QUEEN'S UNIVERSITY BELFAST

Model-based Learning Network for 3-D Localization in mmWave Communications

Yang, J., Jin, S., Wen, C.-K., Guo, J., Matthaiou, M., & Gao, B. (2021). Model-based Learning Network for 3-D Localization in mmWave Communications. *IEEE Transactions on Wireless Communications*. Advance online publication. https://doi.org/10.1109/TWC.2021.3067957

Published in:
IEEE Transactions on Wireless Communications

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2021 IEEE.
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Download date:12. Jul. 2024

Model-based Learning Network for 3-D Localization in mmWave Communications

Jie Yang, *Student Member, IEEE*, Shi Jin, *Senior Member, IEEE*, Chao-Kai Wen, *Senior Member, IEEE*, Jiajia Guo, Michail Matthaiou, *Senior Member, IEEE*, and Bo Gao

***Abstract*—Millimeter-wave (mmWave) cloud radio access networks (CRANs) provide new opportunities for accurate cooperative localization, in which large bandwidths and antenna arrays and increased densities of base stations enhance the delay and angular resolution. This study considers the joint location and velocity estimation of user equipment (UE) and scatterers in a three-dimensional mmWave CRAN architecture. Several existing works have achieved satisfactory results by using neural networks (NNs) for localization. However, the black box NN localization method has limited robustness and accuracy and relies on a prohibitive amount of training data to increase localization accuracy. Thus, we propose a model-based learning network for localization to address these problems. In comparison with the black box NN, we combine NNs with geometric models. Specifically, we first develop an unbiased weighted least squares (WLS) estimator by utilizing hybrid delay and angular measurements, which determine the location and velocity of the UE in only one estimator, and can obtain the location and velocity of scatterers further. The proposed estimator can achieve the Cramér-Rao lower bound under small measurement noise and outperforms other state-of-the-art methods. Second, we establish a NN-assisted localization method called NN-WLS by replacing the linear approximations in the proposed WLS localization model with NNs to learn the higher-order error components, thereby enhancing the performance of the estimator, especially in a large noise environment. The solution possesses the powerful learning ability of the NN and the robustness of the proposed geometric model. Moreover, the ensemble learning is applied to improve the localization accuracy further. Comprehensive simulations show that the proposed NN-WLS is superior to the benchmark methods in terms of localization accuracy, robustness, and required time resources.**

Manuscript received June 3, 2020; revised January 10 and March 14, 2021; accepted March 16, 2021. The work was supported in part by the National Natural Science Foundation of China (NSFC) for Distinguished Young Scholars with Grant 61625106, and the NSFC under Grant 61941104. The work of C.-K. Wen was supported in part by the Ministry of Science and Technology of Taiwan under grants MOST 108-2628-E-110-001-MY3 and Qualcomm through a Taiwan University Research Collaboration Project. The work of M. Matthaiou was supported by the EPSRC, U.K., under Grant EP/P000673/1 and by a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme. The work was supported by the Scientific Research Foundation of Graduate School of Southeast University (YBPY2015). The associate editor coordinating the review of this paper and approving it for publication was J. Gross. *(Corresponding author: Shi Jin.)*

This paper was presented in part at the IEEE VTC-Fall, Honolulu, HI, USA, September 2019 [1]. Jie Yang, Shi Jin, and Jiajia Guo are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China (e-mail: {yangjie;jinshi;jiajiaguo}@seu.edu.cn). Chao-Kai Wen is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, 804, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw). Michail Matthaiou is with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, Belfast, U.K. (e-mail: m.matthaiou@qub.ac.uk). Bo Gao is with the ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen, China (e-mail: gao.bo1@zte.com.cn).

***Index Terms*—Cooperative localization, cloud radio access network, hybrid measurements, millimeter-wave communications, neural network, weighted least squares.**

I. Introduction

Future networks should offer unlimited coverage to any devices anywhere and anytime to stimulate the amalgamation of localization and wireless communications [2]. Millimeter-wave (mmWave) communication is a promising technology for meeting such requirements in future wireless communications. Localization is a highly desirable feature of mmWave communications [3], [4]. The user equipment (UE) location can be used to provide location-based services, such as navigation, mapping, social networking, augmented reality, and intelligent transportation systems. Additionally, location-aware communications can be realized by the obtained location information to improve communication capacity and network efficiency [5].

MmWave bands offer larger bandwidths than the presently used sub-6 GHz bands, hence, higher resolution of the time of arrival (TOA), time difference of arrival (TDOA), and frequency difference of arrival (FDOA) can be consequently achieved. In addition, the penetration loss from mmWave bands is inherently large [6]–[8]. Thus, the difference between the received power of the line-of-sight (LOS) path and the non-LOS (NLOS) path is pronounced, thereby simplifying the elimination of NLOS interference [9]–[11]. To compensate for severe penetration loss and increased path-loss, large antenna arrays and highly directional transmission should be combined to facilitate the acquisition of the angle of arrival (AOA) and the angle of departure (AOD) [12]. Moreover, cloud radio access networks (CRANs) can enhance mmWave communication by improving the network coverage [13]. CRANs provide a cost-effective way to achieve network densification, in which distributed low-complexity remote radio heads (RRHs) are deployed close to the UE and coordinated by a central unit (CU) for joint processing. The obtained location information can be shared with network nodes. Therefore, mmWave CRANs can offer accurate cooperative localization in urban and indoor environments, wherein conventional GPS may fail [14]–[16]. Channel parameters required in localization can be measured accurately [17]–[21] in static and mobile scenarios in the initial access and communication stages owing to the remarkable delay and angular resolution of mmWave communication systems without the need to install additional expensive infrastructure.

Localization has become a popular research topic in recent years. Different localization techniques have been summarized

in [22]. Currently, widespread localization methods apply the principle in which the channel parameters (e.g., AOA, TOA, TDOA, and FDOA) are initially extracted from the received waveform and grouped together as a function of the location parameters, and then different estimators are used to determine the UE locations. The classical linear weighted least squares (WLS) estimators were applied in [23]–[28]. In [23], [24], several closed-form TOA-based WLS estimators have been proposed. A few AOA-based methods were developed in [25] and in the related references. AOA and its combination with ranging estimates are expected to achieve high location accuracy. Reference [26] considered the localization problem of the three-dimensional (3-D) stationary targets in Multiple-Input Multiple-Output (MIMO) radar systems that utilized hybrid TOA/AOA measurements, from which a computationally efficient closed-form algorithm was developed with the WLS estimator, to achieve the Cramér-Rao lower bound (CRLB) under small measurement noise. Comparison shows that less effort has been devoted to the localization of moving targets. Reference [27] estimated location and velocity by introducing two-stage WLS estimators and using the hybrid TDOA/FDOA measurements. Reference [28] developed a WLS estimator to estimate the location and velocity of a moving target with a constant-velocity in a two-dimensional (2-D) scenario. Nevertheless, the aforementioned studies have overlooked the localization of scatterers. Recently, [29] advocated that future communication systems will turn multipath channels "from foe to friend" by leveraging distinguishable multipath components that resulted from unparalleled delay and angular resolution in mmWave systems. Thus, the information from reflected signals can be exploited in the reconstruction of the 3-D map of the surrounding environment. In this study, we consider the joint location and velocity estimation of a moving UE, as well as scatterers, in the 3-D scenario with mmWave communication systems by using hybrid TDOA/FDOA/AOA measurements. Unlike closed-form methods with multistage estimators, the proposed method determines the location and velocity of the UE in only one estimator.

All of the aforementioned localization techniques [23]–[28] are geometric approaches, in which delay and angular measurements are extracted and from which the location and velocity of the UE, as well as the scatterers, are triangulated or trilaterated. A function can be approximated by geometric techniques given the existence of an underlying transfer function between the measurements and the locations. In recent years, artificial intelligence (AI) has received considerable attention because of its promising performance in solving complicated problems. Researchers have utilized neural networks (NNs) to learn underlying transfer functions. Meanwhile, AI-based localization solutions, such as fingerprinting methods [30], [31], have emerged. A deep learning-based indoor fingerprinting system was presented in [30] to achieve meter-level localization accuracy. The experiments in [31] showed the feasibility of using deep learning methods for localization in actual outdoor environments. AI-based fingerprinting methods have alleviated modeling issues and can provide better performance than model-based localization techniques that use geometric relationships by fitting real-life measurements [32], [33]. However, extremely large amounts of training data are required to meet the high requirements of localization accuracy. Purely data-based and model-based, and hybrid data and model-based wireless network designs are discussed in [34]. To overcome the disadvantages of purely data- or model-based localization methods, we conceive hybrid data- and model-based localization methods by building and enhancing our localization estimator on the geometric model with NNs. At present, the literature on localization by combining NNs with geometric models, which is the focus of the current work, is scarce.

This study addresses the 3-D localization of moving UE and scatterers in mmWave communication systems. To our best knowledge, the present study is the first to combine the WLS estimator and NNs in 3-D localization problems. The contributions of this study are presented as follows:

- **Localization Model**: First, we establish a joint location and velocity estimation model by utilizing hybrid TDOA/FDOA/AOA measurements. Then, we develop an efficient closed-form WLS estimator. Unlike other closed-form WLS-based methods [27] with multistage estimators, the proposed method can determine the UEs location and velocity in only one stage. Second, we exploit the single-bounce NLOS paths and the estimated UE location and velocity to build the scatterer localization model. Then, we deduce the closed-form WLS estimator to determine the scatterers' location and velocity. The proposed estimator is proven asymptotically unbiased and able to attain CRLB under small measurement noise through simulations.
- **Learning Network**: Although the proposed WLS estimator performs well, its performance starts deteriorating as the noise level increases. Therefore, we propose a NN-assisted WLS method called NN-WLS to improve the localization accuracy further. The NN-WLS benefits from the powerful learning ability of the NN and the robustness of the geometric model. In addition, the proposed NN-WLS is fast because it can eliminate iterations in the proposed WLS algorithm. Furthermore, we embed ensemble learning into the proposed NN-WLS method to enhance localization accuracy. Simulation results show that NN-WLS outperforms the WLS estimator significantly when the measurement noise has an intrinsic relationship. In addition, the proposed NN-WLS is superior in terms of localization accuracy and robustness based on a comprehensive comparison with benchmark methods.

Notations—Uppercase boldface $\mathbf{A}$ and lowercase boldface $\mathbf{a}$ denote matrices and vectors, respectively. For any matrix $\mathbf{A}$, the superscripts $\mathbf{A}^{-1}$ and $\mathbf{A}^T$ stand for inverse and transpose, respectively. For any vector $\mathbf{a}$, the 2-norm is denoted by $\|\mathbf{a}\|$. $\text{diag}\{\cdot\}$ denotes a diagonal matrix with entries in $\{\cdot\}$, and $\text{blkdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_k)$ denotes a block-diagonal matrix constructed by $\mathbf{A}_1, \ldots, \mathbf{A}_k$. $\mathbb{E}\{\cdot\}$ denotes statistical expectation, whilst $|\cdot|$ denotes the module of a complex value or the cardinality of a set. The notation a° is the true value of the estimated parameter a.

TABLE I
NOTATIONS OF IMPORTANT VARIABLES.

Notation	Definition	Notation	Definition
$\mathbf{b}_n$	location of the n-th RRH	$\mathbb{M}_n$	set of measurements of the n-th RRH
$\mathbf{u}^\circ$	location of the UE	$\mathbb{M}_a$	set of selected LOS measurements
$\dot{\mathbf{u}}^\circ$	velocity of the UE	$\mathbb{M}_{r,n}$	set of remaining measurements of the n-th RRH
$\mathbf{x}^\circ$	$\mathbf{x}^\circ = [\mathbf{u}^{\circ T}, \dot{\mathbf{u}}^{\circ T}]^T$ 6-dimensional state vector of the UE	ϕ_n°	azimuth AOA-related parameter for LOS path of the n-th RRH
$\mathbf{s}_{n,l}^\circ$	location of the l-th scatterer between the n-th RRH and the UE	θ_n°	elevation AOA-related parameter for LOS path of the n-th RRH
$\dot{\mathbf{s}}_{n,l}^\circ$	velocity of the l-th scatterer between the n-th RRH and the UE	$\phi_{n,l}^{s\circ}$	azimuth AOA-related parameter for l-th NLOS path of the n-th RRH
$\mathbf{x}_{n,l}^{s\circ}$	$\mathbf{x}_{n,l}^{s\circ} = [\mathbf{s}_{n,l}^{\circ T}, \dot{\mathbf{s}}_{n,l}^{\circ T}]^T$, 6-dimensional state vector of the l-th scatterer between the n-th RRH and the UE	$\theta_{n,l}^{s\circ}$	elevation AOA-related parameter for l-th NLOS path of the n-th RRH
$\alpha_{n,l}^\circ$	complex gain for the l-th path of the n-th RRH	$\mathbf{m}$	vector of noisy measurements in $\mathbb{M}_a$
$\tau_{n,l}^\circ$	delay for the l-th path of the n-th RRH	$\mathbf{m}^\circ$	true value of measurements corresponding to $\mathbf{m}$
$\phi_{n,l}^\circ$	azimuth AOA for the l-th path of the n-th RRH	$\Delta\mathbf{m}$	Gaussian noise vector corresponding to $\mathbf{m}$ with zero mean and covariance matrix $\mathbf{Q}$
$\theta_{n,l}^\circ$	elevation AOA for the l-th path of the n-th RRH	$\mathbf{m}_{n,l}^s$	vector of noisy measurements in $\mathbb{M}_{r,n}$
$\nu_{n,l}^\circ$	Doppler shift for the l-th path of the n-th RRH	$\mathbf{m}_{n,l}^{s\circ}$	true value of measurements corresponding to $\mathbf{m}_{n,l}^s$
r_{n1}°	TDOA-related parameter for LOS path of the n-th RRH	$\Delta\mathbf{m}_{n,l}^s$	Gaussian noise vector corresponding to $\mathbf{m}_{n,l}^s$ with zero mean and covariance matrix $\mathbf{Q}_{n,l}^s$
$r_{n1,l}^{s\circ}$	TDOA-related parameter for l-th NLOS path of the n-th RRH	N	number of RRHs
$\dot{r}_{n1}^\circ$	FDOA-related parameter for LOS path of the n-th RRH	N_a	number of selected LOS paths
$\dot{r}_{n1}^{s\circ}$	FDOA-related parameter for l-th NLOS path of the n-th RRH		

II. SYSTEM MODEL

We study the moving UE and scatterer localization problems in a mmWave CRAN with N RRHs [13] (Fig. 1). Each RRH is equipped with a large antenna array with K antenna elements and connected to the CU via an individual fronthaul link. We assume that the clocks of different RRHs in the CRAN are synchronized. For ease of expression, we consider the system model with a single UE. The system model can be easily extended to solve the case with multiple UE as long as the pilot signals for different UE are orthogonal in time. The important variables are summarized in Table I.

A. System Geometry

We consider a 3-D space $\mathbb{R}^3 = \{[x, y, z]^T : x, y, z \in \mathbb{R}\}$ with N known RRHs located at $\mathbf{b}_n = [x_n^b, y_n^b, z_n^b]^T$, for $n = 1, 2, \ldots, N$. The geometry between the RRHs and the UE is shown in Fig. 1. We assume that the unknown location and velocity of the UE are represented by $\mathbf{u}^\circ = [x^\circ, y^\circ, z^\circ]^T$ and $\dot{\mathbf{u}}^\circ = [\dot{x}^\circ, \dot{y}^\circ, \dot{z}^\circ]^T$, respectively. Note that $\mathbf{u}^\circ$ is a function of time with $\partial\mathbf{u}^\circ/\partial t = \dot{\mathbf{u}}^\circ$. We only consider the LOS and the single-bounce NLOS paths because of the sparsity and high path loss of the mmWave channel [9], [10]. The unknown location and velocity of the l-th scatterer between the n-th RRH and the UE are represented by $\mathbf{s}_{n,l}^\circ = [x_{n,l}^{s\circ}, y_{n,l}^{s\circ}, z_{n,l}^{s\circ}]^T$ and $\dot{\mathbf{s}}_{n,l}^\circ = [\dot{x}_{n,l}^{s\circ}, \dot{y}_{n,l}^{s\circ}, \dot{z}_{n,l}^{s\circ}]^T$, respectively, with $l = 1, 2, \ldots, L_n$, where L_n is the number of scatterers between the n-th RRH and the UE. Here, $\mathbf{s}_{n,l}^\circ$ is a function of time with $\partial\mathbf{s}_{n,l}^\circ/\partial t = \dot{\mathbf{s}}_{n,l}^\circ$. We aim to determine $\mathbf{u}^\circ$, $\dot{\mathbf{u}}^\circ$, $\mathbf{s}_{n,l}^\circ$, and $\dot{\mathbf{s}}_{n,l}^\circ$, where $l = 1, 2, \ldots, L_n$ and $n = 1, 2, \ldots, N$ by the signals received at the RRHs.

