Joint Beamforming Design and Sensing in Satellite and RIS-Enhanced Terrestrial Networks: A Federated Learning Approach

Sonia Pala, *Member, IEEE*, Keshav Singh, *Member, IEEE*, Chih-Peng Li, *Fellow, IEEE,* Octavia A. Dobre, *Fellow, IEEE*, and Trung Q. Duong, *Fellow, IEEE*

***Abstract*—This paper presents a novel analytical framework for minimizing transmit power in satellite and terrestrial integrated networks using reconfigurable intelligent surface (RIS) technology within integrated sensing and communication systems. We employ a cutting-edge federated deep reinforcement learning approach, utilizing a federated deep deterministic policy gradient (F-DDPG) algorithm, to tackle the complex non-convex power minimization problem effectively. The approach leverages federated learning to dynamically adapt to network changes, ensuring compliance with beamforming designs, multiple target signal-to-interference-plus-noise ratio thresholds, and RIS phase-shift requirements through an effective feedback loop. In particular, we propose an F-DDPG algorithm that outperforms existing benchmarks such as the federated deep Q-network (DQN), centralized DDPG, and conventional DDPG and DQN methods. Through simulations, we demonstrate that integrating RIS significantly lowers base station (BS) power requirements against both random configurations and non-RIS setups. The optimal RIS configuration with 60 elements achieves a 6.3% reduction in BS transmit power compared to the random RIS scenario and a 34.2% reduction compared to the no-RIS setup. Additionally, our results demonstrate that increasing the number of RIS elements markedly improves sensing capabilities while maintaining the same level of transmit power.**

***Index Terms*—Integrated sensing and communication (ISAC), satellite-terrestrial integrated network (STIN), reconfigurable intelligent surfaces (RIS), beamforming design, federated learning, deep reinforcement learning (DRL).**

I. Introduction

SENSING technologies are essential for advancing future wireless network functionalities, such as those projected for 6G, which enable applications including navigation, activity detection, movement tracking, and environmental monitoring [1], [2]. With the continuous advancements in wireless communications and radar sensing, the demand for scarce spectrum resources increasingly outpaces the supply, highlighting their critical scarcity and value. To mitigate these challenges, the integrated sensing and communication (ISAC) paradigm emerged as an effective strategy to efficiently utilize spectral, hardware, and energy resources. The ISAC achieves this by harmonizing signal processing techniques and leveraging a unified hardware infrastructure to concurrently support both sensing and communication functionalities. Despite its advantages, the majority of ISAC research remains focused on terrestrial scenarios, thereby restricting its ability to provide services on a global scale [3].

Satellite communication networks offer a viable solution to overcoming coverage limitations due to their wide coverage capability [4]. In earlier years, the deployment of large-scale satellite networks was hindered by substantial setup costs and a comparative deficiency in capacity against ground-based networks. Nevertheless, the growing demands for global communication services and breakthroughs in technology have recently made the concept of satellite constellations a focal point of interest in both the academic realm and the industry. This shift led to the initiation of various satellite constellation projects aimed at furnishing global coverage, highlighting projects such as Starlink, Telesat, and OneWeb [5]. Integrating satellite with terrestrial networks presents a viable strategy for achieving widespread broadband access, leveraging the extensive coverage of satellite systems alongside the high-capacity infrastructure of terrestrial networks [6]. Notably, the 3rd Generation Partnership Project (3GPP) has explored this synergistic approach in Releases 15, 16, and 17, examining the integration of terrestrial and non-terrestrial networks to extend network services to underserved areas, enhance service continuity, and optimize multicast/broadcast communications [7], [8]. According to the vision for future wireless networks outlined in the 6G White paper, the seamless integration between satellite and terrestrial networks is crucial, signaling a pivotal shift towards a more inclusive and versatile communication network infrastructure [9], [10].

The work of K. Singh and C.-P. Li was supported in part by the National Science and Technology Council of Taiwan under Grants NSTC 113-2218-E-110 -008, NSTC 112-2221-E-110-038-MY3, NSTC 112-2221-E-110-029-MY3 and NSTC 113-2222-E-110-008-MY3, and in part by the Sixth Generation Communication and Sensing Research Center funded by the Higher Education SPROUT Project, the Ministry of Education of Taiwan. The work of T. Q. Duong was supported by the Canada Excellence Research Chairs (CERC) Program CERC-2022-00109. The work of O. A. Dobre was supported by the Canada Research Chairs Program CRC-2022-00187. (*Corresponding author: K. Singh*.)

Sonia Pala, Keshav Singh, and Chih-Peng Li are with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan (Email: sony.pj12@gmail.com, keshav.singh@mail.nsysu.edu.tw, cpli@faculty.nsysu.edu.tw).

O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada (E-mail: odobre@mun.ca).

T. Q. Duong is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada, and is also with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K. (E-mail: tduong@mun.ca).

A. Related Works

In recent years, the concept of satellite-terrestrial integrated networks (STIN) has emerged as a focal point of research, driven by its potential to substantially enhance resource allocation through the integration of terrestrial and satellite networks [11], [12]. This innovative concept has piqued interest in both commercial and academic sectors, leading to a plethora of studies aimed at advancing the functionalities of STIN systems. Among these, significant strides have been made in areas such as resource management [13], [14], cooperative coordination [15], [16], and notably, beamforming techniques. In particular, beamforming stands out for its dual ability to improve signal quality for intended recipients while minimizing interference to others, thereby facilitating effective interference management and allowing the seamless coexistence or collaboration between satellite and terrestrial networks [17], [18]. In [19], researchers explored the challenge of developing a hybrid analog-digital beamforming design for spectrum-sharing between satellite and terrestrial systems, introducing an optimization scheme for analog-digital beamforming. The authors in [20] introduced a beamforming strategy aimed at enhancing the signal-to-interference-plus-noise ratio (SINR) for terrestrial users while minimizing interference to satellite users. Meanwhile, research presented in [21] developed two beamforming approaches to optimize cellular user data rates within the interference limitations of satellite users. These studies leveraged beamforming technology not only to improve overall system performance but also to bolster satellite network security by leveraging green interference from terrestrial networks. Furthermore, the insights from [21] were extended in [22] to propose a joint beamforming strategy designed to secure communications in scenarios involving multiple satellite users and potential eavesdroppers.

On the other hand, a significant body of research has explored the concept of ISAC within terrestrial network frameworks. The adoption of multiple-input multiple-output (MIMO) technology within terrestrial ISAC systems has been leveraged to notably improve the spectral and energy efficiencies of communication systems, a development thoroughly investigated in [23]. Moreover, the research outlined in [24]–[28] focused on the development of hybrid beamforming methods, employing diverse MIMO radar techniques to boost the performance of dual-functional radar communication systems. Recent advancements have also seen the adoption of orthogonal frequency division multiplexing (OFDM) for communications within terrestrial ISAC systems to effectively address inter-symbol interference issues, thereby facilitating enhanced target sensing capabilities as highlighted in [29]. However, terrestrial ISAC systems do not offer global coverage and are limited by inadequate data reception and processing capabilities. In response to these limitations, the idea of integrating ISAC with low Earth orbit (LEO) satellites has gained attention, driven by advancements in satellite onboard processing technologies. A strategy for hybrid beamforming in ISAC-LEO systems is detailed in [30], which also considers the impact of beam squint. Further, a novel ISAC-aided dynamic resource allocation strategy that enhances random access efficiency and system throughput within satellite-terrestrial relay networks (STRNs) is detailed in [31]. Even though ISAC leverages the larger bandwidths of millimeter wave (mmWave) frequencies for enhanced data rates and radar resolution, the higher frequencies introduce significant signal blockage issues, adversely affecting performance. To mitigate this, reconfigurable intelligent surfaces (RISs) can establish effective virtual connections between the ISAC base station (BS) or satellite and sensing targets, offering a promising strategy to overcome the challenge. RIS technology has garnered significant attention as a transformative approach for reconfiguring wireless environments by tuning the phase shifts of low-cost reflecting elements, thereby enhancing system performance in next-generation networks [32].

Recent studies have extensively explored the synergy between RIS and ISAC systems [33], [34], highlighting two predominant approaches in RIS-enhanced ISAC systems [35]. The first approach utilized RIS primarily to enhance communication capabilities while maintaining direct links from the transceiver to the target for sensing purposes [36], [37]. Specifically, the authors in [36] explored designing both transmit and receive beamforming strategies alongside RIS phase adjustments for multi-user settings. Conversely, [37] focused on reducing the transmit power of dual-function radar-communication (DFRC) BS by concurrently optimizing both active and passive beamforming in light of RIS-induced interference. Leveraging the advantages of RIS, the authors in [38] adopted deep reinforcement learning (DRL) to explore the integration of RIS within satellite networks, presenting promising solutions to latency, dynamic channel conditions, and energy constraints in 6G Internet of Things (IoT) environments. In [39], research focused on optimizing beamforming for RIS-enhanced hybrid satellite-terrestrial networks with blocked satellite and BS-user links. The concept of active RIS, incorporating amplifiers to mitigate the double path fading effect, has been explored in general communication systems [40] and ISAC frameworks [41], showing improved performance under optimized power budgets and element numbers. This paper, however, focuses on passive RIS in satellite-terrestrial systems, leaving active RIS integration as a potential direction for future work.

B. Motivation

The potential of ISAC in revolutionizing satellite and RIS-enhanced terrestrial networks is vast, yet a thorough examination of its full potential remains unexplored. Research carried out in [11]–[22] focused on the nuances of STINs without delving into the integration of ISAC or the innovative use of RIS. Further, while studies [24]–[27], [29] have investigated ISAC within terrestrial contexts, their scope does not extend to achieving global coverage or overcoming the data processing and reception challenges inherent to terrestrial networks. Efforts to incorporate ISAC within satellite frameworks, as carried out in [30], [31], have not considered the integration with RIS. Although the synergy between RIS and ISAC was examined in [33]–[37], such research remained limited to terrestrial implementations. However, the works in [38], [39] explored

the benefits of RIS in satellite and STINs, yet overlooked ISAC applications. Even though the authors in [37] made significant contributions to the power minimization problem, their research was confined to the terrestrial ISAC system and did not extend to STIN implementations. Given these gaps, the challenge arises of addressing the power minimization problem through joint beamforming design within ISAC systems that span both satellite and RIS-enhanced terrestrial networks. This underscores the complexity of such tasks, where traditional optimization methods may prove inadequate. To the best of the authors' knowledge, a comprehensive study leveraging federated learning for ISAC in satellite and RIS-enhanced terrestrial networks has not yet been significantly pursued in the existing literature.

C. Contribution

Motivated by the aforementioned discussion, we introduce a novel analytical framework to evaluate the performance of the ISAC-enabled satellite and RIS-enhanced terrestrial integrated network and address the power minimization problem through a federated learning approach. The key contributions are outlined as follows:

- We develop an intricate framework that seamlessly combines ISAC functionalities across both satellite and terrestrial domains, enhanced by RIS technology. This framework is meticulously designed to minimize the transmit power at the BS, while simultaneously guaranteeing the sensing performance through a minimum SINR requirement at the sensing targets and the users.
- We tackle the non-convex resource allocation problem through a federated DRL (F-DRL) approach. This method simplifies optimization by dynamically adapting to real-time changes and employing multi-agent reinforcement learning for effective resource management in both satellite and terrestrial segments. It enables agents to adjust satellite transmit power and meet SINR and RIS phase-shift requirements, thus boosting network performance through strategic feedback.
- The proposed federated deep deterministic policy gradient (F-DDPG) algorithm exhibits enhanced performance compared to benchmarks such as the federated deep Q-network (F-DQN), centralized DDPG (C-DDPG), and conventional DDPG and DQN methods. The integration of RIS within our framework substantially reduces BS power requirements compared to random configurations and without RIS. Additionally, our results demonstrate that increasing the number of RIS elements markedly improves sensing capabilities while maintaining the same level of transmit power.

