Beamforming and Reflection Design for Short Packet ISAC With Non-Ideal RIS: An A3C-Based Approach

Behrad Mahmoudi, Ahmad Khonsari,
Farshad Zeinali, *Student Member, IEEE*,
Mohammad Robat Mili,
Mahdi Boloursaz Mashhadi, *Senior Member, IEEE*,
and Pei Xiao, *Senior Member, IEEE*

***Abstract*—Integrated sensing and communication (ISAC) is a promising solution to mitigate the increasing congestion of the wireless spectrum. In this paper, we investigate the short packet communication regime within an ISAC system assisted by a reconfigurable intelligent surface (RIS) to meet the low latency ultra-reliable requirements in the next-generation wireless networks. We consider a non-ideal RIS model that captures effects of the phase-dependent amplitude variations in the reflection coefficients, and we have incorporated the near-field model into the channels between the RIS and the users or targets. In this setup, we jointly design the transmit beamforming and the RIS phase shifts to maximize the sum rate while satisfying the sensing signal-to-noise ratio (SNR) requirement. The system simultaneously carries out multitarget sensing and multi-user short packet communications with the help of the RIS. Considering the non-convex and dynamic nature of the resulting optimization problem, we propose an asynchronous advantage actor-critic (A3C) based method for beamforming and reflection design in this setup. Numerical results demonstrate the superiority of the proposed scheme over the benchmarks.**

***Index Terms*—Asynchronous advantage actor-critic (A3C), integrated sensing and communication (ISAC), non-ideal reconfigurable intelligent surface (RIS), reinforcement learning (RL), short packet communication.**

I. Introduction

Integrated sensing and communication (ISAC) has emerged as a promising solution to address scarcity of the spectrum by enabling sensing and communication functionalities to share wireless resources on a unified platform. As a key technology in future networks, ISAC should support new traffic types with stringent latency, data rates, and reliability requirements, such as vehicle-to-vehicle (V2V) communication for traffic safety, or real-time video transmission for augmented reality [1], [2]. These applications require short codewords for low latency and high reliability, rendering the traditional Shannon capacity formula inapplicable. Thus, optimizing the ISAC systems in future wireless networks with finite-length codewords is of significant interest [3], [4].

A key research area in ISAC systems is designing dual-function transmit waveforms, where a multi-input multi-output (MIMO) system assisted with the reconfigurable intelligent surface (RIS) technology can be considered to attain beamforming gains and extra degrees of freedom (DoFs) can be exploited by establishing additional non-line-of-sight (NLoS) links through the RIS. An RIS typically consists of a planar meta-surface with cost-effective, hardware-efficient, passive reflecting elements. An overview of signal processing techniques for RIS/IRS-aided wireless systems is provided in [5]. Due to the above benefits, recent works have utilized RISs in ISAC systems [6], [7].

Prior works mostly assumed an ideal phase shift model for the RIS with lossless reflection and uniform amplitude [8], [9], [10], [11], which is impractical due to hardware limitations [12]. Conventional iterative optimization techniques have been applied in these studies for beamforming and reflection design [8], [9], [10], however, such optimization techniques are generally complex and less effective in dynamically changing wireless environments. Furthermore, these prior works did not consider the latency-sensitive short packet communication regime. This is despite the fact that in modern applications, particularly in the realms of Internet of Things (IoT) and Machine-Type Communication (MTC) in industrial automation, there is a huge demand for low-latency and high-reliability communications. Short packet communications is of specific interest in applications such as autonomous vehicles, real-time control in robotics and drones, and industrial automation, where delays can lead to significant consequences. Table I provides an overview of the key differences between this work and the literature on beamforming and reflection design for RIS-assisted ISAC systems. To the best of our knowledge, this work is the first of its kind to address the joint beamforming and reflection design problem using an RL-based technique for the ISAC setup assisted with non-ideal RIS in the short packet regime.

The rest of this paper is organized as follows. In Section II, we provide the system model and problem formulation. In Section III, we present our proposed A3C-based beamforming and reflection design approach. We provide the simulation results in Section IV, and Section V concludes the paper.

Notations: Boldface capital and lower-case letters represent matrices and vectors, e.g. A and a. Matrix transposition is expressed by $(.)^T$, and $\mathbb{E}\{.\}$ is the expectation operator. Tr$(.)$ denotes trace of a matrix, and U$(.)$ and $Q^{-1}(.)$ denote the unit step function, and the inverse complementary Gaussian cumulative distribution function, respectively. $(.)^T$, $(.)^H$ are matrix transposition and complex Hermitian, $\odot$ denotes element-wise multiplication, and $\mathbf{I}_K$ is the $K \times K$ identity matrix.

