

Multi-Agent DRL for QoS-Constrained Active RIS–MA ISAC

Seungseok Sin^a, Jaehyeok Kim^a, Insik Cho^a, Taehong Kim^a, Huaping Liu^b, Sangmi Moon^c, and Intae Hwang^d

^aDept. of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju, South Korea

^bOregon State University, United States of America

^cDept. of IT and Communication Convergence Engineering, Kunsan National University, Kunsan, South Korea

^dSchool of Electronics and Computer Engineering and Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju, South Korea

ssskit7@naver.com, kjhyeok5369@naver.com, dr-cho@jnu.ac.kr, cute0571@jnu.ac.kr,
huaping.liu@oregonstate.edu, moonsm@kunsan.ac.kr, and hit@jnu.ac.kr

Abstract— This paper proposes a QoS-constrained multi-agent deep reinforcement learning (MA-DRL) framework for active RIS- and movable antenna-assisted integrated sensing and communication (ISAC) systems. The framework jointly optimizes BS beamforming, RIS coefficients, and MA positions to maximize radar SINR while satisfying communication QoS. By decomposing the high-dimensional control problem into cooperative agents, the proposed method effectively handles strong coupling. Simulation results show superior radar performance over conventional baselines while maintaining QoS.

Keywords- Active Reconfigurable Intelligent Surface, Communication QoS, Integrated Sensing and Communication, Joint Optimization, Movable Antenna, Multi-Agent Deep Reinforcement Learning.

I. INTRODUCTION

Next-generation 6G systems are expected to support both high-rate communication and environmental sensing, where integrated sensing and communication (ISAC) has emerged as a key technology [1], [2]. However, ISAC inherently faces a trade-off between communication and sensing due to shared resources.

Active reconfigurable intelligent surfaces (RISs) and movable antennas (MAs) provide promising solutions by enabling propagation control and spatial adaptability [3]–[7]. Nevertheless, the joint optimization of BS beamforming, RIS coefficients, and MA positions is highly challenging due to strong coupling and nonconvex constraints under QoS requirements.

To address this, we propose a QoS-constrained multi-agent deep reinforcement learning (MA-DRL) framework that decomposes the control task into cooperative BS, RIS, and MA agents. Simulation results demonstrate improved radar sensing performance over conventional baselines while maintaining communication QoS.

"This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT: Ministry of Science and ICT) (IITP-2026-RS-2022-00156287)." "This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the ICT Challenge and Advanced Network of HRD (ICAN) support program supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) (IITP-2026-RS-2022-00156385)." "This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT: Ministry of Science and ICT) (RS-2024-00333826)." "This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2026-25476685)." "This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Challenge and Advanced Network of HRD (ICAN) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2026-RS-2024-00436837)."

II. SYSTEM AND CHANNEL MODEL

This study considers an ISAC architecture that jointly employs an active RIS and MAs. As illustrated in Fig. 1, the system consists of a BS equipped with M MA elements, an UAV carrying an active RIS, a radar target, and multiple single-antenna users. For notational simplicity, the ISAC BS and the UAV-mounted RIS are hereafter referred to as the BS and the RIS, respectively. The MAs at the BS can relocate within a predefined movement region, enabling the transceiver to adapt its spatial configuration in response to channel variations. With this adaptive array, the BS simultaneously transmits communication signals to K users while emitting sensing waveforms toward the radar target. Since the LoS path between the BS and the target is assumed to be blocked, target sensing is realized via reflection from the UAV-mounted active RIS. The RIS comprises N reflecting elements, each equipped with a low-noise amplifier. By amplifying the incident signal prior to re-radiation, the active RIS compensates for severe attenuation in multi-hop cascaded links, particularly the four-hop BS–RIS–target–RIS–BS path, thereby enhancing both communication and sensing performance.

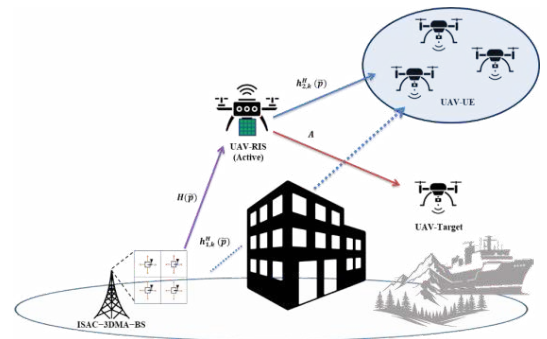


Fig. 1. System model of the active RIS- and MA-aided ISAC system.