In summary, the novelty of our work lies in the integrated satellite-terrestrial ISAC system design, the application of FMA-DRL to tackle non-convex optimization in STIN systems, the development of a federated learning framework addressing privacy and scalability challenges, and a comprehensive benchmarking study of federated DRL paradigms.

D. Structure of the Paper

Section II describes the system model, while Section III details the formulation of the optimization problem. In Section IV, we present the proposed federated multi-agent deep reinforcement learning (FMA-DRL) approach, structured within the Markov decision process (MDP) framework. Section V provides a discussion of the numerical simulation results. Finally, Section VI offers concluding remarks. The key notations used throughout this paper are summarized in Table I.

TABLE I: Key notations.

Notation	Definition	Notation	Definition
$(\mathbf{A})^{\mathrm{H}}$	Hermitian	$(\mathbf{A})^{\mathrm{T}}$	Transpose
$\mathrm{diag}(\cdot)$	Diagonalization operator	$\lvert\mathbf{A}\rvert$	Modulus operator
$\mathbb{C}^{M\times M}$	Complex matrix	$\mathbb{C}^{M\times 1}$	Complex vector
$\lvert\lvert\mathbf{A}\rvert\rvert$	Norm operator	$\mathbf{A}^*$	Optimal $\mathbf{A}$
$\circ$	Hadamard product	$\mathbb{E}\{\cdot\}$	Expectation

II. System Model

In the schematic shown in Fig. 1, we investigate a RIS-aided downlink (DL) ISAC system utilizing a STIN. The ISAC-geostationary orbit (GEO) satellite is equipped with N_t antennas for transmission and N_r antennas for reception, adopting a monostatic configuration. This system provides service to L satellite users (SUs), each equipped with a single antenna, denoted by the set $\mathcal{L} = \{1, ..., L\}$, and engages in the detection of several targets, represented by the set $\mathcal{T}_s = \{1, ..., T_s\}$. The design incorporates uniform linear arrays (ULAs) for all antenna configurations, with a consistent half-wavelength separation between each pair of adjacent antennas.

Expanding upon the previously described GEO satellite-based ISAC system, we also explore a terrestrial counterpart that incorporates a RIS-enhanced ISAC system[1]. In this terrestrial setup, a BS is equipped with two ULAs for enhanced communication and sensing capabilities. The system incorporates RIS with N passive elements designed to reflect the signal from BS towards K single antenna DL cellular users (CUs) indexed as $\mathcal{K} = \{1, ..., K\}$. This setup allows CUs to receive signals directly from the BS as well as via the RIS-reflected path, offering both direct and indirect link connectivity. Furthermore, the terrestrial BS utilizes a M_t-antenna ULA to cater to densely populated regions within the same frequency spectrum. The DL ISAC signal, transmitted by a ULA consisting of M_t elements, is designed for simultaneous communication with K single-antenna DL CUs, and for performing target detection on multiple targets, denoted by the set $\mathcal{T}_b = \{1, ..., T_b\}$, within the terrestrial domain. The target echo signal is received at the BS via the receiving ULA, which comprises M_r elements. This design ensures that the terrestrial base station optimally controls the RIS for cellular communication, while the satellite system effectively serves users and covers targets outside terrestrial BS

[1]While increased satellite beam directivity reduces interference to terrestrial zone users, our research is crucial for advanced interference management, enhancing ISAC capabilities, ensuring scalability, addressing real-world deployment factors, and improving user experience in integrated satellite-terrestrial systems.

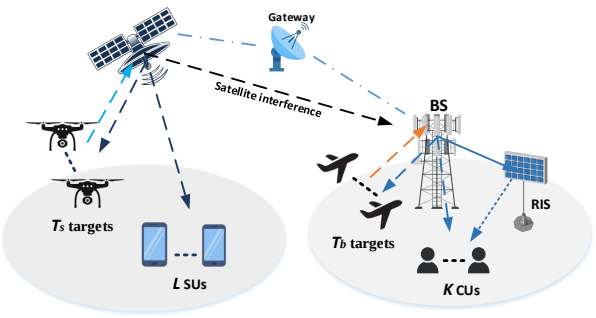

Fig. 1: Illustration of ISAC enabled satellite and RIS-enhanced terrestrial integrated network.

coverage, thereby enhancing the overall system efficiency and effectiveness. Dividing into two sub-ISAC systems optimizes performance and resource allocation for communication and sensing tasks in their respective zones.

A. Satellite Channel Model

The communication channel from the ISAC-GEO satellite to the l^{th} SU encompasses free space loss, the radiation pattern of the antenna, and rain attenuation, resulting in the modeling of the DL channel as [42]

$$\mathbf{g}_l = \mathbf{b}_l \circ \mathbf{q}_l \circ \exp\{j\boldsymbol{\varphi}_l\}, \tag{1}$$

where $\mathbf{b}_l = [b_{l,1}, b_{l,2}, ..., b_{l,N_t}] \in \mathbb{C}^{N_t}$ encapsulates both the radiation pattern of the satellite beam and the losses due to free space. The approximation for the n_t^{th} entry in b_k is as follows

$$b_{l,n_t} = \frac{\sqrt{F_g F_{l,n_t}}}{4\pi \frac{d_l}{\lambda}\sqrt{\kappa T_{sys} B_g}}, \tag{2}$$

where F_g denotes the gain of the antenna at the user's location, d_l represents the distance from the satellite to the l^{th} user, λ signifies the wavelength of the carrier signal, κ stands for Boltzmann's constant, T_{sys} indicates the temperature associated with receive noise, and B_g defines the bandwidth. The beam gain from the n_t^{th} feed to the l^{th} user, denoted as F_{l,n_t}, can be estimated by

$$F_{l,n_t} = F_{max}\left[\frac{J_1(u_{l,n_t})}{2u_{l,n_t}} + 36\frac{J_3(u_{l,n_t})}{u_{l,n_t}^3}\right]^2, \tag{3}$$

where F_{max} represents the maximum beam gain achievable for each beam, with u_{l,n_t} calculated as $2.07123\sin(\theta_{l,n_t})/\sin(\theta_{3\ dB})$. The term θ_{l,n_t} denotes the angle from the l^{th} user to the center of the n_t^{th} beam, while $(\theta_{3\ dB})$ is the angle at which the beam's gain drops by 3 dB compared to its center. J_1 and J_3 correspond to the first and third-order Bessel functions of the first kind, respectively. $q_l = [q_{l,1}, q_{l,2}, ..., q_{l,N_t}] \in \mathbb{C}^{N_t}$ describes the rain attenuation coefficients, with each component defined as $q_{l,n_t} = \xi_{l,n_t}^{1/2}$. The power gain ξ_{l,n_t}, expressed in dB as $\xi_{l,n_t}(\text{dB}) = 20\log_{10}(\xi_{l,n_t})$, typically adheres to a log-normal distribution, such that $\ln(\xi_{l,n_t}(\text{dB})) \sim \mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ. Furthermore, $\boldsymbol{\varphi}_l = [\varphi_{l,1}, \varphi_{l,2}, .., \varphi_{l,N_t}] \in \mathbb{C}^{N_t}$ is the phase vector with uniform distribution, where $\varphi_{l,n_t} \sim \mathcal{U}(0, 2\pi)$. $\mathbf{G} = [\mathbf{g}_1, ..., \mathbf{g}_L] \in \mathbb{C}^{N_t \times L}$ denotes the satellite channels connecting the satellite with all the SUs.

B. Radar and Satellite Communication Model

Initially, we focus on the DL transmission within the ISAC-satellite system. Here, a narrowband ISAC signal, represented as $\mathbf{x}^{sat} \in \mathbb{C}^{M_t \times 1}$, is transmitted to achieve both radar sensing and multiuser communication in the DL, utilizing multi-antenna beamforming. Given the presence of multiple targets, this integrated signal formulation is as follows:

$$\mathbf{x}^{sat} = \sum_{l=1}^{L} \mathbf{w}_l s_l^{sat} + \sum_{t_s=1}^{T_s} \mathbf{s}_{t_s}^{sat}, \tag{4}$$

where the beamformer vector $\mathbf{w}_l \in \mathbb{C}^{N_t \times 1}$ corresponds to the beamforming weights employed for the DL SU l, where $l \in \mathcal{L}$. The data symbol of user l, denoted as $s_l^{sat} \in \mathbb{C}$, has a unit power, indicating that its expected power is normalized to $\mathbb{E}\{|s_l^{sat}|^2\} = 1$. Additionally, $\mathbf{s}_{t_s}^{sat} \in \mathbb{C}^{N_t \times 1}$, signifies the unique radar signal aimed at target t_s, which is characterized by its covariance matrix $\mathbf{W}_{t_s} \triangleq \mathbb{E}\{\mathbf{s}_{t_s}^{sat}\mathbf{s}_{t_s}^{sat\ H}\}$. This definition is pivotal for enhancing the degrees of freedom (DoF) of the transmit signal $\mathbf{x}^{sat}$, facilitating superior sensing accuracy. The formulation of $\mathbf{W}_{t_s}$ allows for the synthesis of $\mathbf{s}_{t_s}$, as indicated in [43]. Furthermore, the system adheres to a total power constraint, expressed as $\sum_{l=1}^{L}\|\mathbf{w}_l\|^2 + \sum_{t_s=1}^{T_s}\text{Tr}(\mathbf{W}_{t_s}) \leq P_{\text{max}}^{sat}$, where P_{max}^{sat} signifies the peak power budget allocated for the satellite operations.

Subsequently, we formulate the echo signal for the multiple targets under consideration. We assume that the radar channel includes line of sight (LoS) components, with both the transmitting and receiving ULAs at the satellite spaced at half the wavelength. The steering vector for the transmit array towards direction ψ is defined as $\mathbf{c}_t(\psi) \triangleq 1/\sqrt{N_t}[1, e^{j\pi\sin(\psi)}, ..., e^{j\pi(N_t-1)\sin(\psi)}]^T$, and similarly, the receive steering vector is given by $\mathbf{c}_r(\psi) \triangleq 1/\sqrt{N_r}[1, e^{j\pi\sin(\psi)}, ..., e^{j\pi(N_r-1)\sin(\psi)}]^T$. Assuming that the target t_s being detected is positioned at an angle denoted as ψ_{t_s}, the reflection from the target can be represented as $\beta_{t_s}\mathbf{c}_r(\psi_{t_s})\mathbf{c}_t^H(\psi_{t_s})\mathbf{x}^{sat}$. Here, $\beta_{t_s} \in \mathbb{C}$ represents the complex amplitude of the target, which is primarily influenced by factors like path loss and radar cross section [44]. We make the assumption that ψ_{t_s} and β_{t_s} of the target are known or previously estimated at the satellite. This information is used to design an appropriate transmit signal that is best suited for detecting the specific target of interest [45]–[49]. With the given targets echo, the received signal at the satellite can be expressed as

$$\begin{aligned}\mathbf{y}_{t_s}^{sat} &= \underbrace{\mathbf{H}_{t_s}\mathbf{x}^{sat}}_{\text{Target reflection}} + \underbrace{\sum_{t'\neq t_s, t'=1}^{T_s}\mathbf{H}_{t'}\mathbf{x}^{sat}}_{\text{Echo signal of interferer targets}} + \tilde{\mathbf{n}}_{tar}, \forall t_s \\ &= \beta_{t_s}\mathbf{C}(\psi_{t_s})\mathbf{x}^{sat} + \sum_{t'\neq t_s, t'=1}^{T_s}\beta_{t'}\mathbf{C}(\psi_{t'})\mathbf{x}^{sat} + \tilde{\mathbf{n}}_{tar}, \forall t_s,\end{aligned} \tag{5}$$

where $\mathbf{H}_{t_s}(t) = \beta_{t_s}\mathbf{c}_r(\psi_{t_s})\mathbf{c}_t^H(\psi_{t_s})$ is the radar channel and $\mathbf{C}(\psi_{t_s}) \triangleq \mathbf{c}_r(\psi_{t_s})\mathbf{c}_t^H(\psi_{t_s})$. The term $\tilde{\mathbf{n}}_{tar} \in \mathbb{C}^{N_r \times 1}$ indicates the additive white Gaussian noise (AWGN) with covariance $\tilde{\sigma}_{tar}^2\mathbf{I}_{N_r}$.