II. System Model and Problem Formulation

We consider an RIS-assisted ISAC system as shown in Fig. 1, where a base station (BS) with M antennas, simultaneously tracks T point targets, and communicates with K single-antenna users with the assistance of an N-element RIS. We denote the information bits transmitted to each user k by b_k, that are encoded by the BS into block codes of length L denoted by $z_k,\ k \in \{1,\ldots,K\}$. The transmission period is divided into T time slots, each with a length of L/T, which is used to simultaneously serve all the users but track only a single target in a time-division manner. We assume that transmission of the sensing signals is highly directional so that there are no echoes or interference from other targets. We assume a quasi-static flat-fading model where the wireless channel remains unchanged over the whole codeword. The

Received 9 August 2024; revised 1 December 2024; accepted 27 December 2024. Date of publication 3 January 2025; date of current version 20 May 2025. The review of this article was coordinated by Dr. Cunhua Pan. *(Corresponding author: Farshad Zeinali.)*

Behrad Mahmoudi is with the Department of Electrical and Computer Engineering, University of Tehran, Tehran 14155-6619, Iran (e-mail: mahmoudi.behrad@ut.ac.ir).

Ahmad Khonsari is with the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 14155-6619, Iran, and also with the School of Computer Science in the Institute for Research in Fundamental Sciences, Tehran 19538-33511, Iran (e-mail: a_khonsari@ut.ac.ir).

Farshad Zeinali, Mahdi Boloursaz Mashhadi, and Pei Xiao are with the 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, GU2 7XH Guildford, U.K. (e-mail: f.zeinali@surrey.ac.uk; m.boloursazmashhadi@surrey.ac.uk; p.xiao@surrey.ac.uk).

Mohammad Robat Mili is with the Pasargad Institute for Advanced Innovative Solutions, Tehran 5613611958, Iran (e-mail: mohammad.robatmili@gmail.com).

Digital Object Identifier 10.1109/TVT.2025.3525525

TABLE I
THE KEY DIFFERENCES BETWEEN THIS WORK AND THE LITERATURE

	Role of RIS	Echo paths	Optimization algorithm	Ideal /Non-ideal RIS	Short packet
[4]	No RIS	LoS	BMM[1] and EPMO[2]	Ideal	Yes
[6]	C&S	LoS & NLoS	Iterative SOCP[3]	Ideal	No
[8]	C&S	LoS	ADMM[4]-MM	Ideal	No
[9]	C	LoS	Manifold/Successive optimization	Ideal	No
[10]	C&S	LoS	SCA[5]+Manifold optimization	Ideal	No
This work	C&S	LoS & NLoS	A3C RL	Non-ideal	Yes

[1]Bisection-based majorization and minimization, [2]Exact penalty-based manifold optimization, [3]Second-order cone programming, [4]Alternating direction method of multipliers, [5]Successive Convex Approximation.

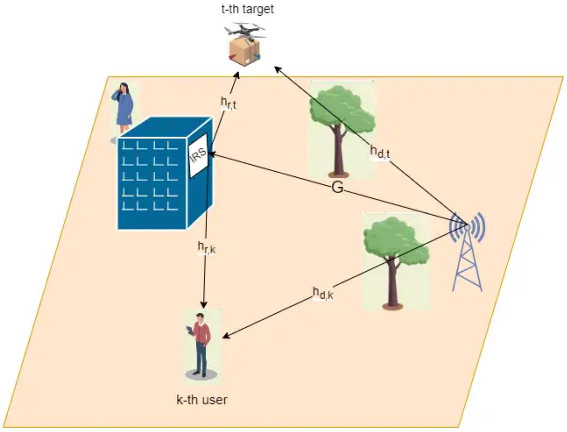


Fig. 1. An RIS-assisted ISAC system model.

transmit signal of the BS is then expressed as

$$\mathbf{x} = \mathbf{W}_c \mathbf{z}_c + \mathbf{W}_s \mathbf{z}_s = \mathbf{W}\mathbf{z}, \tag{1}$$

where $\mathbf{W}_c \in \mathbf{C}^{M\times K}$ and $\mathbf{W}_s \in \mathbf{C}^{M\times M}$ are the communication and sensing beamforming matrices, respectively. Vector $\mathbf{z}_c \in \mathbf{C}^K$ denotes the block code encoded by BS with $\mathbb{E}\{\mathbf{z}_c\mathbf{z}_c^H\} = \mathbf{I}_K$, and vector $\mathbf{z}_s \in \mathbf{C}^M$ denotes the radar probing signal with $\mathbb{E}\{\mathbf{z}_s\mathbf{z}_s^H\} = \mathbf{I}_M$. These vectors are assumed to be statistically independent of each other. We define the overall beamforming matrix and transmitted symbols as follows

$$\mathbf{W} = [\mathbf{W}_c \mathbf{W}_s] \in \mathbf{C}^{M\times(K+M)},$$
$$\mathbf{z} = \left[\mathbf{z}_c^T \mathbf{z}_s^T\right]^T \in \mathbf{C}^{(K+M)}. \tag{2}$$

