Deep Reinforcement Learning for Integrated Sensing and Communication in RIS-Assisted 6G V2X System

Xudong Long, Yubin Zhao, *Senior Member, IEEE*, Huaming Wu, *Senior Member, IEEE*, and Cheng-Zhong Xu, *Fellow, IEEE*

***Abstract*—The recent advancements in integrated sensing and communications (ISACs) technology have introduced new possibilities to address the quality of communication and high-resolution positioning requirements in the next-generation wireless communication network (6G) vehicle-to-everything (V2X). Simultaneously providing high-accurate positioning and high-communication capacity (CC) for the intelligent service of the vehicle target is challenging. In this article, we propose a reconfigurable intelligent surface (RIS)-assisted 6G V2X system to achieve highly accurate positioning of the vehicle target with basic communication requirements. We provide the CC and the 3-D fisher information matrix (FIM) formulations of the vehicle target. We demonstrate the direct impact of phase modulation in the reflector units on joint positioning accuracy and CC performance. Meanwhile, we design a flexible deep deterministic policy gradient (FL-DDPG) algorithm network with an ϵ-greedy strategy to solve the high-dimensional nonconvex optimization problem, achieves minimal positioning error while satisfying various CC requirements. Simulation results demonstrate that the FL-DDPG algorithm enhances positioning accuracy by a minimum of 89% and improves the achievable rate of the vehicle target by nearly 3 times, which outperforms traditional mathematical methods. Compared with classical deep reinforcement learning methods, FL-DDPG achieves better positioning accuracy while satisfying the communication requirements. When confronting imperfect channel, FL-DDPG enables addressing the channel estimation errors effectively on the ISAC system.**

***Index Terms*—6G V2X, deep reinforcement learning (DRL), Fisher information matrix (FIM), integrated sensing and communication (ISAC), reconfigurable intelligent surface (RIS).**

I. Introduction

With the rapid developments of intelligent transportation, autonomous driving, and vehicular network, there is an urgent need for reliable data communication and high-precise positioning information to benefit the safety, efficiency and comfort of driving. Combining the next-generation wireless communication network (6G), vehicle-to-everything (V2X) in smart transportation becomes attractive [1]. The 6G V2X system not only facilitates the exchange of information between vehicles, infrastructure and other roadside units, but also achieves accuracy positioning of vehicles. Thus, 6G V2X system becomes a necessary part for future transportation system.

Due to the expansion of frequency bands in the 6G network, integrated sensing and communication (ISAC) presents new opportunities for intelligent driving in V2X scenarios. Preliminary results indicate that implementing ISAC not only conserves resources and reduces hardware costs [2], but also achieves simultaneous positioning and communication of vehicles for users [3]. However, wireless signal propagation suffers severe propagation losses due to obstacles, which may degrade positioning accuracy and communication performance. Therefore, achieving performance enhancement in the ISAC system with these challenging communication conditions holds significant importance for 6G V2X design.

With the emergence of 6G communication, reconfigurable intelligent surfaces (RISs) stands as the key breakthrough technique for future communications. RIS is a planar structure composed of numerous low-cost passive reflecting units, capable of independently adjusting the amplitude and phase shift of incident signals for each unit [4], [5]. RIS can be easily installed on the exterior surfaces of various objects and can overcome performance degradation in ISAC systems caused by Non-Line-of-Sight (NLoS) obstacles and propagation losses [6], [7], [8]. With the advantages of improving spatial reuse and enhancing signal quality, RIS can be widely applied to 6G V2X systems.

For ISAC applications, the beamforming scheme can be constructed to ensure high-speed, reliable communication and minimize positioning errors by estimating the Angle of Departure (AoD), Angle of Arrival (AoA), and time of arrival (ToA) [9], [10], and by controlling the phase shift parameters of RIS. In this context, the Cramér–Rao lower bound (CRLB) and Shannon capacity serve as the main metrics for beamforming strategy design [11]. However, the passive beamforming design of RIS is a typical nonconvex integer programming problem, which is complicated and hard to solve [12]. Traditional mathematical approaches are

Manuscript received 14 May 2024; revised 16 July 2024; accepted 18 August 2024. Date of publication 30 August 2024; date of current version 6 December 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 62171484, and in part by the China University Research Innovation Fund, Ministry of Education under Grant 2023BD027. *(Corresponding author: Yubin Zhao.)*

Xudong Long and Yubin Zhao are with the School of Microelectronics Science and Technology, Sun Yat-Sen University, Zhuhai 519082, China (e-mail: longxd@mail2.sysu.edu.cn; zhaoyb23@mail.sysu.edu.cn).

Huaming Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China (e-mail: whming@tju.edu.cn).

Cheng-Zhong Xu is with the State Key Laboratory of IoTSC and the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: czxu@um.edu.mo).

Digital Object Identifier 10.1109/JIOT.2024.3449969

difficult to implement on the RIS due to the unavoidable programming limitations, such as the selection of initial points, high-dimension disasters, and continuous actions for real applications. These approaches are easily getting trapped in local optimal solutions in the real applications, degrading overall performance. Additionally, beamforming schemes require real-time adaptation since the targets are moving and the channel varies accordingly in the V2X scenario. Thus, an online and fast calculating beamforming scheme is necessary.

With advancements in deep learning and reinforcement learning, new approaches have been developed for solving the real time optimization problems [13], [14]. Deep reinforcement learning (DRL) which combines both advantages of deep learning and reinforcement learning, has become a promising technique for enhancing optimization performance [15]. On one hand, DRL can rapidly adapt to changes in the communication environment by continually learning and adjusting its strategies in spatial surroundings, thereby discovering the global optimal solution. On the other hand, the neural networks in DRL not only autonomously learn valuable features from high-dimensional input spaces but also effectively address optimization problems involving continuous actions in spatial contexts. Therefore, compared to traditional mathematical methods, DRL is more efficient and reliable for RIS-assisted ISAC systems.

In this article, we propose a RIS-assisted 6G V2X system and a related passive beamforming scheme for jointly optimizing ISAC performance. The main goal is to minimize the target positioning error while ensuring communication requirements. To solve the objective, we develop a flexible deep deterministic policy gradient (FL-DDPG) network for real time adaptation of RIS passive beamforming. In addition, we analyze the robustness of FL-DDPG with imperfect channel model. The major contributions of this work are four folds.

1) In the 6G V2X system, we introduce the RIS as the main component in ISAC applications to address the performance degradation caused by signal blockage and high-propagation loss. The proposed system is capable of minimizing positioning error while simultaneously satisfying diverse communication requirements.
2) We derive a generalized 3-D model for the Fisher information matrix (FIM) and the related CRLB of the RIS-assisted ISAC system. Then, the Shannon capacity of the 6G V2X system with RIS is derived. The formulations indicate the phase modulation impacts of RIS on overall ISAC performance. Thus, we formulate a joint objective for positioning optimization, which is a nonconvex integer programming.
3) To solve this problem, we develop an FL-DDPG network tailored for the RIS-assisted ISAC system. An ϵ-greedy strategy is integrated into the actor network, enhancing the system's ability to cope with fast-fading channels and channel changes through adaptive adjustment of ϵ. This network exhibits strong adaptability to optimization problems involving high-dimensional spaces and continuous actions, enabling it to provide globally optimal solutions for the joint optimization problem of position accuracy and communication capacity (CC).
4) In practical scenarios, due to the existence of channel estimation error, quantization error and other factors, imperfect channel can degrade the overall performance. Therefore, we derive the FIM and the associated CRLB under imperfect channel model condition for RIS-assisted ISAC systems. Furthermore, we analyze the robustness of FL-DDPG network in the presence of imperfect channel, demonstrating its capability to handle the impact of imperfect channel model on joint optimization of system positioning accuracy and CC.

Simulation results indicate that the RIS-assisted ISAC system proposed in our 6G V2X scenario can flexibly control the phase shift of the RIS to minimize positioning error while satisfying diverse communication requirements under various imperfect channel model conditions. In particular, we evaluate the number of RIS reflection units, the bits of RIS, signal transmission power, and the optimization duration, which are the main parameters for FL-DDPG. In the presence of severe signal fading, the proposed FL-DDPG significantly improves the position accuracy of targets by at least 89% and increases the CC at the target by nearly 30%.

The remainder of this article is structured as follows. Section II introduces related works; Section III presents the system model; Section IV derives the position error bound (PEB); Section V presents the problem formulation; Section VI presents the proposed FL-DDPG algorithm; Section VII derives the PEB under imperfect channel model; Section VIII illustrates the simulation results; and Section IX concludes this article.

II. Related Work

A. ISAC in V2X

In recent years, the development of ISAC technology within V2X scenarios has introduced new possibilities for reliable and secure autonomous driving, garnering widespread attention [16]. Many technologies equipped with ISAC functions have been proposed. Chiriyath et al. [17] proposed a joint signal model and the derivation for perception and communication. They defined a radar velocity estimation criterion based on the CRLB, and introduced a theoretical evaluation criterion for joint estimation of perception and communication. González-Prelcic et al. [18] conducted beam alignment experiments in the vehicle-to-infrastructure (V2I) scenario, confirming that perceptual information can be utilized to assist communication. Nartasilpa et al. [19] analyzed the interference of perceptual information on communication systems. Huang et al. [20] proposed an ISAC scheme that integrated frequency and spatial agility, ensuring that the communication performance had no impact on perceptual performance. Liu et al. [2] designed a multiple input–multiple output (MIMO) beamforming method that enhanced the system's positioning and communication performance in the V2X scenario. Wang et al. [21] developed a method for multivehicle tracking and identification association using ISAC signals, enhancing the system's communication performance by associating identification information from different vehicles. Zhang et al. [22] proposed a robust

transceiver design for ISAC systems with bounded channel estimation errors, which maximizes the minimum perceptual performance of multiple objectives while satisfying communication requirements. Liu et al. [23] demonstrated a generalized point-to-point ISAC model for addressing the joint optimization problem of perception and communication in imperfect channel. Zhang et al. [24] introduced a method for coverting ISAC systems with imperfect channel model, which ensures both communication performance and stealthiness while balancing multiple radar objectives. The aforementioned methods are able to effectively enhance ISAC performance in scenarios with direct channels. However, in actual V2X scenarios, obstacles lead to a degradation in both perceptual and communication performance.

B. RIS of ISAC

RIS can effectively address the degradation of positioning and communication performance in the context of ISAC systems caused by blocked channels. In NLoS propagation scenarios, Han et al. [25] validated that RIS had the capability to enhance communication performance. Basar et al. [26] established a mathematical framework for RIS-assisted point-to-point communication and analyzed the communication performance of the ISAC system. He et al. [27] investigated the potential performance improvement of a single RIS assisting the ISAC system in position in NLoS scenarios. Alegría and Rusek [28] also investigated the potential improvement of RIS-assisted position in NLoS scenarios and conducted theoretical analysis on the localization estimation using the CRLB. Wang and Zhang [29] proposed a RIS-assisted joint beamforming method, which improved the system's position accuracy to the centimeter level. Ammous and Valaee [30] proposed a RIS-assisted target position method based on Kalman filtering. He et al. [31] proposed an adaptive phase shifter design based on hierarchical codebooks and feedback from the mobile station to enhance the position accuracy of targets. Decarli et al. [32] derived the bounds of near-field positioning and assessed the role of RIS in V2X sidelink position accuracy. Basar et al. [33] demonstrated that the increase in the user signal-to-noise ratio (SNR) is directly proportional to the square of the number of RIS reflection elements in a single-input single-output (SISO) system. Huang et al. [34] proposed a two-step transmission protocol aimed at improving communication and perceptual performance through RIS-assisted channel estimation. Meng et al. [35] introduced an ISAC scheme in the V2X scenario, enhancing position and communication performance by integrating RIS on the vehicle surface. Liu et al. [36] proposed a RIS-assisted MIMO beamforming design for target localization and multiuser communication in traffic scenarios. Li et al. [37] proposed the first worst-case robust beamforming design problem in the RIS-assisted multiuser multiple input–single output (MU-MISO) system considering the imperfect channel model condition. Luan et al. [38] proposed a conditional value-at-risk (CVaR) method to address the chance constraints caused by imperfect channel in RIS-assisted ISAC systems. Hu et al. [39] demonstrated a multistrategy alternate optimization (MSAO) algorithm that optimizes beamforming vectors, sensor autocorrelation matrices, and RIS phase-shift matrices to mitigate the impact of imperfect channel on ISAC. However, the main challenge of the RIS assisted ISAC is that the coupling effect of SNR, power, positioning accuracy, CRLB, and other data together for phase shift optimization calculation. The phase shift design for the RIS should not only consider the joint optimization problem but also real time adaptation with low complexity.