In practice, a receive beamformer denoted as $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, .., \tilde{\mathbf{u}}_{t_s}] \in \mathbb{C}^{N_r \times t_s}$, is employed to capture the desired reflected signal of the target t_s from the received signal $\mathbf{y}_{t_s}^{sat}$. Subsequently, using this information, the SINR of the target t_s can be expressed as

$$\gamma_{t_s}^{\text{tar}} = \frac{\mathbb{E}\{|\tilde{\mathbf{u}}_{t_s}^H \beta_{t_s}\mathbf{C}(\psi_{t_s})\mathbf{x}^{sat}|^2\}}{\mathbb{E}\{|\tilde{\mathbf{u}}_{t_s}^H\mathbf{D}\mathbf{x}^{sat}|^2\} + \mathbb{E}\{|\tilde{\mathbf{u}}_{t_s}^H\tilde{\mathbf{n}}_{tar}|^2\}}, \forall t_s, \tag{6}$$

where $\mathbf{D} = \sum_{t' \neq t_s} \beta_{t'}\mathbf{C}(\psi_{t'})$. Alternatively, we represent the communication channel between the l^{th} DL SU and the satellite as $\mathbf{g}_l \in \mathbb{C}^{1 \times N_t}$. Further, we assume that the SUs are positioned outside the service area of the BS, which is designated as a terrestrial zone. This placement ensures that the SUs may encounter mutual interference due to remaining SUs in the zone but remain unaffected by any interference originating from the BS [50]. Consequently, the received signal at the l^{th} DL SU is given as

$$y_l^{sat} = \underbrace{\mathbf{g}_l\mathbf{w}_l s_l^{sat}}_{\text{Desired signal}} + \underbrace{\sum_{l'=1, l' \neq l}^{L} \mathbf{g}_l\mathbf{w}_{l'}s_{l'}^{sat}}_{\text{Multi SU interference}} + \underbrace{\sum_{t_s=1}^{T_s} \mathbf{g}_l\mathbf{s}_{t_s}^{sat}}_{\text{Interfering sensing signal}} + n_l^{sat}, \forall l, \tag{7}$$

where n_l^{sat} indicates the AWGN with ${\sigma_l^{sat}}^2$ variance. Based on this, the SINR of the DL l^{th} SU can be formulated by referring to (7) as

$$\gamma_l^{sat} = \frac{|\mathbf{g}_l\mathbf{w}_l|^2}{\sum_{l'=1, l' \neq l}^{L} |\mathbf{g}_l\mathbf{w}_{l'}|^2 + \sum_{t_s=1}^{T_s} \mathbf{g}_l\mathbf{W}_{t_s}\mathbf{g}_l^H + {\sigma_l^{sat}}^2}, \forall l. \tag{8}$$

C. Radar and Terrestrial Communication Model

Initially, the DL transmission involves sending a narrowband ISAC signal, $\mathbf{x}^{bs} \in \mathbb{C}^{M_t \times 1}$, tailored for simultaneous target detection and communication with multiple DL users through multi-antenna beamforming. In this context, considering the presence of multiple sensing targets, the combined signal can be formulated as

$$\mathbf{x}^{bs} = \sum_{k=1}^{K} \mathbf{v}_k s_k + \sum_{t_b=1}^{T_b} \mathbf{s}_{t_b}, \tag{9}$$

where $\mathbf{v}_k \in \mathbb{C}^{M_t \times 1}$ denotes the beamforming vector for the DL CU k, $k \in \mathcal{K}$, and $s_k \in \mathbb{C}$ is the transmitted data symbol for user k, assumed to have unit power, i.e., $\mathbb{E}\{|s_k|^2\} = 1$. Additionally, $\mathbf{s}_{t_b} \in \mathbb{C}^{M_t \times 1}$ is the radar-specific signal intended for target t_b, characterized by a covariance matrix $\mathbf{V}_{t_b} \triangleq \mathbb{E}\{\mathbf{s}_{t_b}\mathbf{s}_{t_b}^H\}$, designed to expand the DoF of the transmitted signal $\mathbf{x}^{bs}$ for improved radar sensing. The generation of $\mathbf{s}_{t_b}$ follows upon determining $\mathbf{V}_{t_b}$.

In modeling the echo signal from the radar system, we consider that the channel of the sensing targets includes direct LoS components. The array steering vector for transmission towards direction ϑ is represented by $\mathbf{a}_t(\vartheta) \triangleq 1/\sqrt{M_t}[1, e^{j\pi\sin(\vartheta)}, \ldots, e^{j\pi(M_t-1)\sin(\vartheta)}]^T$, and the steering vector for reception in the direction ϑ is $\mathbf{a}_r(\vartheta) \triangleq 1/\sqrt{M_r}[1, e^{j\pi\sin(\vartheta)}, ..., e^{j\pi(M_r-1)\sin(\vartheta)}]^T$. When considering the detection of a target t_b, located at the angle ϑ_{t_b}, the corresponding target reflected signal can be represented as $\alpha_{t_b}\mathbf{a}_r(\vartheta_{t_b})\mathbf{a}_t^H(\vartheta_{t_b})\mathbf{x}^{bs}$, where $\alpha_{t_b} \in \mathbb{C}$ signifies the target's complex amplitude, which is influenced by path loss and radar cross-section characteristics [44]. We assume that the BS has prior knowledge or estimations of ϑ_{t_b} and α_{t_b} for the detection of the particular target of interest [45]–[49]. Consequently, the received signal at the BS by taking into account the reflected echoes from the targets can be formulated as

$$\begin{aligned}\mathbf{y}_{t_b}^{bs} &= \underbrace{\bar{\mathbf{H}}_{t_b}\mathbf{x}^{bs}}_{\text{Target reflection}} + \underbrace{\sum_{t' \neq t_b, t'=1}^{T_b} \bar{\mathbf{H}}_{t'}\mathbf{x}^{bs}}_{\text{Echo signal of interferer targets}} + \mathbf{n}_{rad} \\ &= \alpha_{t_b}\mathbf{A}(\vartheta_{t_b})\mathbf{x}^{bs} + \sum_{t' \neq t_b, t'=1}^{T_b} \alpha_{t'}\mathbf{A}(\vartheta_{t'})\mathbf{x}^{bs} + \mathbf{n}_{rad}, \forall t_b,\end{aligned} \tag{10}$$

where $\bar{\mathbf{H}}_{t_b}(t) = \alpha_{t_b}\mathbf{a}_r(\vartheta_{t_b})\mathbf{a}_t^H(\vartheta_{t_b})$ is the radar channel and $\mathbf{A}(\vartheta_{t_b}) \triangleq \mathbf{a}_r(\psi_{t_b})\mathbf{a}_t^H(\psi_{t_b})$. The term $\mathbf{n}_{rad} \in \mathbb{C}^{M_r \times 1}$ indicates the AWGN with covariance $\sigma_{rad}^2\mathbf{I}_{M_r}$.

In practice, a receive beamforming matrix represented as $\mathbf{U} = [\mathbf{u}_1, .., \mathbf{u}_{t_b}] \in \mathbb{C}^{M_r \times t_b}$, is utilized to capture the desired reflected signal from target t_b within the received signal $\mathbf{y}_{t_b}^{bs}$. Following this, the SINR for target t_b is expressed as

$$\gamma_{t_b}^{\text{rad}} = \frac{\mathbb{E}\{|\mathbf{u}_{t_b}^H\alpha_{t_b}\mathbf{A}(\vartheta_{t_b})\mathbf{x}^{bs}|^2\}}{\mathbb{E}\{|\mathbf{u}_{t_b}^H\mathbf{B}\mathbf{x}^{bs}|^2\} + \mathbb{E}\{|\mathbf{u}_{t_b}^H\mathbf{n}_{rad}|^2\}}, \forall t_b, \tag{11}$$

where $\mathbf{B} = \sum_{t' \neq t_b} \alpha_{t'}\mathbf{A}(\vartheta_{t'})$. Further, we assume that each CU encounters mutual interference due to other CUs in the zone and from satellite interference[2] [50]. The interference channel $\mathbf{Z} = [\mathbf{z}_1; \ldots; \mathbf{z}_L] \in \mathbb{C}^{L \times N_t}$ represents the channel gain from the satellite to all CUs. Each element $\mathbf{z}_l \in \mathbb{C}^{1 \times N_t}$ captures the interference effect from the satellite's N_t antennas to the l^{th} CU. Hence, the signal received by the k^{th} CU is described as follows:

$$y_k^{bs} = \underbrace{\mathbf{h}_k\mathbf{v}_k s_k}_{\text{Desired signal}} + \underbrace{\sum_{k'=1, k' \neq k}^{K} \mathbf{h}_k\mathbf{v}_{k'}s_{k'}}_{\text{Multi-CU interference}} + \underbrace{\sum_{t_b=1}^{T_b} \mathbf{h}_k\mathbf{s}_{t_b}}_{\text{Sensing signal}} + \underbrace{\sum_{l=1}^{L} \mathbf{z}_l\mathbf{w}_l s_l^{sat}}_{\text{Satellite interfering channels}} + n_k, \forall k, \tag{12}$$

where n_k indicates the AWGN with variance ${\sigma_k}^2$. The effective communication channel gain from the BS to the k^{th} DL CU is denoted by $\mathbf{h}_k \in \mathbb{C}^{1 \times M_t}$, where $\mathbf{h}_k = \mathbf{h}_{b,k} +$

[2]Satellite communication interferes with ground communication due to higher transmission power and clear LoS propagation, while ground communication does not interfere with satellite communication because terrestrial signals are attenuated by obstacles and spatial separation [50].