In order to characterize the fundamental relationship between the RIS reflection amplitude and phase shift, we use the non-ideal RIS model introduced in [12], which derives the reflection coefficients as $\psi_n = \beta_n(\theta_n)e^{j\theta_n}$, where $\theta_n \in [-\pi, \pi)$ and $\beta_n(\theta_n) \in [0, 1]$ denote the phase shift and the corresponding phase-dependent amplitude for the n'th RIS element, respectively. Then, $\beta_n(\theta_n)$ is derived as

$$\beta_n(\theta_n) = (1-\beta_{\min})\left(\frac{\sin(\theta_n - \phi) + 1}{2}\right)^{\alpha} + \beta_{\min}, \tag{3}$$

where $\beta_{\min} \geq 0$, $\phi \geq 0$, and $\alpha \geq 0$ are constants determined by the equivalent circuit model of the RIS elements. Note that for $\alpha = 0$, (3) reduces to the ideal phase shift model, i.e, $\beta_n = 1$, $\forall n$. In the practical non-ideal RIS, the equivalent parameters of the RIS circuit elements are assumed to remain fixed after fabrication, and thereby the corresponding $(\kappa, \phi, \beta_{\min})$ values can be simply determined, e.g. through standard curve fitting.

From the communications perspective, we aim to maximize the sum-rate of all users. The received signal at the k-th user is represented as

$$y_{c,k} = \left(\mathbf{h}_{d,k}^T + (\mathbf{h}_{r,k} \odot \alpha_R)^T \,\mathbf{\Psi}\mathbf{G}\right)\mathbf{W}\mathbf{z} + n_k. \tag{4}$$

We represent the baseband equivalent channel response from the BS-to-user k, RIS-to-user k, and BS-to-RIS as $\mathbf{h}_{d,k} \in \mathbf{C}^M$, $\mathbf{h}_{r,k} \in \mathbf{C}^N$, and $\mathbf{G} \in \mathbf{C}^{N\times M}$, respectively. Also, we denote the non-ideal reflection coefficient by the diagonal matrix

$$\mathbf{\Psi} = \mathbf{B}(\mathbf{\Theta}) \odot \mathbf{\Theta} = diag\left(\boldsymbol{\beta}(\boldsymbol{\theta})\right) \odot diag\left(\boldsymbol{\theta}\right), \tag{5}$$

as the reflection-coefficients matrix at the RIS, where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$ is the phase shift of the RIS elements, and $\boldsymbol{\beta}(\boldsymbol{\theta}) = [\beta_1(\theta_1), \ldots, \beta_N(\theta_N)]^T$ according to (3). We have also noted that the near-field assumption is more appropriate for channels with short distances. Therefore, we have adopted the near-field model for the channel between the RIS and the users [13].

$$\alpha_R = \left[\alpha_R[1], \alpha_R[2], \ldots \alpha_R[N]\right]^T, \tag{6}$$

where

$$\alpha_R[n] = \exp\left(-j\frac{2\pi}{\lambda}\left(J_n(\omega_k, \varphi_k) + Q_n\left(\omega_k, \varphi_k, d_{R,k}^0\right)\right)\right), \tag{7}$$

In this equation, $J_n(\omega_k, \varphi_k)$ and $Q_n(\omega_k, \varphi_k, d_{R,k}^0)$ are defined as

$$J_n(\omega_k, \varphi_k) = (d_{z,n}\omega_u - d_{y,n}\phi_u), \tag{8}$$

$$Q_n\left(\omega_k, \varphi_k, d_{R,k}^0\right) = \frac{1}{2d_{R,k}^0}\left(\left(d_z^2 + d_y^2\right) - \left(d_{z,n}\omega_k - d_{y,n}\varphi_k\right)^2\right), \tag{9}$$

where $d_{z,n}$ and $d_{y,n}$ indicates the distance from the n^{th} element of the RIS to the center of the RIS along the z and y axes, respectively. Moreover, $d_{R,k}^0$ is the distance from center of the RIS to the user k, and

$$\omega_k = \sin\phi_{k,0}, \tag{10}$$

$$\varphi_k = \sin\theta_{k,0}\cos\phi_{k,0}, \tag{11}$$

where $\theta_{k,0}$ and $\phi_{k,0}$ are vertical and azimuth angles of arrival, respectively.