C. DRL of ISAC

DRL, owing to its distinctive capabilities in addressing complex nonconvex problems, has been employed in optimizing phase-shift design of RIS. As demonstrated in [40] and [41], DRL models have been effective applied in addressing nonconvex optimization problems in the high-dimensional space of RIS-assisted ISAC systems. Faisal et al. [42] proposed an optimization algorithm framework based on DRL in the RIS-assisted wireless transmission system, achieving the upper limit of the received SNR at a relatively low-time cost. Xu et al. [43] proposed a RIS-assisted ISAC system for reducing the interference signals on high-speed rail communication performance, which utilized the DRL to solve the optimization problem of continuous phase shift changes in RIS. Zhong et al. [44] proposed a DRL algorithm that addressed the joint optimization problem of RIS phase shift and power allocation, while maintaining low complexity. Yang et al. [45] utilized DRL to address long-term stochastic optimization problems related to phase shifts. Lin et al. [46] introduced a learning algorithm based on deep Q-network for optimizing pilot interval and pilot power, obtaining the optimal estimation performance while reducing the system costs. Tang et al. [47] and Lei et al. [48] proposed a double Q-network-based DRL method for V2X edge computing. Although the above works are similar to ours, our proposed FL-DDPG mainly focuses on the optimization of the minimum CRLB with communication constraints in 6G V2X instead of just improving the communication quality. In addition, we employ specific strategy in FL-DDPG for rapid channel variations especially in the 6G V2X scenario.

III. System Model

We consider a RIS-assisted 6G V2X system as illustrated in Fig. 1. The system contains a base station (BS), a passive RIS, and several vehicles equipped with multiple antennas, which are considered targets. In addition, the system is implemented in the urban area, which has buildings as obstacles for signal propagation.

The BS is equipped with a uniform linear array (ULA) of N_b antennas, and each target is equipped with N_t antennas. Due to potential obstacles between the BS and the targets, which cause significant signal fading during transmission, RIS is employed to establish a virtual link. The RIS is a $N_x \times N_y$ uniform planar array (UPA) equipped with N_r reflective elements.

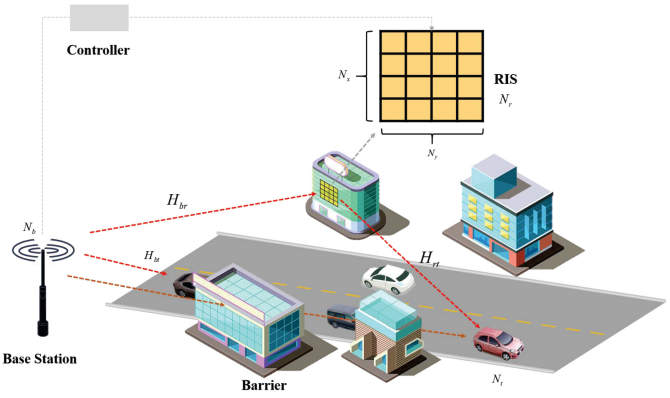


Fig. 1. RIS-assisted 6G V2X system framework. In the 6G V2X scenario, the BS is a multiantenna signal transmitter, communicating and positioning vehicle targets in the scene through wireless signals. The wireless signals from the BS can directly reach the vehicle targets and indirectly via RIS. The RIS is a planar array controlled by a central controller to improve the positioning and communication performance of the ISAC system by adjusting the phase shift and amplitude of the reflective units.

A. Downlink Transmit Signal

In the downlink, BS transmits data through M_b orthogonal frequency-division multiplexing (OFDM) subcarriers. The transmit data on nth subcarrier is $\mathbf{x}[n] = [x_1[n], x_2[n], \ldots, x_{M_b}[n]]^T \in \mathbb{C}^{M_b \times 1}$,which follows the complex Gaussian distribution with zero mean and unit variance, i.e., $\mathbb{E}[\mathbf{x}[n]\mathbf{x}[n]^H] = 1$. Let $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{M_b}] \in \mathbb{C}^{N_b \times M_b}$ represent the beamforming matrix, with $\mathbf{W}_i$ denoting the unit-norm transmitting vector. Then the downlink transmit signal vector is $\mathbf{Wx}[n]$. Thus, the BS transmit power is given by $\mathbb{E}||\mathbf{Wx}[n]||^2 = ||\mathbf{W}^2||$. This signal is transmitted to the target through both a direct channel and an indirect channel using the RIS.

B. Channel Model

We denote the channel matrix from BS to the RIS as $\mathbf{H}_{br} \in \mathbb{C}^{N_r \times N_b}$:

$$\mathbf{H}_{br} = \mathbf{a}_{br}\left(\varphi^a_{br}, \varphi^e_{br}\right)\mathbf{a}^H_{br}(\psi_{br}) \tag{1}$$

where $\mathbf{a}_{br}(\psi_{br})$ and $\mathbf{a}_{br}(\varphi^a_{br}, \varphi^e_{br})$ are the transmitter and receiver antenna response vectors from the BS to RIS, respectively; here, ψ_{br} is the transmission angle of the signal at the BS; φ^a_{br} is the azimuth AOA; and φ^e_{br} is the elevation AOA.

And the channel between RIS and target is $\mathbf{H}_{rt} \in \mathbb{C}^{N_t \times N_r}$:

$$\mathbf{H}_{rt} = \mathbf{a}_{rt}(\psi_{rt})\mathbf{a}^H_{rt}\left(\varphi^a_{rt}, \varphi^e_{rt}\right) \tag{2}$$

where $\mathbf{a}_{rt}(\varphi^a_{rt}, \varphi^e_{rt})$ and $\mathbf{a}_{rt}(\psi_{rt})$ are the transmitter and receiver antenna response vectors from the RIS to target, respectively; φ^a_{rt} and φ^e_{rt} are the azimuth AOD and elevation AOD at the RIS-target link; and ψ_{rt} is the AOA on the target side.

The direct channel from the BS to the target is denoted as $\mathbf{H}_{bt} \in \mathbb{C}^{N_t \times N_r}$

$$\mathbf{H}_{bt} = \mathbf{a}_{bt,\text{in}}(\psi_{tb})\mathbf{a}^H_{bt,\text{out}}(\psi_{bt}) \tag{3}$$

where $\mathbf{a}_{tb,\text{out}}(\psi_{tb})$ and $\mathbf{a}_{bt,\text{in}}(\psi_{bt})$ are the transmitter and receiver antenna response vectors. Angles ψ_{tb} and ψ_{bt} are the AOA and AOD of the signal from the BS to the target. These array response vectors is expressed as

$$\mathbf{a}(\psi) = \frac{1}{\sqrt{N_{\text{ant}}}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin\psi}, \ldots, e^{j\frac{2\pi}{\lambda}(N_{\text{ant}}-1))\sin\psi}\right]^T \tag{4}$$

where N_{ant} is the number of antennas, and λ and d denoted the signal wavelength and antenna spacing, respectively. In addition, we have

$$\mathbf{a}\left(\varphi^a, \varphi^e\right) = \frac{1}{\sqrt{N^2}}[1, e^{j\frac{2\pi}{\lambda}}d_r[m\sin\varphi^a\sin\varphi^e + n\cos\varphi^e, \ldots \\ e^{j\frac{2\pi}{\lambda}}d_r\left[(N_x-1)\sin\varphi^a\sin\varphi^e + \left(N_y-1\right))\cos\varphi^e\right]^T \tag{5}$$

where d_r be the interval of RIS reflect elements, and N_x and N_y represent the number of rows and columns of the RIS. Without loss of generality, we set $d = d_r = (\lambda/2)$.

Then, channel $\mathbf{H}[n]$ is comprised of two distinct components, namely, the LoS channel and the NLoS channel, which is given by

$$\mathbf{H}[n] = \mathbf{H}_{\text{LoS}}[n] + \mathbf{H}_{\text{NLoS}}[n] \tag{6}$$

where $\mathbf{H}_{\text{LoS}}[n]$ represents the direct channel between BS and target, while $\mathbf{H}_{\text{NLoS}}[n]$ corresponds to indirect channel, where signals depart from BS and pass through RIS before being transmitted to the target

$$\mathbf{H}_{\text{LoS}}[n] = \gamma_l h_l \mathbf{H}_{bt} e^{j2\pi B\frac{n}{N}\tau_l} \tag{7}$$

where $\gamma_l = \sqrt{([N_b N_t]/\rho_l)}$; ρ_l is the path loss of the direct reflecting channel; h_l is the small-scale fading propagation process; and τ_l is the corresponding transmission delay between BS and target.

The indirect link $\mathbf{H}_{\text{NLoS}}[n]$ is expressed as follows:

$$\mathbf{H}_{\text{NLoS}}[n] = \gamma_{nl} h_{nl} \mathbf{H}_{rt}\boldsymbol{\Theta}\mathbf{H}_{br} e^{j2\pi B\frac{n}{N}\tau_{nl}} \tag{8}$$

where $\gamma_{nl} = \sqrt{([N_b N_r]/\rho_{nl})}$, with ρ_{nl} is the path loss of the indirect channel, h_{nl} is the small-scale fading propagation process, and τ_{nl} is the corresponding transmission delay between BS-RIS-target link.

The $\boldsymbol{\Theta} = \mathbf{diag}(\mathbf{u}) \in \mathbb{C}^{N_r \times N_r}$ indicates the diagonal complex matrix of the RIS reflection parameter, and $\mathbf{u}$ is given by:

$$\mathbf{u} = \left[\rho_r e^{j\theta_1}, \rho_r e^{j\theta_2}, \ldots, \rho_r e^{j\theta_{N_r}}\right]^T \in \mathbb{C}^{N_r \times 1} \tag{9}$$

where $\rho_r e^{j\theta_i}$ represent the RIS reflection parameter of ith element, and ρ_r and θ represents the reflectivity coefficient and phase shift, respectively. Without loss of generality, we set $\rho_r = 1$. In practical applications, constraints related to costs and hardware limitations necessitate that the phase shift of each reflective unit can only be discretely selected from a limited set of values. Thus, we set $\theta_i \in \mathcal{A} = [0, \Delta\delta, 2\Delta\delta, \ldots, (2^{B_{it}} - 1)\Delta\delta]$, where $\Delta\delta$ is the uniformly distributed phase shift interval and B_{it} is the number of bits of RIS.

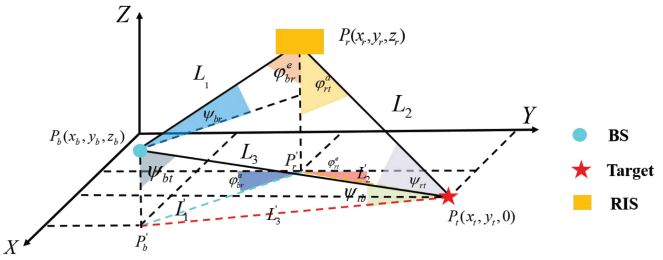


Fig. 2. System geometric model. It simplifies the 6G V2X ISAC system model and demonstrates the geometric relationship between BS, RIS, and the target in 3-D space.

C. Receiver Signal

The BS transmits signals to the target via both direct and indirect links. The received signal at the target is represented as

$$\mathbf{y}[n] = \sqrt{P_b}(\mathbf{H}_{\text{LoS}}[n] + \mathbf{H}_{\text{NLoS}}[n])\mathbf{W}\mathbf{x}[n] + \mathbf{n}_c[n] \quad (10)$$

where P_b is the BS transmitter power and $\mathbf{n}_c \sim \mathcal{N}(0, \sigma_c^2 \mathbf{I}_{N_t})$ is the additive white Gaussian noise. Therefore, the receive SNR of the target is expressed as

$$\gamma_c = \frac{P_b|(\mathbf{H}_{\text{LoS}}[n] + \mathbf{H}_{\text{NLoS}}[n])\mathbf{W}|^2}{\sigma_c^2}. \quad (11)$$

IV. Position Error Bound

We construct the 6G V2X scenario as a generic 3-D spatial model. Utilizing the geometric information among the BS, RIS, and the target, we systematically derive analytical expressions for the FIM and CRLB for RIS-assisted ISAC system.

A. Geometric Model

We simplify the scene in Fig. 1 and use geometric information to represent the spatial relationships among the BS, RIS, and target, which is illustrated in Fig. 2. The central position of the BS is $\mathbf{p_b} = [x_b, y_b, z_b]^T \in \mathbb{R}^3$, the coordinates of the RIS are given by $\mathbf{p_r} = [x_r, y_r, z_r]^T \in \mathbb{R}^3$, and the position of the target is represented as $\mathbf{p}_t = [x_t, y_t, 0] \in \mathbb{R}^3$, where the z-coordinate of $\mathbf{p}_t$ is 0 due to the target being located on the plane. Then, L_1, L_2, and L_3 are the Euclidean distances between the pairs of BS, RIS, and target, respectively. In addition, L_1', L_2', and L_3' are the projections of L_1, L_2, and L_3 on the X-Y plane.

According to Fig. 2, we derive the analytical expressions for the Euclidean distances L_1, L_2, and L_3, as well as the computational equations for L_1', L_2', and L_3'. Leveraging the geometric relationships among BS, RIS, and the target, we also derive analytical expressions for parameters, such as τ_l, τ_{nl}, ψ_{br}, φ_{br}^a, φ_{br}^e, ψ_{rt}, φ_{rt}^a, φ_{rt}^e, ψ_{bt}, and ψ_{tb}, which are detailed in Appendix A.