B. Transmission Model

The UE sends a signal $\sqrt{p_s}s(t)$, in which p_s is the transmitted energy, and $\mathbb{E}\{|s(t)|^2\} = 1$. Given that the mmWave channel is sparse, we assume that $L_n + 1 \leq Q$, where Q is the number of RF chains for each RRH. The received signal $\mathbf{r}_n(t) \in \mathbb{C}^{Q\times 1}$ at RRH n is given by [36]

$$\mathbf{r}_n(t) = \mathbf{A}\left(\sum_{l=0}^{L_n} \alpha_{n,l}^\circ \sqrt{p_s} s(t - \tau_{n,l}^\circ)\mathbf{a}(\phi_{n,l}^\circ, \theta_{n,l}^\circ)e^{j2\pi\nu_{n,l}^\circ t}\right) + \mathbf{n}(t), \tag{1}$$

where $\alpha_{n,l}^\circ$, $\tau_{n,l}^\circ$, $\phi_{n,l}^\circ$, $\theta_{n,l}^\circ$, and $\nu_{n,l}^\circ$ denote the complex gain, delay, azimuth AOA, elevation AOA, and Doppler shift for the l-th path, respectively; $\mathbf{a}(\cdot)$ is the array response vector; $\mathbf{A} \in \mathbb{C}^{Q\times K}$ is the combining matrix in the mmWave hybrid architecture; and $\mathbf{n}(t) \in \mathbb{C}^{Q\times 1}$ is the zero-mean white Gaussian noise with a known power spectrum density. The channel parameters ($\phi_{n,l}^\circ$, $\theta_{n,l}^\circ$, $\tau_{n,l}^\circ$, $\nu_{n,l}^\circ$), for $l = 0, 1, \ldots, L_n$ and $n = 1, 2, \ldots, N$ can be extracted from (1) [17]–[19]. Here, $(l = 0)$ represents the LOS path, and $(l > 0)$ represents the NLOS path. Localization can be embedded in either the initial access stage or data transmission stage without additional overhead.

C. Relationship Between Channel and Location Parameters

In this subsection, we map the channel parameters to the location parameters.

- **TDOA**: For the LOS path, the distance between the UE and the RRH n is

$$r_n^\circ = v_c(\tau_{n,0}^\circ - \omega) = ||\mathbf{u}^\circ - \mathbf{b}_n||, \tag{2}$$

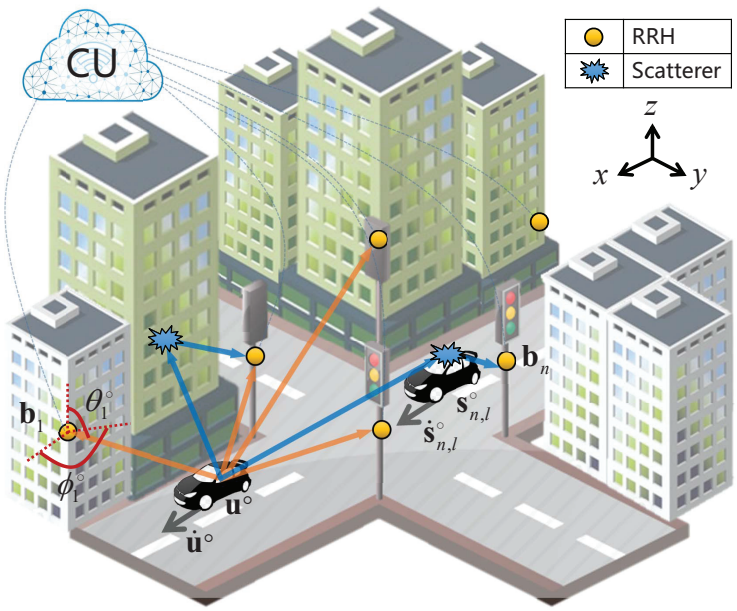


Fig. 1. Illustration of the mmWave CRAN system model, in which RRHs are connected with the CU.

where v_c is the signal propagation speed, and ω is the unknown clock bias between CRAN and UE. Without loss of generality, we define the TOA of LOS path received by the RRH 1 $\tau_{1,0}^\circ$ as the reference time. Then, the TDOA between the LOS path of the RRH n and the reference time is $\tau_{n,0}^\circ - \tau_{1,0}^\circ$. Thus, we define the TDOA-related parameter as

$$r_{n1}^\circ = v_c(\tau_{n,0}^\circ - \tau_{1,0}^\circ) = r_n^\circ - r_1^\circ, \tag{3}$$

where the unknown ω can be eliminated.
For the NLOS path, we have

$$r_{n,l}^{s\circ} = v_c(\tau_{n,l}^s - \omega) = ||\mathbf{u}^\circ - \mathbf{s}_{n,l}^\circ|| + ||\mathbf{s}_{n,l}^\circ - \mathbf{b}_n||. \tag{4}$$

Then, the TDOA between the l-th NLOS path of the n-th RRH and the reference time is $\tau_{n,l}^\circ - \tau_{1,0}^\circ$, and we define the TDOA-related parameter as

$$r_{n1,l}^{s\circ} = v_c(\tau_{n,l}^\circ - \tau_{1,0}^\circ) = r_{n,l}^{s\circ} - r_1^\circ. \tag{5}$$

Therefore, r_{n1}° and $r_{n1,l}^{s\circ}$ are the TDOA-related parameters, which are used in our proposed algorithms and are derived from the TDOA by multiplying with v_c.

- **FDOA**: For the LOS path, we define the time derivative of r_n° in (2) as $\dot{r}_n^\circ$, and we have

$$\dot{r}_n^\circ = \frac{\partial r_n^\circ}{\partial t} = \frac{\dot{\mathbf{u}}^{\circ T}\mathbf{u}^\circ + \mathbf{u}^{\circ T}\dot{\mathbf{u}}^\circ - 2\dot{\mathbf{u}}^{\circ T}\mathbf{b}_n}{2\sqrt{(\mathbf{u}^\circ - \mathbf{b}_n)^T(\mathbf{u}^\circ - \mathbf{b}_n)}} = \frac{\dot{\mathbf{u}}^{\circ T}(\mathbf{u}^\circ - \mathbf{b}_n)}{||\mathbf{u}^\circ - \mathbf{b}_n||}, \tag{6}$$

which is the relative velocity between UE and RRH n. Without loss of generality, we define the FOA or Doppler shift of LOS path received by the RRH 1 $\nu_{1,0}^\circ$ as the reference frequency. Then, the FDOA between the LOS path of the RRH n and the reference frequency is $\nu_{n,0}^\circ - \nu_{1,0}^\circ$. Thus, given the signal wavelength λ_c, we obtain the FDOA-related parameter as

$$\dot{r}_{n1}^\circ = \lambda_c(\nu_{n,0}^\circ - \nu_{1,0}^\circ) = \dot{r}_n^\circ - \dot{r}_1^\circ. \tag{7}$$

For the NLOS path, we define the time derivative of $r_{n,l}^{s\circ}$ in (4) as $\dot{r}_{n,l}^{s\circ}$, and we have

$$\dot{r}_{n,l}^{s\circ} = \frac{(\dot{\mathbf{u}}^\circ - \dot{\mathbf{s}}_{n,l}^\circ)^T(\mathbf{u}^\circ - \mathbf{s}_{n,l}^\circ)}{||\mathbf{u}^\circ - \mathbf{s}_{n,l}^\circ||} + \frac{\dot{\mathbf{s}}_{n,l}^{\circ T}(\mathbf{s}_{n,l}^\circ - \mathbf{b}_n)}{||\mathbf{s}_{n,l}^\circ - \mathbf{b}_n||}. \tag{8}$$

Then, the FDOA between the l-th NLOS path of the n-th RRH and the reference frequency is $\nu_{n,l}^\circ - \nu_{1,0}^\circ$, and we obtain the FDOA-related parameter as

$$\dot{r}_{n1,l}^{s\circ} = \lambda_c(\nu_{n,l}^\circ - \nu_{1,0}^\circ) = \dot{r}_{n,l}^{s\circ} - \dot{r}_1^\circ. \tag{9}$$

Thus, $\dot{r}_{n1}^\circ$ and $\dot{r}_{n1,l}^{s\circ}$ are the FDOA-related parameters, which are used in our proposed algorithms, and they are derived from the FDOA by multiplying with λ_c.

- **AOA**: For the LOS path, we get

$$\begin{aligned} \phi_n^\circ &= \phi_{n,0}^\circ = \arctan \tfrac{y^\circ - y_n^b}{x^\circ - x_n^b}, \\ \theta_n^\circ &= \theta_{n,0}^\circ = \arcsin \tfrac{z^\circ - z_n^b}{||\mathbf{u}^\circ - \mathbf{b}_n||}. \end{aligned} \tag{10}$$

Then, for the NLOS path, we have

$$\begin{aligned} \phi_{n,l}^{s\circ} &= \phi_{n,l}^\circ = \arctan \tfrac{y_{n,l}^{s\circ} - y_n^b}{x_{n,l}^{s\circ} - x_n^b}, \\ \theta_{n,l}^{s\circ} &= \theta_{n,l}^\circ = \arcsin \tfrac{z_{n,l}^{s\circ} - z_n^b}{||\mathbf{s}_{n,l}^\circ - \mathbf{b}_n||}. \end{aligned} \tag{11}$$

Thus, $(\phi_n^\circ, \theta_n^\circ)$ and $(\phi_{n,l}^{s\circ}, \theta_{n,l}^{s\circ})$ are the AOA-related parameters.

Summarizing, the relationships between TDOA/ FDOA/ AOA-related channel and location parameters are given in (3), (5), (7), (9), (10), and (11). In the following sections, we focus on developing effective algorithms to estimate the unknown location and velocity of the UE and scatterers as

accurate as possible by utilizing hybrid TDOA/FDOA/AOA measurements. This task is not trivial given that the relations are nonlinear and nonconvex functions of $\mathbf{u}^\circ$, $\dot{\mathbf{u}}^\circ$, $\mathbf{s}_{n,l}^\circ$, and $\dot{\mathbf{s}}_{n,l}^\circ$.

III. Problem Formulation

A. Measurement Selection

Each RRH obtains a set of measurements, $\mathbb{M}_n = \{(\phi_{n,m}, \theta_{n,m}, \tau_{n,m}, \nu_{n,m}) | m = 1, 2, \ldots, M_n\}$ for $n = 1, 2, \ldots, N$, where M_n is the estimated number of paths, which is $L_n + 1$ for a perfect value. In this subsection, we propose a method to select measurements of N_a LOS paths from all obtained $\sum_{n=1}^{N} M_n$ paths.

First, select the m_n^*-th path in $\mathbb{M}_n$, which has the smallest delay (probably the LOS path) among M_n paths, for $n = 1, 2, \ldots, N$. Second, we define the rough estimate of $\mathbf{u}^\circ$ by the RRH n to further eliminate NLOS paths in the chosen N paths as follows

$$\hat{\mathbf{u}}_n = \mathbf{b}_n + v_c \tau_{n,m_n^*} [\cos\theta_{n,m_n^*} \cos\phi_{n,m_n^*}, \cos\theta_{n,m_n^*} \sin\phi_{n,m_n^*}, \sin\theta_{n,m_n^*}]^T, \quad (12)$$

for $n = 1, 2, \ldots, N$. Points in set $\{\hat{\mathbf{u}}_n | n = 1, 2, \ldots, N\}$ are close and dispersed to one another if they are generated by LOS and NLOS measurements, respectively. Subsequently, we classify $\hat{\mathbf{u}}_n$ for $n = 1, 2, \ldots, N$ into two classes by K-means algorithm and obtain two class centers, namely, $\mathbf{c}_{\text{LOS}}$ and $\mathbf{c}_{\text{NLOS}}$. As the energy of LOS paths is much greater than that of NLOS paths in mmWave frequencies, we can further eliminate the NLOS paths in $\mathbf{c}_{\text{LOS}}$. A threshold is set according to the energy gap between the LOS and NLOS paths to determine the value of N_a. Then, the set of selected LOS measurements is $\mathbb{M}_a$. The set of remaining measurements is $\mathbb{M}_{r,n}$, where $\mathbb{M}_{r,n} \cap \mathbb{M}_a = \emptyset$ for $n = 1, 2, \ldots, N$.

We aim to estimate the unknown $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$ from the measurements in $\mathbb{M}_a$ (LOS measurements) and the unknown $\mathbf{s}_{n,l}^\circ$ and $\dot{\mathbf{s}}_{n,l}^\circ$ from the measurements in $\mathbb{M}_{r,n}$ (nearly all NLOS measurements) for $l = 1, 2, \ldots, |\mathbb{M}_{r,n}|$ and $n = 1, 2, \ldots, N$ as accurately as possible.

B. Possible Solution

After the measurement selection process, the corresponding measurements in sets $\mathbb{M}_a$ and $\mathbb{M}_{r,n}$ can be fed into black box NNs and trained end-to-end using real datasets to learn $\mathbf{x}^\circ = [\mathbf{u}^{\circ T}, \dot{\mathbf{u}}^{\circ T}]^T$ and $\mathbf{x}_{n,l}^{s\circ} = [\mathbf{s}_{n,l}^{\circ T}, \dot{\mathbf{s}}_{n,l}^{\circ T}]^T$ directly, respectively (Fig. 2(a)). However, the localization accuracy of this method is limited, and a prohibitively large amount of training data is required to improve the localization accuracy. To skip this step, our strategy is based on the argument that the model is mathematically well developed with fewer uncertainties [37]. However, the model generally relies on some approximations and ideal assumptions, which worsen the performance when the measurement noise increases. Motivated by the powerful learning ability of the NN, its use to replace the approximate operations in the model can further improve the performance. Therefore, we combine NNs with geometric models in this study. Specifically, we first develop an unbiased model-based WLS localization estimator (Section IV). Then, we establish a NN-assisted WLS localization method (Section V) by introducing NNs into the developed WLS model (or estimator) to learn the higher-order error components, thereby improving the performance of the estimator, especially in a large noisy environment.

IV. Model-based WLS Localization

In this section, we devise a closed-form localization estimator that approximates the maximum likelihood (ML) estimator under small noise conditions. We improve the performance of the traditional multi-stage WLS estimator [27] by exploiting angular information and establishing a one-stage WLS estimator. We further extend the traditional WLS localization estimator that can only be used for UE localization to the scatterers' localization.

Measurements in $\mathbb{M}_a$ are used to estimate the location and velocity of the UE. According to (3), (7), and (10), we denote a noise-free vector of hybrid TDOA/FDOA/AOA parameters as $\mathbf{m}^\circ = [r_{21}^\circ, \dot{r}_{21}^\circ, \ldots, r_{N_a1}^\circ, \dot{r}_{N_a1}^\circ, \phi_1^\circ, \theta_1^\circ, \ldots, \phi_{N_a}^\circ, \theta_{N_a}^\circ]^T$. Then, we model the hybrid measurements by the additive noise model as $\mathbf{m} = \mathbf{m}^\circ + \Delta\mathbf{m}$, where $\mathbf{m} = [r_{21}, \dot{r}_{21}, \ldots, r_{N_a1}, \dot{r}_{N_a1}, \phi_1, \theta_1, \ldots, \phi_{N_a}, \theta_{N_a}]^T$, and $\Delta\mathbf{m} = [\Delta r_{21}, \Delta\dot{r}_{21}, \ldots, \Delta r_{N_a1}, \Delta\dot{r}_{N_a1}, \Delta\phi_1, \Delta\theta_1, \ldots, \Delta\phi_{N_a}, \Delta\theta_{N_a}]^T$ is a Gaussian noise vector with zero mean and covariance matrix $\mathbf{Q}$. Measurements in $\mathbb{M}_{r,n}$ are used to estimate the location and velocity of scatterers for $n = 1, 2, \ldots, N$. According to (5), (9), and (11), we obtain the l-th noise-free vector of hybrid parameters as $\mathbf{m}_{n,l}^{s\circ} = [r_{n1,l}^{s\circ}, \dot{r}_{n1,l}^{s\circ}, \phi_{n,l}^{s\circ}, \theta_{n,l}^{s\circ}]^T$, and hybrid measurements $\mathbf{m}_{n,l}^s = \mathbf{m}_{n,l}^{s\circ} + \Delta\mathbf{m}_{n,l}^s$, where $\mathbf{m}_{n,l}^s = [r_{n1,l}^s, \dot{r}_{n1,l}^s, \phi_{n,l}^s, \theta_{n,l}^s]^T$, and $\Delta\mathbf{m}_{n,l}^s = [\Delta r_{n1,l}^s, \Delta\dot{r}_{n1,l}^s, \Delta\phi_{n,l}^s, \Delta\theta_{n,l}^s]^T$ with zero mean and covariance matrix $\mathbf{Q}_{n,l}^s$, for $l = 1, 2, \ldots, |\mathbb{M}_{r,n}|$.

A. UE Localization

In this subsection, we present a closed-form method for estimating the UE location and velocity. We first establish a set of pseudo-linear TDOA and FDOA equations by nonlinear transformation and AOA exploitation. Subsequently, AOA equations are derived and combined with TDOA and FDOA equations to obtain an accurate estimation.