A. Transmit signal model

The transmit signal of the ISAC BS is expressed as

$$x = W_r s + W_c c = [W_r, W_c] \begin{bmatrix} s^T \\ c^T \end{bmatrix} = W x, \quad (1)$$

where $s \in \mathbb{C}^{M \times 1}$ denotes the radar-only signal, and $c \in \mathbb{C}^{K \times 1}$ represents the communication signal vector intended for K single-antenna users. The matrices $W_r \in \mathbb{C}^{M \times M}$ and $W_c \in \mathbb{C}^{M \times K}$ denote the radar and communication beamforming matrices, respectively. Accordingly, the overall BS beamforming matrix is defined as $W = [W_r, W_c] \in \mathbb{C}^{M \times (M+K)}$. The radar signal s is generated using pseudo-random coding, satisfying $E[ss^H] = I_M$. The information signal c follows a complex Gaussian distribution, $CN(0, I_K)$. Accordingly, the covariance matrix of the transmitted waveform can be expressed as

$$R = E[xx^H] = WW^H = R_c + \sum_{k=1}^K R_k. \quad (2)$$

To characterize the propagation environment, we adopt a planar far-field approximation, following the modeling approach in [5]. For a given propagation path, the transmit and receive links share identical angles of departure (AoDs), angles of arrival (AoAs), and path-gain coefficients; only their signal phases differ. The positions of the MAs at the BS are represented as $\tilde{p} = [p_1, p_2, \dots, p_M] \in \mathbb{R}^{2 \times M}$, where $p_m = [x_m, y_m]^T$ denotes the coordinates of the m -th MA. The transmit and receive elevation and azimuth angles are defined as $\theta_j^t \in [0, \pi]$, $\varphi_j^t \in [0, \pi]$, $(1 \leq j \leq L_t)$ and $\theta_i^r \in [0, \pi]$, $\varphi_i^r \in [0, \pi]$, $(1 \leq i \leq L_r)$, respectively, where L_t and L_r denote the numbers of departure and arrival paths for the BS-RIS channel, respectively. The number of departure paths of the BS-user k link is indicated by L_t^k . Because the channel observed at each MA depends on both its spatial position and the surrounding propagation geometry, the channel response varies with antenna relocation. Following [5], the field-response vector associated with an MA located at $p = [x, y]^T$, or the BS-RIS link is given by

$$g(p) = [e^{j2\pi \rho_t^1(p)}, \dots, e^{j2\pi \rho_t^{L_t}(p)}]^T \in \mathbb{C}^{L_t \times 1}, \quad (3)$$

where $\rho_t^j(p) = x \sin \theta_j^t \cos \varphi_j^t + y \cos \theta_j^t$ represents the additional propagation distance of the j -th transmit path relative to a reference point. The aggregate field responses of the M movable antennas in the BS-RIS link are collected into matrix (4), where each column corresponds to the response of an individual MA at position p_m .

$$G(\tilde{p}) = [g(p_1), g(p_2), \dots, g(p_M)] \in \mathbb{C}^{L_t \times M}. \quad (4)$$

The corresponding matrix of the BS-user k link is

$$G_k(\tilde{p}) = [g_k(p_1), g_k(p_2), \dots, g_k(p_M)] \in \mathbb{C}^{L_t^k \times M}. \quad (5)$$

The field response of the RIS can be modeled analogously to that of the BS array, where each reflecting element introduces a path-dependent phase shift determined by its spatial coordinate. Let $\Sigma \in \mathbb{C}^{L_r \times L_t}$ and $\Sigma_k \in \mathbb{C}^{L_r^k \times L_t^k}$ denote the path-response matrices of the BS-RIS and BS-user k links, respectively. The channel matrix between the BS and the UAV-mounted RIS is then expressed as

$$H(\tilde{p}) = F(r)^H \Sigma G(\tilde{p}) \in \mathbb{C}^{N \times M}. \quad (6)$$

