Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint Beamforming for RIS Assisted ISAC Systems with Discrete Phase Shifts: A DRL Based Approach

zain mustafa

zainmustafajjja@gmail.com

National University of Sciences and Technology

faisal akram

National University of Sciences and Technology

ahsan khan

National University of Sciences and Technology

Research Article

Keywords: Integrated sensing and communications (ISAC), re-configurable intelligent surfaces (RIS), multiple input multiple output (MIMO), deep deterministic policy gradient (DDPG), twin delayed deep deterministic (TD3), deep reinforcement learning (DRL)

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7225809/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Joint Beamforming for RIS Assisted ISAC Systems with Discrete Phase Shifts: A DRL Based Approach

Zain Mustafa, Faisal Akram, Ahsan Khan

***Abstract*—Integrated sensing and communication (ISAC) is regarded as a promising approach to alleviate spectrum congestion in future communication networks. Leveraging recent breakthroughs in programmable meta-materials, re-configurable intelligent surfaces (RIS) enhance communication systems by enabling multiple input multiple output (MIMO) transmission without requiring additional radio frequency (RF) chains. Therefore, integrating RIS into ISAC systems presents a promising avenue to address the fundamental challenge of enhancing ISAC system performance through only active beamforming, especially in adverse channel conditions. This correspondence investigates the joint optimization of transmit beamforming and RIS reflection coefficients to maximize the sum capacity of ISAC systems. The study adopts a more practical approach by constraining each RIS element to use only a discrete set of phase shifts. Resulting non-convex optimization problem has been addressed using the deep deterministic policy gradient (DDPG) and the twin delayed deep deterministic (TD3) policy gradient deep reinforcement learning (DRL) algorithms. The numerical results and analysis demonstrate the superior performance of TD3 compared to DDPG in terms of the improved sum capacity of ISAC system.**

***Index Terms*—Integrated sensing and communications (ISAC), re-configurable intelligent surfaces (RIS), multiple input multiple output (MIMO), deep deterministic policy gradient (DDPG), twin delayed deep deterministic (TD3), deep reinforcement learning (DRL).**

I. Introduction

The frequency spectrum currently allocated for wireless communications is becoming increasingly scarce due to the exponential growth of the wireless industry. However, large portions of the spectrum currently reserved for radar applications can be shared with communication systems [1]. There are two primary techniques in the literature for communication and radar spectrum sharing (CRSS). The first technique, known as radar communication coexistence (RCC), aims to create effective interference control plans so that there is minimal interference between the two systems while they are operating [2]. The second technique, dual functional radar-communication (DFRC), tries to create cooperative systems that are capable of wireless communication and sensing at the same time [3]. DFRC is also known in literature as radar-communications (RadCom), joint radar and communications (JRC), joint communications and radar (JCR) and the more recent term ISAC [4].

MIMO-based ISAC systems are more prevalent in the literature compared to single-antenna ISAC systems due to their spatial multiplexing capability. MIMO-based ISAC studies typically fall into two categories: information embedding [5], where communication signals are embedded into the radar waveform and transmit beamforming [6], which employs separate waveforms for each purpose, resulting in higher data rates and improved radar performance.

A fundamental challenge in MIMO based ISAC systems lies in optimizing transmit beamforming to balance communication and radar sensing functionalities. Despite efforts to enhance performance through active beamforming, conventional approaches encounter limitations, especially in adverse channel conditions [7]. To overcome these constraints, the concept of RIS has surfaced, offering a ground-breaking approach to augment degrees of freedom (DoF) by intelligently manipulating propagation environments [8], [9].

The integration of RIS into ISAC systems has become a focus of recent research. To improve communication quality of service (QoS) while adhering to power constraint and radar sensing requirements for beampattern similarity, the transmit waveform and the reflection coefficients of RIS are jointly designed in [10]. Under user SNR-based QoS constraints for single user, the transmit precoding matrix and the passive reflection matrix of the RIS are concurrently optimized with the goal of maximizing the radar detection performance [11]. The authors in [12] extended the idea to a multi-user scenario, enhancing the system's applicability and robustness. However, these studies consider communication QoS constraint which is otherwise good in high power ISAC systems. Moreover, the authors in [13] looked into joint beamforming for RIS-assisted ISAC systems in order to optimize sum capacity while adhering to the limitations of maximum power, RIS reflection coefficients, and beam pattern similarity. However, this study considered continuous values of RIS phase shifts for simulation which are difficult to achieve in practice due to hardware limitations [14].

