{\varphi}$, $\boldsymbol{\Gamma}$, Γ_{diff}, $\boldsymbol{\chi}$, $\boldsymbol{I}$, $\tau_{\min}$, $\tau_{\max}$, N_s
Output: Dataset $\mathcal{D}$ with corresponding labels $\mathcal{L}$

1: Define empty lists $\mathcal{D}, \mathcal{L}$.
2: **for** $n_\Gamma = 1 : N_\Gamma$ **do**
3: **for** $z = 1 : Z$ **do**
4: Choose I and χ from $\boldsymbol{I}$ and $\boldsymbol{\chi}$, randomly.
5: Generate $I+1$ random unit modulus complex numbers $c_0, \ldots, c_I$ ($|c_i| = 1$).
6: Choose a random delay $\tau_0 \in [\tau_{\min}, \tau_{\max}]$.
7: $\boldsymbol{X}_{\text{LoS}} \leftarrow \sqrt{\Gamma_{n_\Gamma,\text{lin}}} c_0 \boldsymbol{a}(\varphi_z)\boldsymbol{d}^T(\tau_0)$
8: $\Gamma_{\text{NLoS}} \leftarrow \Gamma_{n_\Gamma} - \chi$ (in dB)
9: Choose I random values $\Gamma^{(1)}, \ldots, \Gamma^{(I)}$ from the interval $[\Gamma_{\text{NLoS}} - \frac{\Gamma_{\text{diff}}}{2}, \Gamma_{\text{NLoS}} + \frac{\Gamma_{\text{diff}}}{2}]$, with the sum Γ_{NLoS}, and store them in $\boldsymbol{\Gamma}_{\text{NLoS}}$.
10: Choose I random values $\tau_1, \ldots, \tau_I$ uniformly from $[\tau_{\min}, \tau_{\max}]$, and store them in $\boldsymbol{\tau}_{\text{NLoS}}$.
11: Choose I random values $\varphi^{(1)}, \ldots, \varphi^{(I)}$ from $[\varphi_{\min}, \varphi_{\max}]$, and store them in $\boldsymbol{\varphi}_{\text{NLoS}}$.
12: $\boldsymbol{A} \leftarrow [c_1\boldsymbol{a}(\varphi^{(1)}), \ldots, c_I\boldsymbol{a}(\varphi^{(I)})]$, (based on (36))
 $\boldsymbol{D} \leftarrow [\boldsymbol{d}(\tau_1), \ldots, \boldsymbol{d}(\tau_I)]$, (based on (18))
13: $\boldsymbol{X}_{\text{NLoS}} \leftarrow \boldsymbol{A}\sqrt{\text{diag}(\boldsymbol{\Gamma}_{\text{NLoS,lin}})}\boldsymbol{D}^T$
14: Generate a standard complex Gaussian noise matrix $\boldsymbol{V} \in \mathbb{C}^{M\times N_s}$.
15: $\boldsymbol{V} \leftarrow \sqrt{MN_s}\boldsymbol{V}/\|\boldsymbol{V}\|_F$
16: $\boldsymbol{Y} \leftarrow \boldsymbol{X}_{\text{LoS}} + \boldsymbol{X}_{\text{NLoS}} + \boldsymbol{V}$
17: Obtain $\mathcal{R} \in \mathbb{R}^{M\times M\times 2}$ based on (32) and (33).
18: Define the one-hot label $\boldsymbol{y} \in \{0,1\}^Z$ such that $[\boldsymbol{y}]_z = 1$ and $[\boldsymbol{y}]_i = 0$ for $i \neq z$.
19: Append $\mathcal{R}$ and $\boldsymbol{y}$ to $\mathcal{D}, \mathcal{L}$, respectively.

[2] This formulation enables the network to mitigate the impact of easily classified zero classes in two ways. Initially, adjusting parameters such as $\gamma_+ \geq 0$ and $\gamma_- > 0$ softens the impact of easy classes ($g \ll 0.5$) in the loss function. Subsequently, the introduction of the shift parameter $\epsilon \geq 0$ imposes a stricter threshold, discarding very easy classes ($g \leq \epsilon$) as outlined in [48].

and $I \in \boldsymbol{I}$ be the number of NLoS paths in this data sample.

- Based on the definitions of SNR and LMR in (24), the total SNR of the I NLoS paths can be obtained as $\Gamma_{\mathsf{NLoS}} = \Gamma - \chi$ (in dB). The SNR of each of these I NLoS paths might differ. To handle this, we define Γ_{diff}, and select the SNR of NLoS paths randomly in the interval $[\Gamma - \Gamma_{\text{diff}}/2, \Gamma + \Gamma_{\text{diff}}/2]$ such that their sum equals Γ_{NLoS}.
- let $\tau_{\min}$ and $\tau_{\max}$ denote the minimum and maximum delay between the BS and the UE through RISs.

The proposed training dataset generation for VNet is detailed in Algorithm 1. The notation $\boldsymbol{a}(\varphi)$ refers to the steering vector of a generic subarray with M elements:

$$\boldsymbol{a}(\varphi) = \left[1, e^{jkd\cos(\varphi)}, \ldots, e^{jk(M-1)d\cos(\varphi)}\right]^T. \tag{36}$$

V. POSITIONING AND REFINEMENT

From the 2D-AODs from each RIS, the position can be recovered. With the estimated position, the clock offset can be estimated and all parameters can be refined. We proceed in two stages.