It is assumed that all the channel state information (CSI) is perfectly known at the BS using channel estimation techniques studied in the literature. The scalar $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the AWGN noise with variance σ_k^2 at the k-th user. Then the signal-to-interference-and-noise ratio (SINR) at user k can be expressed as

$$\gamma_k = \frac{\left|\left(\mathbf{h}_{d,k}^T + (\mathbf{h}_{r,k} \odot \alpha_R)^T \left[\mathbf{B}(\boldsymbol{\Theta}) \odot \boldsymbol{\Theta}\right] \mathbf{G}\right) \mathbf{w}_k\right|^2}{\sum_{j \neq k}^{k+N} \left|\left(\mathbf{h}_{d,k}^T + (\mathbf{h}_{r,k} \odot \alpha_R)^T \left[\mathbf{B}(\boldsymbol{\Theta}) \odot \boldsymbol{\Theta}\right] \mathbf{G}\right) \mathbf{w}_j\right|^2 + \sigma_k^2}, \tag{12}$$

where the $\mathbf{w}_j$ is the j-th column of the beamforming matrix $\mathbf{W}, j = 1, \ldots, K + M$. The achievable data rate of each user k in the short packet regime is then given by

$$R_k = \log_2(1 + \gamma_k) - Q^{-1}(\epsilon_k) \sqrt{\frac{D_k}{L}}, \tag{13}$$

where ϵ_k is the decoding error, L denotes the block length and D_k indicates the channel dispersion given by $D_k = c^2(1 - (1 + \gamma_k)^{-2})$, where $c = \log_2(e)$ [14]. This effective capacity takes into account the latency requirement [15].

From the sensing point of view, we need to guarantee that the sensing signal-to-noise ratio (SNR) remains above a certain threshold ζ_{th} to ensure accurate sensing of the targets. As shown in Fig. 1, the sensing probing signal reaches the targets through both the direct and RIS links, and is then reflected back to the BS via both links as well. We do not consider the propagation delay between these links, and thus, the reflected signal from the t-th target is given by

$$\begin{aligned} \mathbf{y}_{s,t} = &\left(\mathbf{h}_{d,t} + \mathbf{G}^T \boldsymbol{\Psi} (\mathbf{h}_{r,t} \odot \alpha'_R)\right) \left(\mathbf{h}_{d,t}^T + (\mathbf{h}_{r,t} \odot \alpha'_R)^T \boldsymbol{\Psi} \mathbf{G}\right) \mathbf{W}\mathbf{z} \\ &+ \mathbf{n}_s, \end{aligned} \tag{14}$$

where $\mathbf{h}_{d,t} \in \mathbf{C}^M$, $\mathbf{h}_{r,t} \in \mathbf{C}^N$ and $\mathbf{G} \in \mathbf{C}^{N \times M}$ represents the baseband channel between the BS and the t-th target, between the RIS and the t-th target and between the BS and the RIS, respectively. Moreover, we have incorporated the near-field model from a sensing perspective as well. Here, the α'_R is defined as:

$$\alpha'_R = \left[\alpha_R[1], \alpha_R[2], \ldots \alpha_R[N]\right]^T, \tag{15}$$

where

$$\alpha_R[n] = \exp\left(-j\frac{2\pi}{\lambda}\left(J_n(\omega_t, \varphi_t) + Q_n(\omega_t, \varphi_t, d_{R,t}^0)\right)\right). \tag{16}$$

All of the above variables are defined for targets in the same manner as they are for users in (8) to (11).

In (14), the Radar Cross-Section (RCS) effects are absorbed in the channel coefficient vectors $\mathbf{h}_{d,t}$ and $\mathbf{h}_{r,t}$ for sensing. For brevity, we define the overall equivalent channel matrix for target t as

$$\begin{aligned} \mathbf{H}_t = &\left(\mathbf{h}_{d,t} + \mathbf{G}^T \left[\mathbf{B}(\boldsymbol{\Theta}) \odot \boldsymbol{\Theta}\right] (\mathbf{h}_{r,t} \odot \alpha'_R)\right) \\ &\times \left(\mathbf{h}_{d,t}^T + (\mathbf{h}_{r,t} \odot \alpha'_R)^T \left[\mathbf{B}(\boldsymbol{\Theta}) \odot \boldsymbol{\Theta}\right] \mathbf{G}\right), \end{aligned} \tag{17}$$

thus, the sensing SNR of the t-th target is given by

$$\zeta_t = \frac{\mathbb{E}\left|\mathbf{H}_t \mathbf{W}\mathbf{z}\right|^2}{{\sigma_s}^2} = \frac{\mathrm{Tr}(\mathbf{W}^H \mathbf{H}_t^H \mathbf{H}_t \mathbf{W})}{{\sigma_s}^2}, \tag{18}$$

where ${\sigma_s}^2$ is the noise variance at the sensing receiver.