B. CRLB

We define $\hat{\boldsymbol{\zeta}}$ as the estimation of the general channel parameters $\boldsymbol{\zeta}$ which is a complex parameter vector related to the channel in the geometric space. Since the coordinates of BS and RIS are fixed, it is not necessary to consider the influence of ψ_{br}, φ_{br}^a, and φ_{br}^e on CRLB derivation. The specific expression is as follows:

$$\boldsymbol{\zeta} = \left[\tau_l, \tau_{nl}, \psi_{rt}, \varphi_{rt}^a, \varphi_{rt}^e, \psi_{bt}, \psi_{tb}\right]. \quad (12)$$

The mean squared error matrix of $\boldsymbol{\zeta}$ is satisfied by the following inequality:

$$\mathbb{E}\left\{\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right)\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right)^H\right\} \geq \mathbf{J}_{\boldsymbol{\zeta}}^{-1} \quad (13)$$

where $\mathbf{J}_{\boldsymbol{\zeta}} \in \mathbb{C}^{7\times 7}$ is the FIM for channel parameters. The entries of $\mathbf{J}_{\boldsymbol{\zeta}}$ can be expressed as follow:

$$\left[\mathbf{J}_{\boldsymbol{\zeta}}\right]_{i,j} = \frac{2P_b}{\sigma_s^2}\sum_{n=1}^{N}\Re e\left\{\frac{\partial \boldsymbol{\mu}^H}{\partial \zeta_i}\frac{\partial \boldsymbol{\mu}}{\partial \zeta_j}\right\} \quad (14)$$

where $\boldsymbol{\mu} = \mathbf{H}[n]\mathbf{W}\mathbf{x}[n]$ and $\boldsymbol{\zeta}_i$ is the ith entry of $\boldsymbol{\zeta}$, since it is an OFDM system. The solution procedure is shown in Appendix B.

The FIM for PEB can be obtained by means of the 2×7 transformation matrix $\mathbf{T}$, which is expressed as

$$\mathbf{J} = \mathbf{T}\mathbf{J}_{\boldsymbol{\zeta}}\mathbf{T}^H \quad (15)$$

where the transformation matrix $\mathbf{T}$ is expressed as

$$\mathbf{T} = \begin{bmatrix} \frac{\partial \tau_l}{\partial \mathbf{p}_x} & \frac{\partial \tau_{nl}}{\partial \mathbf{p}_x} & \frac{\partial \psi_{rt}}{\partial \mathbf{p}_x} & \frac{\partial \varphi_{rt}^a}{\partial \mathbf{p}_x} & \frac{\partial \varphi_{rt}^e}{\partial \mathbf{p}_x} & \frac{\partial \psi_{bt}}{\partial \mathbf{p}_x} & \frac{\partial \psi_{tb}}{\partial \mathbf{p}_x} \\ \frac{\partial \tau_l}{\partial \mathbf{p}_y} & \frac{\partial \tau_{nl}}{\partial \mathbf{p}_y} & \frac{\partial \psi_{rt}}{\partial \mathbf{p}_y} & \frac{\partial \varphi_{rt}^a}{\partial \mathbf{p}_y} & \frac{\partial \varphi_{rt}^e}{\partial \mathbf{p}_y} & \frac{\partial \psi_{bt}}{\partial \mathbf{p}_y} & \frac{\partial \psi_{tb}}{\partial \mathbf{p}_y} \end{bmatrix} \quad (16)$$

where the parameters in the matrix $\mathbf{T}$ are derived in Appendix B.

Finally, a generalized analytical expression for the CRLB is the inverse matrix of $\mathbf{J}_{\boldsymbol{\zeta}}$, and the PEB is defined as the trace of CRLB

$$\mathbf{PEB} = \sqrt{\mathbf{tr}\left(\mathbf{J}^{-1}\right)}. \quad (17)$$

V. Problem Formulation

The main objective of the RIS-assisted ISAC 6G V2X system is to minimize the target positioning errors while meeting various communication requirements. The varies communication requirements are measured according to the CC

$$R_c = B\log_2(1 + \gamma_c). \quad (18)$$

Then, we employ the PEB as the positioning error metric. In order to minimize the PEB while considering the constraints on communication achievable rate, we formulate the following optimization problem:

$$\begin{aligned} (\mathbb{P}_1): \ & \mathbf{PEB} = \arg\min \sqrt{\mathbf{tr}\left(\mathbf{J}^{-1}\right)} \\ \text{s.t.} \quad & R_c > R_{\min} \\ & \boldsymbol{\Theta} = \mathbf{diag}\left(e^{j\theta_1}, \ldots, e^{j\theta_{N_r}}\right) \\ & \theta_i \in \mathcal{A}, i = 1, 2, \ldots, N_r \end{aligned} \quad (19)$$

where $R_{\min}$ represents the minimum achievable capacity constraint that the target can accept in the communication process. For multiple targets, the objective is changed into the minimum sum of CRLBs with the total communication constraints.

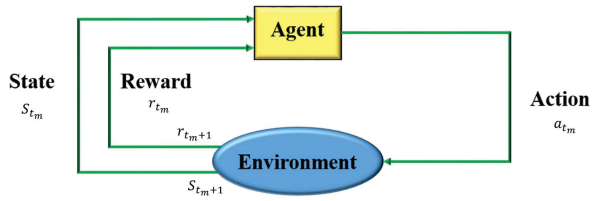


Fig. 3. MDP. At time t_m, the agent executes action a_{t_m} in the environment and then updates the action a_{t_m+1} for the next time step based on the state s_{t_m} and reward r_{t_m}.

Meanwhile, we can also consider the balanced performance among multiple targets, e.g., achieving the minimum of the maximum CRLB or the averaged CRLB, which is beyond the scope of this work.

VI. Deep Reinforcement Learning Algorithm

A. Deep Reinforcement Learning

DRL combines deep learning and reinforcement learning based on Markov decision process (MDP), and contains 4 components, which are agent, action, reward, and a deep neural network (DNN). Agent is a decision-making entity with the goal of learning, through interactions with the environment, to maximize the cumulative reward in a series of decisions. State not only represents the environmental information perceived by the agent and the changes induced by its own behavior, but also serves as the foundation for the agent's decision-making and the estimation of its long-term rewards. Action refers to the behaviors or operations that the agent can choose in a given state. DNN is typically employed for the policy function, determining the probabilities of actions chosen by the agent in a given state. It involves learning and extracting complex relationships among states, actions, and rewards from experience.

B. Markov Decision Process

As illustrated in Fig. 3, the MDP includes an agent, a collection of environment states s_{t_m}, a set of actions a_{t_m}, and a reward function r_{t_m}. Within the framework, the downlink signal is transmitted from the BS to the RIS. The control unit embedded within the RIS executes action while in state according to a specific strategy. Subsequently, the reward mechanism computes the cumulative discounted reward associated with the state subsequent to the execution of the aforementioned action.

Agent: Here, we designate the RIS as the agent of the ISAC system. The decision-making strategy is implemented to solve the objective, which is to flexibly manipulate the phase-shifting matrix $\mathbf{\Theta}$ to minimize the PEB while considering the constraints of the communication achievable rate.

State: During system optimization, phase shift will be adjusted in the each time period. Therefore, a reasonable state space design can not only improve the positioning accuracy of the system but also ensure the communication requirements of targets. Here, we define that the state space $s_{t_m} = (\mathbf{\Theta}_{t_m}, R_{c_{t_m}}, \mathbf{H}_{t_m})$ in t_m time period which is composed of phase shift information $\mathbf{\Theta}_{t_m}$, CC $R_{c_{t_m}}$, and channel information $\mathbf{H}_{t_m}$.

It should be noted that due to the inherent limitations of the neural network architecture employed in DRL, it is unable to process complex numbers as inputs. Hence, during the process of state construction, the complex number presented in the system is decomposed into distinct real and imaginary components, e.g., the channel from BS to target $\mathbf{H}_{bt} = \mathbf{Re}\{\mathbf{H}_{bt}\} + \mathbf{Im}\{\mathbf{H}_{bt}\}$.

Action: In (19), the RIS calculates PEB and the CC simultaneously. The variation in the RIS phase shift can impact the performance of communication and localization in the ISAC system. Therefore, the action space s_{t_m} includes the values of $\mathbf{\Theta}_{t_m}$. At time t_m, the agent takes action a_{t_m} according to the policy ϖ.

Reward: When the agent executes an action based on the policy, it obtains a new state s_{t_m+1}. The reward is then used to assess whether the obtained state s_{t_m+1} satisfies (19). Positive rewards signify the optimization framework's goal, which is to continuously enhance positioning accuracy. Since higher positioning accuracy corresponds to smaller PEB values and larger (1/PEB), we define the penalty function as

$$r_t = \begin{cases} \frac{1}{\mathrm{PEB}} * \varrho & \text{if} \quad C_{\text{on}} = U_n \\ \frac{1}{\mathrm{PEB}} & \text{otherwise} \end{cases} \tag{20}$$

where $\varrho \in (0, 1)$ represents a reward factor $C_{\text{on}} = U_n$ to indicate that the constraints outlined in (19) are not satisfied. This serves to reduce the impact on the positioning accuracy of the system and ensures satisfaction with the communication requirements of the system.

C. FL-DDPG

According to the framework of MDP, we design a FL-DDPG algorithm, which not only improves the system's capability to adapt to fast-fading channels and channel estimation errors but also tends to favor selecting the currently known optimal action during each action choice by incorporating an ϵ-greedy policy. The regulation of RIS phase shifts needs to comprehensively consider multiple factors, such as the positioning errors of multiple targets, CC, and the mutual interference between target users, among other complex conditions. As illustrated in Fig. 4, the FL-DDPG neural network contains the actor network, the critic network, the actor target network, the critic target network, and the replay buffer.

Actor Network: The actor network, also called the policy network with parameter θ_{ϖ}, is dedicated to learning the decision parameter strategy of the entire algorithm. At time t_m, with the environmental state s_{t_m}, the actor network executes the corresponding action a'_{t_m} based on the policy ϖ

$$a_{t_m} = \mu_{t_m}(s_{t_m}|\omega_{\varpi}). \tag{21}$$

To address the impact of rapid channel variations on system performance, we design a greedy strategy to balance the relationship between exploration and exploitation. This enhances the randomness and coverage of the learning process, thereby improving the system's generalization and robustness. Here, we define $a_{t_m\epsilon}$

$$a_{t_m\epsilon} = \mu_{t_m}(s_{t_m}|\omega_{\varpi}) + \epsilon * n_e \tag{22}$$

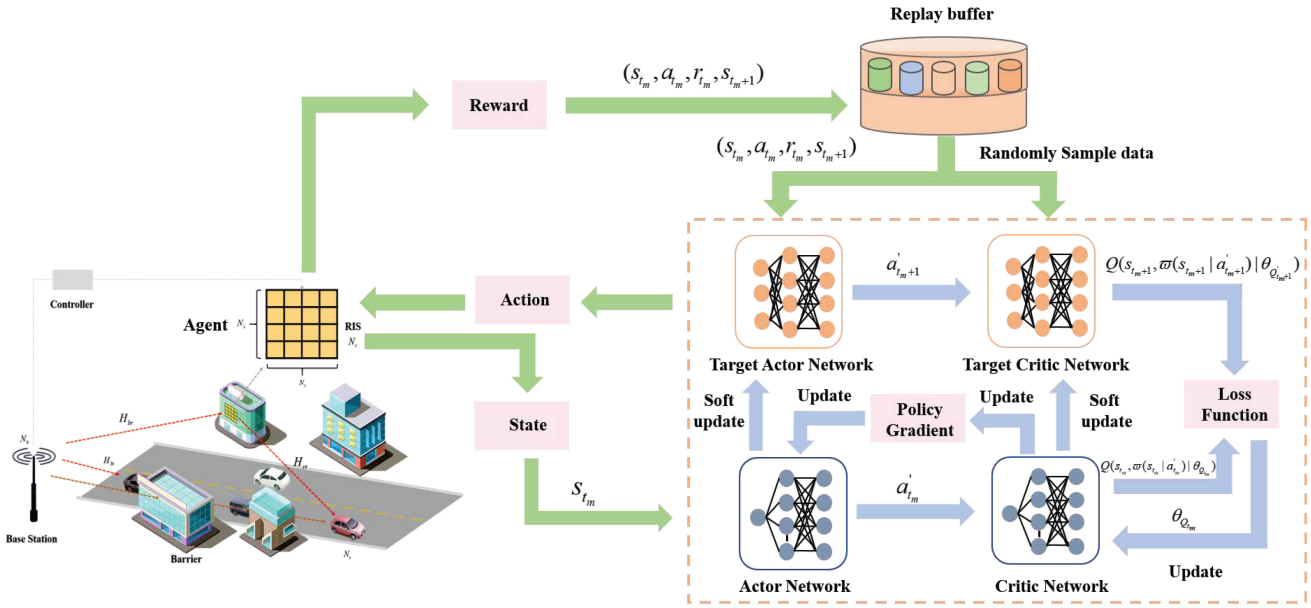


Fig. 4. FL-DDPG Network Framework. The FL-DDPG algorithm consists of an agent, a replay buffer, and four DNN networks. The controller within the RIS controls the operation of the entire FL-DDPG algorithm. The controller first initializes the DNN networks of the FL-DDPG algorithm and the replay buffer. Then, during the interaction between the RIS and the environment, the controller regulates the reflection coefficient of RIS based on the actions generated by the actor network. Meanwhile, the data from each interaction is stored in the replay buffer and small batches of data are regularly sampled from the replay buffer. The DNNs network parameters are updated by the FL-DDPG algorithm. Finally, based on the updated network, the phase of each reflection unit on the RIS is adjusted again.

where ϵ denotes the exploration rate in the greedy strategy; n_e denotes the exploration noise; and ω_{ϖ} is the set of the parameters of actor online network $\mu_{t_m}(\cdot)$. Then, utilizing the ϵ-greedy strategy, an appropriate action is chosen between a_{t_m} and $a_{t_m\epsilon}$ as the new a_{t_m}.