First, we derive $2(N_a - 1)$ pseudo-linear TDOA and FDOA equations. We rewrite (3) as $r_{n1}^\circ + r_1^\circ = r_n^\circ$ and square both sides to yield $(r_{n1}^\circ)^2 + 2r_{n1}^\circ r_1^\circ = (r_n^\circ)^2 - (r_1^\circ)^2$. According to (2), we obtain

$$(r_{n1}^\circ)^2 + 2r_{n1}^\circ r_1^\circ = \mathbf{b}_n^T\mathbf{b}_n - \mathbf{b}_1^T\mathbf{b}_1 - 2(\mathbf{b}_n - \mathbf{b}_1)^T\mathbf{u}^\circ. \quad (13)$$

Equation (13) is pseudo-linear formula with respect to $\mathbf{u}^\circ$ and r_1°. Then, by taking the time derivative of (13), we yield

$$\dot{r}_{n1}^\circ r_{n1}^\circ + \dot{r}_{n1}^\circ r_1^\circ + r_{n1}^\circ \dot{r}_1^\circ = (\mathbf{b}_1 - \mathbf{b}_n)^T\dot{\mathbf{u}}^\circ. \quad (14)$$

Equation (14) is pseudo-linear formula with respect to $\dot{\mathbf{u}}^\circ$, r_1°, and $\dot{r}_1^\circ$. However, r_1° and $\dot{r}_1^\circ$ cannot be obtained directly from the channel measurements TDOA and FDOA. A well-known solution for localization that uses TDOAs and FDOAs is to find $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$ by using multi-stage WLS estimators [27]. The conventional method is based on the estimation of

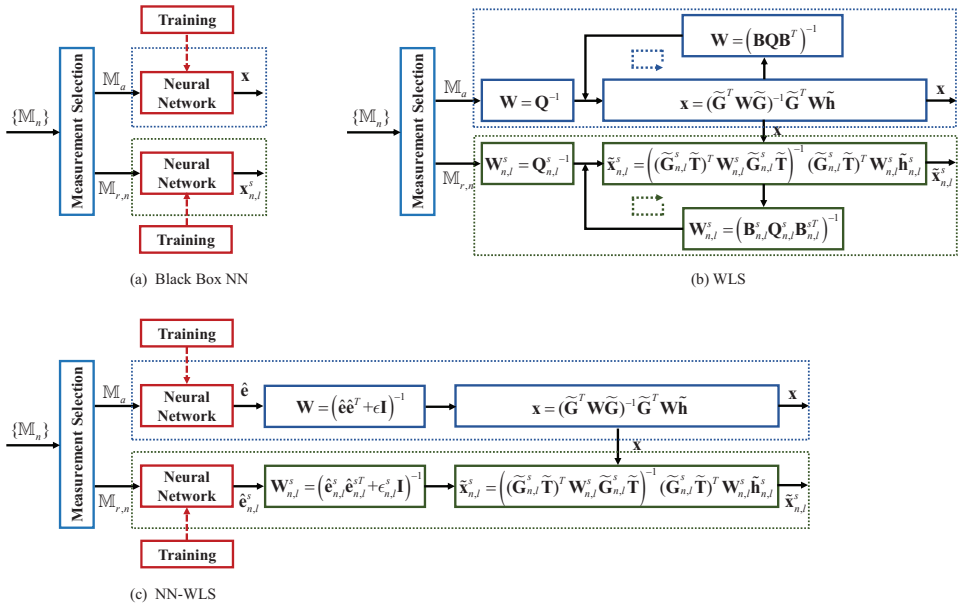


Fig. 2. Block diagrams of (a) Black Box NN; (b) WLS; (c) NN-WLS, where $\mathbf{x} = [\mathbf{u}^T, \dot{\mathbf{u}}^T]^T$ and $\mathbf{x}^s_{n,l} = [\mathbf{s}^T_{n,l}, \dot{\mathbf{s}}^T_{n,l}]^T$ denote the estimated location and velocity of UE and scatterer, respectively.

the redundant parameters, namely, r_1° and $\dot{r}_1^\circ$, together with the UE location and velocity. In the present study, we apply a different approach, that is, we use AOA measurements to eliminate the redundant parameters to estimate $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$ in only one stage.

To eliminate r_1° and $\dot{r}_1^\circ$ in (13) and (14), we define $\mathbf{a}_1^\circ = [\cos\theta_1^\circ \cos\phi_1^\circ, \cos\theta_1^\circ \sin\phi_1^\circ, \sin\theta_1^\circ]^T$, which is a unit-norm angular vector that possesses the properties: $\mathbf{a}_1^{\circ T}\mathbf{a}_1^\circ = 1$ and $\dot{\mathbf{a}}_1^{\circ T}\mathbf{a}_1^\circ = \mathbf{a}_1^{\circ T}\dot{\mathbf{a}}_1^\circ = 0$. Multiplying both sides of (13) by $\mathbf{a}_1^{\circ T}\mathbf{a}_1^\circ$ and utilizing the geometric relationship $\mathbf{u}^\circ - \mathbf{b}_1 = r_1^\circ \mathbf{a}_1^\circ$ yield

$$(r_{n1}^\circ)^2 - 2r_{n1}^\circ \mathbf{a}_1^{\circ T}\mathbf{b}_1 - \mathbf{b}_n^T\mathbf{b}_n + \mathbf{b}_1^T\mathbf{b}_1 = 2[(\mathbf{b}_1 - \mathbf{b}_n)^T - r_{n1}^\circ \mathbf{a}_1^{\circ T}]\mathbf{u}^\circ. \quad (15)$$

Multiplying both sides of (14) by $\mathbf{a}_1^{\circ T}\mathbf{a}_1^\circ$ and utilizing the geometric relationship $\dot{\mathbf{u}}^\circ = \dot{r}_1^\circ \mathbf{a}_1^\circ + r_1^\circ \dot{\mathbf{a}}_1^\circ$ yield

$$\dot{r}_{n1}^\circ r_{n1}^\circ - \dot{r}_{n1}^\circ \mathbf{a}_1^{\circ T}\mathbf{b}_1 = -\dot{r}_{n1}^\circ \mathbf{a}_1^{\circ T}\mathbf{u}^\circ + [(\mathbf{b}_1 - \mathbf{b}_n)^T - r_{n1}^\circ \mathbf{a}_1^{\circ T}]\dot{\mathbf{u}}^\circ. \quad (16)$$

By collecting (15) and (16), for $n = 2, \ldots, N_a$, $2(N_a - 1)$ pseudo-linear TDOA and FDOA equations are obtained. Then, we derive $2N_a$ AOA equations for $n = 1, 2, \ldots, N_a$ according to (10), which is given by

$$\mathbf{c}_n^{\circ T}\mathbf{b}_n = \mathbf{c}_n^{\circ T}\mathbf{u}^\circ, \quad \mathbf{d}_n^{\circ T}\mathbf{b}_n = \mathbf{d}_n^{\circ T}\mathbf{u}^\circ, \quad (17)$$

where $\mathbf{c}_n^\circ = [-\sin\phi_n^\circ, \cos\phi_n^\circ, 0]^T$ and $\mathbf{d}_n^\circ = [-\sin\theta_n^\circ \cos\phi_n^\circ, -\sin\theta_n^\circ \sin\phi_n^\circ, \cos\theta_n^\circ]^T$. Collecting (15), (16) for $n = 2, \ldots, N_a$ and (17) for $n = 1, 2, \ldots, N_a$ yields the following matrix equation,

$$\mathbf{h} = \mathbf{G}\mathbf{x}^\circ, \quad (18)$$

where $\mathbf{x}^\circ = [\mathbf{u}^{\circ T}, \dot{\mathbf{u}}^{\circ T}]^T$ is an unknown six-dimensional vector of location and velocity of the UE, and

$$\mathbf{h} = [\mathbf{q}_2^T, \ldots, \mathbf{q}_{N_a}^T, \mathbf{h}_1^T, \ldots, \mathbf{h}_{N_a}^T]^T, \quad \mathbf{G} = [\mathbf{P}_2^T, \ldots, \mathbf{P}_{N_a}^T, \mathbf{G}_1^T, \ldots, \mathbf{G}_{N_a}^T]^T, \quad (19)$$

$$\mathbf{q}_n = \begin{pmatrix} (r_{n1}^\circ)^2 - 2r_{n1}^\circ \mathbf{a}_1^{\circ T}\mathbf{b}_1 - \mathbf{b}_n^T\mathbf{b}_n + \mathbf{b}_1^T\mathbf{b}_1 \\ \dot{r}_{n1}^\circ r_{n1}^\circ - \dot{r}_{n1}^\circ \mathbf{a}_1^{\circ T}\mathbf{b}_1 \end{pmatrix}, \quad \mathbf{h}_n = \begin{pmatrix} \mathbf{c}_n^{\circ T}\mathbf{b}_n \\ \mathbf{d}_n^{\circ T}\mathbf{b}_n \end{pmatrix}, \quad (20)$$

$$\mathbf{P}_n = \begin{pmatrix} 2[(\mathbf{b}_1 - \mathbf{b}_n)^T - r_{n1}^\circ \mathbf{a}_1^{\circ T}] & \mathbf{0}^T \\ -\dot{r}_{n1}^\circ \mathbf{a}_1^{\circ T} & (\mathbf{b}_1 - \mathbf{b}_n)^T - r_{n1}^\circ \mathbf{a}_1^{\circ T} \end{pmatrix}, \quad \mathbf{G}_n = \begin{pmatrix} \mathbf{c}_n^{\circ T} & \mathbf{0}^T \\ \mathbf{d}_n^{\circ T} & \mathbf{0}^T \end{pmatrix}, \quad (21)$$

where $\mathbf{0}$ is a 3×1 zero vector. Equation (18) is the noise-free matrix representation of the joint location and velocity estimation model.

The noise-free parameters in vector $\mathbf{h}$ and matrix $\mathbf{G}$ in (18) are not available. Let the noisy measurements replace the noise-free parameters in $\mathbf{h}$ and $\mathbf{G}$ (i.e., let $r_{i1} = r_{i1}^\circ + \Delta r_{i1}$, $\dot{r}_{i1} = \dot{r}_{i1}^\circ + \Delta\dot{r}_{i1}$, $\phi_j = \phi_j^\circ + \Delta\phi_j$, and $\theta_j = \theta_j^\circ + \Delta\theta_j$ replace r_{i1}°, $\dot{r}_{i1}^\circ$, ϕ_j°, and θ_j°, for $i = 2, \ldots, N_a$ and $j = 1, \ldots, N_a$), we define the error vector

$$\mathbf{e} = \tilde{\mathbf{h}} - \tilde{\mathbf{G}}\mathbf{x}^\circ, \quad (22)$$

where $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{G}}$ are the noisy counterparts. The WLS solution [38] of $\mathbf{x}^\circ$ can be obtained as

$$\mathbf{x} = (\tilde{\mathbf{G}}^T\mathbf{W}\tilde{\mathbf{G}})^{-1}\tilde{\mathbf{G}}^T\mathbf{W}\tilde{\mathbf{h}}, \quad (23)$$

where the weighting matrix $\mathbf{W} = (\mathbb{E}\{\mathbf{e}\mathbf{e}^T\})^{-1}$. In view of the nonlinearity of $\mathbf{e}$, obtaining the weighting matrix $\mathbf{W}$ is difficult in general. By ignoring the second- and higher-order

noise terms, we approximate $\mathbf{e}$ with its linear terms as

$$\mathbf{e} \approx \mathbf{B}\Delta\mathbf{m}, \tag{24}$$

where

$$\begin{aligned}
&\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{O} & \mathbf{B}_3 \end{bmatrix}, \\
&\mathbf{B}_1 = \text{blkdiag}\left(\begin{bmatrix} 2r_2^\circ & 0 \\ \dot{r}_2^\circ & r_2^\circ \end{bmatrix}, \ldots, \begin{bmatrix} 2r_{N_a}^\circ & 0 \\ \dot{r}_{N_a}^\circ & r_{N_a}^\circ \end{bmatrix}\right), \\
&\mathbf{B}_2 = \begin{bmatrix} \mathbf{B}_{21} & \mathbf{O} \end{bmatrix}, \\
&\mathbf{B}_{21} = \begin{bmatrix} 0 & 0; & a_2 & b_2; & \ldots; & 0 & 0; & a_{N_a} & b_{N_a} \end{bmatrix}, \\
&\mathbf{B}_3 = \text{diag}\left(r_1^\circ \cos\theta_1^\circ, r_1^\circ, \ldots, r_{N_a}^\circ \cos\theta_{N_a}^\circ, r_{N_a}^\circ\right),
\end{aligned} \tag{25}$$

in which the ";" operator separates the rows in a matrix; $a_n = r_1^\circ r_{n1}^\circ \dot{\phi}_1^\circ \cos^2\theta_1^\circ$ and $b_n = r_1^\circ r_{n1}^\circ \dot{\theta}_1^\circ$ for $n = 2, \ldots, N_a$; $\dot{\phi}_1^\circ = \mathbf{c}_1^{\circ T}\dot{\mathbf{u}}^\circ/(r_1^\circ \cos\theta_1^\circ)$ and $\dot{\theta}_1^\circ = \dot{\mathbf{u}}^{\circ T}\mathbf{d}_1^\circ/r_1^\circ$ are the time derivatives of (10) with $n = 1$. The detailed derivations of (24) are listed in Appendix A. As we approximate $\mathbf{e}$ up to its linear noise term $\mathbf{B}\Delta\mathbf{m}$, it follows from the distribution of $\Delta\mathbf{m}$ that $\mathbf{e}$ is a zero-mean Gaussian vector with covariance matrix $\mathbf{BQB}^T$. Therefore, the weighting matrix can be easily calculated as

$$\mathbf{W} = \left(\mathbf{BQB}^T\right)^{-1}, \tag{26}$$

where the weighting matrix $\mathbf{W}$ is dependent on the unknown location $\mathbf{u}^\circ$ and velocity $\dot{\mathbf{u}}^\circ$ via the matrix $\mathbf{B}$. Hence, we initialize $\mathbf{W} = \mathbf{Q}^{-1}$ to provide the initial location and velocity estimates. Updating this initial solution in $\mathbf{B}$ can construct a more accurate weighting matrix by (26) to derive the final solutions of $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$.

B. Scatterer Localization

In this subsection, we present a closed-form method for estimating the scatterers' location and velocity. We take the l-th scatterer between the n-th RRH and the UE for example, where $1 \leqslant l \leqslant |\mathbb{M}_{r,n}|$. First, let $d_{1,n,l}^\circ = ||\mathbf{s}_{n,l}^\circ - \mathbf{b}_n||$, $d_{2,n,l}^\circ = ||\mathbf{u}^\circ - \mathbf{s}_{n,l}^\circ||$, and we have $r_{n,l}^{s\circ} = d_{1,n,l}^\circ + d_{2,n,l}^\circ$. By rewriting (5) as $r_{n1,l}^{s\circ} + r_1^\circ - d_{1,n,l}^\circ = d_{2,n,l}^\circ$, squaring both sides, and making some simplifications, we obtain

$$\begin{aligned}
&(r_{n1,l}^{s\circ} + r_1^\circ)^2 - 2(r_{n1,l}^{s\circ} + r_1^\circ)d_{1,n,l}^\circ \\
&\qquad = \mathbf{u}^{\circ T}\mathbf{u}^\circ - 2\mathbf{u}^{\circ T}\mathbf{s}_{n,l}^\circ + 2\mathbf{b}_n^T\mathbf{s}_{n,l}^\circ - \mathbf{b}_n^T\mathbf{b}_n.
\end{aligned} \tag{27}$$