Aforementioned works used different mathematical optimization tools for solving optimization problems which results in sub-optimal solutions. The emphasis of research has recently switched to machine learning (ML) in order to handle large amounts of data and non-convex optimization issues that arise in wireless communication systems [15]. Through combined optimization of the transmit beamforming and RIS reflection coefficients using DRL for continuous RIS phase shifts, the authors in [16] increased the sum capacity of a multiuser MISO system. Moreover, DRL was used to maximize the total capacity and radar sensing capability of an ISAC system by optimizing active and passive beamforming [17]. However, this work also considered continuous RIS phase shifts and used only communication waveform for transmit beamforming. Use of individual waveform for communication and sensing increases the available DoF for beamforming at the base station (BS) [18].

Motivated by the insights assembled from previous re-

search, this study investigates the joint optimization of transmit beamforming and RIS reflection coefficients for the RIS-assisted ISAC system with constraints of max power, RIS unit modulus, RIS discrete phase shifts, and beam pattern similarity. The resulting problem is non-convex due to multiuser interference and use of classical mathematical optimization tools like iterative search to find optimal solution is difficult for large scale networks. Moreover, use of alternating optimization techniques will result in suboptimal solution. Therefore, this correspondence uses DDPG and TD3 DRL algorithms to jointly optimize the transmit beamforming matrix and RIS phase shift matrix. Contributions of the study include:

- Adoption of a more practical approach by utilizing discrete phase shifts for each RIS element.
- Utilization of independent communication and sensing waveforms for transmit beamforming.
- Application of DDPG and TD3 DRL algorithm to address the optimization problem.

II. System Model and Problem Formulation

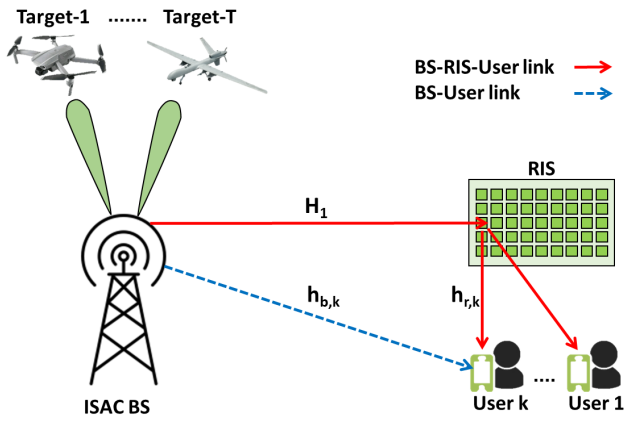


Fig. 1. RIS assisted ISAC system.

We consider an ISAC based BS with M transmit/receive antennas, RIS with N elements and K single antenna users as displayed in Fig. 1. BS is communicating with users through RIS as well as line of sight (LoS) path and simultaneously sensing T targets which are considered to be point-like. It is assumed that RIS is positioned near users and away from targets. Moreover, target echo signals that reflect through RIS and reach receiver are comparatively faint and insignificant since targets have robust LoS linkages with BS. Furthermore, we suppose that the channel matrix from the BS to the RIS $\mathbf{H}_1 \in \mathbb{C}^{N\times M}$, channel vector from RIS to the k-th user $\mathbf{h}_{r,k} \in \mathbb{C}^{N\times 1}$, $\forall k$, and channel vector from BS to k-th user $\mathbf{h}_{b,k} \in \mathbb{C}^{M\times 1}$, $\forall k$, are known at the BS and RIS. It should be noted that CSI acquisition is a complex task and requires installation of dedicated RF chains which is not scope of this paper [16]. Signal received at the k-th user is expressed as

$$y_k = (\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{H}_1 + \mathbf{h}_{b,k}^H)\mathbf{G}\mathbf{x} + w_k, \tag{1}$$