$\mathbf{h}_{r,k}\mathbf{\Phi}\mathbf{H}_{b,r}$. Specifically, $\mathbf{h}_{b,k} \in \mathbb{C}^{1\times M_t}$ signifies the the direct BS-to-k^{th} CU link, $\mathbf{h}_{r,k} \in \mathbb{C}^{1\times N}, \forall k \in \mathcal{K}$ denotes the channel gain between RIS and the k^{th} CU link, and $\mathbf{H}_{b,r} \in \mathbb{C}^{N\times M_t}$ illustrates the channel gain between the BS and the RIS link. Further, we design the RIS phase shift matrix as $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$. It is characterized by the phase-shift vector $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_N]$, delineating the phase changes for each of the N elements in RIS. The phase shift matrix is structured as $\mathbf{\Phi} = \mathrm{diag}(\boldsymbol{\phi})$, where ϕ_n pertaining to the phase shift of the n^{th} element within the interval $[0, 2\pi]$. Based on this, the SINR of the DL k^{th} CU can be formulated by referring to (12) as

$$\gamma_k^{bs} = \frac{|\mathbf{h}_k\mathbf{v}_k|^2}{\sum\limits_{k'=1,k'\neq k}^{K}|\mathbf{h}_k\mathbf{v}_{k'}|^2+\sum\limits_{t_b=1}^{T_b}\mathbf{h}_k^H\mathbf{V}_{t_b}\mathbf{h}_k+\sum\limits_{l=1}^{L}|\mathbf{z}_l\mathbf{w}_l|^2+\sigma_k^2}, \forall k. \tag{13}$$

III. Problem Formulation

In this section, we aim to devise a joint beamforming approach to minimize the total transmit power consumption at the terrestrial BS in the RIS-aided ISAC STIN. This objective entails the optimization of the transmit beamforming vectors $\{\mathbf{w}_l\}_{l=1}^{L}$ and $\{\mathbf{v}_k\}_{k=1}^{K}$, along with $\{\mathbf{W}_{t_s}\}_{t_s=1}^{T_s}$ at the satellite and $\{\mathbf{V}_{t_b}\}_{t_b=1}^{T_b}$ at the BS. Additionally, the optimization extends to the phase shift matrix $\mathbf{\Phi}$ at the RIS within the terrestrial zone and the receive beamformers, represented by $\tilde{\mathbf{U}}$ and $\mathbf{U}$, with the goal of minimizing the transmit power consumption at the BS. Let $\mathcal{J}$ define the set of optimization variables represented by $\mathcal{J} \triangleq \{\{\mathbf{w}_l\}_{l=1}^{L}, \{\mathbf{v}_k\}_{k=1}^{K}, \{\mathbf{W}_{t_s}\}_{t_s=1}^{T_s} \succeq 0, \{\mathbf{V}_{t_b}\}_{t_b=1}^{T_b} \succeq 0, \mathbf{U}, \tilde{\mathbf{U}}, \mathbf{\Phi}\}$. Based on these criteria, the optimization problem is formulated as

$$\underset{\mathcal{J}}{\text{minimize}}\ P = \sum_{k=1}^{K}\|\mathbf{v}_k\|^2 + \sum_{t_b=1}^{T_b}\mathrm{Tr}(\mathbf{V}_{t_b}) \tag{14a}$$

$$\text{s.t} \quad \gamma_{t_s}^{\text{tar}} \geq \tau_{t_s}^{\text{tar}},\ \forall\ t_s \in \mathcal{T}_s, \tag{14b}$$

$$\gamma_{t_b}^{\text{rad}} \geq \tau_{t_b}^{\text{rad}},\ \forall\ t_b \in \mathcal{T}_b, \tag{14c}$$

$$\gamma_l^{sat} \geq \tau_l^{sat},\ \forall\ l \in \mathcal{L}, \tag{14d}$$

$$\gamma_k^{bs} \geq \tau_k^{bs},\ \forall\ k \in \mathcal{K}, \tag{14e}$$

$$\sum_{l=1}^{L}\|\mathbf{w}_l\|^2 + \sum_{t_s=1}^{T_s}\mathrm{Tr}(\mathbf{W}_{t_s}) \leq P_{\max}^{sat}, \tag{14f}$$

$$|\phi_n| = 1, \forall n \in \mathcal{N}, \tag{14g}$$

where the constraints (14b) and (14c) ensure the SINR at the targets meets a minimum threshold, critical for maintaining the quality of sensing operations within the STIN framework. Constraints (14d) and (14e) specify the minimum SINR requirements for the l^{th} SU and k^{th} CU, respectively. (14f) limits the maximum allowed transmit power for the satellite. Additionally, constraint (14g) is the unit modulus constraint of the phase-shift matrix at the RIS. The coupling relationship between the satellite-based and terrestrial RIS-enhanced ISAC systems is crucial for optimal performance, characterized by interference management and resource allocation strategies that ensure seamless integration and coordination.

It is important to note that the resource allocation problem defined in (14) is characterized by its non-convexity [51], making the search for a globally optimal solution difficult with standard polynomial-time algorithms. Furthermore, the coupling of optimization variables adds another layer of complexity, complicating the resolution process and rendering the problem challenging to address. The challenges in addressing the problem extend beyond its non-convex nature. Specifically, the necessity for computationally demanding tasks, such as matrix inversions and singular value decompositions in iterative approaches, complicates their implementation in real-time scenarios. To tackle these issues, we present a solution in the form of a FMA-DRL algorithm, aimed at overcoming the inherent difficulties associated with non-convex optimization. The choice of DRL is motivated by its scalability, ability to handle high-dimensional optimization, and superior empirical performance compared to conventional methods.

To effectively tackle the optimization challenge, we propose a dynamically adaptive strategy suitable for real-time implementation, ensuring optimal performance with only partial environment observations. This strategy redefines the problem within the context of MA-DRL[3], assigning independent agents to both non-terrestrial and terrestrial zones. Within this framework, each agent, represented by BS or a satellite module, navigates its operational parameters within each time slot. This feedback loop enables the dynamic adjustment of transmit power and other operational variables to maintain compliance with the SINR requirements and RIS phase-shift design.

IV. The Proposed Federated Multi-Agent Deep Reinforcement Learning Algorithm

In this approach, we reinterpret the previously mentioned problem within the framework of DRL, with the BS and the satellite functioning as agents. The objective of the satellite agent is to enhance the overall user throughput while reducing interference. In the DRL scenario, base stations are required to develop deep neural network (DNN) models that output either Q-values or direct control measures. A pivotal challenge is the expedited training of these DNN models to align with the dynamic nature of network conditions. Federated learning[4] enhances our system by addressing privacy concerns, enabling localized training, lever aging shared model updates, and optimizing the global objective function across heterogeneous network environments, thus offering significant advantages over independent DRL approaches [53]. To address this, we suggest the adoption of an FMA-DRL strategy. This strategy allows for the collaborative improvement of a predictive model through the mutual exchange of DRL model weights among federated agents. In our framework, dedicated agents are assigned to both the satellite zone and the terrestrial base station area, thereby achieving swift adaptation to network changes while safeguarding user data privacy.

[3]The constraints of problem 14 are strictly enforced through reward function penalties and action space restrictions, ensuring that the proposed solution complies with the problem requirements.

[4]While the proposed framework offers significant potential for improving federated learning in real-world applications, there are several deployment challenges to consider. These include device heterogeneity, communication overhead, data privacy, resource constraints, and scalability [52].

In summary, DRL's ability to effectively handle non-convex optimization problems, coupled with its computational efficiency, flexibility, adaptability, and suitability for real-time implementation, makes it a superior choice over traditional convex optimization methods for our system. In subsequent sections, we initially outline the RL issue by detailing the state space, action space, and reward function relevant to the problem (14). Following this, we introduce two variants of FMA-DRL: the F-DDPG and the F-DQN, both aimed at addressing the challenge of the formulated problem.

A. MDP Formulation

DRL methodologies are typically framed within the structure of MDP, which are characterized by a 5-element tuple: $\{\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \bar{\gamma}\}$. Here, $\mathcal{S}$ denotes the set of all possible states, $\mathcal{A}$ represents the set of all possible actions, $\mathcal{P}$ signifies the probabilities of moving from one state to another state given an action, expressed as $Pr(s^{t+1}|s^t, a^t)$, $r^t(s^t, a^t) \in \mathcal{R}$ is the function that assigns a reward based on the state and action at time t, and $\bar{\gamma} \in [0, 1)$ is the discount factor which adjusts the value of future rewards. At any given time step t, given the current state s^t, the agent selects an action $a^t \in \mathcal{A}$, which leads to a transition to a new state s^{t+1} with a transition probability of $Pr(s^{t+1}|s^t, a^t) \in \mathcal{P}$, and the agent receives a reward r^t. The decision-making strategy of the agent, known as a policy $\pi(s, a)$, defines the likelihood of choosing action $a^t = a$ when in state $s^t = s$, or formally, $\pi(s, a) = Pr(a^t = a|s^t = s)$. In particular, agent, $\mathcal{S}, \mathcal{A}$, and $\mathcal{R}$ are designed as follows:

1) Agent: In our framework, we allot distinct agents to handle operations in the satellite zone as well as the terrestrial zone.

2) State & Observation Space: The aim is to integrate a comprehensive array of environmental information pertaining to the problem within the state space. Denote $\mathcal{S} = \{\mathcal{S}_{sat}, \mathcal{S}_{bs}\}$ as the state space of the system, which includes the overall channel conditions and the behavior of all the agents involved. This state space is devised from the information available to the BS and satellite, whether obtained directly or through inference, and it is pivotal in formulating the reward function. The observation state feature set at ISAC-satellite encompasses all channel data within the satellite area, and the previous beamformers and transmission power [54]–[56]. Therefore, the designated observation state space for the ISAC-satellite is described as

$$\mathcal{S}_{sat} = \{\mathbf{g}_l^t, \mathbf{H}_{t_s}^t, \{\mathbf{w}_l\}^{t-1}, \{\mathbf{W}_{t_s}\}^{t-1}, \tilde{\mathbf{U}}^{t-1}, -P^{t-1}\}. \quad (15)$$

Meanwhile, the observation state space at the BS includes the current state feature set, which comprises all channel information within the terrestrial area, the phase shift matrix at the RIS, and the previous beamformers and transmission power [54], [55]. Therefore, the observation state space feature set at the BS is defined as

$$\mathcal{S}_{bs} = \{\mathbf{h}_k^t, \bar{\mathbf{H}}_{t_b}^t, \mathbf{\Phi}^{t-1}, \{\mathbf{v}_k\}^{t-1}, \{\mathbf{V}_{t_b}\}^{t-1}, \{\mathbf{w}_l\}^{t-1}, \mathbf{U}^{t-1}, -P^{t-1}\}. \quad (16)$$

The observations collected from the agents are saved in a centralized buffer, where each agent retrieves information through its unique control channel. In the training process, updates to the neural network are carried out offline through a random selection of observations from this repository. Following this, agents utilize archived observations to shape their decision-making for forthcoming actions. This process facilitates efficient adaptation and learning in a constantly changing network setting, steadily advancing the decision-making skills of the agents.

3) Action Space: An action results from the policy outputs (either from DQN or the actor-network in actor-critic schemes). Let the action space for the system be denoted as $\mathcal{A} = \{\mathcal{A}_{sat}, \mathcal{A}_{bs}\}$, designed by integrating a policy incorporating both the comprehensive beamforming matrices and the phase-shift matrix at the RIS [54]–[56]. With multiple agents involved, the action space is designed to include the individual actions of each agent. Specifically, the subset of the action space $\mathcal{A}_{sat} \in \mathcal{A}$ that pertains to the satellite area is defined as

$$\mathcal{A}_{sat} = \{\{\mathbf{w}_l\}^t, \{\mathbf{W}_{t_s}\}^t, \tilde{\mathbf{U}}^t\}. \quad (17)$$

Likewise, the subset of the action space $\mathcal{A}_{bs} \in \mathcal{A}$ associated with the terrestrial area is defined as

$$\mathcal{A}_{bs} = \{\{\mathbf{v}_k\}^t, \{\mathbf{V}_{t_b}\}^t, \mathbf{U}^t, \mathbf{\Phi}^t\}. \quad (18)$$

4) Reward Function: The reward function in our study is designed to calculate the instant reward received when an action is performed at state s^t, targeting the objective in (14) by optimizing the action selection comprehensively. In the realm of DRL, the objective is for the agent to identify actions that lead to the maximization of aggregate rewards over time through discrete-time interactions with the environment. For this purpose, we assign to each agent a reward $r_{\mathcal{F}}^t$, where $\mathcal{F} \in \{sat, bs\}$. To elaborate, "$sat$" refers to the satellite zone, and "bs" signifies the terrestrial BS zone. The reward value, r^t, at learning time step t is defined as the negative of the total transmit power, specifically $r^t = -P^t$.