Critic Network: The critic network is a Q-network based on parameter θ_Q, which primarily assesses the strategy ϖ in the actor network and updates the actor network parameters. At time t_m, the input of the critic network is the current state s_{t_m} and action a'_{t_m}, and the output is the corresponding Q-function value $Q(s_{t_m}, \varpi(s_{t_m}|a'_{t_m})|\theta_{Q_{t_m}})$.

Target Actor Network: The target actor network generates the target actions to be executed in the next state. At time t_{m+1}, with the environmental state $s_{t_{m+1}}$, the target actor network executes the corresponding action $a'_{t_{m+1}}$ based on the policy ϖ. Subsequently, the parameters of the actor network are slowly transferred to the target actor network using a soft update strategy. The soft update of the parameter $\theta_{\varpi}(t_{m+1})$ is expressed as

$$\theta_{\varpi'_{t_{m+1}}} = \tau_{\varpi}\theta_{\varpi_{t_m}} + (1-\tau_{\varpi})\theta_{\varpi'_{t_{m+1}}} \tag{23}$$

where $0 < \tau_{\varpi} \ll 1$ represents the soft updating factors.

Target Critic Network: Target critic network is also a Q-network. At time t_{m+1}, the input of the critic network is the next state s_{t_m+1} and generated action a'_{t_m+1}, and the output is the corresponding Q-function value $Q(s_{t_m+1}, \varpi(s_{t_m+1}|a'_{t_m+1})|\theta_{Q'_{t_m+1}})$. Similar to the target actor network, parameters $\theta_Q(t_m)$ are also slowly updated from the critic network parameters using a soft update strategy. The parameter updates can be expressed as follows:

$$\theta_{Q'_{t_m+1}} = \tau_Q\theta_{Q_{t_m}} + (1-\tau_Q)\theta_{Q'_{t_m+1}} \tag{24}$$

where $0 < \tau_Q \ll 1$ represents the soft updating factors.

At each training slot t_m, the agent observes the RIS-assisted positioning accuracy and communication performance, resulting in the acquisition of an environmental state s_{t_m} This state is then input into the actor network, which subsequently produces the corresponding action a_{t_m}. At each time t_m step, a_{t_m} is chosen based on the ϵ-greedy strategy, and the phase shift of the RIS is updated in real-time. To adapt more flexibly to the dynamic changes in the system, ϵ varies during different training epochs. In the early stages of training, more emphasis is placed on exploration, and ϵ is initially set to a larger value. As the learning progresses, ϵ gradually decreases to increase the utilization of known information. Subsequently, the agent carries out the action, and the critic network computes the corresponding reward, allowing the agent to acquire a new state s_{t_m+1}. With each updated state, the agent accumulates a series of experience tuples, referred to as $(s_{t_m}, a_{t_m}, r_{t_m}, s_{t_m+1})$. Each of these experience tuples is stored in the replay memory $\mathcal{D}$ to facilitate the training of the neural network. Ultimately, through the continuous adjustment of the parameters in both the actor and critic networks, the optimal strategy is determined. This optimal strategy ensures that the highest level of positioning accuracy is achieved while also satisfy the target's requirements for communication performance.

During the training process, a minibatch of $\hbar$-size will be randomly sampled from the replay memory $\mathcal{D}$. The parameters of the critic network are then updated using the temporal difference error as a reference. The loss function is defined as follows:

$$L(\theta_Q) = \frac{1}{\hbar}\left[Q'_{t_m+1} - Q(s_{t_m}, \varpi(s_{t_m}|a'_{t_m})|\theta_{Q_{t_m}})\right]^2 \tag{25}$$

Algorithm 1 Training Process of FL-DDPG Algorithm

Initialize experience replay memory $\mathcal{D}$;
Initialize the training critic network θ_Q and the training actor θ_ϖ network separately, random weighs, and bias;
Initialize ISAC system
Input: Channel information;
Output: PEB, Achievable rate;
for each episode e = 1,..., $E_{\max}$ do:
  Initialize state space as s_0 and reset RIS-assisted location system;
  for $t_m = 0, 1, 2, \ldots, T_{\max}$ do:
    Update ϵ exploration rate in the greedy strategy.
    The agent choose action a_{t_m} according to current policy ϖ;
    $a_{t_m\epsilon} = \mu_{t_m}(s_{t_m}|\omega_\varpi) + \epsilon * n_e$
    ϵ-greedy strategy choice a_{t_m}
    Execute action a_{t_m}, observe reward r_{t_m}, and s_{t_m} evolves into next state s_{t_m+1};
    Save $(s_{t_m}, a_{t_m}, r_{t_m}, s_{t_m+1})$ into $\mathcal{D}$;
    Soft update the target networks of central trainer according to (22);
    if stored tuples $\geq (1/3)|\mathcal{D}|$ then
      Randomly sample $\hbar$ transitions form $\mathcal{D}$;
      Update the critic network by minimizing the loss in (25);
      Update the actor network by maximizing the policy gradient in (27);
    end if
  end if
end if

where Q'_{t_m+1} is the target value of the state-value function, which is calculated by the Bellman equation

$$Q'_{t_m+1} = r_{t_m} + \gamma_b \max Q\Big(s_{t_m+1}, \varpi(s_{t_m+1}|a'_{t_m+1})|\theta_{Q'_{t_m+1}}\Big). \tag{26}$$

The critic network parameters is updated by minimizing the loss function (25)

$$\theta_Q \longleftarrow \theta_Q - \tau_{tc} \bigtriangledown_{\theta_Q} L\big(\theta_Q\big). \tag{27}$$

The optimization objective of the actor network is to maximize the state-action function **Q**. Given that **Q** is differentiable and the action space is continuous, the actor network can be updated using the policy gradient with an ascent factor, as demonstrated below

$$\bigtriangledown_{\theta_Q} J(\theta_\varpi) = \frac{1}{\hbar}\sum_{i=1}^{\hbar} \bigtriangledown_s Q(s,a)| \times \theta_Q|_{s(i),a(i)} \bigtriangledown \theta_\varpi \varpi(s|\theta_\varpi)|_{s(i)} \tag{28}$$

where $s = s_{t_m}$ and $a = \varpi(s_{t_m})$.

The complete algorithm flow described above is depicted in Algorithm 1, which can intuitively observe the FL-DDPG training process.

The controller within the RIS controls the operation of the entire FL-DDPG algorithm. The controller first initializes the DNN networks of the FL-DDPG algorithm and the replay buffer. Then, during the interaction between the RIS and the environment, the controller regulates the reflection coefficient of RIS based on the actions generated by the actor network. Meanwhile, the data from each interaction is stored in the replay buffer, and small batches of data are regularly sampled from the replay buffer. The DNNs network parameters are updated by the FL-DDPG algorithm. Finally, based on the updated network, the phase of each reflection unit on the RIS is adjusted again.

D. Complexity Analysis

In the four DNNs of FL-DDPGD, the actor network is composed of an input layer, three hidden layers, and an output layer, with each ith network having l_{a_i} neurons. Rectified linear unit activation functions are applied to the hidden layers and the output layer, while the output layer uses a sigmoid activation function. Similarly, the critic network consists of the same five-layer architecture, with the number of neurons in the ith layer denoted as l_{c_i}.

During the training process, both actor network and critical network participate in the training, and the parameters are updated and iterated through backward propagation. In addition, the training process also involves the prediction results of the target actor network and the target critical network. Then, the algorithm complexity for all single backward propagation training steps is $\mathcal{O}(\sum_{i=0}^{3} 2l_{a_i}l_{a_{i+1}} + \sum_{i=0}^{3} 2l_{c_i}l_{c_{i+1}})$.

During the online application process, data only needs to pass through the actor network. For any V2X traffic environment state s_{t_m}, the trained actor network will output the corresponding action a_{t_m}. Based on the principles of connection and computation in DNNs, the computational complexity can be determined as $\mathcal{O}(\sum_{i=0}^{3} l_{c_i}l_{c_{i+1}})$.

Throughout the entire algorithm execution, the agent initiates subsequent actions only when the number of tuples stored in replay memory $\mathcal{D}$ exceeds $\geq (1/3)|\mathcal{D}|$. Therefore, the overall complexity of the proposed FL-DDPG algorithm in this article is denoted as $\mathcal{O}((E_{\max}T_{\max} - (1/3)|\mathcal{D}|)\hbar(\sum_{i=0}^{3} 2l_{a_i}l_{a_{i+1}} + \sum_{i=0}^{3} 2l_{c_i}l_{c_{i+1}} + \sum_{i=0}^{3} l_{c_i}l_{c_{i+1}}) + (1/3)|\mathcal{D}|(\sum_{i=0}^{3} l_{a_i}l_{a_{i+1}}))$.

After offline training, FL-DDPG can quickly provide corresponding actions based on the current state in real-time communication and control systems by simply propagating forward through the actor network.

VII. Imperfect Channel

Here, we also consider the imperfect channel model for the whole system. We denote the imperfect channel between the BS and the RIS as $\hat{\mathbf{H}}_{br}$

$$\hat{\mathbf{H}}_{br} = \mathbf{H}_{br} + \Delta\mathbf{H}_{br} \tag{29}$$

where $\Delta\mathbf{H}_{br}$ is the random channel error and $||\Delta\mathbf{H}_{br}||_F \leq \jmath_{br}$, where $\jmath_{br}$ represents the radius of the uncertain region where the BS is known. The imperfect channel model characterizes that channel quantization errors naturally belong to a bounded region.

The imperfect channel between the RIS and the target is represented by $\hat{\mathbf{H}}_{rt}$

$$\hat{\mathbf{H}}_{rt} = \mathbf{H}_{rt} + \Delta\mathbf{H}_{rt} \tag{30}$$

where $\Delta\mathbf{H}_{rt}$ is the random channel error and $||\Delta\mathbf{H}_{rt}||_F \leq \jmath_{rt}$, where $\jmath_{rt}$ also represents the radius of the uncertain region where the BS is known.

The imperfect channel between the BS and the target is $\hat{\mathbf{H}}_{bt}$

$$\hat{\mathbf{H}}_{bt} = \mathbf{H}_{bt} + \Delta\mathbf{H}_{bt} \tag{31}$$

where $\Delta\mathbf{H}_{bt}$ is the random channel estimate error and $||\Delta\mathbf{H}_{bt}||_F \leq \jmath_{bt}$, where $\jmath_{bt}$ also represents the radius of the uncertain region where the BS is known.

We derive the FIM with imperfect channel model. In this case, (10) can be rewritten as

$$\hat{\mathbf{y}}[n] = \sqrt{P_b}\left(\hat{\mathbf{H}}_{br}[n] + \hat{\mathbf{H}}_{rt}[n]\boldsymbol{\Theta}\hat{\mathbf{H}}_{br}[n]\right)\mathbf{W}\mathbf{x}[n] + \hat{\mathbf{n}}_c[n] \tag{32}$$

where $\hat{\mathbf{n}}_c \sim \mathcal{N}(0, (\sigma_c^2 + \sigma_e^2)\mathbf{I}_{N_t})$ is the additive white Gaussian noise, σ_e^2 is the error increment.

Similarly, (14) can be rewritten as

$$\left[\hat{\mathbf{J}}_{\boldsymbol{\xi}}\right]_{i,j} = \frac{2P_b}{\sigma_s^2}\sum_{n=1}^{N}\Re e\left\{\frac{\partial\hat{\boldsymbol{\mu}}^H}{\partial\zeta_i}\frac{\partial\hat{\boldsymbol{\mu}}}{\partial\zeta_j}\right\} \tag{33}$$

where $\hat{\boldsymbol{\mu}} = (\hat{\mathbf{H}}_{br}[n] + \hat{\mathbf{H}}_{rt}[n]\boldsymbol{\Theta}\hat{\mathbf{H}}_{br}[n])\mathbf{W}\mathbf{x}[n]$. The solution procedure is shown in Appendix C.