here y_k is the received signal of the k-th user, $\mathbf{G} = [\mathbf{G_c}\mathbf{G_r}] \in \mathbb{C}^{M\times(K+M)}$ is the transmit beamforming matrix which is concatenation of communication beamforming matrix $\mathbf{G}_c \in \mathbb{C}^{M\times K}$ and radar beamforming matrix $\mathbf{G}_r \in \mathbb{C}^{M\times M}$. Similarly, $\mathbf{x} = [\mathbf{x}_c^H \mathbf{x}_r^H]^H \in \mathbb{C}^{(K+M)\times 1}$ shows signal transmitted from the BS which is concatenation of communications signal vector $\mathbf{x}_c \in \mathbb{C}^{K\times 1}$ with $\mathbb{E}\{\mathbf{x}_c\mathbf{x}_c^H\} = \mathbf{I}_K$ and radar probing signal vector $\mathbf{x}_r \in \mathbb{C}^{M\times 1}$ with $\mathbb{E}\{\mathbf{x}_r\mathbf{x}_r^H\} = \mathbf{I}_M$, where $\mathbf{I}$ shows identity matrix. Additive white Gaussian noise at the k-th user is denoted by w_k. RIS phase shifts are given by the diagonal matrix $\mathbf{\Phi} \triangleq \text{diag}[\phi_1, \phi_2, ..., \phi_N] \in \mathbb{C}^{N\times N}$. Here, we assume that each element of RIS changes only phase of reflected signal and this reflection is lossless. Realizing continuous RIS phase shifts is difficult in practice, therefore we consider discrete phase shift values for RIS. We assume $B = 2^b$ as the total discrete phase shift levels or angular resolution of each RIS element, where b shows bits per RIS element. The continuous phase shifts in the range of $[0, 2\pi)$ can then be uniformly quantized and the set of these phase shifts is $S \triangleq \{0, \Delta\omega, \ldots, (B-1)\Delta\omega\}$ with $\Delta\omega = \frac{2\pi}{B}$. For each RIS element n (where $n = 1, 2, \ldots, N$), the continuous phase shift ϕ_n is quantized to the nearest discrete phase shift ϕ_n' as

$$\phi_n' = \left\lfloor \frac{\phi_n \mod 2\pi}{\Delta\omega} + 0.5 \right\rfloor \cdot \Delta\omega, \tag{2}$$

the quantization error for the n-th element is given by the difference between the continuous phase shift and quantized phase shift $\psi_n = \phi_n - \phi_n'$. Given that the phase shifts are uniformly quantized, the mean absolute discretization error (MADE) for a single RIS element can be calculated as the expected value of the absolute quantization error

$$\text{MADE}_n = \mathbb{E}[|\psi_n|] = \frac{\Delta\omega}{4} = \frac{\pi}{2^{b+1}}. \tag{3}$$

In order to determine SINR of the k-th user, Eq. (1) can be written in terms of desired signal and interference signal

$$y_k = (\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{H}_1 + \mathbf{h}_{b,k}^H)\mathbf{g}_k x_k + \sum_{j=1, j\neq k}^{K+M} (\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{H}_1 + \mathbf{h}_{b,k}^H)\mathbf{g}_j x_j + w_k, \tag{4}$$

where $\mathbf{g}_k$ shows k-th column vector of transmit beamforming matrix $\mathbf{G}$ and x_k shows data for k-th user. The first term in Eq. (4) represents actual signal intended for user k and second term shows interference from signals of remaining users and targets. Therefore, the SINR of the k-th user can be written as

$$\rho_k = \frac{\left|(\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{H}_1 + \mathbf{h}_{b,k}^H)\mathbf{g}_k\right|^2}{\sum_{j=1, j\neq k}^{K+M} \left|(\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{H}_1 + \mathbf{h}_{b,k})\mathbf{g}_j\right|^2 + \sigma_k^2}, \tag{5}$$

the achievable capacity of the k-th user for each Hertz of bandwidth is given as

$$R_k = \log_2(1 + \rho_k), \tag{6}$$

and the achievable sum capacity for all the users can be expressed as

$$R(\mathbf{G}, \mathbf{\Phi}, \mathbf{H}_1, \mathbf{h}_{r,k}, \mathbf{h}_{b,k}) = \sum_{k=1}^{K} R_k. \tag{7}$$