B. Federated Learning Model

In this section, we discuss the utilization of DRL algorithms that employ DNNs to determine action probabilities for optimizing returns. The fundamental concept of federated learning is to develop a unified statistical model (here, a DNN) using data collected across numerous devices. Our approach ensures each agent processes and keeps its state data locally, sending periodic updates to the gateway. As the central server for federated learning, the gateway facilitates seamless communication, optimizes resource allocation, and ensures efficient coordination and data aggregation. The objective of this training methodology is to optimize a predefined objective function by minimizing its value which is given by

$$\min_{\varrho} F(\varrho) = \sum\nolimits_{j \in \mathcal{F}} \kappa^j F_j(\varrho^j). \quad (19)$$

The global model is deployed at a central server located in the network gateway, which connects the terrestrial and satellite zones. For optimizing the global model, we aim to fine-tune the global objective function, $F(\varrho)$, alongside its weights, ϱ, and adjust the local loss function, F_j, with their respective weights, ϱ^j, for each agent across different zones j. The share

of each zone, j, in the overarching network is determined by κ^j, defined as $\kappa^j = \frac{\mathcal{U}_j}{\sum_{j\in\mathcal{F}}\mathcal{U}_j}$, where $\mathcal{U}_j$ denotes the number of users within zone j of the network set $\mathcal{F}$. Specifically, for the satellite zone, represented as $j = sat$, $\mathcal{U}_j = L$, whereas for the terrestrial zone, denoted as $j = bs$, $\mathcal{U}_j = K$.

C. Federated Multi-Agent Deep Deterministic Policy Gradient

DDPG enhances the actor-critic framework by incorporating DNNs to model policy and value functions. This approach provides a more sophisticated solution to the challenges of handling extensive state and action spaces, characteristic of high-dimensional scenarios. DDPG stands out for its capability to handle continuous action spaces, making it adept at decision-making in such environments. Essentially, the proposed DDPG employs two key DNN components in its architecture:

1) Critic Network: The model, also known as a Q-network and characterized by the parameter ϱ_c, processes an input comprising a state s and an action a as inputs, subsequently yielding the Q-value, $Q(s^t, a^t; \varrho_c)$. Further, the action-value function, often referred to as the Q-function is defined as

$$Q_\pi\left(s^t, a^t\right) = \mathbb{E}_\pi\left[R^t | s^t = s, a^t = a\right]. \tag{20}$$

This function can be updated using the Bellman expectation equation [57]. Furthermore, the optimal Q-function in (20) can be determined through the Bellman optimality equation, expressed as

$$Q^*\left(s^t, a^t\right) = r^t + \bar{\gamma} \max_{a^{t+1}\in\mathcal{A}} Q^*\left(s^{t+1}, a^{t+1}\right). \tag{21}$$

Accordingly, the optimal action a^* is derived by

$$a^* = \arg\max_{a\in\mathcal{A}} Q^*\left(s, a\right). \tag{22}$$

2) Actor Network: This is also referred to as a policy network, which accepts a state s as input and produces a continuous action a, denoted by $a^t = \pi(s^t; \varrho_\mu)$ while updating the network parameter ϱ_μ. The actor-network undergoes training through (22), intending to optimize the state-value function.

Besides, DDPG employs both a target actor-network, denoted as $\pi(s^t; \varrho'_\mu)$, and a target critic network, represented by $Q(s^t, a^t; \varrho'_c)$, to enhance training stability. Here, ϱ'_μ and ϱ'_c signify the parameters for the target actor and critic networks, respectively. Further, the actor undergoes training to enhance the objective function by employing a policy gradient method

$$\nabla_{\varrho_\mu} J(\varrho_\mu) \approx \mathbb{E}[\nabla_a Q(s^t, a; \varrho_c)|_{a=\pi(s^t;\varrho_\mu)} \nabla_{\varrho_\mu} \pi(s^t; \varrho_\mu)]. \tag{23}$$

Here, $J(\varrho_\mu) = \mathbb{E}_{s\sim\varrho_c, a\sim\varrho_\mu} R(s, a)$ typically represents the expected cumulative return. Meanwhile, the critic undergoes iterative optimization aimed at reducing the loss function, which is characterized as

$$L(\varrho_c) = \mathbb{E}\left[(y^t - Q(s^t, a^t; \varrho_c))^2\right], \tag{24}$$

where $y^t = R^t + \bar{\gamma} Q(s^{t+1}, \pi(s^{t+1}; \varrho'_\mu); \varrho'_c)$ denotes the expected return. The stability of y^t throughout the training is ensured by incrementally adjusting the parameters of the target networks with a minor coefficient, $\zeta \in [0, 1]$, thus updating as $\varrho'_\mu = \zeta\varrho_\mu + (1-\zeta)\varrho'_\mu$ and $\varrho'_c = \zeta\varrho_c + (1-\zeta)\varrho'_c$.

Algorithm 1 Federated DDPG Algorithm for Each Agent

Input: Initialize the parameter settings for the proposed system model, neural networks at $t = 0$
Input: The aggregation frequency ρ, exploration parameter ϵ, learning rate Ω, number of episodes E
Initialize the actor-network, $\pi(s^t; \varrho_\mu)$ and the critic network $Q(s^t, a^t; \varrho_c)$ with the weights ϱ_μ and ϱ_c.
Create the target DNNs by setting $\varrho'_\mu \leftarrow \varrho_\mu$ and $\varrho'_c \leftarrow \varrho_c$
Initialize a replay buffer
Initialization: get initial ϱ_μ from server
for $ep = 1 \rightarrow E$ **do**
    Initialize a random process η for action exploration
    Receive initial observation state s^1
    for $t = 1 \rightarrow T$ **do**
        Obtain action a^t from the actor-network;
        Add exploration noise to a^t as $a^t = a^t + \eta$
        Calculate the instant reward r^t
        Observe the new state s^{t+1}
        Store experiences in the buffer and sample random mini-batches of experiences to train the DNNs
        Set the expected return y^t
        Update the actor policy via (23) and critic via (24)
        Update the target actor ϱ'_μ and the target critic ϱ'_c
    end for
end for
update $\varrho_\mu^{ep+1} = \varrho_\mu^{ep} + \Omega\nabla_{\varrho_\mu} J(\varrho_\mu)$
if ep mod $\rho = 0$ **then**
    send ϱ_μ^{ep} to server for aggregation
    get aggregated ϱ_μ^{ep} from server
end if

Unlike value-based approaches like Q-learning, policy gradient techniques optimize the policy directly, bypassing the need to calculate Q-values. This strategy avoids the overestimation bias inherent in value-based methods. The update of parameters favors actions leading to more rewarding outcomes. During testing, the best policy is identified by choosing the action that has the highest probability in a deterministic manner. Thus, by incorporating the cost function into (19), we obtain the cost associated with the F-DDPG algorithm as

$$\min_{\varrho} J(\varrho_\mu) = \sum\nolimits_{j\in\mathcal{F}} \kappa^j J_j(\varrho_\mu^j). \tag{25}$$

Furthermore, within the context of the MA-DRL system framework, each agent independently employs a F-DDPG based algorithm, enabling personalized policy optimization and adaptation. This specific methodology is described in **Algorithm 1**.

D. Federated Multi-Agent Deep Q Network

Following the DDPG model, we also consider a value-based reinforcement learning approach with F-DQN to estimate expected future rewards by leveraging the action-value function $Q(s, a)$. This function, $Q_\pi(s^t, a)$, defines the expected sum of discounted rewards for a given state-action pair:

$$Q_\pi(s^t, a) = \mathbb{E}_\pi\left\{\sum_{j=1}^{\infty} \bar{\gamma}^{t+j} r^{t+j} | s^t, a\right\} \tag{26}$$

$$= \mathbb{E}_{s^{t+1}, a}\{r^t + \bar{\gamma} Q_\pi(s^{t+1}, a) | s^t, a^t\}. \tag{27}$$

The agent aims to find the optimal action-value function, $Q^*(s^t, a)$, which represents the maximum expected return

Algorithm 2 Federated DQN Algorithm for Each Agent

Input: Initialize the parameter settings for the proposed system model, neural networks at $t = 0$
Input: ρ, ϵ, Ω, E
Initialization: get initial ϱ_n from server
for $ep = 1 \rightarrow E$ **do**
  Receive initial observation state s^1
  for $t = 1 \rightarrow T$ **do**
    Select random $r \in [0, 1]$. Obtain action a^t using

$$a^t \triangleq \begin{cases} \text{argmax}_a \, Q(s^t, a; \varrho_n) & \text{if } r > \epsilon \\ \text{pick uniformly action} & \text{else} \end{cases}$$

    Take action a^t, go to state s^{t+1} and get reward r^{t+1}
    Store the tuple $\mathcal{B} = \{a^t, s^t, r^{t+1}, s^{t+1}\}$
  end for
  update $\varrho_n^{ep+1} = \varrho_n^{ep} - \Omega \nabla_{\varrho_n} L(\varrho_n)$
  if ep mod $\rho = 0$ **then**
    send ϱ_n^{ep} to server for aggregation
    get aggregated ϱ_n^{ep} from server
  end if
end for

from state s^t onward. Using a DNN, we approximate $Q(s, a; \varrho_n)$ to estimate optimal Q-values by minimizing the loss function

$$L(\varrho_n) = (r^t + \bar{\gamma} \max_a Q(s^{t+1}, a; \varrho_n) - Q(s^t, a; \varrho_n))^2. \quad (28)$$

During training, the agent stores experience tuples $(s^{t-1}, a^{t-1}, r^t, s^t)$ in a dataset $\mathcal{B}$, iteratively updating $Q(s, a; \varrho_n)$ by minimizing $L(\varrho_n)$. To balance exploration and exploitation, an adaptive ϵ-greedy strategy is employed, allowing the agent to explore different actions early in training. In the federated multi-agent setting, each agent optimizes its local action-value function $Q(s, a; \varrho_n^j)$, contributing to a shared federated objective

$$\min_{\varrho_n} L(\varrho_n) = \sum\nolimits_{j \in \mathcal{F}} \kappa^j L_j(\varrho_n^j), \quad (29)$$

enabling coordinated learning among agents within the federated framework. The F-DQN process is outlined in **Algorithm 2**.