With imperfect channel, $\hat{\mathbf{J}}$ for $\mathbf{P\hat{E}B}$ is represented as follows:

$$\hat{\mathbf{J}} = \mathbf{T}\hat{\mathbf{J}}_{\zeta}\mathbf{T}^H. \tag{34}$$

Finally, the CRLB is as follows:

$$\mathbf{P\hat{E}B} = \sqrt{\mathbf{tr}\left(\hat{\mathbf{J}}^{-1}\right)}. \tag{35}$$

VIII. Simulation

A. Simulation Parameters

The proposed FL-DDPG is evaluated via extensive RIS assisted 6G V2X simulations. We simulate a 1000 m × 1000 m V2X area. We set that BS is located at (900, 100, 20) and the target is located at (500, 500, 0). The RIS is located at (200, 300, 40). The numbers of transmitter and receiver antennas are $N_b = 4$ and $N_t = 4$, respectively. The number of reflecting elements at the RIS is set as N_r= 64. The carrier frequency is $f_c = 28$ GHz and the number of subcarriers is $N_s = 10$. The bandwidth is $B = 20$ MHz. The path loss exponent of the direct channel is $\alpha_l = 3.2$, the path loss exponent of the RIS is $\alpha_{nl} = 2.2$, and shadow fading parameters of the direct path and reflecting path are, respectively, $\sigma_l = 3$ and $\sigma_{nl} = 4$. The hyperparameters description FL-DDPG in the algorithm are shown in Table I.

TABLE I
FR-DDPG Super Parameters

Parameter	Description	Value
γ_b	Discounted rate	0.95
μ_{tc}	Learning rate for training critic network	0.001
μ_{ta}	Learning rate for training actor network	0.001
τ_{tc}	Learning rate for target critic network	0.001
τ_{ta}	Learning rate for target actor network	0.001
λ_{tc}	Decaying rate for training critic network	0.00001
λ_{ta}	Decaying rate for training actor network	0.00001
$\mathcal{D}$	Buffer size for experience replay	10000
E_{max}	The number of steps in each episode	10000
T_{max}	The number of experiences in the mini-batch	16
Optimizer	Adam	-
Activation function	ReLu	-

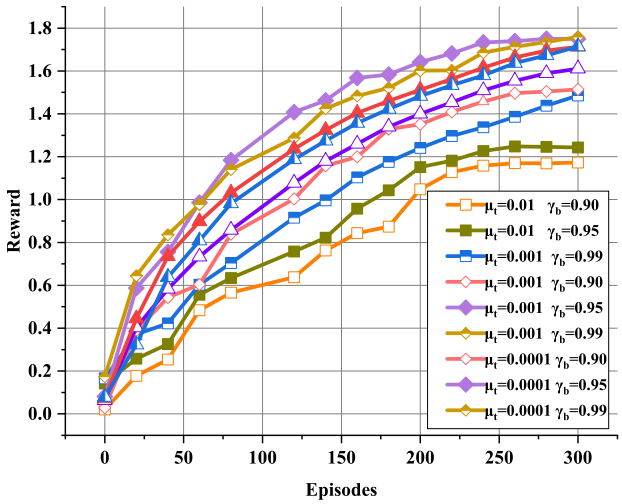


Fig. 5. Evaluation of hyperparameters on FL-DDPG convergence performance. When the replay buffer size is 10 000, $\mu_t = 0.001$, and $\varrho = 0.95$, the FL-DDPG network training can achieve stable and rapid convergence.

B. Hyperparameter Evaluation

We evaluated the impact of different hyperparameters on the convergence performance of the FL-DDPG algorithm training. An over small replay buffer size increases the risk of overfitting, while an over large size increases computation and sampling overhead. After extensive experiments, we set the replay buffer size to 10 000 for simulation experiments, with the learning rate $\mu_t = \mu_{tc} = \mu_{ta} = \tau_{tc} = \tau_{ta}$. As illustrated in Fig. 5, when $\mu_t = 0.01$, the training converges quickly but unstable and prone to divergence. When $\mu_t = 0.0001$, the training process is stable, but the convergence speed is very slow. Appropriately increasing the value of the discount factor γ_b can enhance the training performance. However, if the value of γ_b is too large, it can decrease the convergence speed. In summary, when the replay buffer size is 10 000, $\mu_t = 0.001$, and $\gamma_b = 0.95$, the FL-DDPG network training can achieve stable and rapid convergence.

C. Convergence Evaluation

We evaluate the convergence performance of the FL-DDPG system framework with the ϵ-greedy strategy. As depicted in Fig. 6, the CC requirement (CCR) is set to 60 bit/s/Hz. In the early stages of training, the noise interference during the action

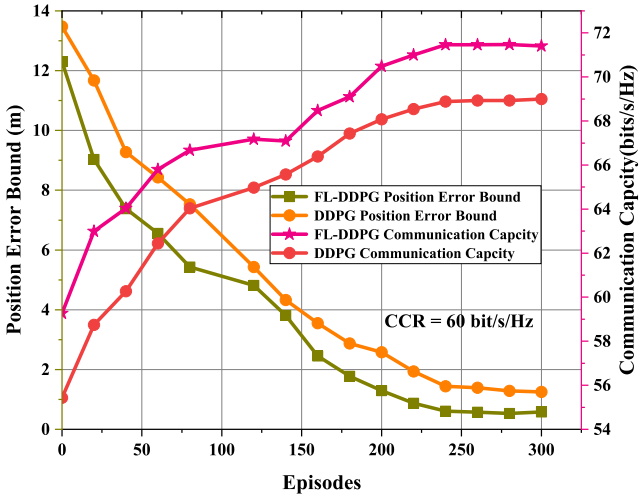


Fig. 6. RIS-assisted 6G V2X system framework. The convergence performance of FL-DDPG and DDPG in terms of positioning accuracy and communication capacity, with a CCR of 60 bit/s/Hz.

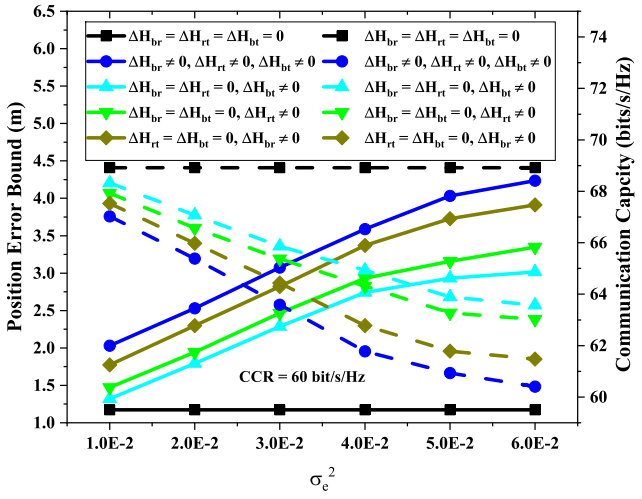


Fig. 7. Imperfect channel model condition. The impact of imperfect channel model on the robustness of RIS-assisted ISAC system, CCR 60 bit/s/Hz, and error increment σ_e^2.

selection process of FL-DDPG and the randomness of the replay buffer will cause fluctuations in positioning accuracy and CC. After 300 training cycles, the PEB converges from 9.43 m to 0.51 m, and the CC increases from 58.43 to 72.43 bit/s/Hz. With the same settings, DDPG without the ϵ-greedy strategy in the actor network converges to a PEB of 1.37 m and a channel capacity of 68.993 bit/s/Hz. Thus, the FL-DDPG network adapts better to rapid channel variations and enhances the system's robustness in 6G V2X scenario.

D. Imperfect Channel Model

We set the number of reflecting elements in the RIS to 64, with a transmission power of 25 dBm, and a CCR of 60 bit/s/Hz. By adjusting the error increment σ_e^2, we investigate the impact of the imperfect channel model on RIS-assisted ISAC systems. As illustrated in Fig. 7, when $\Delta\mathbf{H}_{br} \neq 0$, $\Delta\mathbf{H}_{rt} \neq 0$ and $\Delta\mathbf{H}_{bt} \neq 0$, the PEB increases from 2.03 m to 4.24 m, and the CC decreases from 67.02 to 60.39 bit/s/Hz with the increasing error variance σ_e^2. Additionally, we individually consider the impact of channel errors $\Delta\mathbf{H}_{bt}$, $\Delta\mathbf{H}_{rt}$, and $\Delta\mathbf{H}_{br}$ on system robustness. The PEB increases to 3.01 m, 3.34 m, and 3.91 m, respectively, while the CC decreases to 63.57, 63.02, and 61.47 bit/s/Hz. It is observed that as the increases of σ_e^2, the CC of the ISAC system decreases. However, all results satisfy the CCR of 60 bit/s/Hz, with the worst PEB being 3.91 m, which is within an acceptable range. This demonstrates that the FL-DDPG network is capable of handling the impact of channel estimation errors on ISAC systems, ensuring the normal operation of ISAC systems.

E. Position Accuracy

We compare the FL-DDPG with heuristic particle swarm optimization (PSO), and genetic algorithm (GA) to assess its positioning and communication performance in various scenarios. First, as illustrated in Fig. 8, with different CCR constraint settings, the positioning accuracy using FL-DDPG outperforms PSO and GA. The positioning accuracy is enhanced by a minimum of 89% when compared to the positioning system without RIS. Then, the CC estimation obtained through FL-DDPG outperforms PSO and GA. Regardless of whether the CCR is set at 60, 65, or 70 bit/s/Hz, the final estimated CC is higher than the CCR, meeting the communication requirements of targets. Compared with the system without RIS, the CC is improved by nearly 3 times. With an increase in the number of RIS reflection units, both the positioning accuracy and CC of the system rise. This indicates that the number of RIS reflective units is a dominant factor that affects the positioning accuracy and communication performance.

F. Time Consumption

Here, we compare the time consumptions of using FL-DDPG, PSO, and GA to assess the efficiency. As depicted in Fig. 9, FL-DDPG exhibits lower execution times compared to the PSO and GA. It is worth noticing that the runtime of all three optimization algorithms gradually increases as the number of RIS reflection units changes. This phenomenon is attributed to the relationship between the estimation matrix of the FIM and the number of RIS reflection units. Thus, FL-DDPG outperforms others in terms of efficiency, making it more suitable for the flexible adjustment of communication and perceptual positioning functions.

G. Deep Reinforcement Learning Comparison

We compare FL-DDPG with the state-of-art DRL methods, including proximal policy optimization (PPO), twin delayed deep deterministic policy gradient (TD3), and soft actor-critic (SAC), to assess its positioning and communication performance. As illustrated in Fig. 10, CCR is set to 65 bit/s/Hz. After 300 training cycles, PEB of FL-DDPG and TD3 converges to 0.97 m and 1.433 m, respectively, which is superior to SAC and PPO. Among them, due to the introduction of the ϵ-greedy strategy, FL-DDPG can better deal with the performance impact caused by the rapid change of channel and channel estimation error, so the positioning accuracy performance of FL-DDPG is better than TD3. Meanwhile, the final CCs of the four DRL methods exceeds 70 bit/s/Hz, which