Focusing the signal intensity towards the target and lowering it elsewhere is a common tactic used in radar sensing to enhance target identification and estimate [13]. A beampattern similarity metric is frequently used to compare the given beampattern with the ideal one to evaluate sensing performance. For instance, steering vector for a direction θ is defined as $\mathbf{a}(\theta) = [1, e^{j2\pi\delta/\lambda \sin(\theta)}, ..., e^{j2\pi(M-1)\delta/\lambda \sin(\theta)}]^T \in \mathbb{C}^{M\times 1}$, where δ represents antenna spacing and λ signifies signal wavelength. The designed transmit beampattern can be calculated as [19]

$$P_d(\theta; \mathbf{G}) = \mathbb{E}\{\left|\mathbf{a}^H(\theta)\mathbf{G}\mathbf{x}\right|^2\} = \mathbf{a}^H(\theta)\mathbf{G}\mathbf{G}^H\mathbf{a}(\theta), \quad (8)$$

where $P_d(\theta; \mathbf{G})$ represents designed beampattern for angle θ and beamforming matrix $\mathbf{G}$. The beampattern similarity measure ϵ can be calculated as the mean squared error (MSE) among the ideal and designed beampatterns [13], i.e.,

$$\epsilon(\alpha, \mathbf{G}) = \frac{1}{L}\sum_{l=1}^{L} \left|\alpha P_i(\theta_l) - P_d(\theta_l; \mathbf{G})\right|^2, \quad (9)$$

where θ_l shows the l-th sampled angle, $P_i(\theta_l)$ shows ideal beampattern at angle θ_l, and α is a scaling factor for ideal beampattern $P_i(\theta_l)$. By adjusting the value of α, we can adjust the level of beampattern similarity between the ideal and designed beampatterns.

The goal here is to maximize the sum capacity R by finding the optimal values for the transmit beamforming matrix $\mathbf{G}$ and the RIS phase shifts matrix $\mathbf{\Phi}$. The optimization problem is given as

$$\max_{\mathbf{G},\mathbf{\Phi}} R(\mathbf{G}, \mathbf{\Phi}, \mathbf{H}_1, \mathbf{h}_{r,k}, \mathbf{h}_{b,k}) \quad (10a)$$

$$\text{s.t } \frac{1}{L}\sum_{l=1}^{L} \left|\alpha P_i(\theta_l) - P_d(\theta_l; \mathbf{G})\right|^2 \leq \epsilon, \quad (10b)$$

$$\text{tr}\{\mathbf{G}\mathbf{G}^H\} \leq P, \quad (10c)$$

$$|\phi_n| = 1, \quad \forall n = 1, 2, 3, \ldots, N, \quad (10d)$$

$$\angle\phi_n \in S, \quad \forall n = 1, 2, 3, \ldots, N, \quad (10e)$$

where P shows the total power available at transmitter. The described optimization problem is non-convex due to constraints like multi-user interference and discrete RIS phase shifts.

Conventional mathematical optimization tools often rely on iterative search methods that may converge to suboptimal solutions, especially in scenarios with high-dimensional constraints. Such tools are also computationally intensive, making them impractical for large-scale networks or real-time applications where quick adaptability is crucial. Moreover, the use of alternating optimization techniques will result in suboptimal solution. Large amount of data and non-convex optimization problems inherent in wireless communication can be effectively handled using machine learning tools. Therefore, we use TD3 and DDPG DRL algorithms to address our optimization problem.

III. DRL Based Design

A. DRL Based Modeling of the Problem

Fundamental components of any DRL problem are state, action, and reward. For problem (10), these fundamental components can be defined as:

1) State: State at time t, s_t, represents the condition of the environment at that particular time instant. For problem (10), state at time t consists of channel state matrix from the BS to RIS, $\mathbf{H}_1 \in \mathbb{C}^{N\times M}$, channel state vector from RIS to the k-th user $\mathbf{h}_{r,k} \in \mathbb{C}^{N\times 1}$, $\forall k$, and channel state vector from BS to k-th user $\mathbf{h}_{b,k} \in \mathbb{C}^{M\times 1}$, $\forall k$. It also includes the action from the last time step a_{t-1} (which comprises of both $\mathbf{G} \in \mathbb{C}^{M\times(K+M)}$ and $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$) and beampattern similarity metric ϵ. Thus, dimension of state space $= 2(N \cdot M + N \cdot K + M \cdot K + M(K + M) + N) + 1$. Here, 2 has been multiplied to show that real and imaginary part of each complex entity is given as separate input to the deep neural networks (DNN) and 1 has been added for beampattern similarity metric ϵ.

2) Action: Action at time t a_t, refers to the decision or choice made by the agent in response to the state received from the environment. For problem (10), agent can optimize either transmit beamforming $\mathbf{G} \in \mathbb{C}^{M\times(K+M)}$ or RIS phase shifts $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$. Therefore, the action space dimension $= 2(M(K + M) + N)$.