Then, by taking the time derivative of (27), we have

$$\begin{aligned}
&(r_{n1,l}^{s\circ}+r_1^\circ)(\dot{r}_{n1,l}^{s\circ}+\dot{r}_1^\circ)-(\dot{r}_{n1,l}^{s\circ}+\dot{r}_1^\circ)d_{1,n,l}^\circ-(r_{n1,l}^{s\circ}+r_1^\circ)\dot{d}_{1,n,l}^\circ \\
&\qquad = \dot{\mathbf{u}}^{\circ T}\mathbf{u}^\circ - \dot{\mathbf{u}}^{\circ T}\mathbf{s}_{n,l}^\circ - \mathbf{u}^{\circ T}\dot{\mathbf{s}}_{n,l}^\circ + \mathbf{b}_n^T\dot{\mathbf{s}}_{n,l}^\circ,
\end{aligned} \tag{28}$$

where $\dot{d}_{1,n,l}^\circ$ is the time derivation of $d_{1,n,l}^\circ$. By utilizing the AOA parameters, together with estimated $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$ in Section IV-A, we can eliminate the redundant parameters ($d_{1,n,l}^\circ$ and $\dot{d}_{1,n,l}^\circ$) in (27) and (28) to estimate $\mathbf{s}_{n,l}^\circ$ and $\dot{\mathbf{s}}_{n,l}^\circ$ in one stage. r_1° and $\dot{r}_1^\circ$ are obtained by estimated $\mathbf{u}^\circ$ and $\dot{\mathbf{u}}^\circ$, thus, $r_{n,l}^{s\circ} = r_{n1,l}^{s\circ} + r_1^\circ$ and $\dot{r}_{n,l}^{s\circ} = \dot{r}_{n1,l}^{s\circ} + \dot{r}_1^\circ$ are obtained. By defining $\mathbf{a}_{n,l}^{s\circ} = [\cos\theta_{n,l}^{s\circ}\cos\phi_{n,l}^{s\circ}, \cos\theta_{n,l}^{s\circ}\sin\phi_{n,l}^{s\circ}, \sin\theta_{n,l}^{s\circ}]^T$, eliminating $d_{1,n,l}^\circ$ and $\dot{d}_{1,n,l}^\circ$ in (27) and (28), and combining AOA equations, we obtain the following matrix representation,

$$\mathbf{h}_{n,l}^s = \mathbf{G}_{n,l}^s \mathbf{x}_{n,l}^{s\circ}, \tag{29}$$

where

$$\mathbf{h}_{n,l}^s = \begin{pmatrix} (r_{n,l}^{s\circ})^2 + 2r_{n,l}^{s\circ}\mathbf{a}_{n,l}^{s\circ\,T}\mathbf{b}_n - \mathbf{u}^{\circ T}\mathbf{u}^\circ + \mathbf{b}_n^T\mathbf{b}_n \\ r_{n,l}^{s\circ}\dot{r}_{n,l}^{s\circ} + \dot{r}_{n,l}^{s\circ}\mathbf{a}_{n,l}^{s\circ\,T}\mathbf{b}_n - \dot{\mathbf{u}}^{\circ T}\mathbf{u}^\circ \\ \mathbf{c}_{n,l}^{s\circ\,T}\mathbf{b}_n \\ \mathbf{d}_{n,l}^{s\circ\,T}\mathbf{b}_n \end{pmatrix},$$

$$\mathbf{G}_{n,l}^s = \begin{pmatrix} 2(\mathbf{b}_n - \mathbf{u}^\circ + r_{n,l}^{s\circ}\mathbf{a}_{n,l}^{s\circ})^T & \mathbf{0}^T \\ (\dot{r}_{n,l}^{s\circ}\mathbf{a}_{n,l}^{s\circ} - \dot{\mathbf{u}}^\circ)^T & (r_{n,l}^{s\circ}\mathbf{a}_{n,l}^{s\circ} + \mathbf{b}_n - \mathbf{u}^\circ)^T \\ \mathbf{c}_{n,l}^{s\circ\,T} & \mathbf{0}^T \\ \mathbf{d}_{n,l}^{s\circ\,T} & \mathbf{0}^T \end{pmatrix},$$

$\mathbf{c}_{n,l}^{s\circ} = [-\sin\phi_{n,l}^{s\circ}, \cos\phi_{n,l}^{s\circ}, 0]^T$, $\mathbf{d}_{n,l}^{s\circ} = [-\sin\theta_{n,l}^{s\circ}\cos\phi_{n,l}^{s\circ}, -\sin\theta_{n,l}^{s\circ}\sin\phi_{n,l}^{s\circ}, \cos\theta_{n,l}^{s\circ}]^T$, and $\mathbf{x}_{n,l}^{s\circ} = [\mathbf{s}_{n,l}^{\circ T}, \dot{\mathbf{s}}_{n,l}^{\circ T}]^T$. However, four measurements are not enough for six unknowns. We assume that the moving scatterers are vehicles that move along the same road as the UE and we can regard the road as straight within a short distance. Thus, the direction of the scatterer velocity is aligned with the UE within a short distance. Let a unit vector $\mathbf{n}_v = \dot{\mathbf{u}}^\circ/\|\dot{\mathbf{u}}^\circ\|$ denote the direction of the UE velocity. When $\dot{\mathbf{u}}^\circ$ is estimated in Section IV-A, $\mathbf{n}_v$ is obtained. We have $\dot{\mathbf{s}}_{n,l}^\circ = \dot{s}_{n,l}^\circ \mathbf{n}_v$, where $\dot{s}_{n,l}^\circ$ represents the magnitude of velocity. With a transformation matrix,

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_{3\times3} & \mathbf{0} \\ \mathbf{O}_{3\times3} & \mathbf{n}_v \end{pmatrix}, \tag{30}$$

we obtain

$$\mathbf{h}_{n,l}^s = \mathbf{G}_{n,l}^s \mathbf{T}\tilde{\mathbf{x}}_{n,l}^{s\circ}, \tag{31}$$

where $\tilde{\mathbf{x}}_{n,l}^{s\circ} = [\mathbf{s}_{n,l}^{\circ T}, \dot{s}_{n,l}^\circ]^T$ is an unknown four-dimensional vector of location and velocity magnitude of the scatterer. [1] Replacing the noise-free parameters $\{r_{n1,l}^{s\circ}, \dot{r}_{n1,l}^{s\circ}, \phi_{n,l}^{s\circ}, \theta_{n,l}^{s\circ}, \mathbf{u}^\circ, \dot{\mathbf{u}}^\circ\}$ in (31) by the noisy measurements $\{r_{n1,l}^s, \dot{r}_{n1,l}^s, \phi_{n,l}^s, \theta_{n,l}^s\}$ and estimated $\{\mathbf{u}, \dot{\mathbf{u}}\}$ results in the error vector

$$\mathbf{e}_{n,l}^s = \tilde{\mathbf{h}}_{n,l}^s - \tilde{\mathbf{G}}_{n,l}^s \mathbf{T}\tilde{\mathbf{x}}_{n,l}^{s\circ}. \tag{32}$$

By approximating $\mathbf{e}_{n,l}^s$ up to the linear noise term, we have $\mathbf{e}_{n,l}^s \approx \mathbf{B}_{n,l}^s \Delta\mathbf{m}_{n,l}^s$, where

$$\mathbf{B}_{n,l}^s = \begin{pmatrix} 2d_{2,n,l}^\circ & 0 & 0 & 0 \\ \dot{d}_{2,n,l}^\circ & d_{2,n,l}^\circ & -r_{n,l}^{s\circ}d_{1,n,l}^\circ\dot{\phi}_{n,l}^{s\circ}\cos^2\phi_{n,l}^{s\circ} & -r_{n,l}^{s\circ}d_{1,n,l}^\circ\dot{\theta}_{n,l}^{s\circ} \\ 0 & 0 & d_{n1}^\circ\cos\theta_n^{s\circ} & 0 \\ 0 & 0 & 0 & d_{n1}^\circ \end{pmatrix}, \tag{33}$$

and $\dot{\phi}_{n,l}^{s\circ} = \mathbf{c}_{n,l}^{s\circ T}\dot{\mathbf{s}}_{n,l}^\circ/(d_{1,n,l}^\circ\cos\theta_{n,l}^{s\circ})$, $\dot{\theta}_{n,l}^{s\circ} = \dot{\mathbf{s}}_{n,l}^{\circ T}\mathbf{d}_1^\circ/d_{1,n,l}^\circ$. The derivations of (33) are similar to those in Appendix A, and we omit these details because of lack of space in this

[1] A scatterer can be in the opposite direction of a UE, because $\dot{s}_{n,l}^\circ$ can be negative. Moreover, we can judge whether the assumption that the scatterer is on the same road as the UE is met by comparing the estimated scatterer location with a rough offline map. If the assumption is satisfied, then we will believe the corresponding velocity estimate.

paper. Thus, the WLS solution of $\tilde{\mathbf{x}}_{n,l}^{s\circ}$ is given by

$$\tilde{\mathbf{x}}_{n,l}^{s} = \left((\tilde{\mathbf{G}}_{n,l}^{s}\tilde{\mathbf{T}})^T\mathbf{W}_{n,l}^{s}\tilde{\mathbf{G}}_{n,l}^{s}\tilde{\mathbf{T}}\right)^{-1}(\tilde{\mathbf{G}}_{n,l}^{s}\tilde{\mathbf{T}})^T\mathbf{W}_{n,l}^{s}\tilde{\mathbf{h}}_{n,l}^{s}, \tag{34}$$

where $\mathbf{W}_{n,l}^{s} = \left(\mathbf{B}_{n,l}^{s}\mathbf{Q}_{n,l}^{s}\mathbf{B}_{n,l}^{sT}\right)^{-1}$. The weighting matrix $\mathbf{W}_{n,l}^{s}$ is dependent on $\mathbf{s}_{n,l}^{\circ}$ and $\dot{\mathbf{s}}_{n,l}^{\circ}$ through $\mathbf{B}_{n,l}^{s}$. At the beginning, we can use $\mathbf{W}_{n,l}^{s} = (\mathbf{Q}_{n,l}^{s})^{-1}$ in (34) to produce a solution from which to generate a better $\mathbf{W}_{n,l}^{s}$ to yield a more accurate solution.

C. Discussion

The proposed model-based localization method is summarized in Algorithm 1 and Fig. 2(b). Repeating the solution computation one to two times in Algorithm 1 (b) and (c) is sufficient to yield an accurate solution that reaches the CRLB for small Gaussian noise.

Remark 1. *Ignoring the second- and higher-order noise terms, we yield* $\mathbb{E}\{\mathbf{x}\} \approx \mathbf{x}^{\circ}$ *and* $\mathbb{E}\{\tilde{\mathbf{x}}_{n,l}^{s}\} \approx \tilde{\mathbf{x}}_{n,l}^{s\circ}$. *Thus, the presented estimator is asymptotically unbiased. The covariance matrices are given by* $\mathrm{cov}(\mathbf{x}) \approx \left((\mathbf{B}^{-1}\mathbf{G})^T\mathbf{Q}^{-1}\mathbf{B}^{-1}\mathbf{G}\right)^{-1}$ *and* $\mathrm{cov}(\tilde{\mathbf{x}}_{n,l}^{s}) \approx \left((\mathbf{B}_{n,l}^{s}{}^{-1}\mathbf{G}_{n,l}^{s}\mathbf{T})^T\mathbf{Q}_{n,l}^{s}{}^{-1}\mathbf{B}_{n,l}^{s}{}^{-1}\mathbf{G}_{n,l}^{s}\mathbf{T}\right)^{-1}$, *which approach to their corresponding CRLB under small Gaussian noise levels. Refer to Appendix* B.

Remark 2. *The weighting matrix in WLS provides the relative importance of the components of an error vector to be minimized [38]. In the proposed method, the derived weighting matrices ignore the second- and higher-order error terms, which are non-negligible when the noise is large. To increase the robustness of the algorithm, the weighting matrices should include the second- and higher-order error components. An additional refinement mechanism is proposed in the following section to learn higher-order noise terms in a large noise environment by embedding NNs.*

V. NN-Assisted WLS Localization

The model-based WLS estimator proposed in Section IV is proven asymptotically unbiased and effective in achieving the CRLB under small noise conditions. The general assumption is that the measurement noise follows a Gaussian distribution. However, in reality, the measurement errors are not completely random. Moreover, an underlying relationship exists between them. Thus, by utilizing the powerful learning ability of NNs, this underlying relationship can be learned to further improve the localization performance of the proposed WLS estimator, especially at high noise levels.

In this section, we design a NN-assisted WLS (coined as NN-WLS) localization method that embeds NNs into the proposed WLS estimators in (23) and (34), thereby improving the localization performance. Different from treating the NN as a black box (Black Box NN) that directly learns location and velocity, the NNs in our approach are used to learn the residual vectors $\mathbf{e}$ in (22) and $\mathbf{e}_{n,l}^{s}$ in (32), respectively. Then, the estimated $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}_{n,l}^{s}$ are used to construct the weighting matrices $\mathbf{W}$ and $\mathbf{W}_{n,l}^{s}$ in (23) and (34) and then estimate $\mathbf{x}^{\circ}$ and $\tilde{\mathbf{x}}_{n,l}^{s\circ}$, respectively (Fig. 2(c)). The proposed NN-WLS method can derive more accurate results than the model-based WLS estimator and the Black Box NN method by learning the residual vectors. We also apply ensemble learning to improve the performance of the proposed NN-WLS method further.

Algorithm 1 : Pseudocode of the Proposed Model-based Localization Method

(a) Measurement Selection (Separate LOS and NLOS Measurements):

Require: $\mathbb{M}_n$ for $n = 1, 2, \ldots, N$.
Ensure: $\mathbb{M}_a$ and $\mathbb{M}_{r,n}$ for $n = 1, 2, \ldots, N$.

Choose the m_n^*-th path in $\mathbb{M}_n$ which has the smallest delay among M_n paths, for $n = 1, 2, \ldots, N$.
Calculate $\hat{\mathbf{u}}_n$ for $n = 1, 2, \ldots, N$, according to (12). Classify $\hat{\mathbf{u}}_n$ for $n = 1, 2, \ldots, N$ into two classes by K-means algorithm, and obtain two class centers $\mathbf{c}_{\mathrm{LOS}}$ and $\mathbf{c}_{\mathrm{NLOS}}$.
Calculate distance $d_n = ||\mathbf{c}_{\mathrm{LOS}} - \hat{\mathbf{u}}_n||$, and sort d_n for $n = 1, 2, \ldots, N$ in ascending order. Choose measurements of N_a paths corresponding to the first N_a smallest distances, and the set of selected measurements is $\mathbb{M}_a$. The set of remaining measurements is $\mathbb{M}_{r,n}$, for $n = 1, 2, \ldots, N$.

(b) UE Localization (Use LOS Measurements):

Require: $\mathbb{M}_a$.
Ensure: $\mathbf{x} = [\mathbf{u}^T, \dot{\mathbf{u}}^T]^T$.

Find $\mathbf{x}$ from (23) with $\mathbf{W} = \mathbf{Q}^{-1}$.
repeat
    Calculate the matrix $\mathbf{B}$ in (25) by the obtained $\mathbf{x}$.
    Update the weighting matrix $\mathbf{W}$ in (26) by the obtained $\mathbf{B}$.
    Find $\mathbf{x}$ from (23) with new $\mathbf{W}$.
until convergence

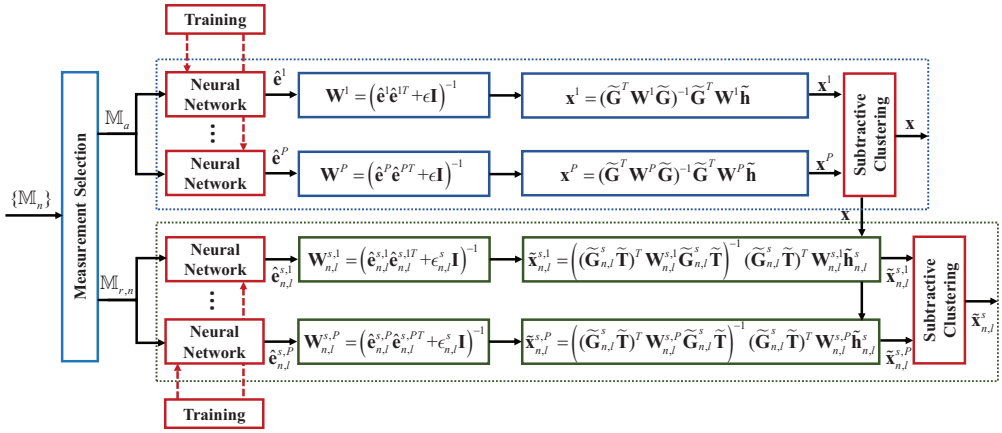
(c) Scatterer Localization (Use NLOS Measurements):

Require: $\mathbf{u}$, $\dot{\mathbf{u}}$, and $\mathbb{M}_{r,n}$, for $n = 1, 2, \ldots, N$.
Ensure: $\tilde{\mathbf{x}}_{n,l}^{s} = [\mathbf{s}_{n,l}^T, \dot{s}_{n,l}]^T$, for $l = 1, 2, \ldots, |\mathbb{M}_{r,n}|$ and $n = 1, 2, \ldots, N$.