The channel between the BS and user k is given by $h_{1,k}^H(\tilde{p}) = 1^H \Sigma_k G_k(\tilde{p}) \in \mathbb{C}^{1 \times M}$, while the channel between the UAV-RIS and user k is denoted as $h_{2,k}^H \in \mathbb{C}^{1 \times N}$. Assuming that the radar target has a sufficiently small spatial extent (i.e., it can be approximated as a point target), the RIS-target channel can be expressed as

$$A = \beta_r a(\theta_A) a(\theta_A)^H \in \mathbb{C}^{N \times N}, \quad (7)$$

where θ_A denotes the AoA of the reflected signal with respect to the RIS. The RIS steering vector is given by:

$$a(\theta) = [1, e^{j2\pi \frac{d_{RIS}}{\lambda} \sin \theta}, \dots, e^{j2\pi (N-1) \frac{d_{RIS}}{\lambda} \sin \theta}]^T, \quad (8)$$

where d_{RIS} denotes the inter-element spacing of the RIS. In this system, the active RIS reflects and amplifies the signals received from the BS, directing them toward both the radar target and the communication users. The first reflected signal (corresponding to the amplified BS-RIS path) is given by

$$y_1 = QH(\tilde{p})x + Qv_1, \quad (9)$$

where $Q \in \mathbb{C}^{N \times N}$ denotes the reflection-amplification matrix of the active RIS, and $v_1 \sim CN(0, \sigma^2 I_N)$ represents additive Gaussian noise introduced during the first reflection. The second reflected signal, corresponding to the echo from the target that is re-reflected by the RIS and received at the BS, is expressed as

$$y_2 = Q^H A Q H(\tilde{p})x + Q^H A Q v_1 + Q^H v_2, \quad (10)$$

where $v_2 \sim CN(0, \sigma^2 I_N)$ denotes the additive Gaussian noise introduced during the second reflection. To ensure stable operation of the active RIS, its total transmit power is constrained by a maximum power budget P_{RIS}^{max} , satisfying

$$E[\|y_1\|_2^2 + \|y_2\|_2^2] \leq P_{RIS}^{max}. \quad (11)$$

III. PROPOSED ALGORITHM

The radar echo signal received at the BS can be expressed as

$$\tilde{y}_r = H(\tilde{p})^H (y_1 + y_2) + z_r. \quad (12)$$

where z_r denotes AWGN at the radar receiver. After standard self-interference (SI) cancellation, the effective received radar signal can be written as [5]

$$y_r = H^H Q^H A Q H x + H^H Q^H A Q v_1 + H^H Q^H v_2 + \eta H^H Q^H H + z_r, \quad (13)$$

$$\text{SINR} = \text{Tr}(B R B^H J^{-1}), \quad (14)$$

For the communication link, each user k receives the superposition of the direct BS–UE and the reflected BS–RIS–UE signals, expressed as

$$y_{c,k} = h_{1,k}^H x + h_{2,k}^H Q H(\tilde{p}) x + h_{2,k}^H Q v_1 + z_k, \quad (15)$$

where $z_k \sim \mathcal{CN}(0, \sigma_z^2 \mathbf{I}_M)$ denotes the AWGN at user k . Define the equivalent composite channel as $h_k^H = h_{1,k}^H + h_{2,k}^H Q H(\tilde{p})$. Then, the communication SINR for user k is expressed as

$$\text{SINR}_k = \frac{h_k^H R_k h_k}{h_k^H (R - R_k) h_k + \sigma^2 h_{2,k}^H Q Q^H h_{2,k} + \sigma_z^2}. \quad (16)$$

This paper considers the following joint optimization problem for an active-RIS- and movable-antenna-assisted ISAC system:

$$(P1) \quad \max_{W, Q, \tilde{p}} \quad \text{SINR}_{\text{radar}} \quad (17a)$$

$$s. t. \quad \text{SINR}_{k_k} \geq \Gamma, \forall k, \quad (17b)$$

$$\text{Tr}(R) \leq P_{BS}, \quad (17c)$$

$$E[\|y_1^r\|_2^2 + \|y_2^r\|_2^2] \leq P_{RIS}^{\max} \quad (17d)$$

$$\|q_n\|^2 \leq a_{RIS}, \forall n, \quad (17e)$$

$$\|p_k - p_q\|_2 \geq D, \forall k \neq q \quad (17f)$$

$$\tilde{p} \in C. \quad (17g)$$

where the objective is to maximize the radar SINR subject to communication QoS and physical constraints, including BS and RIS power limits, RIS amplification bounds, minimum inter-antenna spacing, and the feasible antenna movement region. Problem P1 is highly nonconvex due to the strong coupling among the BS beamforming matrix, RIS reflection–amplification coefficients, and MA positions, as well as the nonlinear constraints. Moreover, repeatedly solving P1 in dynamic wireless environments incurs high computational complexity. To address this issue, we reformulate P1 as a multi-agent deep reinforcement learning (MA-DRL) task for efficient adaptive control.