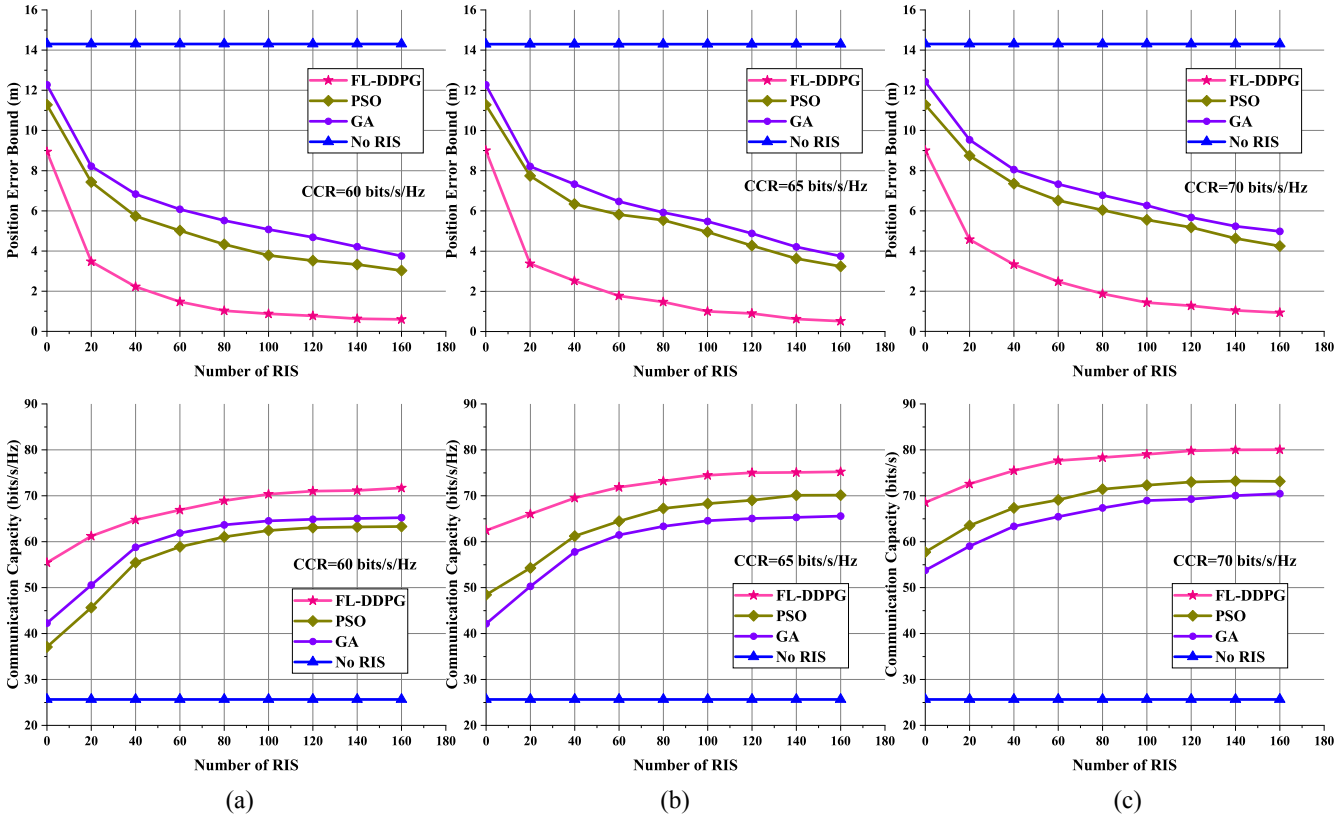


Fig. 8. Positioning and communication performance. When CCR= 60, 65, and 70 bit/s/Hz, the convergence performance of positioning accuracy and communication capacity of FL-DDPG is compared with PSO,GA and No RIS. (a) CCR = 60 bit/s/Hz. (b) CCR = 65 bit/s/Hz. (c) CCR = 70 bit/s/Hz.

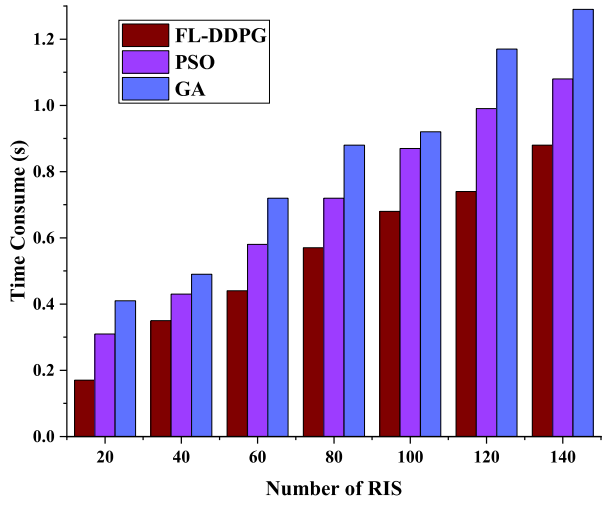


Fig. 9. Time consumption with different number of RIS. The time consumption of FL-DDPG, PSO, and GA algorithms under different numbers of RIS units.

meets the minimum CC requirements. This analysis indicates that FL-DDPG can achieve better positioning accuracy while meeting the communication requirements compared to other DRL methods.

H. RIS Bits

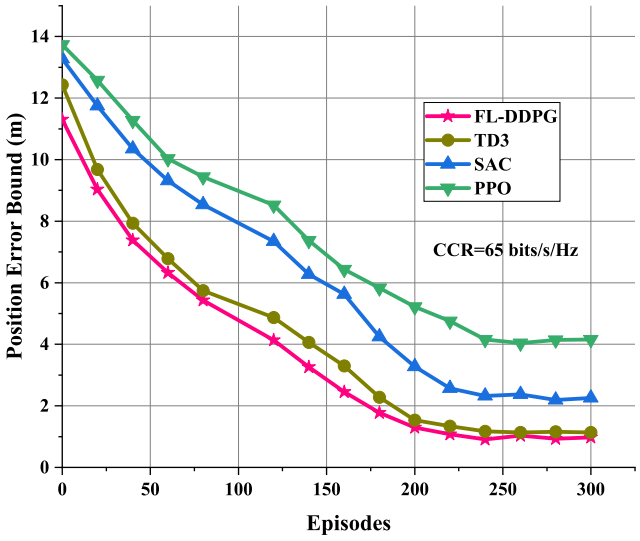
The performance of the RIS reflection unit is influenced by the number of bits during phase modulation. Thus, we evaluate the impact of the number of bits on the performance of communication and positioning. As depicted in Fig. 11, regardless of the number of bits, the positioning accuracy of the FL-DDPG is consistently higher than others. Then, with an increased number of bits, the positioning accuracy of FL-DDPG shows improvement. However, this improvement is limited. This analysis indicates that the bits are not the primary parameter in the RIS assisted 6G V2X system.

I. Transmitted Power Value

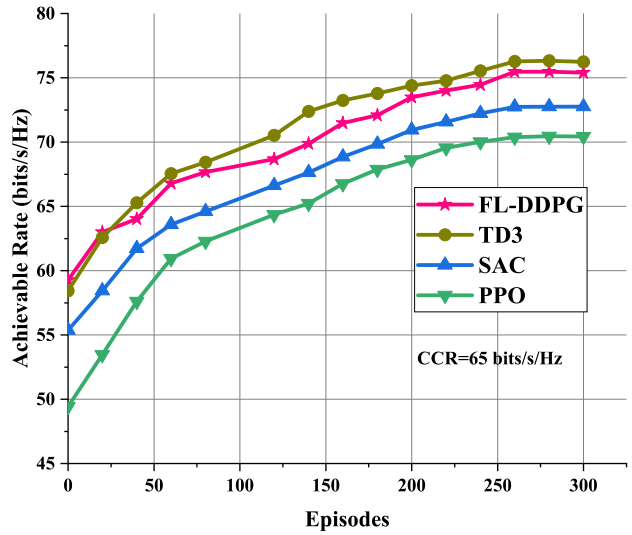
To evaluate the impact of transmitted power of BS, we gradually increase the power from 5 dBm to 30 dBm, and the results are presented in Fig. 12. FL-DDPG consistently outperforms TD3, SAC, PPO, PSO, and GA in terms of positioning accuracy, while its CC is slightly lower than that of TD3. Meanwhile, as the BS transmission power increases, both positioning accuracy and CC are improved using FL-DDPG. In addition, such improvements are gradually converged with the increase of power level.

IX. Conclusion

In this article, we propose the FL-DDPG algorithm to optimize ISAC performance for the RIS-assisted 6G V2X system. We derive the generalized FIM and construct non-convex optimization problem of high-dimensional phase shift control actions for RIS. We also analyze the robustness of the FL-DDPG algorithm with imperfect channel. Simulation



(a)



(b)

Fig. 10. DRL analysis.The convergence performance of positioning accuracy and CC of FL-DDPG is compared with classic DRL algorithms TD3, SAC, and PPO, with a CCR of 65 bit/s/Hz. (a) Position accuracy. (b) CC.

results indicate that the proposed ISAC system can flexibly adjust the phase shift, enabling it to minimize positioning accuracy while satisfying various communication requirements. Compared to other methods, the FL-DDPG method exhibits higher positioning accuracy and CC. Additionally, when compared to scenarios without RIS, our proposed system improves positioning accuracy by at least 89% and enhances the CC at the receiver end by nearly 3 times. In future work, we will apply our algorithm to actual hardware devices. By incorporating a feedback mechanism, we will continuously update and adjust the algorithm using real measurement data, aiming to enhance the performance of the hardware system. Subsequently, the approximated CRLB derived from the validated real-world system will serve as a benchmark, guiding the design and optimization of other positioning algorithms.

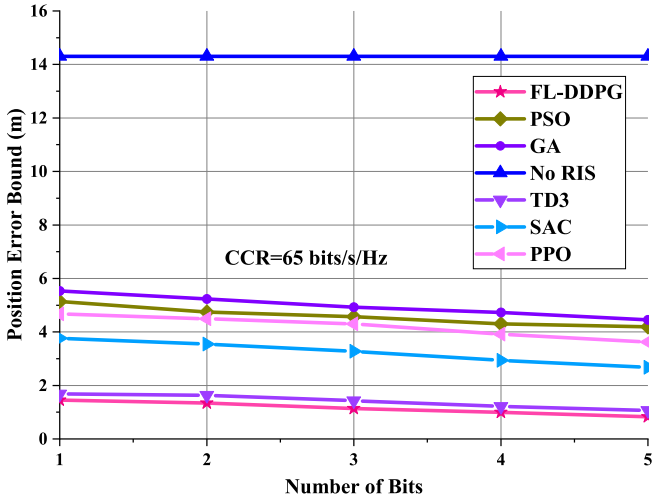


Fig. 11. PEB with different number of bits. Comparative analysis of the PEB of FL-DDPG, TD3, SAC, PPO, PSO, GA, and No RIS under different number of bits.

APPENDIX A
GEOMETRICAL PARAMETER

Based on the next-generation wireless communication network (6G) vehicle-to-everything (V2X) geometric model depicted in Fig. 2, it is observed that the spatial positions of the BS and RIS remain constant. These positions are denoted by coordinates $\mathbf{p}_b = [x_b, y_b, z_b]^T$ and $\mathbf{p}_r = [x_r, y_r, z_r]^T$, respectively. Additionally, the target's spatial coordinates are represented by $\mathbf{p}_t = [x_t, y_t, 0]^T$. We can estimate the Euclidean distances L_1, L_2 and L_3 of BS to RIS, RIS to target, and BS to target, which is expressed as

$$\begin{cases} L_1 = \sqrt{(x_b - x_r)^2 + (y_b - y_r)^2 + (z_b - z_r)^2} \\ L_2 = \sqrt{(x_r - x_t)^2 + (y_r - y_t)^2 + (z_r)^2} \\ L_3 = \sqrt{(x_b - x_t)^2 + (y_b - y_t)^2 + (z_b)^2}. \end{cases} \tag{36}$$

Then, L_1', L_2', and L_3' are the projections of L_1, L_2, and L_3 on the X-Y plane, respectively, which is expressed as:

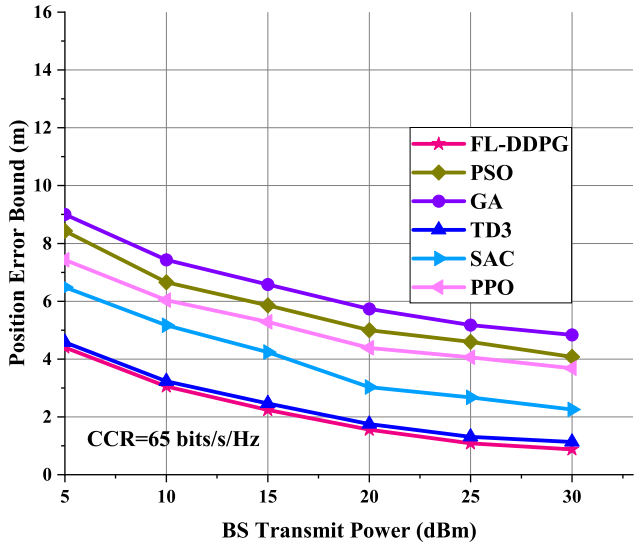
$$\begin{cases} L_1' = \sqrt{(x_b - x_r)^2 + (y_b - y_r)^2} \\ L_2' = \sqrt{(x_r - x_t)^2 + (y_r - y_t)^2} \\ L_3' = \sqrt{(x_b - x_t)^2 + (y_b - y_t)^2}. \end{cases} \tag{37}$$

In the 6G V2X system, the presence of both line-of-sight links and NLoS links necessitates the signal delays. The downlink signal delays are denoted as τ_l and τ_{nl}

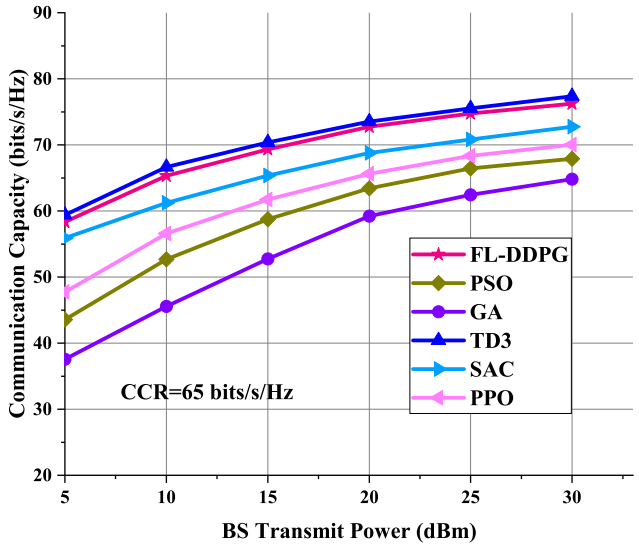
$$\begin{cases} \tau_l = \frac{L_3}{c} \\ \tau_{nl} = \frac{L_1}{c} + \frac{L_2}{c}. \end{cases} \tag{38}$$

The AoD of the downlink signal sent from BS to RIS is ψ_{br}, and the azimuth angle and elevation angle in the RIS response angles are φ_{br}^a and φ_{br}^e

$$\begin{cases} \psi_{br} = \arcsin \frac{z_b - z_r}{L_1} \\ \varphi_{br}^a = \arcsin \frac{x_b - x_r}{L_1'} \\ \varphi_{br}^e = \arccos \frac{z_b - z_r}{L_1}. \end{cases} \tag{39}$$



(a)



(b)

Fig. 12. Transit power value. The convergence performance of positioning accuracy and CC of TD3, SAC, PPO, PSO, and GA under different BS transmission power. (a) Position accuracy. (b) CC.