3) Reward: Indicates the feedback or response of the environment to the agent based on its action, thus facilitating learning and adaptation. For a particular value of state at time t, s_t and corresponding action obtained from the actor network a_t, reward has been defined as the sum capacity R of the system at that time step.

B. Algorithm Description

We use both DDPG and TD3 DRL algorithms to achieve the optimization objective. DDPG is an actor-critic based DRL algorithm designed for continuous action spaces. It uses a deterministic policy, where the actor network generates specific actions, and a critic network evaluates them through Q-value predictions [20].

TD3 is an advanced DRL algorithm designed to address overestimation bias in deterministic policy gradients [21]. The key components of TD3 are an actor network, which represents the policy and outputs actions based on current states, and two critic networks, which estimate Q-values to prevent overestimation by using the minimum value. Additionally, target networks for both the actor and critics are updated slowly to stabilize the training.

IV. Numerical Results and Analysis

This section evaluates the efficiency of the proposed TD3 algorithm against the DDPG algorithm for both discrete and continuous values of RIS phase shifts. For simulations, we use a Rayleigh distribution to randomly create the channel matrix $\mathbf{H}_1$, vector $\mathbf{h}_{r,k}$, and vector $\mathbf{h}_{b,k}$. We presume that the environment's shadowing effects and large-scale path loss have been taken into account. Consider an ISAC based BS with $M = 8$ antennas communicating with $K = 8$ users through RIS which has $N = 8$ elements and simultaneously sensing $T = 2$ targets. Total power available at transmitter $P = 10$ dB, number of bits per RIS element $b = 5$, and angular resolution

per RIS element $B = 2^5 = 32$. The ideal beampattern $P_i(\theta_l)$ for sensing is given as

$$P_i(\theta_l) = \begin{cases} 1, & \theta_l \in \left[\theta_t - \frac{\Delta\theta}{2}, \theta_t + \frac{\Delta\theta}{2}\right], t = 1, \ldots, T \\ 0, & \text{otherwise}, \end{cases} \quad (11)$$

where θ_t represents the angle of the t-th target. Angle of target-1 $\theta_1 = -35°$, angle of target-2 $\theta_2 = 35°$ and beam width $\Delta\theta = 10°$. The hyperparameters for the TD3 and DDPG algorithms are detailed in Table I. Results are averaged over 200 episodes, with each episode lasting 8,000 time steps.

TABLE I
DESCRIPTION OF HYPERPARAMETERS

Hyperparameter	TD3	DDPG
Discount factor	0.99	0.99
Replay buffer size	10^6	10^6
Batch size	100	64
Learning rate	10^{-3}	10^{-3}
Policy update frequency	2	1

Figure 2 shows comparison of proposed TD3 DRL algorithm with DDPG algorithm. The graph clearly shows that after nearly identical starting point, the TD3 DRL algorithm attains higher average rewards compared to the DDPG algorithm for both discrete and continuous RIS phase shifts. Steeper learning curve for TD3-based approach displays its higher convergence rate than DDPG. The TD3 algorithm requires approximately 4,000 time steps to converge, whereas the DDPG algorithm requires around 6,000 time steps. Moreover, performance of both algorithms for the discrete phase shifts scenario is lower as compared to continuous phase shifts scenario due to reduced DoF in terms of selection of optimal phase shifts at RIS.

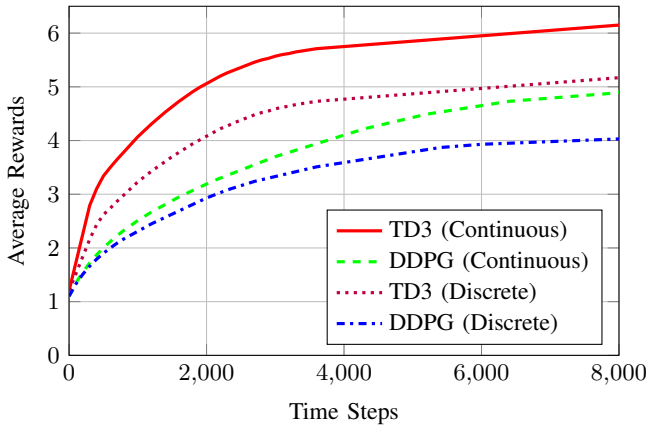


Fig. 2. Average rewards vs. time steps ($M = 8, N = 8, K = 8, P = 10$ dB).