In this paper, we jointly design the beamforming matrix at the BS, as well as the RIS phase shifts $\mathbf{W}, \boldsymbol{\Theta}$, to maximize the sum communication rate while satisfying the target sensing SNR requirements, considering the non-ideal phase-dependent amplitude of the RIS reflection coefficients, under the constraint of the total transmit power budget. The resulting optimization problem is thereby formulated as

$$\begin{aligned} &\max_{\mathbf{W}, \boldsymbol{\theta}} \sum_{k=1}^{K} R_k \\ &\text{s.t.} \quad \sum_{j=1}^{K+M} [\mathbf{W}]_j^H [\mathbf{W}]_j \le P_T, \\ &\zeta_t \ge \zeta_{th}, \forall t \in \{1, \ldots, T\}, \\ &\beta_n(\theta_n) = (1 - \beta_{\min}) \left(\frac{\sin(\theta_n - \phi) + 1}{2}\right)^\alpha + \beta_{\min}, \forall n, \end{aligned} \tag{19}$$

in which $[\mathbf{W}]_j$ denotes the j'th column of $\mathbf{W}$, ζ_{th} is a pre-defined target sensing SNR requirement, and P_{T} indicates the total transmit power budget. We note that the objective function and constraints in (19) are highly non-linear and non-convex with regard to the optimization variables. To address this non-convex problem, intricate transformations based on convex optimization can iteratively obtain a locally optimal solution, albeit at the cost of computational efficiency. Moreover, dynamic changes in the wireless environment, e.g. changes in the CSI, lead to a non-trivial performance loss when using such time-consuming, iterative solutions. In the following, we propose an inherently adaptive beamforming and reflection design approach based on A3C RL to optimize the variables $\mathbf{W}, \boldsymbol{\theta}$ in a computationally affordable manner.

Algorithm 1: The Proposed A3C-Based Algorithm for Non-Ideal RIS-Assisted ISAC With Short Packet Length.

```
Initialize the neural networks, actor and critic parameters
Initialize the environment parameters
for each episodes do
    Synchronize the target actor and target critic parameters, i.e., $\varrho' = \varrho$ and $\xi' = \xi$.
    for each actor network do
        Obtain the initial state $\mathbf{s}^0$.
        for each step $i \in [0, I]$ do
            Execute $\mathbf{a}^i = \left[\mathbf{W}^i, \boldsymbol{\Theta}^i\right]$ according to policy $\mu(\mathbf{s}^i, \mathbf{a}^i; \varrho)$ in current state $\mathbf{s}^i$.
            Receive reward $r^i$ and transmit to next state $\mathbf{s}^{i+1} = \left[h_{c,r,k}^{i+1}, h_{s,r,t}^{i+1}, R_k^{i+1}, \zeta_t^{i+1}, I_k^i\right]$.
            Update the actor network parameters $\varrho$.
        Update the critic network parameters $\xi$.
```

III. A3C-Based Beamforming and Reflection Design

In this section, we provide details of our proposed beamforming and reflection design algorithm for solving the non-convex optimization problem (19) based on an A3C RL framework. Here, we discuss the main elements of the A3C RL algorithm, and introduce the corresponding state space, action space, and reward function to solve (19) efficiently.

A. A3C Algorithm

DRL algorithms have recently shown great success in tackling complex optimization problems arising in wireless communication networks [16], [17], [18]. The actor-critic algorithm is a powerful DRL solution for continuous action space problems, where the actor decides which action to take, and the critic evaluates the action. The A3C algorithm enables the simultaneous operation of multiple actors, each training its neural network asynchronously. This asynchronous training notably accelerates the convergence speed. To exploit the above benefits, we here adopt an A3C RL approach, which consists of three main components: the state space, the action space, and the reward.

1) State Space: The agent will take actions by observing the state at each time step i which consists of five components: (i) $h^i_{c,r,k}$, which represents the instant channel between user k and BS if $r = 0$, and the instant channel between RIS and user k if $r = 1$; (ii) $h^i_{s,r,t}$, which indicates the instant channel between target t and BS if $r = 0$, and instant channel between RIS and target t if $r = 1$; (iii) the data rate of each user k, R^i_k; (iv) the sensing SNR of target t, ζ^i_t; and (v) the previous interference caused by sensing or communication to user k, I^{i-1}_k. The state is given by $\mathbf{s}^i = [h^i_{c,r,k}, h^i_{s,r,t}, R^i_k, \zeta^i_t, I^{i-1}_k]$, and the set of all states in all time steps is called state space $\mathcal{S} = \{\mathbf{s}^i\}_{i\in\mathcal{I}}$, where $\mathcal{I} = \{1, 2, \ldots I\}$ stands for the set of time steps.

2) Action Space: In each time step i, the agent takes action based on the observed state $\mathbf{s}^i$. The action $\mathbf{a}^i$ includes the design of beamforming matrix of the BS, and the phases of the RIS elements, i.e. $\mathbf{a}^i = [\mathbf{W}^i, \mathbf{\Theta}^i]$.