Then, φ_{rt}^a and φ_{rt}^e are the azimuth AOD and elevation AOD at the RIS-target link, respectively, and ψ_{rt} is the AoA of at the target

$$\begin{cases} \psi_{rt} = \arcsin \frac{z_r}{L_2} \\ \varphi_{rt}^a = \arccos \frac{y_r - y_t}{L_2'} \\ \varphi_{rt}^e = \arccos \frac{z_r}{L_2}. \end{cases} \tag{40}$$

In addition, ψ_{bt} and ψ_{tb} are the transmitting and receiving angles in the BS-target link, respectively

$$\begin{cases} \psi_{bt} = \arccos \frac{z_b}{L_3} \\ \psi_{tb} = \arcsin \frac{z_b}{L_3}. \end{cases} \tag{41}$$

APPENDIX B
CRLB DERIVATION

The estimated parameter in our system is $\zeta = [\tau_l, \tau_{nl}, \psi_{rt}, \varphi_{rt}^a, \varphi_{rt}^e, \psi_{bt}, \psi_{tb}]^T$. And the FIM $\mathbf{J}$ is obtained by transformation matrix $\mathbf{T}$ and $\mathbf{J}_\zeta$

$$\mathbf{J} = \mathbf{T}\mathbf{J}_\zeta \mathbf{T}^{\mathbf{H}} \tag{42}$$

where the transformation matrix $\mathbf{T}$ is

$$\mathbf{T} = \begin{bmatrix} \frac{\partial \tau_l}{\partial \mathbf{p}_{x_t}} & \frac{\partial \tau_{nl}}{\partial \mathbf{p}_{x_t}} & \frac{\partial \psi_{rt}}{\partial \mathbf{p}_{x_t}} & \frac{\partial \varphi_{rt}^a}{\partial \mathbf{p}_{x_t}} & \frac{\partial \varphi_{rt}^e}{\partial \mathbf{p}_{x_t}} & \frac{\partial \psi_{bt}}{\partial \mathbf{p}_{x_t}} & \frac{\partial \psi_{tb}}{\partial \mathbf{p}_{x_t}} \\ \frac{\partial \tau_l}{\partial \mathbf{p}_{y_t}} & \frac{\partial \tau_{nl}}{\partial \mathbf{p}_{y_t}} & \frac{\partial \psi_{rt}}{\partial \mathbf{p}_{y_t}} & \frac{\partial \varphi_{rt}^a}{\partial \mathbf{p}_{y_t}} & \frac{\partial \varphi_{rt}^e}{\partial \mathbf{p}_{y_t}} & \frac{\partial \psi_{bt}}{\partial \mathbf{p}_{y_t}} & \frac{\partial \psi_{tb}}{\partial \mathbf{p}_{y_t}} \end{bmatrix}$$

and the elements of the matrix $\mathbf{T}$ are calculated by (43), shown at the bottom of the page.

Then, $\mathbf{J}_\zeta$ is a 7×7 matrix, which is express as

$$[\mathbf{J}_\zeta]_{i,j} = \frac{2P_b}{\sigma_s^2} \sum_{n=1}^{N} \Re e \left\{ \frac{\partial \boldsymbol{\mu}^H}{\partial \zeta_i} \frac{\partial \boldsymbol{\mu}}{\partial \zeta_j} \right\} \tag{44}$$

where $\boldsymbol{\mu} = (\mathbf{H}_{\text{LoS}}[n] + \mathbf{H}_{\text{NLoS}}[n])\mathbf{W}\mathbf{x}[n]$, P_b is the transmitting power, σ_s is the variance, and ζ_i is the ith entry of ζ.

The elements of the matrix $([\partial \boldsymbol{\mu}]/\zeta_i)$, where $\zeta_i \in [\tau_l, \tau_{nl}, \psi_{rt}, \varphi_{rt}^a, \varphi_{rt}^e, \psi_{bt}, \psi_{tb}]^T$, is formulated in (45), shown at the top of the next page.

In (45), A_1, A_2, A_3, A_4 are the complex coefficients obtained in the derivation calculation, which is

$$\begin{cases} A_1 = \gamma_l h_l j 2\pi B \frac{n}{N} e^{j2\pi B \frac{n}{N} \tau_l} \\ A_2 = \gamma_{nl} h_{nl} j 2\pi B e^{j2\pi B \frac{n}{N} \tau_{nl}} \\ A_3 = \gamma_{nl} h_{nl} e^{j2\pi B \frac{n}{N} \tau_{nl}} \\ A_4 = \gamma_l h_l e^{j2\pi B \frac{n}{N} \tau_l}. \end{cases} \tag{47}$$

And A_{rt}, A_{bt}, and A_{tb} are all diagonal matrices obtained in the derivation calculation

$$\begin{cases} \mathbf{a}_{rt} = j\frac{2\pi}{\lambda} \cos(\psi_{rt}) \mathbf{diag}(0, 1, \ldots, (N_t - 1)) \\ \mathbf{a}_{bt} = j\frac{2\pi}{\lambda} \cos(\psi_{bt}) \mathbf{diag}(0, 1, \ldots, (N_b - 1)) \\ \mathbf{a}_{tb} = j\frac{2\pi}{\lambda} \cos(\psi_{bt}) \mathbf{diag}(0, 1, \ldots, (N_t - 1)) \end{cases} \tag{48}$$

where λ is the wavelength, and N_b and N_t are the number of antennas equipped by BS and target, respectively. In addition, $\mathbf{a}_{rt_{(N_x,N_y)}}^a$ and $\mathbf{a}_{rt_{(N_x,N_y)}}^e$ are represented as

$$\begin{cases} \mathbf{a}_{rt_{(N_x,N_y)}}^a = j\frac{2\pi}{\lambda} d_r((N_x - 1)\cos(\varphi_{rt}^a)\sin(\varphi_{rt}^e)) \\ \mathbf{a}_{rt_{(N_x,N_y)}}^e = j\frac{2\pi}{\lambda} d_r((N_x - 1) \\ \sin(\varphi_{rt}^a)\cos(\varphi_{rt}^e) - (N_y - 1)\sin(\varphi_{rt}^e)) \end{cases} \tag{49}$$

$$\mathbf{T} = \begin{bmatrix} \frac{(x_t - x_b)}{cL_3} & \frac{(x_t - x_r)}{cL_2} & \frac{z_r(x_t - x_r)}{L_2^3\sqrt{1 - \frac{z_r^2}{L_2^2}}} & \frac{(y_r - y_t)(x_r - x_t)}{L_2^3\sqrt{1 - \frac{(y_r - y_t)^2}{L_2^2}}} & \frac{z_r(x_r - x_t)}{L_2^3\sqrt{1 - \frac{z_r^2}{L_2^2}}} & \frac{z_b(x_t - x_b)}{L_3^3\sqrt{1 - L_3^2}} & \frac{z_b(x_t - x_b)}{L_3^3\sqrt{1 - L_3^2}} \\ \frac{(y_t - y_b)}{cL_3} & \frac{(y_t - y_r)}{cL_2} & \frac{z_r(y_t - y_r)}{L_2^3\sqrt{1 - \frac{z_r^2}{L_2^2}}} & \frac{L_2'^2 + (y_r - y_t)(x_r - x(t))}{L_2^3\sqrt{1 - \frac{(y_r - y_t)^2}{L_2^2}}} & \frac{z_r(y_r - y_t)}{L_2^3\sqrt{1 - \frac{z_r^2}{L_2^2}}} & \frac{z_b(y_t - y_b)}{L_3^3\sqrt{1 - L_3^2}} & \frac{z_b(y_t - y_b)}{L_3^3\sqrt{1 - L_3^2}} \end{bmatrix} \tag{43}$$

$$\begin{bmatrix} \frac{\partial \boldsymbol{\mu}[n]}{\partial \tau_l} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \tau_{nl}} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \psi_{rt}} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \varphi^a_{rt}} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \varphi^e_{rt}} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \psi_{bt}} \\ \frac{\partial \boldsymbol{\mu}[n]}{\partial \psi_{tb}} \end{bmatrix} = \begin{bmatrix} A_1 \mathbf{a}_{bt,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}}(\psi_{tb}) \mathbf{W}\mathbf{x}[n] \\ A_2 \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}\left(\varphi^a_{rt}, \varphi^e_{rt}\right) \boldsymbol{\Theta} \mathbf{a}_{br}\left(\varphi^a_{br}, \varphi^e_{br}\right) \mathbf{a}^H_{br}(\psi_{br}) \mathbf{W}\mathbf{x}[n] \\ A_3 \mathbf{a}_{rt} \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}\left(\varphi^a_{rt}, \varphi^e_{rt}\right) \boldsymbol{\Theta} \mathbf{a}_{br}\left(\varphi^a_{br}, \varphi^e_{br}\right) \mathbf{a}^H_{br}(\psi_{br}) \mathbf{W}\mathbf{x}[n] \\ A_3 \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}\left(\varphi^a_{rt}, \varphi^e_{rt}\right) \mathbf{diag}\left(\mathbf{a}^a_{rt}\right) \boldsymbol{\Theta} \mathbf{a}_{br}\left(\varphi^a_{br}, \varphi^e_{br}\right) \mathbf{a}^H_{br}(\psi_{br}) \mathbf{W}\mathbf{x}[n] \\ A_3 \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}\left(\varphi^a_{rt}, \varphi^e_{rt}\right) \mathbf{diag}\left(\mathbf{a}^e_{rt}\right) \boldsymbol{\Theta} \mathbf{a}_{br}\left(\varphi^a_{br}, \varphi^e_{br}\right) \mathbf{a}^H_{br}(\psi_{br}) \mathbf{W}\mathbf{x}[n] \\ A_4 \mathbf{a}_{tb,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}} \mathbf{a}^H_{bt}(\psi_{tb}) \mathbf{W}\mathbf{x}[n] \\ A_4 \mathbf{a}_{tb} \mathbf{a}_{bt,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}}(\psi_{tb}) \mathbf{W}\mathbf{x}[n] \end{bmatrix} \tag{45}$$

$$\begin{bmatrix} \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \tau_l} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \tau_{nl}} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \psi_{rt}} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \varphi^a_{rt}} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \varphi^e_{rt}} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \psi_{bt}} \\ \frac{\partial \hat{\boldsymbol{\mu}}[n]}{\partial \psi_{tb}} \end{bmatrix} = \begin{bmatrix} A_1 \left(\mathbf{a}_{bt,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}}(\psi_{tb}) + \Delta \mathbf{H}_{bt} \right) \mathbf{W}\mathbf{x}[n] \\ A_2 \left(\mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) + \Delta \mathbf{H}_{rt} \right) \boldsymbol{\Theta} \left(\mathbf{a}_{br}(\varphi^a_{br}, \varphi^e_{br}) \mathbf{a}^H_{br}(\psi_{br}) + \Delta \mathbf{H}_{br} \right) \mathbf{W}\mathbf{x}[n] \\ A_3 \mathbf{a}_{rt} \left(\mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \boldsymbol{\Theta} \mathbf{a}_{br}(\varphi^a_{br}, \varphi^e_{br}) \mathbf{a}^H_{br}(\psi_{br}) + \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \boldsymbol{\Theta}(\Delta \mathbf{H}_{br}) \right) \mathbf{W}\mathbf{x}[n] \\ A_3 \left(\mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \mathbf{diag}(\mathbf{a}^a_{rt}) \boldsymbol{\Theta} \mathbf{a}_{br}(\varphi^a_{br}, \varphi^e_{br}) \mathbf{a}^H_{br}(\psi_{br}) + \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \mathbf{diag}(\mathbf{a}^a_{rt}) \boldsymbol{\Theta}(\Delta \mathbf{H}_{br}) \right) \mathbf{W}\mathbf{x}[n] \\ A_3 \left(\mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \mathbf{diag}(\mathbf{a}^e_{rt}) \boldsymbol{\Theta} \mathbf{a}_{br}(\varphi^a_{br}, \varphi^e_{br}) \mathbf{a}^H_{br}(\psi_{br}) + \mathbf{a}_{rt}(\psi_{rt}) \mathbf{a}^H_{rt}(\varphi^a_{rt}, \varphi^e_{rt}) \mathbf{diag}(\mathbf{a}^e_{rt}) \boldsymbol{\Theta}(\Delta \mathbf{H}_{br}) \right) \mathbf{W}\mathbf{x}[n] \\ A_4 \mathbf{a}_{tb,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}} \mathbf{a}^H_{bt}(\psi_{tb}) \mathbf{W}\mathbf{x}[n] \\ A_4 \mathbf{a}_{tb} \mathbf{a}_{bt,\text{in}}(\psi_{bt}) \mathbf{a}^H_{bt,\text{out}}(\psi_{tb}) \mathbf{W}\mathbf{x}[n] \end{bmatrix} \tag{46}$$

where d_r represents the gap between RIS reflection units, N_x denotes the horizontal coordinate of the RIS reflection plane, and N_x signifies the vertical coordinate of the RIS reflection plane.