E. Complexity Analysis

For the computation at the t^{th} iteration, we categorize the dimensions of the action and state spaces with the notations $|a^t|$ and $|s^t|$, correspondingly. **Algorithm 1** breaks down the process into two primary segments: 1) Calculation of rewards, which holds a computational complexity of $\mathcal{O}(|s^t|)$. 2) Selection of actions, where the actor and critic networks' complexity is determined by the neuron count in each layer and the total number of layers. For the actor network, let the number of neurons in its m^{th} layer be represented by W^m, and the total layer count by L^a. Consequently, the complexity for a single layer m is given by $\mathcal{O}(W^{m-1}W^m + W^m W^{m+1})$, leading to a total actor network computational complexity of $C_t^a = \mathcal{O}(|s^t| \cdot W^1 + \sum_{m=2}^{L^a-1}(W^{m-1}W^m + W^m W^{m+1}) + W^{L^a-1} \cdot |a^t|)$. In the critic network, with V^q as the neuron count for layer q and L^c for the total layers, the complexity for layer q is $\mathcal{O}(V^{q-1}V^q + V^q V^{q+1})$, leading to a total critic network computational complexity of $C_t^c = \mathcal{O}(|s^t| \cdot V^1 + \sum_{q=2}^{L^c-1}(V^{q-1}V^q + V^q V^{q+1}) + V^{L^c-1})$. Thus, the overall complexity for choosing actions is denoted by $C_t = C_t^a + C_t^c$. The overarching computational complexity of the algorithm across all iterations is therefore expressed as $\mathcal{O}(E \cdot T \cdot C_t)$.

Next, the computational complexity of the DQN-based learning algorithm specified in **Algorithm 2** is quantified as $\mathcal{O}(E \cdot T(\sum_{m=0}^{L^d} F_d^m F_d^{m+1}))$, where L^d represents the number of hidden layers in the DNN, and F_d denotes the number of neurons in each layer.

TABLE II: Simulation Parameters [50], [58]

Parameters	**Value**
Carrier frequency	28 GHz (Ka-band)
Bandwidth	500 MHz
3 dB angle	0.4°
Height of GEO satellite	35786 km
Maximum beam gain	52 dBi
User terminal antenna gain F_g	42.7 dBi
UPA inter element spacing	$\lambda/2$
Rain fading parameters	$(\mu^{rain}, \sigma^{rain}) = (-2.6, 1.63)$
Boltzmann's constant, κ	1.38×10^{-23}J/m
Noise temperature of system, $Tsys$	517K

V. Numerical Simulations and Discussion

A. Parameter Setup

In this segment, we delve into the performance analysis of our proposed federated learning algorithms, leveraging PyTorch for model development and employing the Adam optimization technique for model training. The architectural setup for both the proposed F-DDPG and the F-DQN across MA systems is similar, employing two hidden layers each with 256 neurons [59], [60]. Furthermore, the neural network parameters are updated using the Adam optimizer and the activation function used is ReLU. Moreover, we set the hyper-parameters as $\bar{\gamma} = 0.9$, both critic and actor-network learning rates are given as $\Omega = 0.001$, memory buffer $W = 10000$, size of minibatch = 32, episodes $E = 5000$, each episode encompasses a horizon of $T = 10$ time slots, aggregate frequency $\rho = 100$.

Moreover, Table II outlines the parameters used in the simulation [61]. The satellite is configured with $N_t = N_r = \bar{N} = 4$ antennas for transmission and reception. It includes $L = 4$ SUs and $\mathcal{T}_s = 2$ targets within the satellite zone, with these targets positioned at angles $\psi_1 = -20°$ and $\psi_2 = 30°$. On the terrestrial side, the BS features $M_t = M_r = M = 4$ transmit and receive antennas, serving $K = 4$ CUs and integrating $N = 64$ elements of a RIS. We consider $\mathcal{T}_b = 2$ targets in the terrestrial zone, located at angles $\vartheta_1 = 0°$ and $\vartheta_2 = 20°$. For each CU channel, we model the transmission path using a LoS approach [62], characterized by the channel representation $\mathbf{h}_k = \sqrt{\xi_k M_r}\mathbf{a}_r(\tilde{\vartheta_k}), \forall k$. This notation includes ξ_k to denote the path loss and $\tilde{\vartheta_k}$ for the angular direction of the user. A standard path loss value of -103.6 dB is applied to model the link between each CU and the BS. The directional angles for the DL CUs, listed as $\{\tilde{\vartheta}_1, \tilde{\vartheta}_2\}$ and $\{\tilde{\vartheta}_3, \tilde{\vartheta}_4\}$ are configured to $\{-40°, 60°\}$ and $\{45°, -65°\}$, respectively. The simulation does not incorporate any form of user grouping. Both the

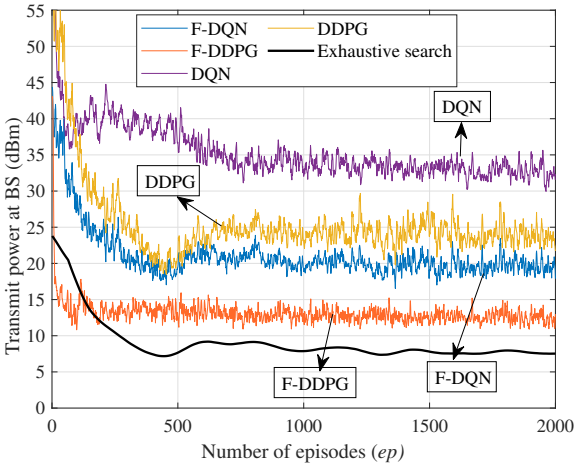

Fig. 2: Convergence behavior.

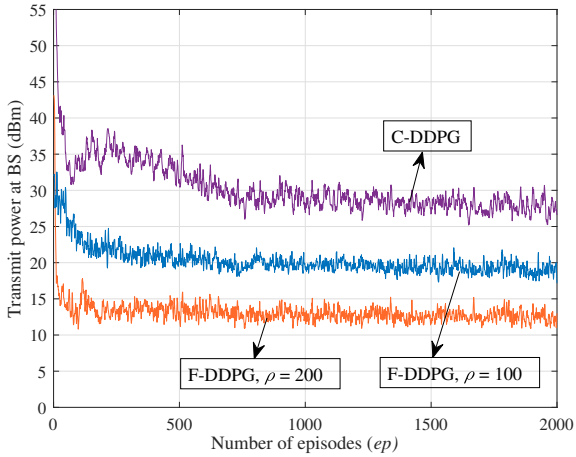

Fig. 3: Convergence behavior of F-DDPG and C-DDPG at different aggregate frequencies.

satellite and BS adhere to a maximum transmit power limit of $P_{max}^{sat} = 40$ dBm. For the satellite channel model, the noise power is normalized by $\kappa T_{sys} B_g$ which results in the noise variance being set to $\sigma_l^{sat\,2} = {\sigma_k}^2 = 1, \forall l, \forall k$ [50]. For simplicity, we consider $\tau_{t_s}^{\text{tar}} = \tau^{\text{tar}} = 12$ dB, $\forall t_s$ and $\tau_{t_b}^{\text{rad}} = \tau^{\text{rad}} = 12$ dB, $\forall t_b$. The numerical results presented are the average outcomes from 100 distinct channel realizations. Unless stated otherwise, the parameter settings adhere to the aforementioned specifications.

B. Benchmark Schemes

For comparative analysis, we incorporate the following benchmark schemes.

1) **Centralized scheme**: In the centralized DRL framework, each iteration requires agents to share data with a central server for immediate control actions, facilitating the development of a unified model. This setup is implemented for DDPG, denoted as C-DDPG in the simulation [63].
2) **Communication - only scheme**: This benchmark omits the sensing SINR requirements within a RIS-enhanced STIN, labeled as "w/o sensing" in the simulation. This strategy assists in determining the effects of incorporating sensing capabilities on communication performance [64].
3) **Random RIS scheme**: In this scheme, while the satellite and BS utilize our proposed beamforming approach, the RIS adopts a random passive beamforming approach [28], [65]–[68].
4) **No-RIS scheme**: This scheme examines the performance of our ISAC-STIN setup without RIS intervention in the terrestrial domain, marked as "w/o RIS" in the figures. underscores the significance of integrating RIS into our network design [65]–[68].

Fig. 2 illustrates the minimum transmit power at the BS for the F-DDPG algorithm implemented in a federated setting across multiple MA systems with $\rho = 200$. The training process spanned 2000 episodes, with each episode consisting of a sequence of $T = 10$ time slots. For comparison, the F-DQN, traditional DDPG, and DQN algorithms are also considered in the same setting. Additionally, an exhaustive search benchmark is included to evaluate the proposed algorithm's performance against a method that guarantees optimal solutions but with significantly higher computational complexity and slower convergence. F-DDPG demonstrates smoother and more consistent convergence compared to F-DQN, attributed to their different exploration strategies. This phenomenon is likely due to the distinct exploration mechanisms employed by the two algorithms. On one hand, DQN relies on the ϵ-greedy method, introducing more randomness until convergence is attained. On the other hand, DDPG leverages a policy gradient framework that prioritizes updates along state-action paths associated with superior average rewards, which seems to benefit more directly from the aggregation process. Compared to the exhaustive search, F-DDPG achieves near-optimal solutions with considerably faster convergence and lower complexity, demonstrating its practical applicability for real-time systems. While exhaustive search ultimately reaches the optimal solution, its computational overhead underscores the advantages of DRL-based approaches in complex and dynamic settings.

Fig. 3 illustrates the minimum transmit power required at the BS when using the DDPG algorithm in both federated and benchmark centralized settings across MA systems. The figure compares the performance of the system over different aggregate intervals for F-DDPG along with C-DDPG. The depicted curves show the impact of server-agent aggregation intervals, specifically at $\rho = 100$ and $\rho = 200$. Specifically, it dictates the frequency of model update exchanges between the server and agents, affecting the convergence rate and effectiveness of the federated learning algorithms. With shorter aggregation intervals, the system tends to require more power due to more frequent updates. In C-DDPG, updates occur every iteration, requiring frequent model exchanges that indirectly increase BS transmit power to support continuous synchronization. While this dependency is not explicitly modeled in the objective function, it results from C-DDPG's frequent update structure. In contrast, F-DDPG employs predetermined aggregation intervals, reducing synchronization events and thereby lowering

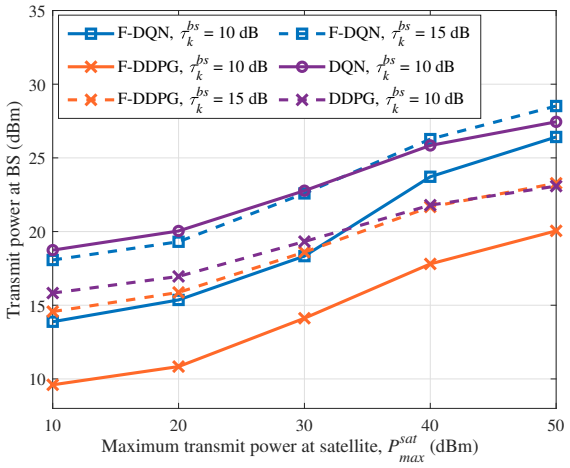

Fig. 4: Impact of P_{max}^{sat} with different τ_k^{bs}.

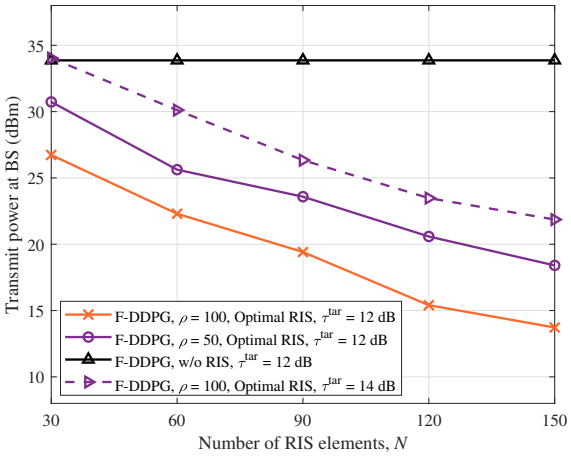

Fig. 6: Impact of N at different τ^{tar} and aggregate frequencies.