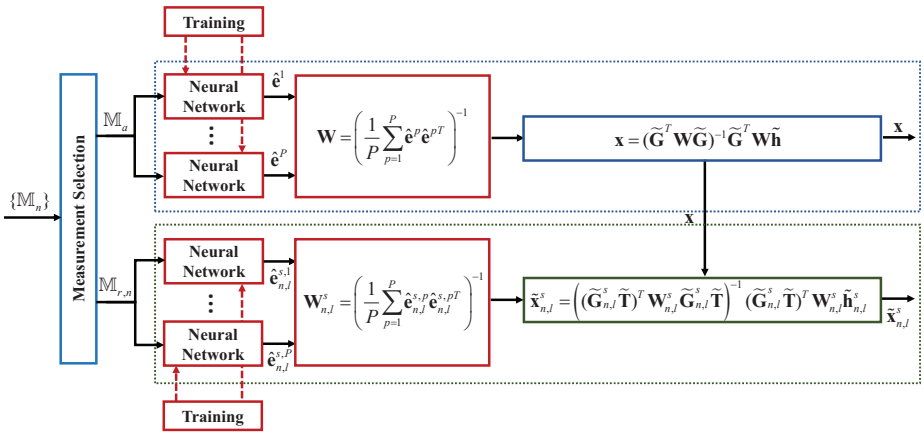
for $n = 1$ to N **do**
    for $l = 1$ to $|\mathbb{M}_{r,n}|$ **do**
        Find $\tilde{\mathbf{x}}_{n,l}^{s}$ from (34) with $\mathbf{W}_{n,l}^{s} = (\mathbf{Q}_{n,l}^{s})^{-1}$.
        repeat
            Calculate the matrix $\mathbf{B}_{n,l}^{s}$ in (33) by the obtained $\tilde{\mathbf{x}}_{n,l}^{s}$.
            Update the weighting matrix $\mathbf{W}_{n,l}^{s}$ by the obtained $\mathbf{B}_{n,l}^{s}$.
            Find $\tilde{\mathbf{x}}_{n,l}^{s}$ from (34) with new $\mathbf{W}_{n,l}^{s}$.
        until convergence
    end
end

A. NN-WLS

As shown in Fig. 2(c), the NN-WLS method is a revised version of the WLS estimator derived by introducing learnable vectors $\mathbf{e}$ and $\mathbf{e}_{n,l}^{s}$. We provide a general introduction here by taking $\mathbf{e}$ as an example. According to [38], the weighting matrix is given by $\mathbf{W} = (\mathbb{E}\{\mathbf{e}\mathbf{e}^T\})^{-1}$. In the WLS estimator proposed in Section IV, the vector $\mathbf{e}$ is approximated by the



(a) ENN-A-WLS



(b) ENN-B-WLS

Fig. 3. Block diagram of two ensemble learning-based NN-WLS localization methods.

linear term. Thus, the approximation error increases with the noise level, thereby compromises the algorithm performance. Therefore, we propose the NN-WLS method, in which we learn the vector $\mathbf{e}$ by a NN. The input of the NN is a measurement vector $\mathbf{m} \in \mathbb{R}^{2(N_a-1)+2N_a}$, which is generated by measurements in set $\mathbb{M}_a$ and given by

$$\mathbf{m} = [r_{21}, \dot{r}_{21}, \ldots, r_{N_a 1}, \dot{r}_{N_a 1}, \phi_1, \theta_1, \ldots, \phi_{N_a}, \theta_{N_a}]^T. \quad (35)$$

Here, the measurement noise is not necessarily Gaussian distributed. The output of the NN is the estimated residual vector $\hat{\mathbf{e}}$. Then, the estimated $\hat{\mathbf{e}}$ is used to construct $\mathbf{W}$ by

$$\mathbf{W} = (\hat{\mathbf{e}}\hat{\mathbf{e}}^T + \epsilon \mathbf{I})^{-1}, \quad (36)$$

where ϵ is a value to ensure that the inverse of $(\hat{\mathbf{e}}\hat{\mathbf{e}}^T + \epsilon \mathbf{I})$ exists. Finally, we obtain the estimate $\mathbf{x}$ by using the model in (23). In practice, the training dataset is constructed during an offline phase, in which a site survey is conducted to collect the vectors of the received signals of all RRHs from different UEs at numerous reference points of known locations, as given in (1). Then, the channel parameters are extracted from the received signals with signal processing methods. The extracted channel parameters construct measurement vector $\mathbf{m}$. Hence, we obtain label $\mathbf{e}$ corresponding to the known location with (22). Finally, training is performed on the basis of the T_{train} samples, with the structure of each sample as $(\mathbf{m}, \mathbf{e})$. For simulations, the location and velocity of the UE are randomly generated for each sample, then the measurement vector $\mathbf{m}$ is obtained accordingly by (35), and $\mathbf{e}$ is generated by (22). We consider the fully connected (FC) NN, and the input and output layers both have $4N_a - 2$ neurons. The input $(4N_a - 2)$-dimensional real-valued vector is initially normalized with the value of the element in $[0, 1]$. As for the rectified linear unit (ReLU), $\mathrm{ReLU}(x) = \max(x, 0)$ is used as the activation function for middle layers. The sigmoid function $\sigma(x) = 1/(1 + e^{-x})$ is used as the activation function in the final layer because the output is the normalized vector that has elements scaled within the $[0, 1]$ range. We generate the final estimation $\hat{\mathbf{e}}$ by rescaling. The set of parameters is updated by the ADAM algorithm. The loss function refers to the mean square error (MSE), which is given by

$$L(\Theta) = \frac{1}{T_{train}} \sum_{t=1}^{T_{train}} \|\hat{\mathbf{e}}_t - \mathbf{e}_t\|^2. \quad (37)$$

Similarly, for the learnable vector $\mathbf{e}^s_{n,l}$, the input of the NN is a measurement vector $\mathbf{m}^s_{n,l}$, which is generated by

measurements in set $\mathbb{M}_{n,l}^s$ and given by

$$\mathbf{m}_{n,l}^s = [r_{n1,l}^s, \dot{r}_{n1,l}^s, \phi_{n,l}^s, \theta_{n,l}^s]^T. \tag{38}$$

The output of the NN is the estimated residual vector $\hat{\mathbf{e}}_{n,l}^s$. Then, the estimated $\hat{\mathbf{e}}_{n,l}^s$ is used to construct $\mathbf{W}_{n,l}^s$ by

$$\mathbf{W}_{n,l}^s = (\hat{\mathbf{e}}_{n,l}^s \hat{\mathbf{e}}_{n,l}^{sT} + \epsilon_{n,l}^s \mathbf{I})^{-1}, \tag{39}$$

where $\epsilon_{n,l}^s$ is a value to ensure the existence of the inverse of $(\hat{\mathbf{e}}_{n,l}^s \hat{\mathbf{e}}_{n,l}^{sT} + \epsilon_{n,l}^s \mathbf{I})$. The previously predicted UE location and velocity in vector $\mathbf{x}$ are also used to construct the estimation model (34), by which we obtain the estimate $\tilde{\mathbf{x}}_{n,l}^s$. This part must be executed $\sum_{n=1}^{N} M_n - N_a$ times in parallel to localize all possible scatterers. The similar FC-NN architecture and training process are considered, except that the input and the output layers have four neurons.

Remark 3. *The proposed NN-WLS combines the NNs with the geometric model, thereby consolidating the powerful computing ability of NNs and the robustness of models. The particular advantages are presented as follows. First, the NNs can provide a more accurate estimation of* $\mathbf{e}$ *and* $\mathbf{e}_{n,l}^s$ *than the first-order approximation in the previously proposed WLS algorithms. Thus, in some practical scenarios, the NN-WLS can achieve good performance and can be executed even without knowing the covariance matrix* $\mathbf{Q}$ *and* $\mathbf{Q}_{n,l}^s$*, whereas the* $\mathbf{Q}$ *and* $\mathbf{Q}_{n,l}^s$ *in the WLS algorithms are assumed to be known to initialize the weighting matrix* $\mathbf{W}$ *and* $\mathbf{W}_{n,l}^s$*, respectively. Moreover, the WLS algorithm is iterative, which implies slow reconstruction, whereas the NN-WLS does not need any iterations, thereby reducing the required time resources.*

B. Ensemble Learning-based NN-WLS

Training the NN with the loss function defined in (37) cannot guarantee that the NN-WLS outputs the globally optimal estimator, even for sufficient data. According to [39], [40], the ensemble learning methods often provide additional performance enhancement. Ensemble methods correspond to learning algorithms that construct a set of learners and generate a new prediction by taking a vote of the predictions, which may be weighted. In the backpropagation algorithm for training the NNs, the initial weights of the networks are set randomly. If the algorithm is applied to the same training dataset but with different initial weights, then the resulting predictions may vary. NNs that are independently trained with the same training dataset have high probabilities of not making the same prediction error. Therefore, we can improve the performance of the NN-assisted WLS algorithm further by introducing an ensemble of P-independently trained NNs.

In this study, we propose two ensemble learning-based NN-WLS localization methods, namely, ENN-A-WLS and ENN-B-WLS, as illustrated in Fig. 3(a) and Fig. 3(b), respectively. The following instructions use the localization of UE $\mathbf{x}^\circ$ as an example. The similarity between the ENN-A-WLS and ENN-B-WLS lies in that both of them consist of P independently trained NNs by the same training dataset. The input of each NN is a measurement vector $\mathbf{m}$ given in (35) generated by measurements in set $\mathbb{M}_a$, and the output of each NN is the estimated $\hat{\mathbf{e}}^p$, for $p = 1, \ldots, P$. The difference is described as follows: As depicted in Fig. 3(a), the ENN-A-WLS repeats the NN-WLS P times, because P NNs are trained independently and in parallel, such that output of each NN-WLS is an independent prediction $\mathbf{x}^p$, for $p = 1, \ldots, P$. Accurate predictions of UE location are clustered together, and the wrong predictions are located far apart; such approach is also applied in UE velocity. We implement the core part of the ENN-A-WLS, which determines the voting mechanism, by the subtractive clustering. Unlike the simple averaging method, the performance of which seriously deteriorates by the effect of extremely abnormal predictions. The subtractive clustering method is based on a density measure. The density measure for the p-th location prediction is defined as

$$D_p = \sum_{j=1}^{P} \exp\left(-\|\mathbf{u}^p - \mathbf{u}^j\|^2/(r_a/2)^2\right), \tag{40}$$

where r_a is a positive value to denote the radius. The data points outside this radius only contribute slightly to the density measure. Therefore, by setting a proper r_a, the subtractive clustering method can find the point where the predicted values are most clustered. The point with the highest density measure is selected as the final estimate of UE location. UE velocity is obtained in the same way. As shown in Fig. 3(b), the ENN-B-WLS combines the output $\hat{\mathbf{e}}^p$ of each NN, for $p = 1, \ldots, P$, to construct the weighting matrix as

$$\mathbf{W} = \left(\frac{1}{P}\sum_{p=1}^{P} \hat{\mathbf{e}}^p \hat{\mathbf{e}}^{pT}\right)^{-1}, \tag{41}$$

which uses the average of finite P samples to approximate statistical $\mathbf{W} = (\mathbb{E}\{\mathbf{e}\mathbf{e}^T\})^{-1}$. Then, we obtain the estimate $\mathbf{x}$ by using the model in (23) with the constructed $\mathbf{W}$ in (41). Scatterers are localized in a similar way and further details are omitted.

VI. Numerical Results

A. Model-based WLS Localization

In this subsection, we analyze the performance of the proposed WLS estimator. We consider a scenario with $N = 18$ RRHs, and their locations are given in Table II. [2] The UE is located at $\mathbf{u}^\circ = [250, 450, 0]^T$ m with the velocity $\dot{\mathbf{u}}^\circ = [-10, 2, 5]^T$ m/s. The CU selects N_a LOS paths from RRHs to locate the UE. Although the presented algorithm is derived for Gaussian noise model with general covariance matrix, we consider the following form of the covariance matrix of the noise terms $\Delta\mathbf{m}$ for simplicity,

$$\mathbf{Q} = \text{blkdiag}(\overbrace{\mathbf{Q}_d, \ldots, \mathbf{Q}_d}^{(N_a-1)}, \overbrace{\mathbf{Q}_a, \ldots, \mathbf{Q}_a}^{N_a}), \tag{42}$$

where $\mathbf{Q}_d = \text{diag}(\delta_d^2, (0.1\delta_d)^2)$, $\mathbf{Q}_a = \text{diag}(\delta_a^2, \delta_a^2)$, and δ_d, $0.1\delta_d$, and δ_a are the standard deviations of TDOA, FDOA,

[2]Our proposed method (i.e., TDOA/FDOA/AOA) can work with either a randomized selection of z-axis coordinates for different RRHs or the same value of z-axis coordinates for all RRHs. To compare the performance of different methods, we choose a randomized selection of the z-axis coordinates for the first six RRHs.

TABLE II
LOCATIONS OF THE RRHS IN METERS.

	1	2	3	4	5	6	7	8	9
x	235.5042	287.5042	235.5042	287.5042	235.5042	287.5042	235.5042	287.5042	235.5042
y	389.5038	389.5038	489.5038	489.5038	589.5038	589.5038	851.5038	851.5038	651.5038
z	26	32	10	40	14	50	26	26	26

	10	11	12	13	14	15	16	17	18
x	287.5042	235.5042	287.5042	235.5042	287.5042	235.5042	287.5042	235.5042	287.5042
y	651.5038	751.5038	751.5038	851.5038	851.5038	951.5038	951.5038	1051.5038	1051.5038
z	26	26	26	26	26	26	26	26	26

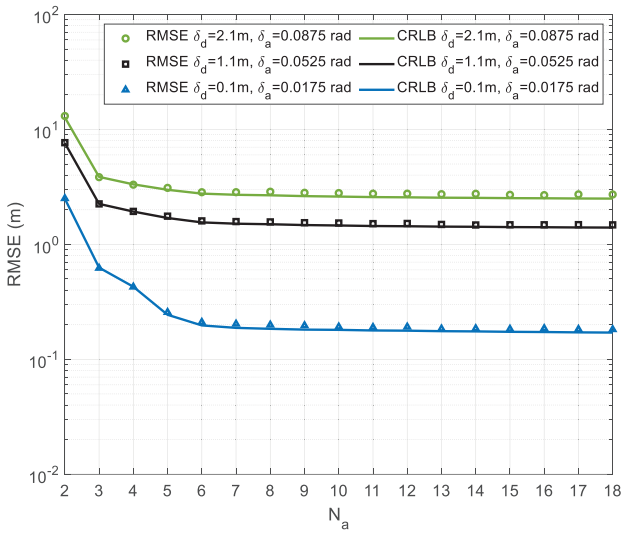


Fig. 4. RMSE performance of the proposed algorithm in location estimation with different numbers of selected LOS paths.

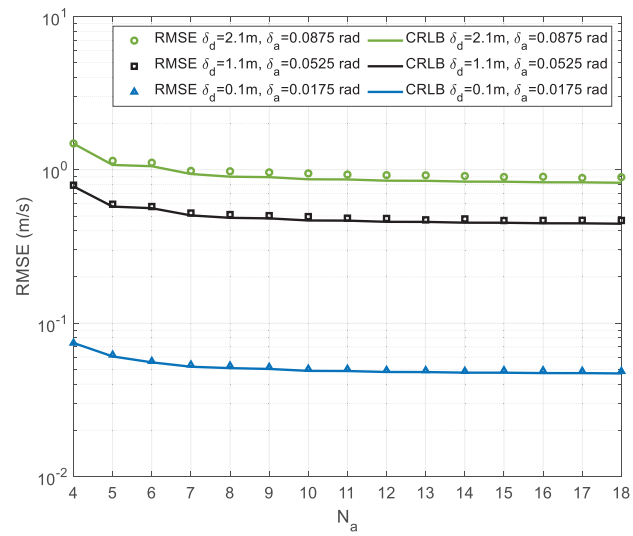


Fig. 5. RMSE performance of the proposed algorithm in velocity estimation with different numbers of selected LOS paths.

and AOA measurements. The order of the elements in (42) is the same as that in $\Delta\mathbf{m}$, in which the first $(N_a - 1)$ pairs are TDOA and FDOA pairs (the covariance matrix for each pair is $\mathbf{Q}_d$), and the last N_a pairs are AOA pairs (the covariance matrix for each pair is $\mathbf{Q}_a$). Similarly, we consider the covariance matrix of the $\Delta\mathbf{m}^s_{n,l}$ for the (n,l)-th scatterer in the form of $\mathbf{Q}^s_{n,l} = \text{diag}(\delta_d^2, (0.1\delta_d)^2, \delta_a^2, \delta_a^2)$. The localization accuracy is assessed via the root mean square error (RMSE), e.g., $\text{RMSE}(\mathbf{u}) = \sqrt{\sum_{t=1}^{T_{MC}} ||\mathbf{u}_t - \mathbf{u}^\circ||^2 / T_{MC}}$, where $\mathbf{u}_t$ is the estimate of $\mathbf{u}^\circ$ at the t-th Monte Carlo simulation.

In the first simulation scenario, we evaluate the performance of the proposed UE localization algorithm with different numbers of selected LOS paths. Fig. 4 and Fig. 5 depict the RMSEs versus N_a. Here, the numerical results are obtained from $T_{MC} = 5000$ independent Monte Carlo simulations. Note that having a larger number of LOS paths is beneficial to achieve localization accuracy. For location estimation (Fig. 4), the proposed WLS algorithm requires $N_a \geqslant 2$ LOS paths. The localization accuracy is significantly enhanced as N_a increases to 3 and is saturated when $N_a \geqslant 6$. For velocity estimation (Fig. 5), the proposed WLS estimator requires $N_a \geqslant 4$ LOS paths. The performance improves gradually with N_a and reaches saturation for $N_a \geqslant 6$. In all cases, the CRLBs can be attained, and the bounds are tighter for smaller N_a, δ_d, and δ_a.

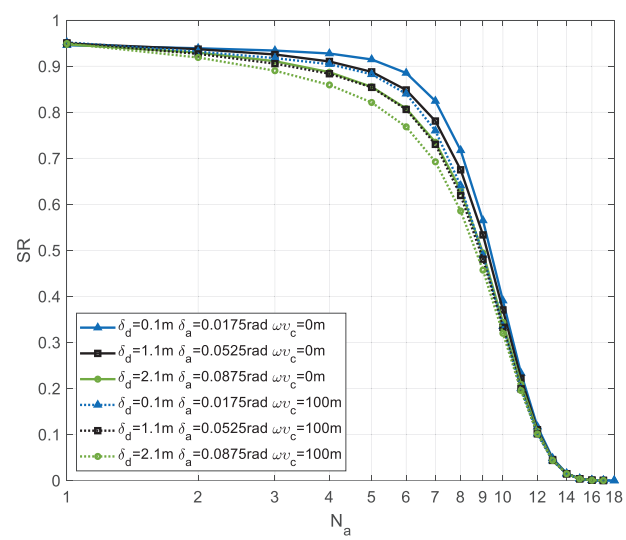


Fig. 6. SR performance of the proposed measurement selection method.