A. STAGE 1 – COARSE ESTIMATION

1) POSITION ESTIMATION

Let $\{\check{\boldsymbol{\Phi}}^d_{r_g,u}\}_{g=1}^G$ denote the estimated elevation and spatial AODs by the VNet. For $g = 1, \ldots, G$, define ℓ_g to represent the line passing through the center $\boldsymbol{p}_{r_g}$ of the g-th RIS with unit direction $\boldsymbol{u}_g \in \mathbb{R}^{3\times 1}$ specified by the estimated AODs $\check{\boldsymbol{\Phi}}^d_{r_g,u}$. For an arbitrary point $\boldsymbol{v} \in \mathbb{R}^{3\times 1}$, the sum of squared distances from the lines $\ell_1, \ldots, \ell_G$ is obtained as:

$$\begin{aligned} S(\boldsymbol{v}) &= \sum_{g=1}^{G} \|\boldsymbol{v} - \boldsymbol{p}_{r_g}\|_2^2 - \sum_{g=1}^{G} \left(\boldsymbol{v} - \boldsymbol{p}_{r_g}\right)^T \boldsymbol{u}_g \boldsymbol{u}_g^T \left(\boldsymbol{v} - \boldsymbol{p}_{r_g}\right) \\ &= \sum_{g=1}^{G} \left(\boldsymbol{v} - \boldsymbol{p}_{r_g}\right)^T \left(\boldsymbol{I}_3 - \boldsymbol{u}_g \boldsymbol{u}_g^T\right) \left(\boldsymbol{v} - \boldsymbol{p}_{r_g}\right). \end{aligned} \tag{37}$$

After setting the gradient of $S(\boldsymbol{v})$ to zero, the point $\check{\boldsymbol{p}}_u$ that minimizes $S(\boldsymbol{v})$ can be obtained with complexity $O(G)$ as:

$$\check{\boldsymbol{p}}_u = \left[\sum_{g=1}^{G} \left(\boldsymbol{I}_3 - \boldsymbol{u}_g \boldsymbol{u}_g^T\right)\right]^{\dagger} \sum_{g=1}^{G} \left(\boldsymbol{I}_3 - \boldsymbol{u}_g \boldsymbol{u}_g^T\right) \boldsymbol{p}_{r_g}. \tag{38}$$

2) CLOCK OFFSET ESTIMATION

Next, all the delays are estimated. Using (38), the delay between the BS and the UE without considering the clock offset, can be obtained as:

$$\hat{\tau}_{r_g,s} = \frac{\|\boldsymbol{p}_b - \boldsymbol{p}_{r_g}\| + \|\check{\boldsymbol{p}}_u - \boldsymbol{p}_{r_g}\|}{c}, \tag{39}$$

where the subscript s emphasizes not taking clock offset into account. Based on (39), the uncertainty regions $\{\mathbb{I}_g\}_{g=1}^G$ for the delays $\{\tau_{r_g}\}_{g=1}^G$ can be defined as $\mathbb{I}_g = [\hat{\tau}_{r_g,s} + \delta_{\min} - \tau_\epsilon, \hat{\tau}_{r_g,s} + \delta_{\max} + \tau_\epsilon]$, where $\delta_{\min}$ and $\delta_{\max}$ are lower and upper bounds of the clock offset δ. Moreover, τ_ϵ denotes a small constant to determine the margin of error. Subsequently, we obtain estimates of the delay steering vectors $\{\boldsymbol{d}(\tau_{r_g})\}_{g=1}^G$ using (31),

$$\hat{\boldsymbol{d}}(\tau_{r_g}) = \hat{\boldsymbol{X}}_g^c \boldsymbol{a}_{r_g}^*\left(\check{\boldsymbol{\Phi}}^d_{r_g,u}\right), \tag{40}$$

which has a complexity $O(GN_sM)$. Next, using the beamforming technique, the delays $\{\tau_{r_g}\}_{g=1}^G$ are estimated as:

$$\hat{\tau}_{r_g} = \underset{\tau \in \mathbb{I}_g}{\operatorname{argmax}}\, \boldsymbol{d}^H(\tau) \hat{\boldsymbol{d}}(\tau_{r_g}), \tag{41}$$

which is solved by a simple line search with complexity $O(GN_{\mathbb{I}_g}N_s)$, considering $N_{\mathbb{I}_g}$ grid points. Using the estimated UE position $\hat{\boldsymbol{p}}_u$ and the estimated delays $\{\hat{\tau}_{r_g}\}_{g=1}^G$ in the first stage, based on (14), we estimate the clock-offset δ as follows:

$$\hat{\delta} = \frac{1}{G} \sum_{g=1}^{G} \left(\hat{\tau}_{r_g} - \frac{\|\boldsymbol{p}_{r_g} - \boldsymbol{p}_b\| + \|\boldsymbol{p}_{r_g} - \hat{\boldsymbol{p}}_u\|}{c}\right). \tag{42}$$

3) FINE GRID 2D AOD ESTIMATION FOR IMPROVED COARSE POSITIONING

Since the proposed VNet uses a relatively coarse grid for computational complexity reasons, the positioning performance in (37) will be limited by this coarse grid. To refine the AOD estimation and the corresponding position estimate, we proceed as follows.

First, from (31), the LS estimate of the array steering vector $\boldsymbol{a}_{r_g}(\boldsymbol{\Phi}^d_{r_g,u})$ is obtained as

$$\hat{\boldsymbol{a}}_{r_g}\left(\boldsymbol{\Phi}^d_{r_g,u}\right) = \hat{\boldsymbol{X}}_g^{c^T} \boldsymbol{d}^*(\hat{\tau}_{r_g}), \tag{43}$$

with complexity $O(GN_sM)$, which is then used to refine the estimations of the elevation and azimuth AODs $\{\boldsymbol{\Phi}^d_{r_g,u}\}_{g=1}^G$

$$\begin{aligned} \hat{\theta}^d_{r_g,u} &= \underset{\theta}{\operatorname{argmax}}\, \boldsymbol{a}^H(\theta) \left[\hat{\boldsymbol{a}}_{r_g}\left(\boldsymbol{\Phi}^d_{r_g,u}\right)\right]_{1:M}, \\ \hat{\phi}^d_{r_g,u} &= \underset{\psi}{\operatorname{argmax}}\, \boldsymbol{a}^H(\psi) \left[\hat{\boldsymbol{a}}_{r_g}\left(\boldsymbol{\Phi}^d_{r_g,u}\right)\right]_{(M+1):2M}, \end{aligned} \tag{44}$$

which are again 1D searches with overall complexity $O(G(N_\Theta + N_\Psi)M)$, considering N_Θ and N_Ψ grid points for θ and ψ, respectively.

Next, the estimated UE location $\check{\boldsymbol{p}}_u$ is updated by employing (38) once again, this time using the refined AODs $\{\hat{\boldsymbol{\phi}}^d_{r_g,u}\}_{g=1}^G$. This yields the updated location estimate $\hat{\boldsymbol{p}}_u$.