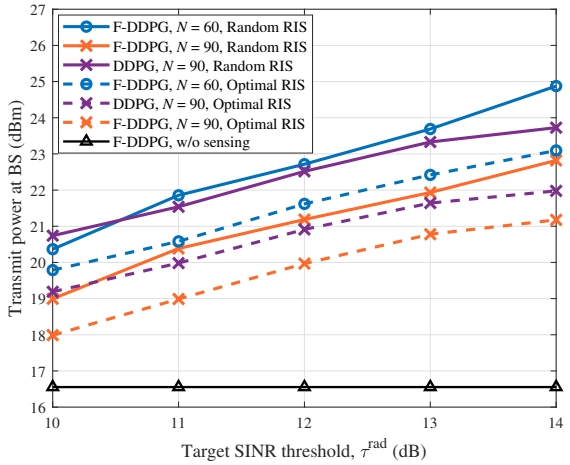

Fig. 5: Impact of τ^{rad} with different N.

BS power requirements for more power-efficient performance. Additionally, the less frequent updates in F-DDPG facilitate a smoother and more consistent convergence path, reducing potential oscillations caused by continuous updates. As a result, F-DDPG exhibits steadier performance and lower power consumption than C-DDPG across all tested frequencies.

Fig. 4 demonstrates the impact of P_{max}^{sat} on the minimum transmit power requirements at the BS. As P_{max}^{sat} increases, system performance improves but simultaneously elevates interference in the terrestrial zone, necessitating higher power outputs at the BS to mitigate this interference. Additionally, this figure also examined the impact of different SINR thresholds at the CUs on the corresponding power demands at the BS. Our findings in Fig. 4 show that as the SINR requirement τ_k^{bs} for CUs becomes more demanding, the BS is forced to allocate even more power to meet these higher quality thresholds. This dynamic underscores a complex trade-off unique to STIN systems, where managing the interplay between satellite and terrestrial power is essential to optimizing overall performance. Further, the proposed F-DDPG demonstrates superior performance compared to F-DQN, traditional DDPG, and DQN methods. This analysis underscores the intricate balance required between efficient power management at the BS and optimal resource allocation within the system.

Fig. 5 examines the influence of the target SINR threshold in the terrestrial zone on the minimum transmit power requirements at the BS. As the τ^{rad} threshold rises, necessitating better sensing performance, the system must allocate increased power for enhanced sensing capabilities. For a clearer comparison, a reference scenario focusing solely on communication excluding target sensing SINR requirements ("w/o sensing") is also analyzed. The "w/o sensing" scenario consumes less power for communication than systems with sensing targets, as validated in Fig. 5. Additionally, the analysis also explores the role of RIS in the network. Increasing the RIS elements increases the number of phase shifters, which boosts channel gain diversity for users and decreases the transmission power necessary to achieve the desired quality of service. Additionally, we compare these effects against the random RIS phase shift allocations. The results show that optimized RIS configurations significantly lower BS power requirements compared to random setups, as demonstrated through the proposed F-DDPG and benchmark DDPG methods. Consequently, the optimal RIS configuration with $N = 60$ elements and the initial parameter settings achieves a 5.8% reduction in transmit power requirements at the BS compared to the random RIS scenario.

Fig. 6 illustrates the impact of increasing the number of elements in an RIS on the minimum transmit power at the BS, indicating that a higher number of reflecting elements leads to improved system performance even with a reduced transmit power at the BS due to the additional spatial DoF. This improvement is especially notable when compared to scenarios without an RIS, labeled as "w/o RIS". The figure also examines different target SINR thresholds τ^{tar} in the satellite zone. As the τ^{tar} increases, the transmit power at the satellite also increases to meet the more stringent sensing requirements of higher thresholds within the ISAC system. This, in turn, leads to increased interference in the terrestrial zone, necessitating more power at the BS to compensate.

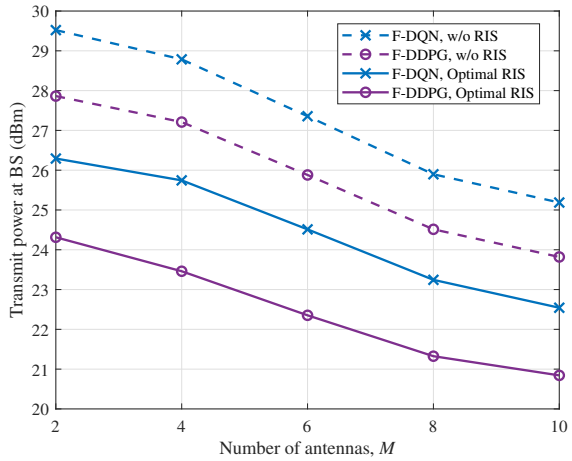

Fig. 7: Impact of M and RIS on transmit power.

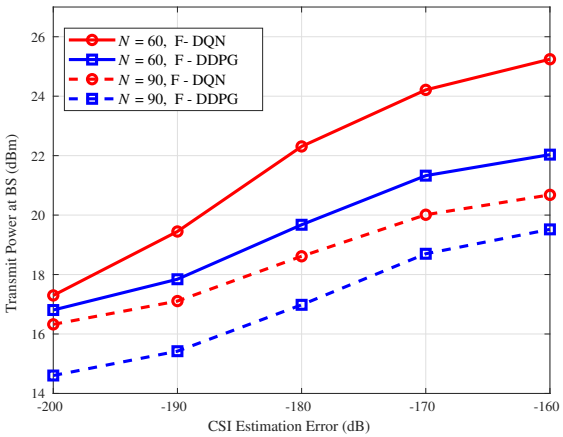

Fig. 9: Transmit power at the BS versus channel estimation error variance (σ_e^2) with different N.

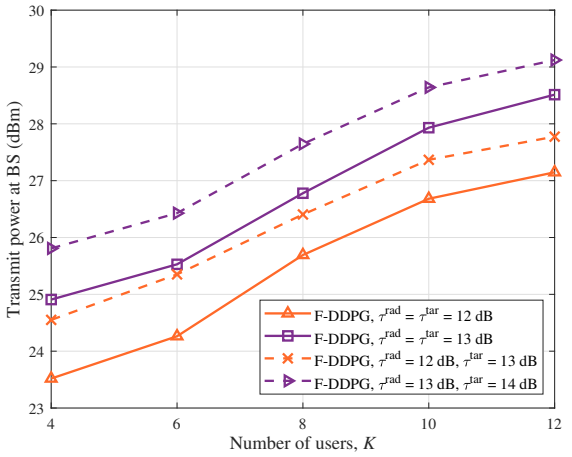

Fig. 8: Impact of K with different τ^{rad} and τ^{tar}.

Moreover, the data indicates that the minimum transmit power required for $\tau^{\text{tar}} = 12$ dB with 90 elements is comparable to that for $\tau^{\text{tar}} = 14$ dB with 120 elements. This comparison indicates that increasing the number of RIS elements can significantly enhance sensing performance while maintaining the same level of transmit power. Additionally, the results validate that the optimal RIS configuration with $N = 60$ elements and the initial parameter settings achieve a 34.2% reduction in transmit power requirements at the BS compared to the scenario without RIS in the system.

Fig. 7 illustrates the impact of number of transmit antennas at the BS and the minimum required transmit power. As anticipated, increasing the number of antennas at the BS leads to a reduction in the transmit power needed for information transfer. This decrease is due to the additional DoFs provided by more antennas, which enhances the efficiency of spatial multiplexing and subsequently lowers power consumption at the BS. Additionally, the impact of integrating an RIS into the system is also investigated. The comparison between systems with and without RIS reveals that RIS contributes additional DoFs, thereby further reducing the transmit power requirements and enhancing system efficiency compared to the no RIS setup.

Fig. 8 delineates the relationship between the increasing number of users and the minimum transmit power at the BS. The analysis reveals a direct correlation: as the number of users increases, there is a corresponding increase in BS transmit power required to sustain optimal communication system performance. This figure further details the influence of increasing target SINR thresholds in both satellite and terrestrial zones. Notably, as the τ^{rad} and τ^{tar} thresholds increase, there is a corresponding increase in the transmit power at the BS. This adjustment is necessary to meet the enhanced sensing demands imposed by these increased SINR thresholds within the ISAC system, ensuring that the system can adequately support both its communication and sensing functionalities under more stringent conditions.

The impact of channel state information (CSI) uncertainty on the performance of the STIN is analyzed in Fig. 9, focusing on the relationship between transmit power at the BS and CSI estimation error variance (σ_e^2). The results demonstrate that as CSI estimation errors increase, system performance degrades significantly for both F-DDPG and F-DQN algorithms. However, the benefits of deploying additional RIS elements remain evident even under CSI uncertainty [69]. For instance, scenarios with $N = 90$ RIS elements consistently outperform those with $N = 60$, owing to greater channel diversity and improved system robustness. This analysis underscore the importance of considering CSI uncertainty in the STIN and validates the effectiveness of the proposed approach in addressing practical challenges like robustness.

Next, we detail the achieved beampattern gain for target functionality realized through **Algorithm 1**. By employing the optimized receive beamformer $\mathbf{u}_{t_b}^*$, which is normalized to ensure $\left\|\mathbf{u}_{t_b}^*\right\| = 1$, along with the optimized transmit signal $\mathbf{x}^{bs*}$, we define the beampattern directed towards the target as

$$p(\vartheta_{t_b}) = |\mathbf{u}_{t_b}^* \mathbf{a}_r(\vartheta_{t_b}) \mathbf{a}_t^H(\vartheta_{t_b}) \mathbf{x}^{bs*}|. \tag{30}$$

Further, we illustrate the beampattern gains achieved for specific target locations in Fig. 10. We consider a case with

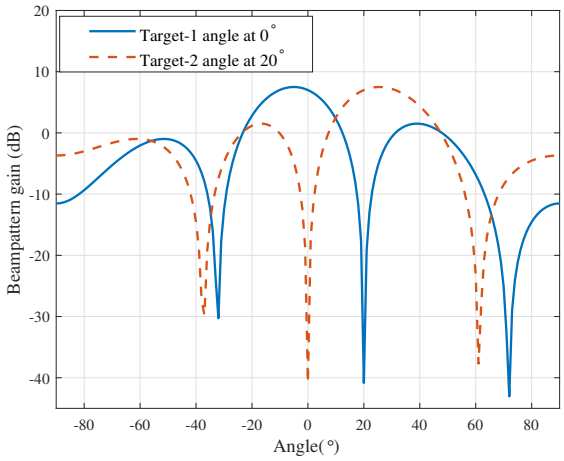

Fig. 10: Beampattern gain of terrestrial zone targets.

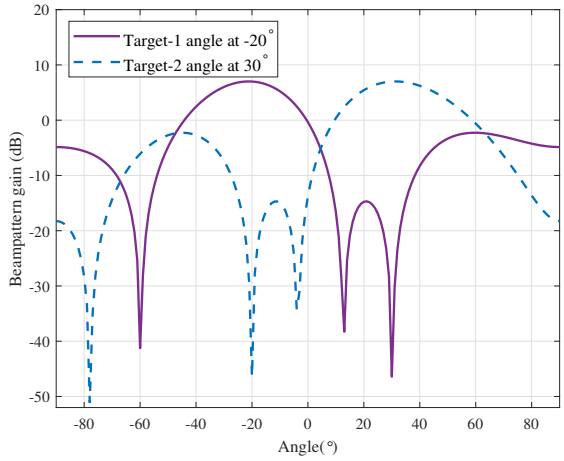

Fig. 11: Beampattern gain of satellite zone targets.