These results demonstrate that as long as 4-6 LOS paths are available, the proposed algorithm can realize UE localization with acceptable performance.

In the second simulation scenario, we want to reveal that four to six LOS paths can be obtained in mmWave

CRAN communication systems by densely deploying RRHs and designing appropriate measurement selection methods. We analyze the performance of the proposed measurement selection method by using the following simulation settings. The detection probability P_d for each RRH is set to 0.5. Each scatterer is distributed uniformly in a 3-D space $\{[x, y, z]^T : 240 \leqslant x \leqslant 280, 450 \leqslant y \leqslant 850, 0 \leqslant z \leqslant 20\}$ in meters, whilst the magnitude of velocity follows $\mathcal{U}[0, 10]$ m/s. Successful selection is the phenomenon in which all of the selected N_a paths are LOS paths. Thus, the success rate (SR) T_{SR}/T_{MC} signifies that T_{SR} times successful selection out of T_{MC} Monte Carlo simulations, and we set $T_{MC} = 100000$. TOA and AOA measurements used in this study follow Gaussian distributions with mean given by (2) and (10), respectively, and standard deviations given by δ_d and δ_a, respectively. Fig. 6 shows the SR performance versus N_a by setting (1) $\delta_d = 0.1$ m and $\delta_a = 0.0175$ rad, (2) $\delta_d = 1.1$ m and $\delta_a = 0.0525$ rad, (3) $\delta_d = 2.1$ m and $\delta_a = 0.0875$ rad. (0.0175 rad $= 1°$, 0.0525 rad $= 2°$, 0.0875 rad $= 3°$, respectively). We also considered clock bias setting $\omega v_c = 0$ m and $\omega v_c = 100$ m for each noise level configuration. In all cases, the SR achieves 85% when $N_a = 4$, and the SR is larger than 80% for $N_a \leqslant 6$ in most cases. The SR can be further improved by increasing the detection probability of RRH and by considering the energy gap between the LOS and NLOS paths [3]. However, this topic is not the focus of this study, hence, we will not go into further details.

In the third simulation scenario, we evaluate the performance of the proposed TDOA/FDOA/AOA based WLS estimator by comparing it with AOA-only, TDOA-only [24], TDOA/AOA [26], TDOA/FDOA [27] WLS estimators, and the corresponding CRLBs. We set $N_a = 6$, $T_{MC} = 5000$, $\delta_d = 0.22\rho$, and $\delta_a = 0.0175\rho$, where ρ is a noise scaling factor (See Table III for specific values). The RMSEs and CRLBs of different estimators are shown in Fig. 7 and Fig. 8 as functions of the noise scaling factor. The results in Fig. 7 show that the proposed TDOA/FDOA/AOA based WLS estimator has the best performance, followed by TDOA/FDOA, TDOA/AOA, TDOA-only, and AOA-only. For velocity estimation in Fig. 8, we only compare the proposed estimator with the TDOA/FDOA WLS estimator because velocity cannot be obtained without FDOA measurements. Fig. 8 shows that the performance of the proposed estimator is slightly remarkable. Furthermore, the proposed TDOA/FDOA/AOA localization can achieve the CRLB for small noise level. Increasing the noise level results in a slow deviation from the CRLB for both location and velocity estimations because the nonlinear terms in $\mathbf{e}$ in the derivation of the proposed algorithm have been ignored. TDOA/FDOA algorithm uses two-stage WLS estimators and has larger deviation from CRLB than the proposed estimator as the noise level increases. The proposed scatterer localization performance is also depicted in Fig. 7 and Fig. 8. The unknown scatterer is located at $[240, 600, -19]^T$ in meters. The velocity direction of the scatterer is the same as $\dot{\mathbf{u}}^{\circ}$

[3]For example, let P_{max} denote the energy of the strongest path, set the threshold as $P_{thre} = P_{max}/2$, and the paths with energy less than P_{thre} are filtered out. Note that $P_{max}/2$ is an empirical setting based on a general ray-tracing dataset for mmWave massive MIMO [35].

TABLE III
MEASUREMENT NOISE SETTINGS.

ρ	0.1	1	10
δ_d (m)	0.022	0.22	2.2
δ_a (rad)	0.00175	0.0175	0.175

TABLE IV
MAE PERFORMANCE COMPARISON.

Method	Location (m)	Velocity (m/s)
Black Box	0.1782	0.2109
WLS	0.0200	0.0143
NN-WLS	**0.0104**	**0.0054**

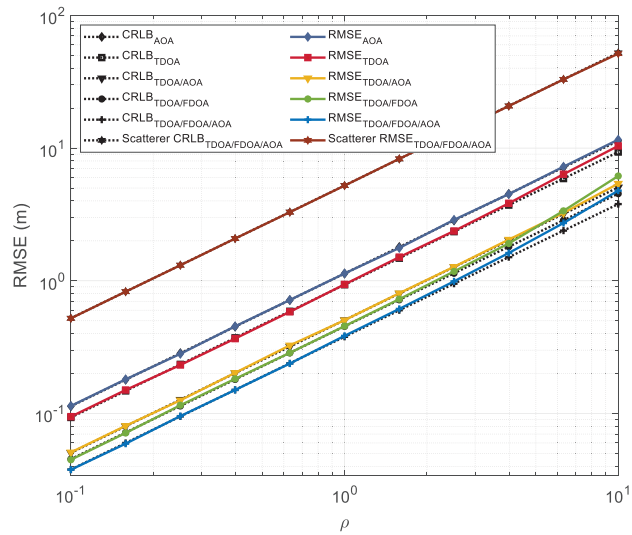


Fig. 7. Comparison of the RMSE of the proposed algorithm in location estimation with that of the AOA-only, TDOA-only, TDOA/AOA, TDOA/FDOA algorithms, and the corresponding CRLBs.

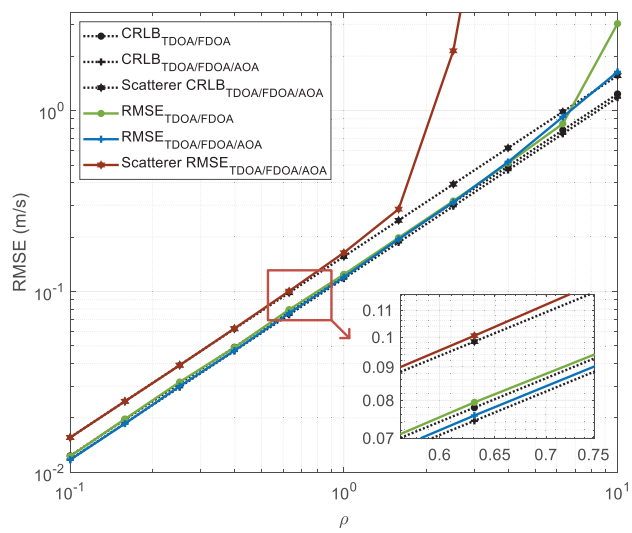


Fig. 8. Comparison of the RMSE of the proposed algorithm in velocity estimation with that of the AOA-only, TDOA-only, TDOA/AOA, TDOA/FDOA algorithms, and the corresponding CRLBs.

TABLE V
MAE (m) PERFORMANCE COMPARISON OF UE LOCATION ESTIMATION.

Method	$\delta_d = 3$ m, $\delta_a = 0.0525$ rad			$\delta_d = 3$ m, $\delta_a = 0.0175$ rad			$\delta_d = 0.1$ m, $\delta_a = 0.0525$ rad		
	Ratio=0.1	Ratio=0.01	Ratio=0.001	Ratio=0.1	Ratio=0.01	Ratio=0.001	Ratio=0.1	Ratio=0.01	Ratio=0.001
WLS	11.42	11.23	11.17	3.86	3.85	3.85	2.12	2.10	2.10
Black Box	3.44	1.72	2.92	3.38	2.06	1.54	3.03	1.23	1.42
NN-WLS	**2.58**	**0.55**	**0.16**	**2.24**	**0.55**	**0.09**	**0.25**	**0.05**	**0.03**

TABLE VI
MAE (m/s) PERFORMANCE COMPARISON OF UE VELOCITY ESTIMATION.

Method	$\delta_d = 3$ m, $\delta_a = 0.0525$ rad			$\delta_d = 3$ m, $\delta_a = 0.0175$ rad			$\delta_d = 0.1$ m, $\delta_a = 0.0525$ rad		
	Ratio=0.1	Ratio=0.01	Ratio=0.001	Ratio=0.1	Ratio=0.01	Ratio=0.001	Ratio=0.1	Ratio=0.01	Ratio=0.001
WLS	3.12	2.98	2.83	2.88	2.80	2.79	0.18	0.18	0.18
Black Box	1.29	1.39	1.20	1.27	1.12	1.11	1.18	1.00	1.09
NN-WLS	**0.68**	**0.25**	**0.22**	**0.59**	**0.15**	**0.13**	**0.12**	**0.07**	**0.06**

with a magnitude of 5 (m/s). The results in Fig. 7 demonstrate that, for the scatterer location, the RMSE can achieve the CRLB. However, the CRLB of the scatterer is higher than that of the UE because the number of measurements used in scatterer localization is less than that in the UE localization. For the scatterer velocity observed in Fig. 8, the RMSE can achieve the CRLB when $\rho \leqslant 1$. Since velocity is mainly determined by the FDoA measurements, and only one measurement can be used for each scatterer, the proposed algorithm can ensure good performance with relatively small noise and is greatly affected by large noise caused by insufficient measurements.

B. NN-assisted WLS Localization

In this subsection, we explore the performance of the proposed NN-assisted WLS localization methods. First, we utilize a general dataset for mmWave massive MIMO constructed on the basis of the ray-tracing data from Remcom Wireless InSite [35] to verify the effectiveness of the proposed algorithms because this approach can simulate real-world scenarios accurately. Specifically, we evaluate the performance of the Black Box NN (Section III-B), the proposed WLS (Section IV), and the proposed NN-WLS (Section V-A) through the same revised ray-tracing dataset.[4] The training, validation, and testing datasets contain 60000, 20000, and 20000 samples, respectively. All testing samples are excluded from the training and validation samples. The inner architecture of the networks used for the Black Box NN and proposed NN-WLS is identical and consists of a three-layer FC-NN. The first two FC layers use 32 neurons, and the third FC layer uses 22 neurons. The localization accuracy is assessed via the mean absolute error (MAE), e.g., $\text{MAE}(\mathbf{u}) = \sum_{t=1}^{T_{test}} ||\mathbf{u}_t - \mathbf{u}_t^{\circ}||/T_{test}$, where $\mathbf{u}_t$ is the estimation of $\mathbf{u}_t^{\circ}$ in the test dataset, and T_{test} is the size of test dataset. The MAE results of the Black Box NN, WLS, and NN-WLS are given in Table IV. The result shows that the NN-WLS is more accurate in terms of location and velocity estimation than the WLS algorithm. The Black Box NN is the simplest to operate but has the worst accuracy. The results verify that the measurement errors are not completely random, that is, an underlying relationship exists between them, and this relationship can be learned by the NN, which motivates our research. For an in-depth analysis, we explore the performance of the proposed NN-assisted WLS localization methods with different noise levels and training dataset sizes in the following.

1) Localization Accuracy to Noise Level: We increase the noise level of measurements to analyze the performance of the NN-assisted WLS algorithm. By observing the ray-tracing dataset, we find that the measurement errors include a dominant part and a fluctuating part. We define the dominant part as the unknown fixed error and the fluctuating part as the Gaussian random error. We define three dominant error settings: (1) $\delta_d = 3$ m, $\delta_a = 0.0525$ rad; (2) $\delta_d = 3$ m, $\delta_a = 0.0175$ rad; (3) $\delta_d = 0.1$ m, $\delta_a = 0.0525$ rad. Three radios are available for each setting, and the standard deviation of the fluctuating error are 0.1, 0.01, and 0.001 times of that of the dominant error. Therefore, nine noise settings have been identified. Training and testing are conducted under the same noise setting. The training, validation, and testing sets contain 12000, 4000, and 4000 samples, respectively. All testing samples are excluded from the training and validation samples. The MAE results for different methods are shown in Table V and Table VI. The performance of the proposed NN-WLS outperforms the WLS algorithm and the black box NN in the given simulation scenarios. Moreover, by decreasing the ratio of the error standard deviation of the random part to that of the fixed part, the MAE of NN-WLS and black box NN decreases. That is, as the proportion of the random part decreases, the ability of the NNs increases. This is due to the fact that the NNs can learn the dominant error and the correlation between measurement errors, but WLS algorithm cannot.

2) Network Performance to Training Dataset Size: We reduce the size of training dataset from 12000 to 1200, and the performance of the NN-WLS and Black Box is shown

[4]We utilize the first 6 RRHs in the ray-tracing dataset, and each sample is generated with a different UE location distributed in a 3D space $\{[x, y, z]^T : 240 \leqslant x \leqslant 280, 410 \leqslant y \leqslant 740, z = 2\}$ in meters. There are no FDoA measurements given in the ray-tracing dataset. For each UE, we generate its velocity in a random way, and then calculate its corresponding FDoA measurements.

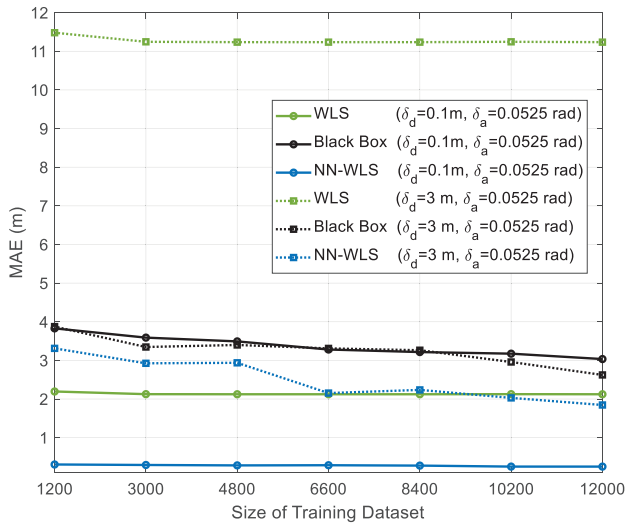


Fig. 9. MAE performance comparison of location estimation between the black box and NN-WLS algorithms for varying training dataset sizes.

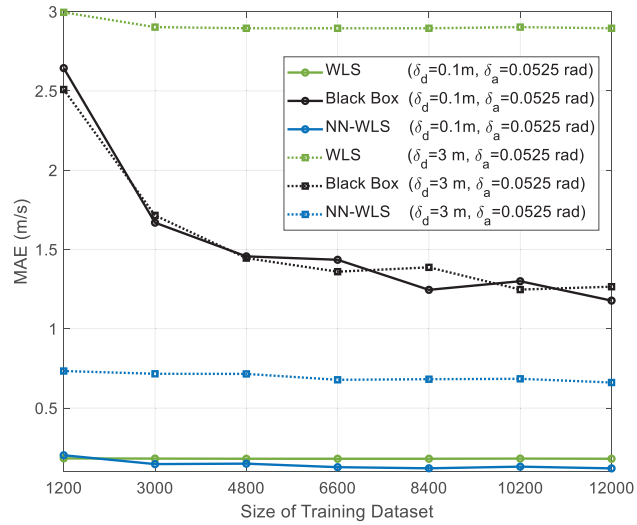


Fig. 10. MAE performance comparison of velocity estimation between the black box and NN-WLS algorithms for varying training dataset sizes.

in Fig. 9 and Fig. 10. In all simulations, the ratio is set to 0.1. For a relatively large noise level, where $\delta_d = 3$ m and $\delta_a = 0.0525$ rad, the performance of NN-WLS saturates with 6600 and 1200 training data for location and velocity estimation, respectively, whereas the black box NN requires more training data to increase accuracy. In addition, reducing the noise level can bring gains to NN-WLS, but not to the black box NN, since the latter is purely data driven and lacks the assistance of the geometric model. When $\delta_d = 0.1$ m and $\delta_a = 0.0525$ rad, NN-WLS adds 1200 and 3000 training samples on the basis of WLS, the estimation accuracy of UE location and velocity can be improved by 86% and 19%, respectively. When $\delta_d = 3$ m and $\delta_a = 0.0525$ rad, NN-WLS adds 6600 and 1200 training samples, the estimation accuracy of UE location and velocity can be improved by 88% and 76%, respectively. The performance of the proposed WLS estimator is enhanced by the NN, especially in a large noise environment.