Sum capacity for different number of RIS elements is depicted in Fig. 3. An increase in the number of RIS elements allow higher spatial DoF which results in an increase of sum capacity for both algorithms. However, this increase in sum capacity is at the cost of higher computational resources to learn optimal RIS phase shifts. Moreover, the steeper slope of the TD3-based approach compared to DDPG indicates that the proposed TD3 algorithm performs better in complex environments resulting from an increase in RIS elements. Consequently, TD3 in the discrete setting surpasses DDPG in the continuous setting once the RIS elements number exceeds 45.

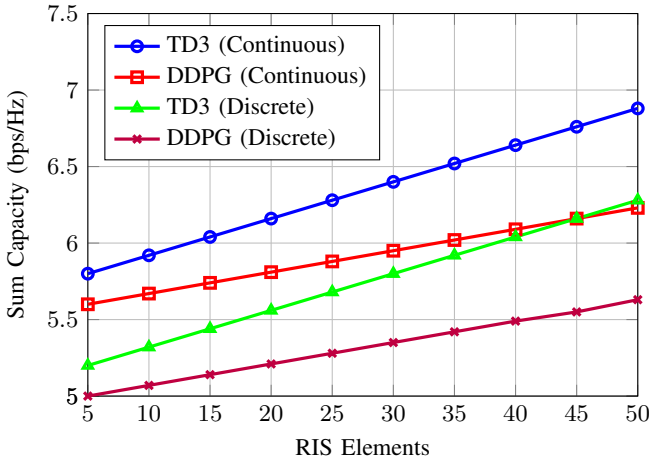


Fig. 3. Sum capacity vs. RIS elements ($M = 8, K = 8, P = 10$ dB).

Figure 4 displays the relationship between sum capacity and transmit power as angular resolution B of each RIS element is increased by increasing number of bits per RIS element b. An increase in the transmit power results in higher sum capacity of ISAC system for all values of b. Moreover, an increase in the angular resolution further enhances sum capacity of ISAC system due to higher DoF in terms of choosing optimal phase shift at RIS. However, increase of sum capacity with angular resolution gradually slows down and the ideal continuous phase shift scenario is nearly achieved when $b = 10$.

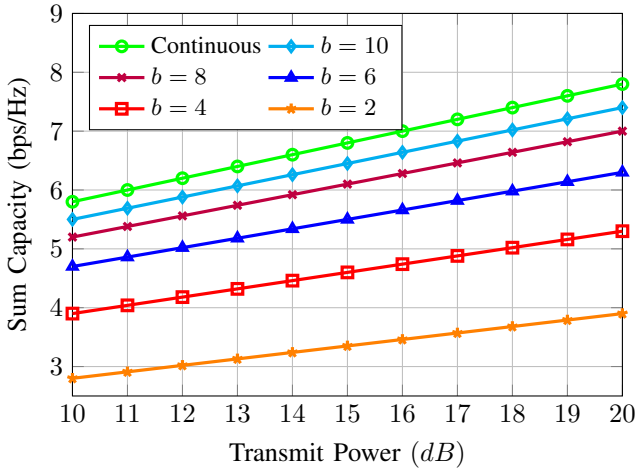


Fig. 4. Sum capacity vs. transmit power($M = 8, N = 8, K = 8$).

Figure 5 illustrates the trade-off between communication performance, measured by sum capacity versus radar sensing performance, measured by the MSE between ideal and designed beampatterns (ϵ) for different system settings. An ϵ value of 0 signifies perfect similarity between the ideal and designed beampatterns, indicating optimal radar performance, and vice versa. The figure shows that as ϵ increases from

0 to 1, communication performance improves in terms of sum capacity. This improvement occurs because a higher MSE between the ideal and designed beampatterns allows for greater DoF to direct beams towards users. Moreover, increase in transmit antennas, RIS elements or users further improves the sum capacity of the system. However, while increase in transmit antennas leads to a steeper capacity plot, increase in RIS elements and users results in a more gradual slope. This occurs because the MSE between the ideal and designed transmit beampattern is closely related to the number of transmit antennas.