We reformulate problem (P1) as a Markov decision process (MDP) and address the joint control of the active RIS– and MA-assisted ISAC system using an MA-DRL framework. The proposed framework formulates the joint control problem as an MDP and decomposes it into three cooperative agents corresponding to the BS, RIS, and MA. As illustrated in Algorithm 1, each agent observes the shared state s_t and

generates partial actions, which are combined to update beamforming, RIS coefficients, and MA positions.

A global reward is computed based on radar SINR, QoS satisfaction, and constraint penalties, and the resulting transitions are stored for off-policy learning. Each agent independently updates its policy using SAC, DDPG, and TD3, respectively. Through iterative interaction with the environment, the agents cooperatively learn to maximize radar performance while satisfying QoS constraints.

```

1. Input: total training steps T, episode length N, replay buffer D, discount factor gamma  $\gamma$ , exploration noise sigma.
2. Initialize:  $\text{Agent}_{BS}(\text{SAC})$ ,  $\text{Agent}_{RIS}(\text{DDPG})$ ,  $\text{Agent}_{MA}(\text{TD3})$ , and their target networks.
3. for episode = 1, ..., E do
4.   Reset environment and obtain initial state  $s_0$  (channels, Q, MA positions,  $\text{SINR}_{\text{radar}}$ ,  $\text{SINR}_k$ ).
5.   for t = 1, ..., N do
6.     Observe current state  $s_t$ .
7.     Select actions:
8.        $a_{BS,t} \leftarrow \pi_{BS}(s_t) + \text{noise}$ 
9.        $a_{RIS,t} \leftarrow \pi_{RIS}(s_t) + \text{noise}$ 
10.       $a_{MA,t} \leftarrow \pi_{MA}(s_t)$  (stochastic sampling)
11.     Apply joint action  $(a_{BS,t}, a_{RIS,t}, a_{MA,t})$  to update W, Q, and  $\tilde{p}$ .
12.     Compute reward:  $r_t = \text{radar\_score} - w_{\text{qos}} * \text{violation} - \text{penalty\_terms}$ .
13.     Store transition  $(s_t, a_t, r_t, s_{t+1}, \text{done})$  in D.
14.     Sample mini-batch from D and update policies:
15.     SAC.update(BS), DDPG.update(RIS), TD3.update(MA).
16.     Set  $s_t \leftarrow s_{t+1}$ .
17.   end for
18. end for
19. Output: learned policies  $\{\pi_{BS}, \pi_{RIS}, \pi_{MA}\}$  that maximize radar SINR under QoS constraints.

```

Fig. 2. Algorithm 1: Multi-agent DRL training procedure for QoS-constrained ISAC joint optimization (BS = SAC, RIS = DDPG, MA = TD3).

At time step t , the MDP is characterized by the state (observation) s_t , action a_t , reward r_t , and next state s_{t+1} . All agents interact with the same environment and learn continuous control policies. To reduce the complexity associated with high-dimensional continuous action spaces, we decompose the overall control variables into three components: BS beamforming, RIS control, and MA position control. Each component is assigned to an independent agent. Given the observed state from the environment, the BS, RIS, and MA agents output partial actions corresponding to W_r , W_c , Q, and $\tilde{p}$, respectively. These partial actions are combined to update the system variables. A global reward—integrating radar SINR, communication QoS satisfaction, and penalties for constraint violations—is then computed. The resulting transition is stored in replay buffers and used to update each agent’s policy in an off-policy manner. The state s_t is constructed to summarize both radar and communication performance, as well as the key control variables. To ensure learning stability and computational efficiency, the state includes the users’ SINRs, radar SINR, MA positions, channel-strength summary statistics, and a QoS violation indicator.