$\mathcal{T}_b = 2$ targets positioned at angles 0° and 20°. Keeping other parameter settings constant, we address the challenge of maximizing the power at BS subject to radar SINR constraints for $\mathcal{T}_b$ targets in the terrestrial zone using **Algorithm 1**. The target beampattern gains showcased in Fig. 10 is the result of employing the optimized receive beamformer $\mathbf{u}_{t_b}^*$, as outlined in (30). Similarly, the beampattern gain for targets in the satellite zone, positioned at angles of -20° and 30° is illustrated in Fig. 11. The illustrations confirm that the main lobes are directed toward the intended targets, underscoring the capability of the proposed beamforming design to detect multiple targets effectively.

VI. Conclusions and Future Work

In this study, we introduced a cutting-edge analytical framework to minimize transmit power at the BS in STINs through the use of RIS within the ISAC framework. By leveraging federated learning, our method dynamically adapted to network changes, ensuring compliance with beamforming designs, multiple target SINR thresholds, and RIS phase-shift requirements via an effective feedback loop. The proposed F-DDPG algorithm across MA systems outperformed existing models, including F-DQN, centralized DDPG, and conventional DDPG and DQN methods. Simulation results have demonstrated that integrating RIS significantly lowers base station power requirements against both random and without RIS configurations. In particular, the optimal RIS configuration with 60 elements achieved a 6.3% reduction in BS transmit power compared to the random RIS scenario and a 34.2% reduction compared to the no-RIS setup. Furthermore, the simulations results validated that increasing the number of RIS elements markedly improves sensing capabilities while maintaining the same level of transmit power.

References

[1] A. Yazar, S. Dogan-Tusha, and H. Arslan, "6G vision: An ultra-flexible perspective," *ITU J. Future Evolving Technol.*, vol. 1, no. 1, pp. 121–140, 2020.

[2] 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. Surv. & Tut.*, vol. 24, no. 1, pp. 306–345, Firstquart. 2022.

[3] L. Yin, Z. Liu, M. R. Bhavani Shankar, M. Alaee-Kerahroodi, and B. Clerckx, "Integrated sensing and communications enabled low earth orbit satellite systems," *IEEE Netw.*, pp. 1–1, 2024.

[4] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, "Broadband LEO satellite communications: Architectures and key technologies," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 55–61, Apr. 2019.

[5] I. del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, 2019.

[6] J. P. Choi and C. Joo, "Challenges for efficient and seamless space-terrestrial heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 156–162, May 2015.

[7] "(Release 15) study on new radio (NR) to support non-terrestrial networks," *3GPP Sophia Antipolis, France, Rep. TR38.811 V15.3.0. Release 15*, Jul. 2020.

[8] J. Peisa, P. Persson, S. Parkvall, E. Dahlman, A. Grovlen, C. Hoymann, and D. Gerstenberger, "5G evolution: 3GPP releases 16 & 17 overview," *Ericsson Technology Review*, vol. 2020, pp. 2–13, 03 2020.

[9] B. Aazhang *et.al*, *Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper)*, 09 2019.

[10] L. Kuang, C. Jiang, Y. Qian, and J. Lu, *Terrestrial-Satellite Communication Networks: Transceivers Design and Resource Allocation*. Springer, 2017.

[11] K. An *et al.*, "Outage performance of cognitive hybrid satellite–terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397–9404, Nov. 2016.

[12] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2253–2267, Oct. 2017.

[13] L. Kuang, X. Chen, C. Jiang, H. Zhang, and S. Wu, "Radio resource management in future terrestrial-satellite communication networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 81–87, Oct. 2017.

[14] B. Li, Z. Fei, X. Xu, and Z. Chu, "Resource allocations for secure cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 78–81, Feb. 2018.

[15] F. Guidolin *et al.*, "A cooperative scheduling algorithm for the coexistence of fixed satellite services and 5g cellular network," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 1322–1327.

[16] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Energy efficient resource allocation in cloud based integrated terrestrial-satellite networks," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.

[17] M. Lin *et.al*, "Joint beamforming and power control for device-to-device communications underlaying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 138–150, Jan. 2016.

[18] M. Lin, L. Yang, W.-P. Zhu, and M. Li, "An open-loop adaptive space-time transmit scheme for correlated fading channels," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 147–158, Apr. 2008.

[19] M. A. Vazquez *et.al*, "Hybrid analog-digital transmit beamforming for spectrum sharing satellite-terrestrial systems," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2016, pp. 1–5.

[20] B. Li *et al.*, "Robust chance-constrained secure transmission for cognitive satellite–terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4208–4219, May. 2018.

[21] K. An, M. Lin, J. Ouyang, and W.-P. Zhu, "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025–3037, Nov. 2016.

[22] M. Lin, Z. Lin, W.-P. Zhu, and J.-B. Wang, "Joint beamforming for secure communication in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 1017–1029, May 2018.

[23] F. Liu *et al.*, "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.

[24] S. Biswas *et.al*, "Design and analysis of FD MIMO cellular systems in coexistence with MIMO radar," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4727–4743, Jul. 2020.

[25] A. Kaushik, C. Masouros, and F. Liu, "Hardware efficient joint radar-communications with hybrid precoding and RF chain optimization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.

[26] F. Liu *et.al*, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, Aug. 2018.

[27] K. Singh, S. Biswas, T. Ratnarajah, and F. A. Khan, "Transceiver design and power allocation for full-duplex MIMO communication systems with spectrum sharing radar," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 556–566, Sep. 2018.

[28] S. Pala, O. Taghizadeh, M. Katwe, K. Singh, C.-P. Li, and A. Schmeink, "Secure RIS-assisted hybrid beamforming design with low-resolution phase shifters," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.

[29] T. Tian, T. Zhang, L. Kong, and Y. Deng, "Transmit/receive beamforming for MIMO-OFDM based dual-function radar and communication," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4693–4708, May 2021.

[30] L. You, X. Qiang, C. G. Tsinos, F. Liu, W. Wang, X. Gao, and B. Ottersten, "Beam squint-aware integrated sensing and communications for hybrid massive MIMO LEO satellite systems," *IEEE J. Sel. Areas Communications*, vol. 40, no. 10, pp. 2994–3009, 2022.

[31] B. Zhao, M. Wang, Z. Xing, G. Ren, and J. Su, "Integrated sensing and communication aided dynamic resource allocation for random access in satellite terrestrial relay networks," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 661–665, Feb. 2023.

[32] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Sig. Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[33] A. Fascista *et.al*, "RIS-aided joint localization and synchronization with a single-antenna receiver: Beamforming design and low-complexity estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 1141–1156, 2022.

[34] S. P. Chepuri, N. Shlezinger, F. Liu, G. C. Alexandropoulos, S. Buzzi, and Y. C. Eldar, "Integrated sensing and communications with reconfigurable intelligent surfaces," *arXiv preprint arXiv:2211.01003*, 2022.

[35] R. P. Sankar, S. P. Chepuri, and Y. C. Eldar, "Beamforming in integrated sensing and communication systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, 2023.

[36] R. Liu, M. Li, and A. L. Swindlehurst, "Joint beamforming and reflection design for RIS-assisted ISAC systems," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*. IEEE, 2022, pp. 997–1001.

[37] M. Hua, Q. Wu, C. He, S. Ma, and W. Chen, "Joint active and passive beamforming design for IRS-aided radar-communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2278–2294, 2022.

[38] M. Wu *et al.*, "Optimization design in RIS-assisted integrated satellite-UAV-served 6G IoT: A deep reinforcement learning approach," *IEEE Internet Things Mag.*, vol. 7, no. 1, pp. 12–18, Jan. 2024.

[39] Z. Lin *et.al*, "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717–3724, Aug. 2022.

[40] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[41] 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.

[42] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4695–4707, Sep. 2015.

[43] P. Stoica, J. Li, and Y. Xie, "On probing signal design for MIMO radar," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4151–4161, Aug. 2007.

[44] J. Pritzker, J. Ward, and Y. C. Eldar, "Transmit precoding for dual-function radar-communication systems," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1065–1070.

[45] L. Chen, Z. Wang, Y. Du, Y. Chen, and F. R. Yu, "Generalized transceiver beamforming for DFRC with MIMO radar and MU-MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1795–1808, Jun. 2022.

[46] C. G. Tsinos, A. Arora, S. Chatzinotas, and B. Ottersten, "Joint transmit waveform and receive filter design for dual-function radar-communication systems," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1378–1392, Nov. 2021.

[47] C.-Y. Chen and P. P. Vaidyanathan, "MIMO radar waveform optimization with prior information of the extended target and clutter," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3533–3544, Sep. 2009.

[48] L. Wu, P. Babu, and D. P. Palomar, "Transmit waveform/receive filter design for MIMO radar with multiple waveform constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1526–1540, Mar. 2018.

[49] G. Cui, H. Li, and M. Rangaswamy, "MIMO radar waveform design with constant modulus and similarity constraints," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 343–353, Jan. 2014.

[50] L. Yin and B. Clerckx, "Rate-splitting multiple access for satellite-terrestrial integrated networks: Benefits of coordination and cooperation," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 317–332, Jan. 2023.

[51] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[52] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[53] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–36, 2021.

[54] W. Xu, J. An, C. Huang, L. Gan, and C. Yuen, "Deep reinforcement learning based on location-aware imitation environment for RIS-aided mmwave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1493–1497, Jul. 2022.

[55] Z. Peng, Z. Zhang, L. Kong, C. Pan, L. Li, and J. Wang, "Deep reinforcement learning for RIS-aided multiuser full-duplex secure communications with hardware impairments," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21 121–21 135, Nov. 2022.

[56] 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.

[57] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[58] Z. Liu, L. Yin, W. Shin, and B. Clerckx, "Max-min fair energy-efficient beam design for quantized isac leo satellite systems: A rate-splitting approach," *arXiv preprint arXiv:2402.09253*, 2024.

[59] R. Zhang, K. Xiong, Y. Lu, P. Fan, D. W. K. Ng, and K. B. Letaief, "Energy efficiency maximization in RIS-assisted SWIPT networks with RSMA: A PPO-based approach," *IEEE J. Sel. Areas in Commun.*, vol. 41, no. 5, pp. 1413–1430, May 2023.

[60] S. Pala *et al.*, "Robust design of RIS-aided full-duplex RSMA system for V2X communication: A DRL approach," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2023, pp. 2420–2425.

[61] Z. Lin, M. Lin, J.-B. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 657–670, Jun. 2019.

[62] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[63] F. D. Calabrese *et al.*, "Learning radio resource management in RANs: Framework, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 138–145, 2018.

[64] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896–4910, Sep. 2014.

[65] S. Pala, K. Singh, M. Katwe, and C.-P. Li, "Joint optimization of URLLC parameters and beamforming design for multi-RIS-aided MU-MISO URLLC system," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 148–152, Jan. 2023.

[66] L. Chai, L. Bai, T. Bai, J. Shi, and A. Nallanathan, "Secure RIS-aided MISO-NOMA system design in the presence of active eavesdropping," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19 479–19 494, Nov. 2023.

[67] P. Saikia, S. Pala, K. Singh, S. K. Singh, and W.-J. Huang, "Proximal policy optimization for RIS-assisted full duplex 6G-V2X communications," *IEEE Trans. Intell. Veh.*, pp. 1–16, 2023.