3) Immediate Reward: In RL algorithms, the agent learns how to act by receiving reward from the environment. In this case, in order to maximize the data rate of the users with some conditions on sensing SNR of the targets, we define the immediate reward at each time step i as

$$r^i = C_r \left(\prod_{t=1}^{T} \mathrm{U}\left(\zeta^i_t - \zeta_{th}\right)\right) \mathrm{U}\left(P_T - \sum_{j=1}^{K+M} \left\|[\mathbf{W}^i]_j\right\|^2\right) \left(\sum_{k=1}^{K} R^i_k\right) + C_p \left(\left(\prod_{t=1}^{T} \mathrm{U}\left(\zeta_{th} - \zeta^i_t\right)\right) + \mathrm{U}\left(\sum_{j=1}^{K+M} \left\|[\mathbf{W}^i]_j\right\|^2 - P_T\right)\right), \tag{20}$$

where ζ_{th} is the minimum acceptable SNR value for sensing, C_r is the reward coefficient, C_p is the punishment coefficient, and ζ^i_t is the resulting sensing SNR value for the t'th target at the i'th time step.

The goal of the agent is to maximize its cumulative reward by interacting with the environment, utilizing actions, states, and rewards. Specifically, at each time step i, the agent selects an action $\mathbf{a}^i$ based on the estimated value of the current state $\mathbf{s}^i$ through the state value function $V(\mathbf{s}^i; \boldsymbol{\xi})$, guided by the policy $\mu(\mathbf{s}^i, \mathbf{a}^i; \boldsymbol{\varrho})$. Additionally, $\boldsymbol{\xi}$ and $\boldsymbol{\varrho}$ represent the parameters of the actor network and critic networks, respectively. The agent transitions to the subsequent state $\mathbf{s}^{i+1}$ and receives the reward r^i. Notably, the A3C algorithm is applicable to both continuous and discrete spaces. The state value function can be mathematically described as

$$V\left(\mathbf{s}^i; \boldsymbol{\xi}\right) = \mathbb{E}\left[\Lambda^i \left|\mathbf{s}^i, \mu\right.\right]. \tag{21}$$

In this context, $\Lambda^i = \sum_{u=0}^{\infty} \delta^u r^{i+u}$ represents the anticipated cumulative reward discounted over time starting from state $\mathbf{s}^i$, with $\delta \in (0, 1]$ as the discount factor. The A3C algorithm utilizes the b-step reward for updating its parameters, which can be formulated as

$$\Upsilon^i = \sum_{u=0}^{b-1} \delta^u r^{i+u} + \delta^b V\left(\mathbf{s}^{i+b}; \boldsymbol{\xi}\right), \tag{22}$$

TABLE II
PARAMETER SETTINGS

Parameter	Value	Parameter	Value
Num. Antennas, M	16	$(\alpha, \beta_{min}, \phi)$	$(1.6, .2, .43\pi)$
Num. Users, K	4	ϵ_k	3×10^{-10}
Num. Targets, T	3	Noise Power @ Users/Sensing RX.	-80 dBm
Bandwidth, B	1MHz	BS-Target Pathloss, α_{Bt}	2.7
BS-RIS Dist.	30m	BS-User Pathloss, α_{Bu}	3.3
RIS-Target/User Dist.	3m	RIS-target/user and BS-RIS Pathloss, α_{BR}	2.3
BS-Target/User Dist.	20m	(C_r, C_p)	$(100, -50)$

where I is an upper limit for the index b. Furthermore, once I actions are executed or the final step is reached, the value and policy functions are updated. The advantage function A^i is introduced to reduce the estimation variance defined as

$$A^i\left(\mathbf{s}^i, \mathbf{a}^i; \boldsymbol{\varrho}, \boldsymbol{\xi}\right) = \Upsilon^i - V\left(\mathbf{s}^i; \boldsymbol{\xi}\right). \tag{23}$$

Utilizing the advantage function can enhance the agent's learning by ensuring that actions are neither underestimated nor overestimated, thereby improving decision-making. Details of the proposed algorithm are delineated in Algorithm 1.