B. STAGE 2 – REFINED POSITION ESTIMATION

The first stage, although providing a consistent estimation, does not fully exploit the location information embedded in the estimated delays $\{\hat{\tau}_{r_g}\}_{g=1}^G$ and the clock offset $\hat{\delta}$. To address this issue, first, we estimate the distance R_g between the center of the g-th RIS and the UE as $\hat{R}_g = c(\hat{\tau}_{r_g} - \hat{\delta}) - \|\boldsymbol{p}_{r_g} - \boldsymbol{p}_b\|$, where $\hat{\tau}_{r_g}$ and $\hat{\delta}$ are the estimated values for the delay through the g-the RIS and the clock-offset, respectively. Similar to Section V-A, the lines $\{\ell_g\}_{g=1}^G$ are

defined using the refined AODs $\{\hat{\boldsymbol{\phi}}^{d}_{r_g,u}\}_{g=1}^{G}$. Moreover, let $\mathcal{S}_g$ denote the sphere with the center $\boldsymbol{p}_{r_g}$ and radius $\hat{R}_g$ for $g = 1, \ldots, G$. In the absence of noise, all the lines $\{\ell_g\}_{g=1}^{G}$ and spheres $\{\mathcal{S}_g\}_{g=1}^{G}$ pass through the UE's position. Motivated by this fact, we exploit this geometric model by using the estimated AODs, delays, and the clock offset to refine the UE's position. To this end, we find a point whose sum of squared distances from the lines and the nearest point on the surface of spheres is minimum. The optimization problem can be written as follows:

$$\tilde{\boldsymbol{p}}_u = \underset{\boldsymbol{p}_u \in \mathbb{R}^3}{\operatorname{argmin}} \left[\sum_{g=1}^{G} \left(\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\right)^T \left(\boldsymbol{I}_3 - \boldsymbol{u}_g \boldsymbol{u}_g^T\right) \left(\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\right) + \sum_{g=1}^{G} \left(\|\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\| - \hat{R}_g \right)^2 \right], \tag{45}$$

where the first term is accounted for the distances from the lines, explained in Section V-A, and the term $|\|\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\| - \hat{R}_g|$ denotes the minimum distance between the point $\boldsymbol{p}_u$ and the surface of the sphere $\mathcal{S}_g$. In general, the problem (45) is non-convex, and thus it may have multiple optimal solutions, and the specific solution obtained depends on the initial guess of the optimization algorithm. To address this issue, we use the previously estimated UE location $\hat{\boldsymbol{p}}_u$ in the first stage as the initial point for gradient descent, with

$$\nabla_{\boldsymbol{p}_u} f(\boldsymbol{p}_u) = 2 \sum_{g=1}^{G} \left(\boldsymbol{I}_3 - \boldsymbol{u}_g \boldsymbol{u}_g^T\right) \left(\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\right) + 2 \sum_{g=1}^{G} \left[1 - \frac{\hat{R}_g}{\|\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\|} \right] \left(\boldsymbol{p}_u - \boldsymbol{p}_{r_g}\right), \tag{46}$$

where $f(.)$ denotes the objective function in (45). The resulting order complexity is $O(GN_{\text{SG2}})$, where N_{SG2} is the number of gradient descent iterations until convergence.

VI. SIMULATION RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed joint 3D localization and synchronization scheme for RIS-assisted mmWave systems through numerical simulations.

A. SIMULATION SETUP

The considered RIS-assisted mmWave system consists of one BS, two RISs, and a single UE with default system parameters presented in Table 1. Some of these parameters may change in different simulations. Moreover, to ensure that the FF condition is met, according to (1) and Table 1, and the fact that the maximum distance between two elements in the considered L-shaped RIS configuration is $D = \sqrt{M^2 + (M-1)^2}\lambda/2$ we have $D_F \approx 2.25\,m$. The UE position is selected to ensure a minimum distance of D_F from the RIS centers, satisfying this condition.

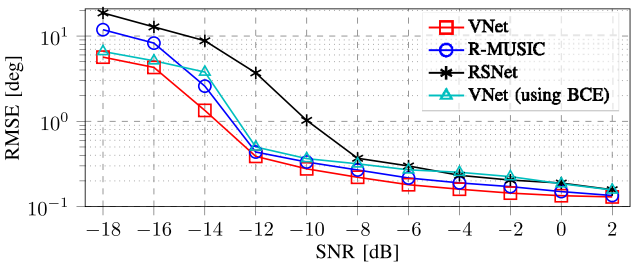
The parameters for training the proposed VNet (Algorithm 1) and the constants in the two-stage proposed localization method are presented in Table 2. Despite the selected SNR and LMR range specified in Table 2 and a fixed value of $N_s = 200$ for training VNet, it will be observed that the VNet generalizes for higher SNRs and LMRs, as well as a wide range of N_s.

TABLE 1. System parameters.

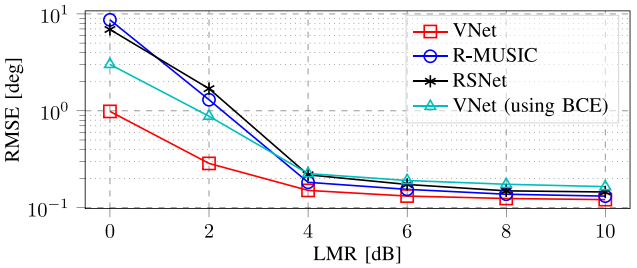
Default System Parameters	Symbol	Value
Carrier frequency	f_c	28 GHz
Environment Dimensions	$x_0 \times y_0 \times z_0$	$100 \times 100 \times 10$ [m]
UE position	$\boldsymbol{p}_u$	$[48, 51, 1]^T$ [m]
BS position	$\boldsymbol{p}_b$	$[0, 50, 7.5]^T$ [m]
BS subarray length	M_b^{θ}, M_b^{ψ}	50, 50
RIS 1 position	$\boldsymbol{p}_{r_1}$	$[50, 0, 10]^T$ [m]
RIS 2 position	$\boldsymbol{p}_{r_2}$	$[50, 100, 5]^T$ [m]
RIS subarrays length	M	15
Clock Offset	δ	10 ns

TABLE 2. Proposed method parameters.