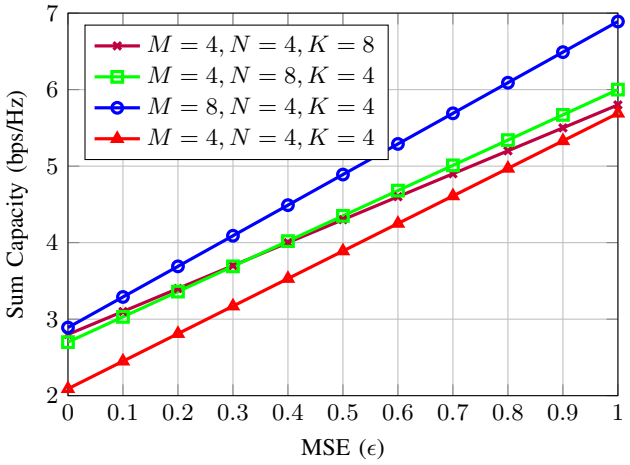


Fig. 5. Sum capacity vs. MSE ($P = 10$ dB).

V. Conclusion

In this correspondence, we explored the joint optimization of transmit beamforming and RIS reflection coefficients in a RIS assisted ISAC system, considering constraints of maximum power, RIS unit modulus, RIS discrete phase shifts, and beam pattern similarity. We considered the setting where a MIMO-based BS detects several targets as well as communicates with multiple single-antenna users using RIS. The proposed approach leverages the DDPG and TD3 DRL algorithms to achieve the optimization objective. The numerical results and analysis demonstrate the superior performance of TD3 compared to DDPG in terms of the improved sum capacity of ISAC system.

References

[1] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, Aug. 2018.

[2] B. Paul, A. R. Chiriyath, and D. W. Bliss, "Survey of RF communications and sensing convergence research," *IEEE Access*, vol. 5, pp. 252–270, 2017.

[3] 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.

[4] J. A. Zhang, M. L. Rahman, K. Wu, X. Huang, Y. J. Guo, S. Chen, and J. Yuan, "Enabling joint communication and radar sensing in mobile networks—A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, pp. 306–345, 2023.

[5] X. Wang, A. Hassanien, and M. G. Amin, "Sparse transmit array design for dual-function radar communications by antenna selection," *Digit. Signal Process.*, vol. 83, pp. 223–234, Dec. 2018.

[6] C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018.

[7] R. Liu, M. Li, H. Luo, Q. Liu, and A. L. Swindlehurst, "Integrated sensing and communication with reconfigurable intelligent surfaces: Opportunities, applications, and future directions," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 50–57, 2023.

[8] Q. Wu, S. Zhang, B. Xiong, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[9] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. D. Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, 2020.

[10] R. Liu, M. Li, Y. Liu, Q. Wu, and Q. Liu, "Joint transmit waveform and passive beamforming design for RIS-aided DFRC systems," *IEEE J. Sel. Topics Signal Process.*, no. 5, pp. 995–1010, Aug. 2022.

[11] X. Wang, Z. Fei, Z. Zheng, and J. Guo, "Joint waveform design and passive beamforming for RIS-assisted dual-functional radar-communication system," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 5131–5136, May 2021.

[12] Z.-M. Jiang, M. Rihan, P. Zhang, L. Huang, Q. Deng, J. Zhang, and E. M. Mohamed, "Intelligent reflecting surface aided dual-function radar and communication system," *IEEE Syst. J.*, vol. 16, no. 1, pp. 475–486, 2022.

[13] H. Luo, R. Liu, M. Li, Y. Liu, and Q. Liu, "Joint beamforming design for RIS-assisted integrated sensing and communication systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13 393–13 397, 2022.

[14] W. Xu, L. Gan, and L. Gan, "A robust deep learning-based beamforming design for RIS-assisted multiuser MISO communications with practical constraints," *IEEE Trans. Cogn Commun. Netw*, vol. 8, no. 2, pp. 694–706, Nov. 2021.

[15] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts*, vol. 21, no. 4, pp. 3072–3108, 2019.

[16] 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.

[17] Y. Eghbali, S. Faramarzi, S. K. Taskou, M. R. Mili, M. Rasti, and E. Hossain, "Beamforming for STAR-RIS-aided integrated sensing and communication using meta DRL," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 919–923, 2024.

[18] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, 2020.

[19] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[21] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International Conference on Machine Learning*, 2018, [Online]. Available: https://api.semanticscholar.org/CorpusID:3544558.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AuthorBiographies.docx