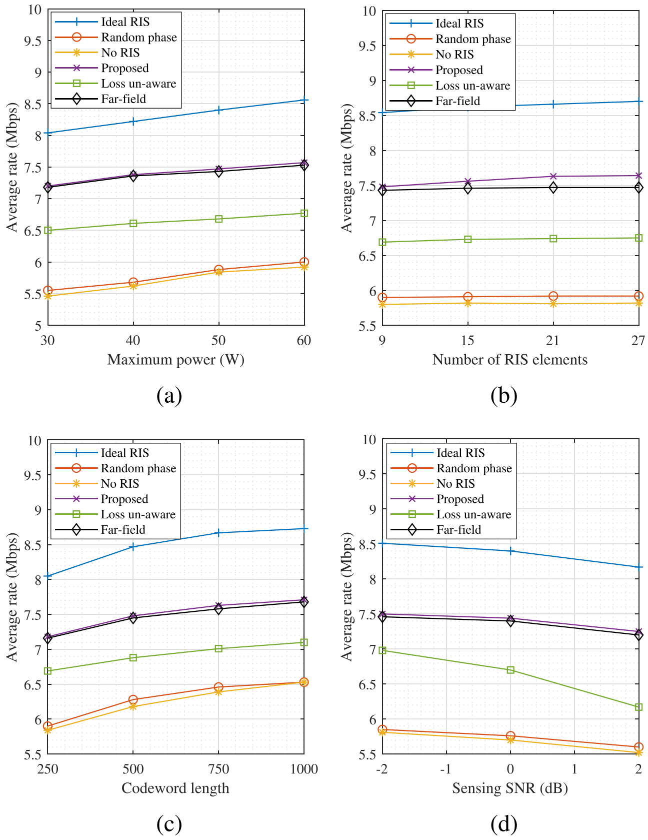
IV. SIMULATION RESULTS

In this section, we numerically evaluate the effectiveness of our proposed algorithm for beamforming and reflection design in the short packet ISAC systems assisted by the non-ideal RIS. We consider a BS equipped with 16 antennas operating at 60 GHz, that serves 4 single antenna users while simultaneously sensing 3 point-like targets, with the aid of a non-ideal RIS. We have considered 5 different RIS patterns with 5, 9, 15, 21, and 27 elements. We considered the RIS with these number of elements to be arranged in rectangular configurations, specifically: 1×5, 3×3, 3×5, 3×7, and 3×9. It is also assumed that vertical and azimuth angles of arrival are derived from a uniform distribution in the range $[\frac{-\pi}{2}, \frac{\pi}{2}]$.

The bandwidth B is 1 MHz, and the distances between the BS-target, BS-user, BS-RIS, RIS-user, and RIS-target are 20 m, 20 m, 30 m, 3 m, 3 m, respectively. We utilize the commonly used distance-dependent path-loss model and assume no dominant line-of-sight component in the channel. The parameter settings are summarized in Table II. We benchmark our proposed approach against the results obtained in the following 5 setups

1) *Ideal RIS:* This is the performance upperbound that would be achievable only with an ideal lossless RIS.
2) *Proposed:* This is the performance obtained with the non-ideal RIS using our proposed scheme.
3) *Loss un-aware:* This is the performance obtained with the non-ideal RIS, when the optimizer wrongly adopts the ideal lossless RIS assumption.
4) *Far-field assumption:* This is the performance achieved when the optimizer adopts the far-field model for calculating the rewards and states.
5) *Random Phase:* In this setup, the phase shifts of the non-ideal RIS are configured randomly.
6) *No RIS:* In this setup, there are no links through RIS, i.e. only a direct link exists for sensing/communication.

In developing our benchmarks, we recognize the importance of accounting for the non-ideal characteristics of RIS. Unlike the Ideal RIS benchmark, where rates and SNRs are calculated without considering



(a)

(b)

(c)

(d)

Fig. 2. Numerical comparisons with benchmarks.

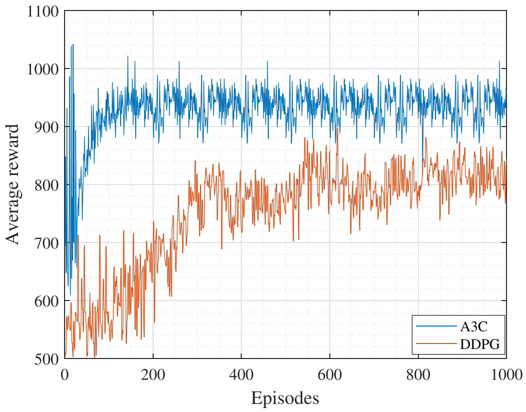


Fig. 3. Convergence performance.

these imperfections, our proposed algorithm incorporates a non-ideal model of RIS. This approach allows the optimizer to better understand the environment and enhance performance in terms of reward maximization. Specifically, we calculate both the states and rewards with reflection coefficients that incorporate a phase-dependent amplitude term, reflecting the non-ideal nature of the RIS. However, in the Ideal RIS benchmark, the states and rewards are calculated without accounting for the phase-dependent amplitude. In all scenarios, the actual data rates and SNRs are derived based on the non-ideal characteristics of the RIS. Although the constraints of non-ideal RIS are not explicitly included in the reward calculations, they are inherently accounted for in the reflection coefficients, ensuring that our optimization process operates under realistic conditions from the outset. Similarly, all the considerations discussed above also apply to the far-field assumption benchmark, where the optimizer does not incorporate the near-field model in its calculations.