Parameter	Value	Parameter	Value
Q	3	$\varphi_{\min}$	45°
r	0.5°	$\varphi_{\max}$	135°
$\boldsymbol{\Gamma}$	$[-18, -17, \ldots, -5]^T$	Γ_{diff}	5 dB
$\boldsymbol{\chi}$	$[0, 1, \ldots, 10]^T$	$\tau_{\min}$	0 ns
$\boldsymbol{I}$	$[1, \ldots, 10]^T$	$\tau_{\max}$	1000 ns
N_s	200	N_g	500
$\delta_{\min}$	0 ns	τ_{ϵ}	5 ns
$\delta_{\max}$	30 ns	θ_0	0.01°
θ_{ϵ}	1°	ψ_0	0.01°
ψ_{ϵ}	1°	$\gamma_-, \gamma_+, \epsilon$	0, 3, 0.4



(a) LMR = 3 dB.



(b) SNR = −5 dB.

FIGURE 4. RMSE of the VNet and the state-of-the-art methods versus SNR and LMR.

B. BASELINES AND METRICS

We evaluate the 2D-AOD estimation performance in terms of the root mean squared error (RMSE) of the proposed VNet against several baselines, namely:

- Classical R-MUSIC [49]: This method utilizes the roots of the characteristic equation derived from the eigenvalues of the SCM to estimate the AODs.
- Roots-spectrum network (RSNet) [30]: As explained in Section I, this method utilizes FTMR followed by FBSS technique, and then performs EVD to extract noise subspace features for its proposed network's input, which are used to estimate the MUSIC-like spatial spectrum.
- VNet using the classical BCE loss function: To demonstrate the effectiveness of ASL, we employ the same VNet structure with BCE loss function.

We evaluate the localization performance of the SG1 estimator (from Section V-A) and SG2 estimator (from Section V-B) in terms of the RMSE against the following baselines:

- The approach from [11], [12], which relies on R-MUSIC AOD estimation and LS position estimation.
- The convergent iterative method (CIM) from [4], which only performs positioning from AOD estimations. Hence, for a fair comparison, we combine it with the proposed VNet.
- RML and JML proposed in [17], which are direct positioning methods as they compute the position estimate directly from the observed waveforms. RML is a grid-based approach, while JML is based on a gridless Nelder-Mead algorithm.

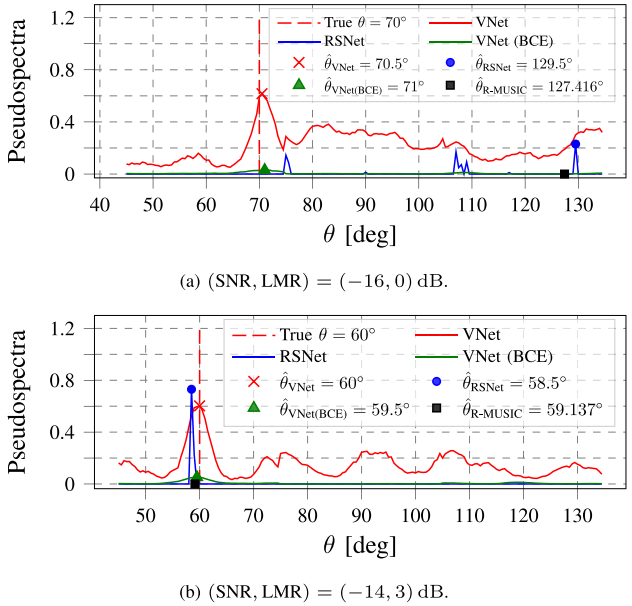
The system models used in these baselines may differ from our considered system model. Therefore, for a fair comparison, certain modifications and generalizations have been made to adapt these methods to our system.

C. RESULTS AND DISCUSSION

1) RESULTS OF THE PROPOSED VNET

Fig. 4(a) compares the RMSE results for a fixed LMR $=$ 3 dB, and Fig. 4(b), shows the results in terms of LMR for a fixed SNR $=-5$ dB. For Fig. 4(a), at each SNR the number of 100 data samples are generated for every angle $\theta \in [45°, 135°]$ (with step 0.5°), resulting the total of 18000 Monte Carlo samples for each SNR. A similar process is done for Fig. 4(b). It can be seen that the proposed VNet outperforms the state-of-the-art methods such as R-MUSIC, especially in low-SNR and low-LMR regimes. The results also highlight the benefit of the proposed loss function over BCE. This superiority stems from the VNet's ability to mitigate interference effectively.

The pseudospectra of the proposed VNet, RSNet, and VNet trained with BCE are depicted in Fig. 5 for varying SNR and LMR values, with the number of subcarriers set to $N_s = 200$. Due to the ASL function (defined in (35)), the proposed method exhibits nonzero outputs at other AODs, especially in low-SNR and low-LMR scenarios. This behavior, driven by the ASL, enables VNet to focus more on the dominant VLoS path, resulting in sharper peaks at the true AODs compared to when trained using BCE loss.



(a) $(\mathrm{SNR}, \mathrm{LMR}) = (-16, 0)$ dB.

(b) $(\mathrm{SNR}, \mathrm{LMR}) = (-14, 3)$ dB.

FIGURE 5. Spectrum plots for two test data samples for different SNR and LMR values.

Notably, R-MUSIC and RSNet underperform in low-SNR and low-LMR regimes.

2) COMPARISON WITH UPA

To demonstrate the superiority of L-shaped arrays over UPAs, we compare their RMSE results in this section. For 2D-AOD estimation with UPAs, we utilize 2D-MUSIC. To this end it is sufficient to utilize $\hat{\boldsymbol{X}}_g^c$ from (31) to derive the SCM, following (32), for input generation. Subsequently, EVD is performed to obtain MUSIC spectra. Let M_t represent the total number of elements in the corresponding RIS. In this experiment, we consider two scenarios with $M_t = 16$ and $M_t = 36$, respectively. We assume the true elevation and azimuth AOD pair to be $(\theta, \phi) = (99.416°, 120.528°)$. For each scenario, we employ 1D-MUSIC twice for an L-shaped array with $M = M_t/2$ for 2D-AOD estimation. For UPAs, we utilize 2D-MUSIC for direct 2D-AOD estimation. Fig. 6 illustrates the RMSE results across SNR $\in [-14, \mathrm{dB}, 2, \mathrm{dB}]$ for a fixed LMR $= 10, \mathrm{dB}$ using 1000 data samples at each SNR. As anticipated, the larger aperture of L-shaped arrays results in more precise estimations compared to UPAs. This highlights the effectiveness of L-shaped arrays in decomposing 2D-AOD estimation into two simpler 1D-AOD tasks while achieving superior accuracy. Consequently, L-shaped arrays offer enhanced AOD estimation performance, albeit requiring larger subarray widths.