3) Network Robustness to Measurement Noise: We study the robustness of the proposed NN-WLS to the varying measurement noise conditions. For comparison, we define the NN-LS algorithm. In particular, after obtaining the estimated residual vector $\hat{\mathbf{e}}$ from the NN (the same way as that implemented in the NN-WLS), we deduct $\hat{\mathbf{e}}$ from (22). Then, by directly applying the LS algorithm, we obtain $\mathbf{x} = (\tilde{\mathbf{G}}^T\tilde{\mathbf{G}})^{-1}\tilde{\mathbf{G}}^T(\tilde{\mathbf{h}}-\hat{\mathbf{e}})$. The black box NN, NN-WLS, and NN-LS are executed using the same datasets. Fig. 11 illustrates the performance of the black box NN, NN-WLS, and NN-LS trained for a specific noise level and deployed in different noise levels. We have five different measurement noise settings: (1) $\delta_d = 0.1$ m and $\delta_a = 0.0175$ rad, (2) $\delta_d = 0.6$ m and $\delta_a = 0.035$ rad, (3) $\delta_d = 1.1$ m and $\delta_a = 0.0525$ rad, (4) $\delta_d = 1.6$ m and $\delta_a = 0.07$ rad, (5) $\delta_d = 2.1$ m and $\delta_a = 0.0875$ rad. In all simulations, the ratio is set to 0.1. The black box NN, NN-WLS, and NN-LS in Figs. 11(a1) and (a2) are trained in the noise setting (1) and are tested in noise settings (1) to (5), where Fig. 11(a1) shows the MAE performance of location estimation and Fig. 11(a2) shows the MAE performance of velocity estimation. Figs. 11(b1) and (b2), Figs. 11(c1) and (c2), Figs. 11(d1) and (d2), and Figs. 11(e1) and (e2) are trained in noise settings (2), (3), (4), and (5), respectively. The size of the training dataset is 1200 in Figs. 11(a1) and (a2) to Figs. 11(e1) and (e2). Moreover, the training dataset in Figs. 11(f1) and (f2) includes all (1)-(5) measurement noise settings, and the size of the training dataset is 8000.

The results indicate that NN-WLS is robust for small noise settings and outperforms the black box NN in most cases in terms of location estimation. In addition, NN-WLS is robust for all the noise settings in terms of velocity estimation, whereas the black box NN shows great performance fluctuations. When the noise setting of test dataset is the same as that of the training dataset, the performance of NN-LS is comparable to NN-WLS. However, in terms of both location and velocity estimation, NN-LS performs poorly when tested by using a different noise setting from the training dataset. NN-LS requires the estimated $\hat{\mathbf{e}}$ to be highly accurate, so that the LS algorithm can be used to derive good results. By contrast, in NN-WLS, the weighting matrix is $\mathbf{W} = (\hat{\mathbf{e}}\hat{\mathbf{e}}^T + \epsilon\mathbf{I})^{-1}$, which contains the information of the dominant ($\hat{\mathbf{e}}$ is the learned mean of the dominant error) and the random error parts ($\epsilon\mathbf{I}$ is the covariance matrix of the Gaussian random error).[5] Under a test dataset with measurement noise setting different from the training dataset, it is difficult for NN to predict a very accurate $\hat{\mathbf{e}}$, but it can predict a relatively accurate weighting matrix $\mathbf{W}$, which makes the NN-WLS more robust than the NN-LS. Therefore, the robustness of the proposed NN-WLS outperforms the NN-LS and the black box NN in the given simulation scenarios.

4) Ensemble Learning-based NN-WLS Methods: We analyze the MAE performance of the proposed ensemble learning-based NN-WLS methods by setting $\delta_d = 3$ m and $\delta_a = 0.0175$

[5]In the simulations, ϵ is an adjustable parameter for ensuring that the matrix is invertible, which should be as small as possible. For example, we set $\epsilon = 0.1$ for measurement noise setting (1).

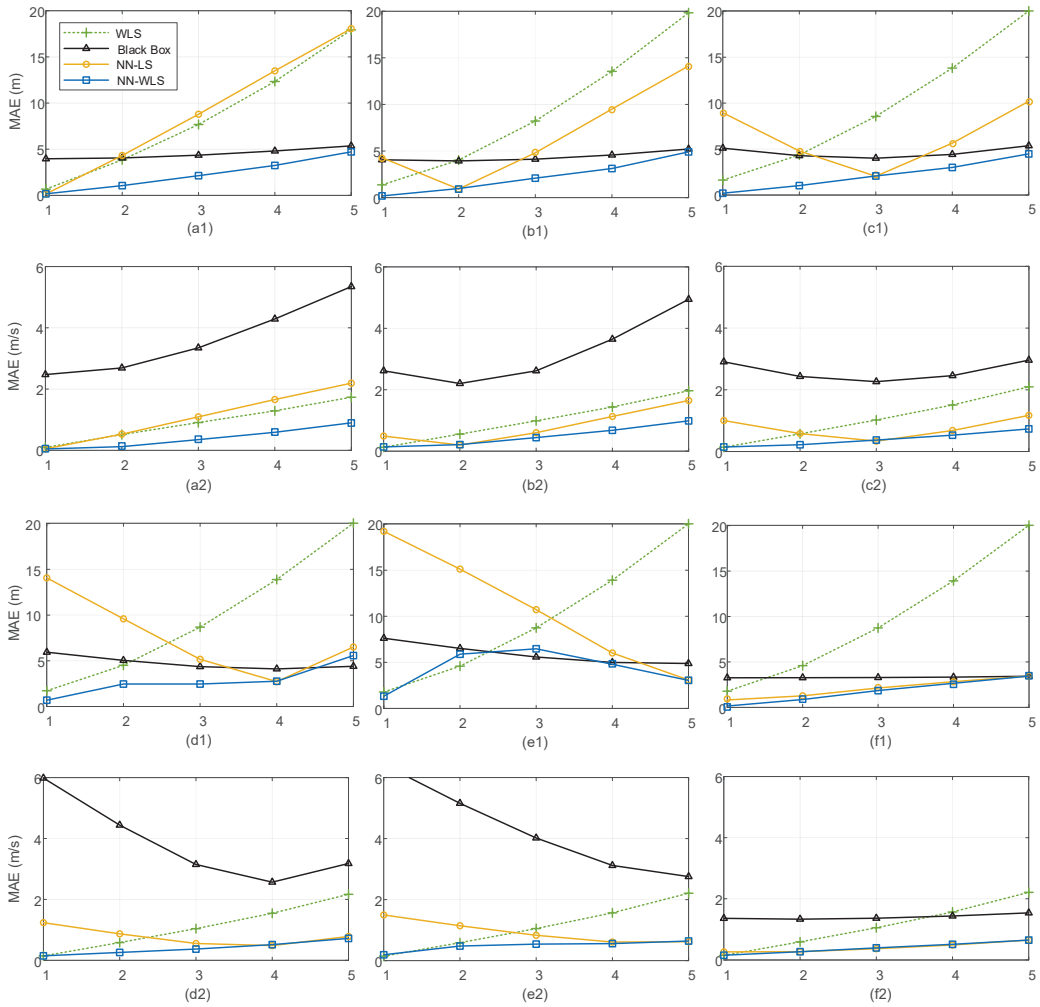


Fig. 11. MAE performance comparison among the WLS, black box, NN-WLS, and NN-LS algorithms for various noise conditions.

rad. The radius r_a for the subtractive clustering algorithm in ENN-A-WLS is set to 0.1, 0.01, and 0.001 for ratios 0.1, 0.01, and 0.001, respectively. The number of emsembled NNs is $P = 100$. The size of the training dataset for NN-WLS, ENN-M-WLS, ENN-A-WLS, and ENN-B-WLS is $12,000$. Here, ENN-M-WLS has the same structure as ENN-A-WLS but uses a simpler averaging method to replace the subtractive clustering algorithm in ENN-A-WLS. The MAE results are presented in Table VII and Table VIII. The proposed ENN-B-WLS has the best performance in terms of location estimation. The reason is that the combination of the predictions of multiple NNs by (41) can approximate the statistical characteristic of $\mathbf{W}$ remarkably. However, the ensemble learning-based NN-WLS method has no evident advantages over NN-WLS in terms of velocity estimation. Not much space for improvement is needed because the values of the velocity in simulations are relatively small, indicating that the estimation error of NN-WLS is also small.

5) Time Resources: We compare the time resources consumed by different localization methods. The model-based WLS estimator needs 0.06 seconds when executed on a desktop computer with a 3.3 GHz Intel(R) Xeon(R) W-2155

TABLE VII
MAE (m) PERFORMANCE COMPARISON OF UE LOCATION ESTIMATION.

Method	$\delta_d = 3$ m, $\delta_a = 0.0175$ rad		
	Ratio=0.1	Ratio=0.01	Ratio=0.001
WLS	3.86	3.85	3.85
NN-WLS	2.24	0.55	0.09
ENN-M-WLS	2.41	0.53	0.13
ENN-A-WLS	1.99	0.53	0.11
ENN-B-WLS	**0.60**	**0.24**	**0.05**

TABLE VIII
MAE (m/s) PERFORMANCE COMPARISON OF UE VELOCITY ESTIMATION.

Method	$\delta_d = 3$ m, $\delta_a = 0.0175$ rad		
	Ratio=0.1	Ratio=0.01	Ratio=0.001
WLS	2.88	2.80	2.79
NN-WLS	**0.59**	0.15	**0.13**
ENN-M-WLS	0.59	0.20	0.17
ENN-A-WLS	**0.59**	**0.14**	**0.13**
ENN-B-WLS	0.71	0.18	0.15

CPU and 64 GB of RAM, using Windows 10 and MATLAB 2018b (64-bit). The time needed by the NN-WLS consists of two parts. The test time of the NN is 1.6×10^{-7} seconds when tested on the 1080 Ti GPU, and the time required to input the results of the NN into the WLS estimator and obtain the final estimation is 0.011 seconds when using MATLAB. Thus, the total time needed by the NN-WLS is 0.011 seconds, which is 17% of the time needed by the model-based WLS estimator. The ensemble learning-based NN-WLS takes more time than the NN-WLS because the ensemble time is 1.6×10^{-3} seconds when using MATLAB. Thus, the total time needed by the ensemble learning based NN-WLS is 0.013 seconds, which is 22% of the time needed by the model-based WLS estimator. This result is reasonable because the WLS algorithm requires initialization and several update processes, which are time consuming, whereas the NN-WLS and the ensemble learning based NN-WLS do not need to execute such processes.

VII. Conclusion

This study considered the joint location and velocity estimation problem in a 3-D mmWave CRAN architecture. First, we embedded the cooperative localization into communications and established the joint location and velocity estimation model with hybrid TDOA/FDOA/AOA measurements. Then, an efficient closed-form WLS solution, which was subsequently proven asymptotically unbiased under small noise levels, was deduced. Second, we built the scatterer localization model by exploiting the single-bounce NLOS paths and the estimated UE location and deduced the closed-form WLS solution. The simulation results indicated that the WLS-based joint estimation algorithm can achieve the CRLB and outperform the benchmarks.

Furthermore, the NN-WLS algorithm was proposed by embedding the NNs into the proposed WLS estimators to replace linear approximation. This study is the first to combine the WLS estimator and NN in 3-D localization methods in the existing literature. The combination harnesses both powerful learning ability of the NN and the robustness of the proposed geometric model. In addition, ensemble learning was introduced to improve performance. A revised ray-tracing dataset was used in the simulations to test the performance of the NN-WLS algorithm. Simulation results showed that NN-WLS is fast because it can eliminate iterations in the proposed WLS algorithm, and significantly outperforms the WLS algorithm when the measurement error vector exhibits some correlation pattern. In addition, through a comprehensive comparison with the black box NN and the NN-LS method, the proposed NN-WLS is more excellent in terms of localization accuracy and robustness.

Appendix A

In this section, we approximate $\mathbf{e}$ up to the linear noise term in (24). For the differentiable function $f(x_1,\ldots,x_n)$ on the variables $x_1,\ldots,x_n$, there holds

$$\begin{aligned} &f(x_1+\Delta x_1,\ldots,x_n+\Delta x_n)-f(x_1,\ldots,x_n)\\ &\qquad=\frac{\partial f}{\partial x_1}\Delta x_1+\ldots+\frac{\partial f}{\partial x_n}\Delta x_n+o(\eta), \end{aligned} \tag{43}$$

where $\eta=\sqrt{(\Delta x_1)^2+\ldots+(\Delta x_n)^2}\to 0$. According to (18) and (22), we get

$$\mathbf{e}=(\tilde{\mathbf{h}}-\tilde{\mathbf{G}}\mathbf{x}^\circ)-(\mathbf{h}-\mathbf{G}\mathbf{x}^\circ). \tag{44}$$

Applying (43) with (44), firstly, for $i=2,\ldots,N_a$, we yield the $(2i-3)$-th entry in $\mathbf{e}$ as

$$\begin{aligned} \mathbf{e}(2i-3)\approx&[2r_{i1}^\circ+2\mathbf{a}_1^{\circ T}(\mathbf{u}^\circ-\mathbf{b}_1)]\Delta r_{i1}\\ &+\left[-2r_{i1}^\circ\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\phi_1^\circ}\mathbf{b}_1+2r_{i1}^\circ\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\phi_1^\circ}\mathbf{u}^\circ\right]\Delta\phi_1\\ &+\left[-2r_{i1}^\circ\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\theta_1^\circ}\mathbf{b}_1+2r_{i1}^\circ\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\theta_1^\circ}\mathbf{u}^\circ\right]\Delta\theta_1, \end{aligned}$$

where $\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\phi_1^\circ}(\mathbf{u}^\circ-\mathbf{b}_1)=0$, $\frac{\partial\mathbf{a}_1^{\circ T}}{\partial\theta_1^\circ}(\mathbf{u}^\circ-\mathbf{b}_1)=0$, and $\mathbf{a}_1^{\circ T}(\mathbf{u}^\circ-\mathbf{b}_1)=r_1^\circ$, hence, we have

$$\mathbf{e}(2i-3)\approx 2r_i^\circ\Delta r_{i1}. \tag{45}$$

Similarly, we have

$$\mathbf{e}(2i-2)\approx\dot{r}_i^\circ\Delta r_{i1}+r_i^\circ\Delta\dot{r}_{i1}+r_1^\circ r_{i1}^\circ\cos^2\theta_1^\circ\dot{\phi}_1^\circ\Delta\phi_1+r_1^\circ r_{i1}^\circ\dot{\theta}_1^\circ\Delta\theta_1. \tag{46}$$

For $j=1,\ldots,N_a$, we have

$$\mathbf{e}(2N_a-3+2j)\approx\left(\frac{\partial\mathbf{c}_j^{\circ T}}{\partial\phi_j^\circ}\mathbf{b}_j-\frac{\partial\mathbf{c}_j^{\circ T}}{\partial\phi_j^\circ}\mathbf{u}^\circ\right)\Delta\phi_j=r_j^\circ\cos\theta_j^\circ\Delta\phi_j. \tag{47}$$

and

$$\begin{aligned} \mathbf{e}(2N_a-2+2j)\approx&\frac{\partial\mathbf{d}_j^{\circ T}}{\partial\phi_j^\circ}(\mathbf{b}_j-\mathbf{u}^\circ)\Delta\phi_j\\ &+\frac{\partial\mathbf{d}_j^{\circ T}}{\partial\theta_j^\circ}(\mathbf{b}_j-\mathbf{u}^\circ)\Delta\theta_j=r_j^\circ\Delta\theta_j. \end{aligned} \tag{48}$$

Finally, transforming the expressions (45), (46), (47), and (48) for $i=2,\ldots,N_a$ and $j=1,\ldots,N_a$ into matrix representation, we obtain the first-order approximation of $\mathbf{e}$ as $\mathbf{e}\approx\mathbf{B}\Delta\mathbf{m}$ in (24).