In Fig. 2(a), we plot the average communication rate versus the maximum power budget. It is obvious that as the maximum power budget increases, the average rate also increases. Next, we plot the average communication rate in relation to the number of RIS elements in Fig. 2(b). The average communication rate increases for all benchmarks as the number of RIS elements rises, but the rate of increase diminishes with a larger number of elements. This observation is attributed to the fact that as the number of RIS elements increases, the distance between those farther from the center of the RIS also increases, amplifying the near-field effects and resulting in a higher level of uncertainty in the environment. Furthermore, in Fig. 2(c) we plot the average communication rate versus codeword length. One can observe that the codeword length has a significant impact on the average rate. However, as the codeword length increases, the impact becomes less pronounced. This is where the trade-off between communication average rate and communication delay becomes apparent. Finally, we plot the tradeoff between the average communication rate versus the sensing SNR requirement in Fig. 2(d). As the sensing SNR requirement increases, the average communication rate achieved by the considered ISAC system decreases, because more resources need to be allocated to sensing. In all the above figures, our proposed scheme aided by a non-ideal RIS outperforms the loss-unaware, far-field assumption, random phase, and No RIS benchmarks. Although the non-ideal RIS shows performance degradation in comparison with the ideal RIS upperbound, i.e. due to its lossy reflection amplitudes, by optimizing its phase shifts, we can improve the performance of both sensing and communication. In this case, optimizing the phase shifts without being aware of the lossy RIS model, i.e. assuming an ideal RIS when there are actual losses, will also cause performance degradation in communication and sensing, which is equivalently around 10% reduction of the average achievable communication rate considering different setups. In all cases, the RIS introduces additional NLoS links, enabling passive beamforming gains. Finally, the average reward curves versus the training episodes demonstrate a relatively quick convergence, within approximately 75 iterations. Fig. 3 illustrates the convergence behaviour underpinning our proposed beamforming strategy. Notably, the A3C beamforming scheme demonstrates superior performance compared to the conventional DDPG approach [19]. This enhancement is mainly due to the advanced adaptability provided by multiple actors of A3C. Quantitatively, it is evident that after 1000 episodes, around 33% gain and less fluctuation were achieved for A3C, compared to DDPG, which depicts the effectiveness of applying the A3C algorithm.

V. Conclusion

In this paper, we investigated the joint transmit beamforming and reflection design for an ISAC system assisted by a non-ideal RIS, considering a short packet communication regime. The practical non-ideal RIS was deployed to enhance both multi-target sensing and multi-user communication simultaneously in a realistic setup. The communication sum rate of the users was maximized under the sensing constraints of multi-target sensing SNR requirements, using an A3C RL-based algorithm. Simulation results showcased the benefits of deploying a non-ideal RIS in delay sensitive ISAC systems, while also unveiling the corresponding performance trade-offs.

References

[1] 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, Feb. 2023.

[2] 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.

[3] J. Jalali, A. Khalili, A. Rezaei, R. Berkvens, M. Weyn, and J. Famaey, "IRS-based energy efficiency and admission control maximization for IoT users with short packet lengths," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12379–12384, Sep. 2023.

[4] J. Singh, B. Naveen, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Pareto optimal hybrid beamforming for short-packet millimeter-wave integrated sensing and communication," Jun. 2024, *arXiv:2406.01945*.

[5] C. Pan et al., "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[6] H. Luo, R. Liu, M. Li, and Q. Liu, "RIS-aided integrated sensing and communication: Joint beamforming and reflection design," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9626–9630, Jul. 2023.

[7] Z. Yu et al., "Active RIS-aided ISAC systems: Beamforming design and performance analysis," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1578–1595, Mar. 2024.

[8] 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.*, vol. 16, no. 5, pp. 995–1010, Aug. 2022.

[9] X. Wang, Z. Fei, J. Huang, and H. Yu, "Joint waveform and discrete phase shift design for RIS-assisted integrated sensing and communication system under Cramer-Rao bound constraint," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1004–1009, Jan. 2022.

[10] H. Zhang, "Joint waveform and phase shift design for RIS-assisted integrated sensing and communication based on mutual information," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2317–2321, Oct. 2022.

[11] S. Aghashahi, Z. Zeinalpour-Yazdi, A. Tadaion, M. B. Mashhadi, and A. Elzanaty, "Single antenna tracking and localization of RIS-enabled vehicular users," *IEEE Trans. Veh. Technol.*, vol. 74, no. 3, pp. 4362–4375, Mar. 2025.

[12] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.

[13] Y. Pan, C. Pan, S. Jin, and J. Wang, "RIS-aided near-field localization and channel estimation for the terahertz system," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 4, pp. 878–892, Jul. 2023.

[14] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[15] W. Yu, L. Musavian, A. U. Quddus, Q. Ni, and P. Xiao, "Low latency driven effective capacity analysis for non-orthogonal and orthogonal spectrum access," in *Proc. 2018 IEEE Globecom Workshops*, 2018, pp. 1–6.

[16] N. C. Luong et al., "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, Fourthquarter 2019.

[17] 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.

[18] M. Azizi, F. Zeinali, M. R. Mili, and S. Shokrollahi, "Efficient AoI-aware resource management in VLC-V2X networks via multi-agent RL mechanism," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 14009–14014, Sep. 2024.

[19] F. Zeinali, S. Norouzi, N. Mokari, and E. A. Jorswieck, "AI-based radio resource and transmission opportunity allocation for 5G-V2X hetnets: NR and NR-U networks," *Int. J. Electron. Commun. Eng.*, vol. 17, no. 9, pp. 217–224, 2023.