3) RESULTS OF THE PROPOSED TWO-STAGE METHOD

In order to show that the proposed method can attain optimal performance, we will first consider results without NLoS paths. This allows us to compare it to the CRLB. Then, we will add multipath to show the robustness of the proposed method.

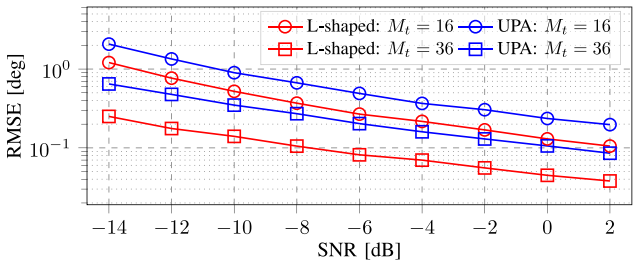


FIGURE 6. Comparing AOD estimation performance of L-shaped arrays with UPAs with the same number of array elements M_t.

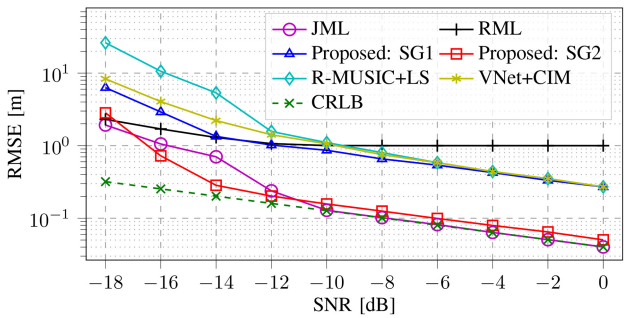


FIGURE 7. Localization RMSE versus SNR without NLoS interference.

Fig. 7 shows localization RMSE values of the UE compared to the state-of-the-art methods. In this experiment, we assume that there are no NLoS paths (i.e., $I_{r_g,u} = 0, I_{b,u} = 0$). Note that since the CIM algorithm assumes that the estimates of AODs are available at the receiver, we use the AODs obtained from the proposed VNet. We can make several observations. First of all, SG2 operates close to the CRLB for SNRs greater than -14 dB, and as the SNR increases, the gap between the RMSE and the CRLB decreases. In the absence of multipath effect, JML eventually reaches the theoretical bound CRLB at SNRs exceeding -10 dB. Consistent with its nature as a grid-based approach, the RML method saturates at -12 dB. In the low-SNR regime, SG1 demonstrates superior performance compared to the LS and CIM methods.

In the next experiment, we compare the performance of the methods in the presence of multipath in Fig. 8. Consider a challenging scenario with $I_{r_1,u} = I_{r_2,u} = 4$ SPs between RISs and UE with same[3] positions $\boldsymbol{p}^{(1)}_{r_1,u} = \boldsymbol{p}^{(1)}_{r_2,u} = [40, 50, 0]^T$, $\boldsymbol{p}^{(2)}_{r_1,u} = \boldsymbol{p}^{(2)}_{r_2,u} = [60, 50, 0]^T$, $\boldsymbol{p}^{(3)}_{r_1,u} = \boldsymbol{p}^{(3)}_{r_2,u} = [60, 45, 0]^T$, $\boldsymbol{p}^{(4)}_{r_1,u} = \boldsymbol{p}^{(4)}_{r_2,u} = [50, 50, 5]^T$, selected close to the true position of the UE. Moreover, we assume $I_{b,u} = 3$ SPs between BS and UE with positions $\boldsymbol{p}^{(1)}_{b,u} = [35, 55, 2]^T$, $\boldsymbol{p}^{(2)}_{b,u} = [30, 40, 0]^T$ and $\boldsymbol{p}^{(3)}_{b,u} = [25, 50, 1]^T$. We observe that across two LMR values (LMR $\in 5, 10$, dB) the proposed SG2 outperforms existing methods, especially in low-SNR scenarios. In the presence of NLoS paths, the performance of the ML-based JML method degrades with lower LMR

[3]In general, SPs may have different locations for each RIS, and some of them may be common among multiple RISs. Here, for simplicity, we assume the same set of SPs for both RISs.

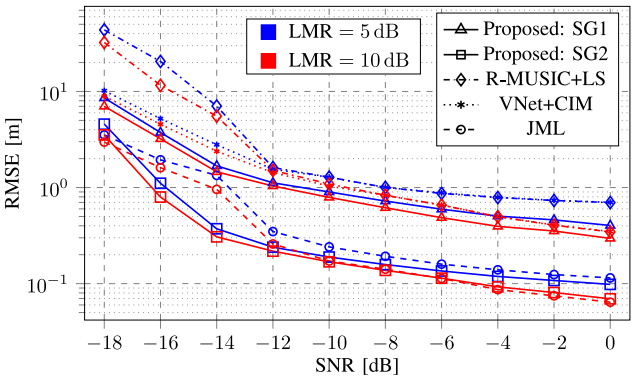


FIGURE 8. Localization RMSE results in the presence of NLoS paths.

values, but as LMR increases, its results converge toward the scenario without multipath. SG1 consistently outperforms LS and CIM methods across the entire SNR range.

As shown in Figs. 7 and 8, the proposed approach outperforms existing methods under the challenging conditions of low-SNR and low-LMR regimes, achieving accuracy within a minimal gap from the theoretically optimal solution. Importantly, as detailed in Section VI-D, this superior performance is accompanied by a significant reduction in computational complexity, particularly when contrasted with ML-based methodologies.

4) PROPOSED METHOD SCALABILITY

In this section, we conduct simulations to demonstrate the scalability of the proposed method in different scenarios. First, we analyze the performance with respect to the number of subcarriers N_s and the number of elements M per RIS subarray. It is noteworthy that the same VNet is used to obtain the result across N_s. However, since the input depends on M, we train a separate network for each value of M to obtain the results. The training process utilizes the same parameters as outlined in Table 2.