Appendix B

In this section, we take the state of UE $\mathbf{x}^\circ$ as an example. We first calculate the partial derivatives required for CRLB. According to [38], the CRLB of $\mathbf{x}^\circ$ for the Gaussian noise model can be defined as

$$\mathrm{CRLB}(\mathbf{x}^\circ)=(\mathbf{D}^T\mathbf{Q}^{-1}\mathbf{D})^{-1}, \tag{49}$$

where $\mathbf{D}=\partial\mathbf{m}^\circ/\partial\mathbf{x}^{\circ T}$. The partial derivatives are given by

$$\begin{aligned} \frac{\partial\mathbf{m}^\circ}{\partial\mathbf{x}^{\circ T}}=&\Bigg[(\frac{\partial r_{21}^\circ}{\partial\mathbf{x}^{\circ T}})^T,(\frac{\partial\dot{r}_{21}^\circ}{\partial\mathbf{x}^{\circ T}})^T,\ldots,(\frac{\partial r_{N_a1}^\circ}{\partial\mathbf{x}^{\circ T}})^T,(\frac{\partial\dot{r}_{N_a1}^\circ}{\partial\mathbf{x}^{\circ T}})^T,\\ &(\frac{\partial\phi_1^\circ}{\partial\mathbf{x}^{\circ T}})^T,(\frac{\partial\theta_1^\circ}{\partial\mathbf{x}^{\circ T}})^T,\ldots,(\frac{\partial\phi_{N_a}^\circ}{\partial\mathbf{x}^{\circ T}})^T,(\frac{\partial\theta_{N_a}^\circ}{\partial\mathbf{x}^{\circ T}})^T\Bigg]^T, \end{aligned} \tag{50}$$

and

$$\frac{\partial r_{i1}^\circ}{\partial\mathbf{x}^{\circ T}}=\left[\frac{\partial r_{i1}^\circ}{\partial\mathbf{u}^{\circ T}},\frac{\partial r_{i1}^\circ}{\partial\dot{\mathbf{u}}^{\circ T}}\right],\frac{\partial\dot{r}_{i1}^\circ}{\partial\mathbf{x}^{\circ T}}=\left[\frac{\partial\dot{r}_{i1}^\circ}{\partial\mathbf{u}^{\circ T}},\frac{\partial\dot{r}_{i1}^\circ}{\partial\dot{\mathbf{u}}^{\circ T}}\right], \tag{51}$$

$$\frac{\partial\phi_j^\circ}{\partial\mathbf{x}^{\circ T}}=\left[\frac{\partial\phi_j^\circ}{\partial\mathbf{u}^{\circ T}},\frac{\partial\phi_j^\circ}{\partial\dot{\mathbf{u}}^{\circ T}}\right],\frac{\partial\theta_j^\circ}{\partial\mathbf{x}^{\circ T}}=\left[\frac{\partial\theta_j^\circ}{\partial\mathbf{u}^{\circ T}},\frac{\partial\theta_j^\circ}{\partial\dot{\mathbf{u}}^{\circ T}}\right], \tag{52}$$

where $i = 2, \ldots, N_a$ and $j = 1, \ldots, N_a$. Firstly, from (2) and (3), we obtain

$$\frac{\partial r_{i1}^{\circ}}{\partial \mathbf{u}^{\circ T}} = \frac{(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{r_i^{\circ}} - \frac{(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{r_1^{\circ}}, \quad \frac{\partial r_{i1}^{\circ}}{\partial \dot{\mathbf{u}}^{\circ T}} = \mathbf{0}. \quad (53)$$

Secondly, from (6) and (7), we get

$$\frac{\partial \dot{r}_{i1}^{\circ}}{\partial \mathbf{u}^{\circ T}} = \frac{\dot{r}_1^{\circ}(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{(r_1^{\circ})^2} - \frac{\dot{r}_i^{\circ}(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{(r_i^{\circ})^2} + \frac{\dot{\mathbf{u}}^{\circ T}}{r_i^{\circ}} - \frac{\dot{\mathbf{u}}^{\circ T}}{r_1^{\circ}}, \quad (54)$$

$$\frac{\partial \dot{r}_{i1}^{\circ}}{\partial \dot{\mathbf{u}}^{\circ T}} = \frac{(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{r_i^{\circ}} - \frac{(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{r_1^{\circ}}. \quad (55)$$

Thirdly, according to (17), we have $(\mathbf{b}_j - \mathbf{u}^{\circ})^T \partial \mathbf{c}_j^{\circ} / \partial \mathbf{u}^{\circ T} = \mathbf{c}_j^{\circ T}$. Since $\mathbf{a}_j^{\circ T}[\cos\phi_j^{\circ}, \sin\phi_j^{\circ}, 0]^T = \cos\theta_j^{\circ}$, we yield $(\mathbf{b}_j - \mathbf{u}^{\circ})^T \partial \mathbf{c}_j^{\circ} / \partial \mathbf{u}^{\circ T} = -r_j^{\circ} \mathbf{a}_j^{\circ T} \partial \mathbf{c}_j^{\circ} / \partial \mathbf{u}^{\circ T} = r_j^{\circ} \cos\theta_j^{\circ} \partial \phi_j^{\circ} / \partial \mathbf{u}^{\circ T}$, that is,

$$\frac{\partial \phi_j^{\circ}}{\partial \mathbf{u}^{\circ T}} = \frac{\mathbf{c}_j^{\circ T}}{r_j^{\circ} \cos\theta_j^{\circ}}, \quad \frac{\partial \phi_j^{\circ}}{\partial \dot{\mathbf{u}}^{\circ T}} = \mathbf{0}, \quad (56)$$

for $j = 1, \ldots, N_a$. Similarly, from (17), we obtain $(\mathbf{u}^{\circ} - \mathbf{b}_j)^T \partial \mathbf{d}_j^{\circ} / \partial \mathbf{u}^{\circ T} + \mathbf{d}_j^{\circ T} = \mathbf{0}$, that is, $(\mathbf{u}^{\circ} - \mathbf{b}_j)^T [\frac{\partial \mathbf{d}_j^{\circ}}{\partial \theta_j^{\circ}} \frac{\partial \theta_j^{\circ}}{\partial \mathbf{u}^{\circ T}} + \frac{\partial \mathbf{d}_j^{\circ}}{\partial \phi_j^{\circ}} \frac{\partial \phi_j^{\circ}}{\partial \mathbf{u}^{\circ T}}] = -\mathbf{d}_j^{\circ T}$. Since $(\mathbf{u}^{\circ} - \mathbf{b}_j)^T = r_j^{\circ} \mathbf{a}_j^{\circ T}$, $\mathbf{a}_j^{\circ T} \partial \mathbf{d}_j^{\circ} / \partial \theta_j^{\circ} = -1$ and $\mathbf{a}_j^{\circ T} \partial \mathbf{d}_j^{\circ} / \partial \phi_j^{\circ} = 0$, we get

$$\frac{\partial \theta_j^{\circ}}{\partial \mathbf{u}^{\circ T}} = \frac{\mathbf{d}_j^{\circ T}}{r_j^{\circ}}, \quad \frac{\partial \theta_j^{\circ}}{\partial \dot{\mathbf{u}}^{\circ T}} = \mathbf{0}. \quad (57)$$

Next, we prove that $\mathrm{cov}(\mathbf{x}) \approx \mathrm{CRLB}(\mathbf{x}^{\circ})$ under small noise levels. The proof relies on the following two key identities, for $i = 2, \ldots, N_a$,

$$(a): r_i^{\circ} \left[\frac{(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{r_i^{\circ}} - \frac{(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{r_1^{\circ}} \right] = (\mathbf{b}_1 - \mathbf{b}_i)^T - r_{i1}^{\circ} \mathbf{a}_1^{\circ T}, \quad (58)$$

$$\begin{aligned}(b): \dot{r}_i^{\circ} &\left[\frac{(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{r_i^{\circ}} - \frac{(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{r_1^{\circ}} \right] \\ &+ r_i^{\circ} \left[\frac{\dot{r}_1^{\circ}(\mathbf{u}^{\circ} - \mathbf{b}_1)^T}{(r_1^{\circ})^2} - \frac{\dot{r}_i^{\circ}(\mathbf{u}^{\circ} - \mathbf{b}_i)^T}{(r_i^{\circ})^2} + \frac{\dot{\mathbf{u}}^{\circ T}}{r_i^{\circ}} - \frac{\dot{\mathbf{u}}^{\circ T}}{r_1^{\circ}} \right] \\ &+ r_{i1}^{\circ} \dot{\phi}_1^{\circ} \cos\theta_1^{\circ} \mathbf{c}_1^{\circ T} + r_{i1}^{\circ} \dot{\theta}_1^{\circ} \mathbf{d}_1^{\circ T} = -\dot{r}_{i1}^{\circ} \mathbf{a}_1^{\circ T}. \quad (59)\end{aligned}$$

Since $(\mathbf{u}^{\circ} - \mathbf{b}_j)^T = r_j^{\circ} \mathbf{a}_j^{\circ T}$, $\dot{\mathbf{u}}^{\circ} = \dot{r}_1^{\circ} \mathbf{a}_1^{\circ} + r_1^{\circ} \dot{\mathbf{a}}_1^{\circ}$, and $\dot{\phi}_1^{\circ} \partial \mathbf{a}_1^{\circ T} / \partial \phi_1^{\circ} + \dot{\theta}_1^{\circ} \partial \mathbf{a}_1^{\circ T} / \partial \theta_1^{\circ} = \dot{\mathbf{a}}_1^{\circ T}$, by some tedious derivation, we can prove that (a) and (b) hold.

References

[1] J. Yang, S. Jin, Y. Han, M. Matthaiou, and Y. Zhu, "3-D position and velocity estimation in 5G mmWave CRAN with lens antenna arrays," in *Proc. IEEE VTC-Fall*, Sept. 2019, pp. 1-6.

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

[3] F. Lemic, J. Martin, C. Yarp, D. Chan, V. Handziski, R. Brodersen, G. Fettweis, A. Wolisz, and J. Wawrzynek, "Localization as a feature of mmWave communication," in *Proc. IEEE IWCMC*, Sep. 2016, pp. 1033-1038.

[4] H. Wymeersch, G. Seco-Granados, G. Destino, D. Dardari, and F. Tufvesson, "5G mmWave positioning for vehicular networks," *IEEE Wireless Commun. Mag.*, vol. 24, no. 6, pp. 80-86, Dec. 2017.

[5] R. D. Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102-112, Nov. 2014.

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

[7] Y. Han, H. Zhang, S. Jin, X. Li, R. Yu, and Y. Zhang, "Investigation of transmission schemes for millimeter-wave massive MU-MIMO systems," *IEEE Syst. J.*, vol. 11, no. 1, pp. 72-83, Mar. 2017.

[8] M. Xiao *et al.* , "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909-1935, Sept. 2017.

[9] H. Deng and A. M. Sayeed, "Mm-wave MIMO channel modeling and user localization using sparse beamspace signatures," in *Proc. IEEE SPAWC*, Jun. 2014, pp. 130-134.

[10] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164-1179, Jun. 2014.

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

[12] J. Yang, C. K. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmWave systems via cosparse image reconstruction technique," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4767-4782, Oct. 2018.

[13] R. G. Stephen and R. Zhang, "Uplink channel estimation and data transmission in millimeter-wave CRAN with lens antenna arrays," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6542-6555, Jul. 2018.

[14] S. Jeong, O. Simeone, A. Haimovich, and J. Kang, "Positioning via direct localization in C-RAN systems," *IET Commun.*, vol. 10, no. 16, pp. 2238-2244, Jun. 2016.

[15] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multisensor localization for the internet of things: Exploring a new class of scalable localization algorithms," *IEEE Signal Process. Mag.* vol. 35, no. 5, pp. 153-167, Sept. 2018.

[16] S. Safavi, U. A. Khan, S. Kar, and J. M. F. Moura, "Distributed localization: A linear theory," *Proc. IEEE*, vol. 106, no. 7, pp. 1204-1223, Jul. 2018.

[17] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 434-450, Mar. 1999.

[18] M. A. Badiu, T. L. Hansen, and B. H. Fleury, "Variational Bayesian inference of line spectra," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2247-2261, May 2017.

[19] Y. Han, T. Hsu, C. Wen, K. Wong and S. Jin, "Efficient downlink channel reconstruction for FDD multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3161-3176, Jun. 2019.

[20] L. Gaudio, M. Kobayashi, G. Caire and G. Colavolpe, "Joint radar target detection and parameter estimation with MIMO OTFS," in *Proc IEEE Radar Conference (RadarConf20)*, Florence, Italy, 2020, pp. 1-6.

[21] L. Xiang, et al. "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," in *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, pp. 1-22.

[22] F. Wen, H. Wymeersch, B. Peng, W. P. Tay, H. C. So, and D. Yang, "A survey on 5G massive MIMO localization," *Digit. Signal Process.*, vol. 94, pp. 21-28, Nov. 2019.

[23] M. Einemo and H. C. So, "Weighted least squares algorithm for target localization in distributed MIMO radar," *Signal Process.*, vol. 115, no. C, pp. 144-150, Oct. 2015.

[24] R. Amiri, F. Behnia, and H. Zamani, "Asymptotically efficient target localization from bistatic range measurements in distributed MIMO radars," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 299-303, Jan. 2017.

[25] 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, Jul. 2015.

[26] R. Amiri, F. Behnia, and H. Zamani, "Efficient 3-D positioning using time-delay and AoA measurements in MIMO radar systems," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2614-2617, Aug. 2017.

[27] K. C. Ho and W. W. Xu, "An accurate algebraic solution for moving source location using TDoA and FDoA measurements," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2453-2463, Aug. 2004.

[28] N. H. Nguyen and K. Dogancay, "Multistatic pseudolinear target motion analysis using hybrid measurements," *Signal Process.*, vol. 130, no. C, pp. 22-36, Jan. 2017.

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

[30] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. IEEE WCNC*, Mar. 2015, pp. 1666-1671.

[31] A. Decurninge, L. G. Ordóñez, P. Ferrand, G. He, B. Li, W. Zhang, and M. Guillaud, "CSI-based outdoor localization for massive MIMO: Experiments with a learning approach," in *Proc. IEEE ISWCS*, Aug. 2018, pp. 1-6.

[32] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive MIMO-OFDM system with deep convolutional neural networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10846-10857, Nov. 2019.

[33] P. Ferrand, A. Decurninge, and M. Guillaud, "DNN-based localization from channel estimates: Feature design and experimental results," [Online]. Available: https://arxiv.org/abs/2004.00363

[34] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331-7376, Oct. 2019.

[35] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. IEEE ITA*, Feb. 2019, pp. 1-8.

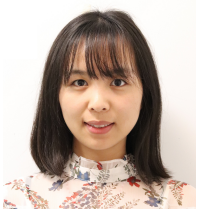
[36] J. Mo, P. Schniter, and R. W. Heath. Jr, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141-1154, Jan. 2018.

[37] H. He, S. Jin, C. K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77-83, Oct. 2019.

[38] S. M. Kay, *Fundamentals of statistical signal processing, Estimation Theory.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[39] T. G. Dietterich, *Ensemble learning, The handbook of brain theory and neural networks.* Cambridge, MA: MIT Press, 2002.

[40] J. Guo, J. Wang, C. -K. Wen, S. Jin, and G. Y. Li, "Compression and acceleration of neural networks for communications," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 110-117, Aug. 2020.



Jie Yang (S'18) received the B.S. degree in communication engineering from Nanjing University of Science and Technology, Nanjing, China, in 2015, the M.S. degree in information and communications engineering from Southeast University, Nanjing, China, in 2018. She is currently working towards the Ph.D. degree in information and communications engineering with Southeast University, Nanjing, China. Her current research interests include signal processing for wireless communications, massive MIMO and millimeter-wave wireless communications.

Shi Jin (S'06-M'07-SM'17) received the B.S. degree in communications engineering from Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from the Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He serves as an Associate Editor for the IEEE Transactions on Wireless Communications, and IEEE Communications Letters, and IET Communications. Dr. Jin and his co-authors have been awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society.

Chao-Kai Wen (S'00-M'04-SM'20) received the Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University, Taiwan, in 2004. He was with Industrial Technology Research Institute, Hsinchu, Taiwan and MediaTek Inc., Hsinchu, Taiwan, from 2004 to 2009. Since 2009, he has been with National Sun Yat-sen University, Taiwan, where he is Professor of the Institute of Communications Engineering. His research interests center around the optimization in wireless multimedia networks.

Jiajia Guo received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2016, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2019. He is currently pursuing the Ph.D. degree in information and communications engineering with Southeast University, Nanjing, China. His current research interests include deep learning for physical layer and massive MIMO.

Michail Matthaiou (S'05–M'08–SM'13) was born in Thessaloniki, Greece in 1981. He obtained the Diploma degree (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece in 2004. He then received the M.Sc. (with distinction) in Communication Systems and Signal Processing from the University of Bristol, U.K. and Ph.D. degrees from the University of Edinburgh, U.K. in 2005 and 2008, respectively. From September 2008 through May 2010, he was with the Institute for Circuit Theory and Signal Processing, Munich University of Technology (TUM), Germany working as a Postdoctoral Research Associate. He is currently a Professor of Communications Engineering and Signal Processing and Deputy Director of the Centre for Wireless Innovation (CWI) at Queens University Belfast, U.K. after holding an Assistant Professor position at Chalmers University of Technology, Sweden. His research interests span signal processing for wireless communications, beyond massive MIMO, intelligent reflecting surfaces, mm-wave/THz systems and deep learning for communications.

Dr. Matthaiou and his coauthors received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2017. He currently holds the ERC Consolidator Grant BEATRICE (2021-2026) focused on the interface between information and electromagnetic theories. He was awarded the prestigious 2018/2019 Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship and also received the 2019 EURASIP Early Career Award. His team was also the Grand Winner of the 2019 Mobile World Congress Challenge. He was the recipient of the 2011 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East and Africa Region and a co-recipient of the 2006 IEEE Communications Chapter Project Prize for the best M.Sc. dissertation in the area of communications. He has co-authored papers that received best paper awards at the 2018 IEEE WCSP and 2014 IEEE ICC and was an Exemplary Reviewer for IEEE COMMUNICATIONS LETTERS for 2010. In 2014, he received the Research Fund for International Young Scientists from the National Natural Science Foundation of China. He is currently the Editor-in-Chief of Elsevier Physical Communication, a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and an Associate Editor for the IEEE JSAC SERIES ON MACHINE LEARNING FOR COMMUNICATIONS AND NETWORKS.

Bo Gao received his Ph.D. degree in information and communication engineering from Tsinghua University, Beijing, China, and his B.S. degree in China University of Geosciences, Wuhan, China, in 2015, and 2010, respectively. He joined the 5G/NR standard team of ZTE Corporation in 2015. He is actively involved in research and standardization activities in wireless communication. Most recently, he served as the Feature Lead for 3GPP 5G/NR UL power control. His current interest topics include mmWave communications, beam management, AI applications on MIMO, and UL power control technologies.