APPENDIX C
CRLB DERIVATION WITH IMPERFECT CHANNEL MODEL

Under imperfect channel model condition, the FIM $\hat{\mathbf{J}}$ is obtained by transformation matrix $\mathbf{T}$ and $\hat{\mathbf{J}}_\zeta$

$$\hat{\mathbf{J}} = \mathbf{T} \hat{\mathbf{J}}_\zeta \mathbf{T}^{\mathbf{H}} \tag{50}$$

the elements of the matrix $\mathbf{T}$ are obtained from (43) on bottom of the previous page.

Then, $\hat{\mathbf{J}}_\zeta$ is a 7×7 matrix, which is express as

$$\left[\hat{\mathbf{J}}_\zeta \right]_{i,j} = \frac{2P_b}{\sigma_s^2} \sum_{n=1}^{N} \Re e \left\{ \frac{\partial \hat{\boldsymbol{\mu}}^H}{\partial \zeta_i} \frac{\partial \hat{\boldsymbol{\mu}}}{\partial \zeta_j} \right\} \tag{51}$$

where $\hat{\boldsymbol{\mu}} = (\hat{\mathbf{H}}_{br}[n] + \hat{\mathbf{H}}_{rt}[n] \boldsymbol{\Theta} \hat{\mathbf{H}}_{br}[n]) \mathbf{W}\mathbf{x}[n]$ is the transmitting power, σ_s is the variance, and ζ_i is the ith entry of ζ.

The elements of the matrix $([\partial \hat{\boldsymbol{\mu}}]/\zeta_i)$, where $\zeta_i \in [\tau_l, \tau_{nl}, \psi_{rt}, \varphi^a_{rt}, \varphi^e_{rt}, \psi_{bt}, \psi_{tb}]^T$, is formulated in (46), shown at the top of the page.

REFERENCES

[1] M. Noor-A-Rahim et al., "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities," *Proc. IEEE*, vol. 110, no. 6, pp. 712–734, Jun. 2022.

[2] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér–Rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, Dec. 2021.

[3] N. Su, F. Liu, Z. Wei, Y.-F. Liu, and C. Masouros, "Secure dual-functional radar-communication transmission: Exploiting interference for resilience against target eavesdropping," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7238–7252, Sep. 2022.

[4] A. Elzanaty, A. Guerra, F. Guidi, and M.-S. Alouini, "Reconfigurable intelligent surfaces for localization: Position and orientation error bounds," *IEEE Trans. Signal Process.*, vol. 69, pp. 5386–5402, Aug. 2021.

[5] C. Pan et al., "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, Jun. 2021.

[6] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE 20th Int. Workshop signal Process. Adv. Wireless Commun. (SPAWC)*, 2019, pp. 1–5.

[7] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[8] Q. Wu, X. Guan, and R. Zhang, "Intelligent reflecting surface-aided wireless energy and information transmission: An overview," *Proc. IEEE*, vol. 110, no. 1, pp. 150–170, Jan. 2021.

[9] Y. Liu, S. Zhang, F. Gao, J. Tang, and O. A. Dobre, "Cascaded channel estimation for RIS assisted mmWave MIMO transmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 2065–2069, Sep. 2021.

[10] T. Ma, Y. Xiao, X. Lei, W. Xiong, and M. Xiao, "Distributed reconfigurable intelligent surfaces assisted indoor positioning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 47–58, Jan. 2023.

[11] Y. Liu, S. Hong, C. Pan, Y. Wang, Y. Pan, and M. Chen, "Cramér–Rao lower bound analysis of multiple-RIS-aided mmWave positioning systems," in *Proc. IEEE 33rd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2022, pp. 1110–1115.

[12] A. Hassanien, M. G. Amin, E. Aboutanios, and B. Himed, "Dual-function radar communication systems: A solution to the spectrum congestion problem," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 115–126, Sep. 2019.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 84–90.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[15] J. Fan, Z. Wang, Y. Xie, and Z. Yang, "A theoretical analysis of deep *Q*-learning," in *Proc. Learn. Dyn. Control*, 2020, pp. 486–489.

[16] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[17] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, "Inner bounds on performance of radar and communications co-existence," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 464–474, Jan. 2016.

[18] N. González-Prelcic, R. Méndez-Rial, and R. W. Heath, "Radar aided beam alignment in mmWave V2I communications supporting antenna diversity," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2016, pp. 1–7.
[19] N. Nartasilpa, A. Salim, D. Tuninetti, and N. Devroye, "Communications system performance and design in the presence of radar interference," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4170–4185, Sep. 2018.
[20] T. Huang, N. Shlezinger, X. Xu, Y. Liu, and Y. C. Eldar, "MAJoRCom: A dual-function radar communication system using index modulation," *IEEE Trans. signal Process.*, vol. 68, pp. 3423–3438, May 2020.
[21] Z. Wang, K. Han, J. Jiang, F. Liu, and W. Yuan, "Multi-vehicle tracking and ID association based on integrated sensing and communication signaling," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1960–1964, Sep. 2022.
[22] Y. Zhang, W. Ni, W. Tang, Y. C. Eldar, and D. Niyato, "Robust transceiver design for ISAC with imperfect CSI," in *Proc. IEEE Glob. Commun. Conf.*, 2023, pp. 1320–1325.
[23] Y. Liu, M. Li, A. Liu, J. Lu, and T. X. Han, "Information-theoretic limits of integrated sensing and communication with correlated sensing and channel states for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10161–10166, Sep. 2022.
[24] Y. Zhang et al., "Robust transceiver design for covert integrated sensing and communications with imperfect CSI," *IEEE Trans. Commun.*, early access, Apr. 12, 2024, doi: 10.1109/TCOMM.2024.3387869.
[25] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.
[26] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.
[27] J. He, H. Wymeersch, L. Kong, O. Silvén, and M. Juntti, "Large intelligent surface for positioning in millimeter wave MIMO systems," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC)*, 2020, pp. 1–5.
[28] J. V. Alegría and F. Rusek, "Cramér–Rao lower bounds for positioning with large intelligent surfaces using quantized amplitude and phase," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 10–14.
[29] W. Wang and W. Zhang, "Joint beam training and positioning for intelligent reflecting surfaces assisted millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6282–6297, Oct. 2021.
[30] M. Ammous and S. Valaee, "Positioning and tracking using reconfigurable intelligent surfaces and extended Kalman filter," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC)*, 2022, pp. 1–6.
[31] J. He, H. Wymeersch, T. Sanguanpuak, O. Silvén, and M. Juntti, "Adaptive beamforming design for mmWave RIS-aided joint localization and communication," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, 2020, pp. 1–6.
[32] N. Decarli, A. Guerra, C. Giovannetti, F. Guidi, and B. M. Masini, "V2X sidelink Localization of connected automated vehicles," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 1, pp. 120–133, Jan. 2024.
[33] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.
[34] Z. Huang, B. Zheng, and R. Zhang, "Transforming fading channel from fast to slow: Intelligent refracting surface aided high-mobility communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4989–5003, Jul. 2022.
[35] K. Meng, Q. Wu, W. Chen, and D. Li, "Sensing-assisted communication in vehicular networks with intelligent surface," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 876–893, Jan. 2024.
[36] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7704–7719, Nov. 2020.
[37] Z. Li, W. Chen, Q. Wu, H. Cao, K. Wang, and J. Li, "Robust beamforming design and time allocation for IRS-assisted wireless powered communication networks," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2838–2852, Apr. 2022.
[38] M. Luan, B. Wang, Z. Chang, T. Hämäläinen, and F. Hu, "Robust beamforming design for RIS-aided integrated sensing and communication system," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6227–6243, Jun. 2023.
[39] L. Hu et al., "RIS-assisted integrated sensing and covert communication design," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 16505–16516, May 2024.
[40] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May 2020.
[41] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Aug. 2020.
[42] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.
[43] J. Xu, B. Ai, and T. Q. Quek, "Toward interference suppression: RIS-aided high-speed railway networks via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 4188–4201, Jun. 2023.
[44] R. Zhong, Y. Liu, X. Mu, Y. Chen, and L. Song, "AI empowered RIS-assisted NOMA networks: Deep learning or reinforcement learning?" *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 182–196, Jan. 2022.
[45] Z. Yang, Y. Liu, Y. Chen, and J. T. Zhou, "Deep reinforcement learning for RIS-aided non-orthogonal multiple access downlink networks," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.
[46] X. Lin, A. Liu, C. Han, X. Liang, and Y. Li, "Joint pilot spacing and power optimization scheme for Nonstationary wireless channel: A deep reinforcement learning approach," *IEEE Wireless Commun. Lett.*, vol. 12, no. 3, pp. 540–544, Mar. 2023.
[47] H. Tang, H. Wu, G. Qu, and R. Li, "Double deep *Q*-network based dynamic framing offloading in vehicular edge computing," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 3, pp. 1297–1310, May/Jun. 2023.
[48] S. Lei et al., "Dissatisfaction feedback and Stackelberg game-based task offloading mechanism for parked vehicle edge computing," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4383–4388, Mar. 2024.

Xudong Long received the master's degree from Shanghai University Of Engineering Science, Shanghai, China, in 2021. He is currently pursuing the Ph.D. degree in electronic science and technology with the School of Microelectronics Science, Sun Yat-Sen University, Zhuhai, China.

His main research interests include reconfigurable intelligent surface and its applications.

Yubin Zhao (Senior Member, IEEE) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in computer science from Freie Universität Berlin, Berlin, Germany, in 2014.

He was an Associate Professor with the Center for Cloud Computing, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, in 2014. He is currently the Associate Professor with the School of Microelectronics Science and Technology, Sun Yat-Sen University, Zhuhai, China. His current research interest includes wireless power transfer, indoor localization, and target tracking.

Dr. Zhao was a recipient of the IEEE Distinguished Service Award in IEEE SmartData 2023 and the Excellent Teacher Award of Collage Computing Science in China in 2023. He serves as the Guest Editor and a reviewer for several journals. He serves as the Vice Technical Chair for IEEE SmartData 2023, and IEEE ScalCom 2022, the Publicity Chair for IEEE NFV-SDN 2019, and the Tutorial Chair for IEEE NFV-SDN 2020.

Huaming Wu (Senior Member, IEEE) received the B.E. and M.S. degrees in electrical engineering from Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively, and the Ph.D. degree (with Highest Hons.) in computer science from Freie University Berlin, Berlin, Germany, in 2015.

He is currently an Associate Professor with the Center for Applied Mathematics, Tianjin University, Tianjin, China. His research interests include model-based evaluation, wireless and mobile network systems, mobile cloud computing, and deep learning.

Cheng-Zhong Xu (Fellow, IEEE) received the Ph.D. degree from The University of Hong Kong, Hong Kong, in 1993.

He is currently a Chair Professor of Computer Science with the University of Macau, Macau, China. Prior to that, he was with the Faculty of Wayne State University, Detroit, MI, USA, and Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Beijing, China. He published two research monographs and more than 400 journal and conference papers and received more than 13 000 citations. His recent research interests are in cloud and distributed computing, systems support for AI, smart city, and autonomous driving.

Prof. Xu was a Best Paper Awardee or a Nominee of conferences, including HPCA'2013, HPDC'2013, Cluster'2015, ICPP'2015, GPC'2018, UIC'2018, and HPBD&IS'2019. He was also a Co-Inventor of more than 120 patents and a Co-Founder of Shenzhen Institute of Baidou Applied Technology. He was a recipient of the Faculty Research Award, the Career Development Chair Award, and the President's Award for Excellence in Teaching of WSU. He was also a recipient of the "Outstanding Oversea Scholar" Award of NSFC. He serves or served for a number of journal editorial boards, including IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, JOURNAL OF PARALLEL AND DISTRIBUTED COMPUTING, *Science China*, and *ZTE Communication*. He was the Chair of IEEE Technical Committee on Distributed Processing from 2015 to 2020.