Fig. 9 illustrates the localization RMSE results across different numbers of subcarriers $N_s \in [50, 400]$ and Fig. 10 for different number of elements $M \in [10, 30]$ for different SNR values. The results across N_s are obtained using a fixed number of antennas $M = 15$, while the results for different values of M are obtained using a fixed value of $N_s = 200$. 5000 Monte Carlo trials are used for each N_s and M. The position of the UE is fixed at $\boldsymbol{p}_u = [48, 51, 1]^T$, and we provide the CRLB values, assuming that there are no NLoS paths. From Fig. 9, it is evident that the proposed VNet, although trained only for $N_s = 200$, demonstrates excellent generalization across different values of N_s. In addition, both stages SG1 and SG2 in the proposed method showcase improvement with higher N_s and SNR. The results of SG2 approach the CRLB values, and the gap between them diminishes for higher values of N_s or SNR. Similar observations can be drawn from Fig. 10. The RMSE value also improves with a higher number of elements deployed in

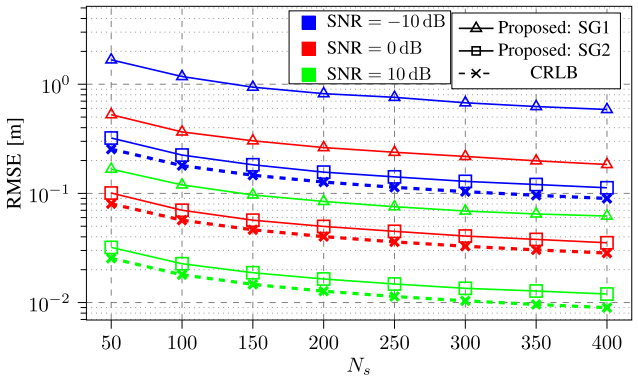


FIGURE 9. Localization RMSE for different numbers of subcarriers.

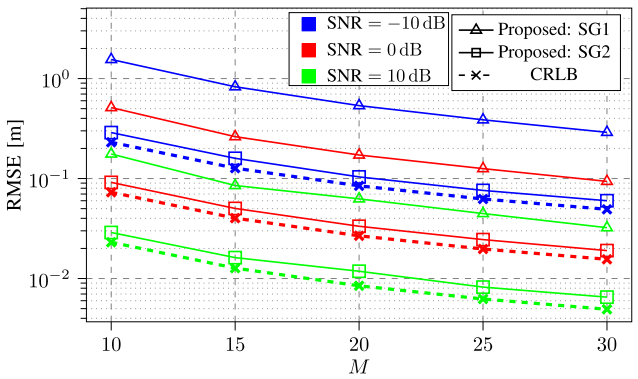


FIGURE 10. Localization RMSE for different numbers of RIS elements.

TABLE 3. Complexity comparison of AOD estimators.

	Preprocessing	Flops	Avg. Est. time (CPU) [ms]
VNet	$O(M^2N_s)$	645330	0.471
RSNet	$O(M^2N_s)+O(M^4)$	161400	1.38
R-MUSIC	$O(M^2N_s)$	-	1.48

each subarray of the RISs, demonstrating robustness across different values of N_s and M.

D. COMPLEXITY ANALYSIS

In this section, we provide a complexity analysis of the proposed methods and compare them to the state-of-the-art methods. Our focus will be on inference only, not training.

1) COMPLEXITY OF AOD ESTIMATORS

Table 3 summarizes the complexities of the proposed VNet and existing methods. It includes the order of complexity for signal pre-processing (column 2), exact FLOPs calculated (column 3), and average total time for angle estimation on the central processing unit (CPU) (column 4). The results demonstrate the computational advantage of our VNet.

2) DETAILED COMPLEXITY ANALYSIS OF THE PROPOSED TWO-STAGE METHOD

The order complexities of various components of our method are detailed in Table 4, with corresponding explanations

TABLE 4. Complexity analysis of the proposed method.

Component	Complexity
Channel Estimation	$O(GN_sM^2)$
Calibration	$O(GN_sM)$
SG1	$O(GN_sM^2)+O(FM^2)+O(G)+O(GN_sM)$ $+O(GN_{\mathrm{L}_g}N_s)+O(G(N_\Theta+N_\Psi)M)$
SG2	$O(GN_{\mathrm{SG2}})$

provided in relevant sections. The term $O(FM^2)$ denotes the simplified complexity of the VNet, where F is a constant derived from the network architecture.[4]

Hence, by focusing on the dominant part, the overall complexity of the proposed method is given by $O(GN_sM^2)$, which refers to the LS channel estimation and the computations of the SCMs in (32). In all simulations, the number of iterations N_{SG2} required was between 3 and 11 iterations with an average of 6.35.

3) OVERALL COMPLEXITY COMPARISON

Table 5 compares the order of complexities of the proposed method with the state-of-the-art ones and also the average amount of time required for each method. It should be noted that only the proposed method and the ML-based RML and JML can estimate the clock offset. Unlike RML and JML, the proposed method is environment-independent. Since RML is grid-based, it requires a 3D search in the uncertainty environment. In all simulations in this paper, we considered an environment where the horizontal coordinates (x, y) are between 35 and 65 and the vertical (z) are between 0 and 6 for these two methods. We divide each dimension with grid length of 1.5 meter resulting a grid mesh of dimensions $(N_x, N_y, N_z) = (21, 21, 5)$ for RML. We use $N_{\mathrm{FFT}} = 512$ for the RML method's FFT computation, as in [17]. In our 3D scenario, N_{NM}, representing the Nelder-Mead algorithm's iterations in JML, averages 235.65 iterations, peaking at 287. This contrasts with the simpler 2D scenario in [17], where only up to 30 iterations were reported.