IV. SIMULATION RESULTS

A. Simulation setup

In this section, we evaluate the proposed MA-DRL-based joint control framework in terms of MA effectiveness, radar sensing performance, and QoS satisfaction. First, we compare the radar-SINR learning curves of single-agent DRL algorithms (SAC, DDPG, and TD3) to identify the impact of MA control. Next, we compare the proposed multi-agent framework with a single-agent scheme to verify its performance gain. Finally, we examine the minimum user SINR to confirm that communication QoS is maintained.

B. Performance evaluation

Fig. 3 presents the average radar SINR achieved by the single-agent and multi-agent DRL approaches. The single-agent scheme attains an average radar SINR of 13.23 dB, whereas the multi-agent framework achieves 15.52 dB. This performance gain confirms that decomposing the control task and training agents independently yields a more effective policy for maximizing radar sensing performance in high-dimensional settings.

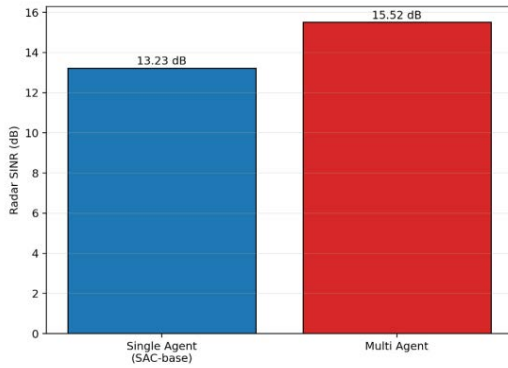


Fig. 3. Average radar SINR comparison: single-agent (SAC-based) vs. multi-agent DRL.

V. CONCLUSION

This paper investigated a joint control problem for active-RIS- and movable-antenna-assisted ISAC systems with the objective of maximizing radar sensing performance under communication QoS constraints. To handle the strong coupling among BS beamforming, RIS coefficients, and MA positions, we proposed a multi-agent deep reinforcement learning (MA-DRL) framework in which the BS, RIS, and MA are modeled as cooperative agents sharing a global reward. Simulation results showed that the proposed framework improves radar SINR compared with baseline schemes while maintaining the required communication QoS. These results demonstrate the effectiveness of MA-DRL for joint control in active-RIS- and MA-assisted ISAC systems.

REFERENCES

- [1] G. Liu et al., "Cooperative sensing for 6G ISAC: Concept, key technologies, performance evaluation, and field trial," *Engineering*, 2025, doi: 10.1016/j.eng.2025.08.033.
- [2] S. Moon, C.-G. Lee, and I. Hwang, "Secrecy rate maximization for RIS-assisted UAV-ISAC systems," *Journal of the Institute of Electronics and Information Engineers*, vol. 61, no. 11, pp. 3–10, Nov. 2024, doi: 10.5573/iee.2024.61.11.3.
- [3] S. Rubab, G. E. M. Abro, H. Mustafa, S. K. Baloch, S. A. Memon, and N. Saeed, "RIS-assisted UAV communications: A review of system models, frameworks and outage performance," *ICT Express*, 2025, doi: 10.1016/j.ict.2025.10.006.
- [4] Md. Saïam, M. Z. Chowdhury, S. R. Hasan, and Y. M. Jang, "Reconfigurable intelligent surface assisted BackCom: An overview, analysis, and future research directions," *ICT Express*, vol. 9, no. 5, pp. 927–940, Oct. 2023, doi: 10.1016/j.ict.2023.07.004.
- [5] Z. Yu, H. Ren, C. Pan, G. Zhou, B. Wang, M. Dong, and J. Wang, "Active RIS-aided ISAC systems: Beamforming design and performance analysis," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1578–1592, Mar. 2024, doi: 10.1109/TCOMM.2023.3332856.
- [6] X. Gao, J. Xu, Y. Liu, D. W. K. Ng, and M. D. Renzo, "Movable antenna-enabled RIS-aided integrated sensing and communication," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 10540–10555, Aug. 2024, doi: 10.1109/TWC.2024.3352953.
- [7] B. Lyu, Y. Bi, Y. Liu, Y. Feng, and H. Hu, "Sum-rate optimization for movable antenna enabled wireless powered communication network," *ICT Express*, 2025, doi: 10.1016/j.ict.2025.10.005. [1] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannis, E. Bjornson, K. Yang, C. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.