VII. CONCLUSION

In this study, we investigated the problem of 3D localization and synchronization of a UE in the presence of a single BS and multiple L-shaped RISs, in a multipath environment. A novel 2D-AOD estimator was proposed, leveraging the RISs' high angular resolution. The proposed method relies on deep learning and can effectively operate in multipath environments while being agnostic of the geometric configuration of RISs, UE, and BS. In addition, a novel hybrid

[4]By solely considering the number of multiplications, a general approximation for the count of floating-point operations (FLOPs) [50] can be derived as $F = O(QM_0\sum_{i=1}^{N_c} n_c^{(i-1)} n_c^{(i)} q^{(i)^2} + Q\sum_{j=1}^{N_f} L_{j-1}L_j)$. which applies to an arbitrary network comprising Q identical parallel subnetworks, each consisting of N_c convolution layers and N_f fully-connected layers. Here, M_0 denotes the length of the input and $L_0 = M_0 q^{(N_c)^2}$ is the length of the output of the flatten layer. Also, L_j represents the number of cells in the j-th fully connected layer.

TABLE 5. Overall computational complexity comparison.

Method	Order of Complexity	Description	Avg. Est. [ms]
Proposed SG1	$O(GN_sM^2)$	-	4.13
Proposed SG2	$O(GN_{\text{SG2}})$	$N_{\text{SG2}} \approx 6.35$	2.45
Overall Proposed	$O(GN_sM^2)$	-	6.58
RML	$O(GN_xN_yN_zN_sM^3)$ $+ O(N_{\text{FFT}} \log N_{\text{FFT}})$	$N_x = 21, N_y = 21,$ $N_z = 5, N_{\text{FFT}} = 512$	1164.3
JML	$O(N_{\text{NM}}N_sM(M^2 + M^{(b)}))$	$N_{\text{NM}} \approx 235.65$	964.7
R-MUSIC+LS	$O(GM^3) + O(GN_sM^2)$	-	6.1
R-MUSIC+CIM	$O(GM^3) + O(GN_sM^2)$ $+ O(GN_{\text{CIM}})$	$N_{\text{CIM}} \approx 311.76$	18.6

AOD/TOA-based two-stage method was proposed, which effectively utilizes the 2D-AOD estimates and combines them to efficiently obtain TOA estimates as well as a UE position estimate. Extensive simulations showed that the proposed method significantly outperforms the state-of-the-art methods, and closely approaches the CRLB values in various scenarios. Moreover, the results revealed that the proposed method demonstrates more robustness against NLoS paths. Furthermore, the complexity analysis reveals that the proposed method necessitates considerably fewer computations compared to previous studies.

REFERENCES

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

[2] A. Bochem and H. Zhang, "Robustness enhanced sensor assisted Monte Carlo localization for wireless sensor networks and the Internet of Things," *IEEE Access*, vol. 10, pp. 33408–33420, 2022.

[3] Y. Wang and K. C. Ho, "An asymptotically efficient estimator in closed-form for 3-D AOA localization using a sensor network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6524–6535, Dec. 2015.

[4] Y. Zou, L. Wu, J. Fan, and H. Liu, "A convergent iteration method for 3-D AOA localization," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8267–8271, Jun. 2023.

[5] Y. Zheng, M. Sheng, J. Liu, and J. Li, "Exploiting AoA estimation accuracy for indoor Localization: A weighted AoA-based approach," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 65–68, Feb. 2019.

[6] T. Jia, H. Liu, K. C. Ho, and H. Wang, "Mitigating sensor motion effect for AOA and AOA-TOA localizations in underwater environments," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6124–6139, Sep. 2023.

[7] T. Wang, H. Xiong, H. Ding, and L. Zheng, "TDOA-based joint synchronization and localization algorithm for asynchronous wireless sensor networks," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3107–3124, May 2020.

[8] K. You, W. Guo, T. Peng, Y. Liu, P. Zuo, and W. Wang, "Parametric sparse Bayesian dictionary learning for multiple sources localization with propagation parameters uncertainty," *IEEE Trans. Signal Process.*, vol. 68, pp. 4194–4209, 2020.

[9] Y. Duan et al., "Data rate fingerprinting: A WLAN-based indoor positioning technique for passive localization," *IEEE Sensors J.*, vol. 19, no. 15, pp. 6517–6529, Aug. 2019.

[10] Y. Liu, Y. Wang, Y. Shen, and X. Shi, "Hybrid TOA-AOA WLS estimator for aircraft network decentralized cooperative localization," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9670–9675, Jul. 2023.

[11] G. C. Alexandropoulos, I. Vinieratou, and H. Wymeersch, "Localization via multiple reconfigurable intelligent surfaces equipped with single receive RF chains," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 1072–1076, May 2022.

[12] J. He, A. Fakhreddine, C. Vanwynsberghe, H. Wymeersch, and G. C. Alexandropoulos, "3D localization with a single partially-connected receiving RIS: Positioning error analysis and algorithmic design," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13190–13202, Oct. 2023.

[13] K. Keykhosravi et al., "Leveraging RIS-enabled smart signal propagation for solving infeasible localization problems: Scenarios, key research directions, and open challenges," *IEEE Veh. Technol. Mag.*, vol. 18, no. 2, pp. 20–28, Jun. 2023.

[14] R. Ghazalian, H. Chen, G. C. Alexandropoulos, G. Seco-Granados, H. Wymeersch, and R. Jäntti, "Joint user localization and location calibration of a hybrid reconfigurable intelligent surface," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 1435–1440, Jan. 2024.

[15] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[16] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[17] A. Fascista, M. F. Keskin, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "RIS-aided joint localization and synchronization with a single-antenna receiver: Beamforming design and low-complexity estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 1141–1156, Aug. 2022.

[18] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Towards ubiquitous positioning by leveraging reconfigurable intelligent surface," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 284–288, Jan. 2021.

[19] K. Keykhosravi, M. F. Keskin, G. Seco-Granados, and H. Wymeersch, "SISO RIS-enabled joint 3D downlink localization and synchronization," in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.

[20] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Channel estimation and user localization for IRS-assisted MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2320–2335, Apr. 2022.

[21] H. Chen, P. Zheng, M. F. Keskin, T. Al-Naffouri, and H. Wymeersch, "Multi-RIS-enabled 3D sidelink positioning," *IEEE Trans. Wireless Commun.*, early access, Jan. 22, 2024, doi: 10.1109/TWC.2024.3353387.

[22] M. Hua, Q. Wu, W. Chen, Z. Fei, H. C. So, and C. Yuen, "Intelligent reflecting surface-assisted localization: Performance analysis and algorithm design," *IEEE Wireless Commun. Lett.*, vol. 13, no. 1, pp. 84–88, Jan. 2024.

[23] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 471–487, 2021.

[24] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6181–6193, May 2023.

[25] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.

[26] Z.-M. Liu, C. Zhang, and S. Y. Philip, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, Dec. 2018.

[27] W. Liu, "Super resolution DOA estimation based on deep neural network," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020.

[28] A. Fadakar, A. Jafari, P. Tavana, R. Jahani, and S. Akhavan, "Deep learning based 2D-DOA estimation using L-shaped arrays," *J. Frankl. Inst.*, vol. 361, no. 6, 2024, Art. no. 106743.

[29] J. Ma, M. Wang, Y. Chen, and H. Wang, "Deep convolutional network-assisted multiple direction-of-arrival estimation," *IEEE Signal Process. Lett.*, vol. 31, pp. 576–580, 2024.

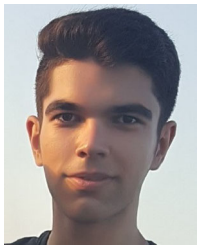
[30] K. Lee et al., "Deep learning-aided coherent direction-of-arrival estimation with the FTMR algorithm," *IEEE Trans. Signal Process.*, vol. 70, pp. 1118–1130, 2022.

[31] G. K. Papageorgiou, M. Sellathurai, and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low SNR," *IEEE Trans. Signal Process.*, vol. 69, pp. 3714–3729, 2021.

[32] S. Zheng et al., "Deep learning-based DOA estimation," *IEEE Trans. Cogn. Commun. Netw.*, early access, Jan. 31, 2024, doi: 10.1109/TCCN.2024.3360527.

[33] A. Fadakar, A. Mansourian, and S. Akhavan, "Deep learning aided multi-source passive 3D AOA wireless positioning using a moving receiver: A low complexity approach," *Ad Hoc Netw.*, vol. 154, Mar. 2024, Art. no. 103382.

[34] W. Wang and W. Zhang, "Joint beam training and positioning for intelligent reflecting surfaces assisted Millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6282–6297, Oct. 2021.
[35] Y. Lin, S. Jin, M. Matthaiou, and X. Yi, "Circular RIS-enabled channel estimation and localization for multi-user ISAC systems," *IEEE Trans. Wireless Commun.*, early access, Jan. 23, 2024, doi: 10.1109/TWC.2024.3353858.
[36] M. Mizmizi, R. A. Ayoubi, D. Tagliaferri, K. Dong, G. G. Gentili, and U. Spagnolini, "Conformal metasurfaces: A novel solution for vehicular communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2804–2817, Apr. 2023.
[37] J.-F. Gu, W.-P. Zhu, and M. Swamy, "Joint 2-D DOA estimation via sparse L-shaped array," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1171–1182, Mar. 2015.
[38] X. Wu and W.-P. Zhu, "On efficient gridless methods for 2-D DOA estimation with uniform and sparse L-shaped arrays," *Signal Process.*, vol. 191, Feb. 2022, Art. no. 108351.
[39] J.-F. Gu and P. Wei, "Joint SVD of two cross-correlation matrices to achieve automatic pairing in 2-D angle estimation problems," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 553–556, 2007.
[40] N. Tayem and H. M. Kwon, "L-shape 2-dimensional arrival angle estimation with propagator method," *IEEE Trans. Antennas Propag.*, vol. 53, no. 5, pp. 1622–1630, May 2005.
[41] S. Kikuchi, H. Tsuji, and A. Sano, "Pair-matching method for estimating 2-D angle of arrival with a cross-correlation matrix," *IEEE Antennas Wireless Propag. Lett.*, vol. 5, pp. 35–40, 2006.
[42] C. Ozturk, M. F. Keskin, V. Sciancalepore, H. Wymeersch, and S. Gezici, "RIS-aided Localization under pixel failures," *IEEE Trans. Wireless Commun.*, early access, Jan. 8, 2024, doi: 10.1109/TWC.2023.3348421.
[43] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2nd Quart., 2022.
[44] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.
[45] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2020, pp. 5000–5004.
[46] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 898–912, Feb. 2022.
[47] A. M. Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, Apr. 2020.
[48] T. Ridnik et al., "Asymmetric loss for multi-label classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 82–91.
[49] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP83)*, 1983, pp. 336–339.
[50] R. Hunger, "Floating point operations in matrix-vector calculus," Inst. Circuit Theory Signal Process, Munich Univ. Technol., München, Germany, Rep. 2007, 2005.



ALIREZA FADAKAR received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 2020 and 2023, respectively, where he is currently a Research Assistant in wireless communication. His research interests lie in the application of machine learning for wireless communications, channel estimation, MIMO networks, RIS-assisted communications, signal processing, digital twin, and optimization.

MARYAM SABBAGHIAN (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1999 and 2001, respectively, and the Ph.D. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2007. From 2008 to 2010, she was a Postdoctoral Fellow with the School of Engineering and Applied Sciences, Harvard University. In 2010, she joined the School of Electrical and Computer Engineering, University of Tehran. Her current research interests include 5G and beyond technologies, neural communications, and applications of information theory in interdiciplinary studies. She is a recipient of the Natural Sciences and Engineering Council NSERC-PDF of Canada, Ontario Graduate Scholarship, and Nortel Scholarship. She was a Technical Program Committee for several IEEE Telecommunication Conferences.

HENK WYMEERSCH (Fellow, IEEE) received the Ph.D. degree in electrical engineering/applied sciences from Ghent University, Belgium, in 2005. He is currently a Professor of communication systems with the Department of Electrical Engineering, Chalmers University of Technology, Sweden. Prior to joining Chalmers, he was a Postdoctoral Researcher with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology from 2005 to 2009. His current research interests include the convergence of communication and sensing in a 5G and Beyond 5G context. He served as an Associate Editor for IEEE COMMUNICATION LETTERS from 2009 to 2013, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2013, and the IEEE TRANSACTIONS ON COMMUNICATIONS from 2016 to 2018. He is currently a Senior Member of the IEEE SIGNAL PROCESSING MAGAZINE Editorial Board. From 2019 to 2021, he was an IEEE Distinguished Lecturer with the Vehicular Technology Society.