QoS Improvement in Multi User Cellular-Symbiotic Radio Network Assisted by Active-STAR-RIS

Rahman Saadat Yeganeh [1], Mohammad Javad Omidi [1,2], Farshad Zeinali [3], *Student Member, IEEE*, Mohammad Robat Mili [4], and Mohammad Ghavami [5], *Senior Member, IEEE*

***Abstract*—In this article, we employ active simultaneously transmitting and reflecting reconfigurable intelligent surfaces (ASRIS) to enhance the quality of 6G cellular network services. The network integrates commensal symbiotic radio (CSR) subsystems to facilitate communication between passive Internet of Things (IoT) users and active users, referred to as symbiotic backscatter devices (SBDs) and symbiotic user equipments (SUEs), respectively. Since the SBDs are passive, transmitting information to the SUEs poses significant challenges. To overcome this challenge, we harness the capabilities of massive multiple input multiple output (MIMO) antennas within the base station (BS) to relay the information transmitted by SBDs with greater power. This scheme uses the non-orthogonal multiple access (NOMA) technique for multiple access among all users, and potential interferences are eliminated using successive interference cancellation (SIC). The primary objective is to maximize the throughput between SBDs and SUEs. To achieve this, we formulate an optimization problem involving variables such as active beamforming coefficients at the BS and ASRIS, phase adjustments of ASRIS, and scheduling parameters between CSR and cellular networks. To solve this optimization problem, we used three deep reinforcement learning (DRL) methods: proximal policy optimization (PPO), twin delayed deep deterministic policy gradient (TD3), and asynchronous advantage actor critic (A3C). These methods were simulated, and the results demonstrate that A3C, TD3, and PPO have the best convergence speeds and achieve the highest increases in network throughput, respectively. Finally, the proposed scheme was evaluated using passive simultaneously transmitting and reflecting RIS (STAR-RIS), which demonstrated poorer performance compared to ASRIS.**

***Index Terms*—Symbiotic radio, active STAR-RIS, sum throughput maximization, IoT, NOMA.**

I. Introduction

A. Background

The sixth-generation (6G) wireless networks need to be capable of providing coverage for billions of IoT, cellular, and other wireless devices [1]. The IoT stands out as a pivotal technology in shaping the future of wireless communications, embracing a diverse range of applications, including smart transportation, smart homes, smart grids, and smart agriculture [2]. However, the extensive deployment of IoT devices faces two critical challenges: energy efficiency and spectrum scarcity issues [3]–[5]. The challenge of spectrum scarcity arises from the limited availability of spectrum resources for IoT applications, as a significant portion has already been allocated to various radio systems, including cellular networks, television broadcasts, and other communication services. This constraint underscores the urgency of resolving these challenges to optimize spectrum utilization. Additionally, the energy efficiency issue presents a significant obstacle, given the high cost associated with regularly charging devices or replacing batteries [6]. We propose a solution to address the aforementioned challenges by employing symbiotic radio (SR) systems and reconfigurable intelligent surfaces (RIS) structures.

SR stands out as an innovative technology that not only preserves the merits of prior systems like cognitive radio [7]–[9] and ambient backscatter communication (AmBC) [10], [11] but also addresses and eliminates their drawbacks. It currently ranks among the captivating subjects in both scientific and industrial domains [12]–[14]. The SR network can be classified into two types based on the relationship between the symbol periods of the SBDs and the BS: parasitic SR (PSR) and commensal SR (CSR) [11], [15]. In PSR, SBDs can exchange information at a high rate, but they also suffers from interference between the signals of SBDs and BS in the receiver, necessitating complex interference cancellation techniques. Furthermore, PSR requires synchronization between the BS, SBDs, and receiver. On the other hand, CSR is suitable for IoT networks with low data rates, and addresses the drawbacks of the PSR setup [16]. By reducing interference between different network components, the receiver can perform joint decoding of information from both the BS and SBDs, enabled by transmit collaboration between them [7].

On the other hand, RIS is a new technology composed of two-dimensional surface that can reduce the need for expensive BS active antennas with complex hardware in the network and provide complete coverage over a specific area. The initial type of these surfaces only had the capability to reflect signals in one direction [17], and due to inadequate coverage, a more advanced type called STAR-RIS was introduced in subsequent designs [18], [19]. STAR-RISs are equipped with a large number of low-cost passive elements whose transmission and reflection coefficients can be controlled. It is worth emphasizing that utilizing STAR-RIS in passive mode necessitates a substantial number of elements on its surface to attain optimal operational gains, resulting in its enlargement. This presents a formidable challenge to the expansion of these systems. As a

[1]Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, IRAN (emails: r.saadat@ec.iut.ac.ir, omidi@iut.ac.ir).

[2]Department of Electronics and Communication Engineering, Kuwait College of Science and Technology (KCST), Doha, Kuwait.

[3]The Pasargad Institute for Advanced Innovative Solutions (PIAIS), Tehran, Iran (email: farshad.zeinali@piais.ir).

[4]The Pasargad Institute for Advanced Innovative Solutions (PIAIS), Tehran, Iran (email: Mohammad.Robatmili@gmail.com).

[5]Electrical and Electronic Engineering Department, London South Bank University, London SE1 0AA, U.K. (email: ghavamim@lsbu.ac.uk).

result, recent research endeavors have introduced the concept of ASRIS to skillfully address and resolve this issue [20].

B. Related Works

1) Reconfigurable Intelligent Surfaces: In [21], a STAR-RIS assisted MIMO system, in which the sum rate maximization problem was investigated in both unicasting and broadcasting signal models. Also, in [22], an investigation into the secrecy performance of STAR-RIS assisted NOMA networks took place. In [23], the proposition of active RISs and the investigation of joint optimization for reflecting phase shift and receive beamforming were introduced. Subsequently, [24] presented a validated signal model for active RISs. In the article [20], a hardware model is introduced for active STAR-RISs incorporating both coupled and independent transmission and reflection phase-shifts. More precisely, this paper demonstrates the utilization of reflection-type amplifiers and quadrature hybrid couplers to achieve amplification of the transmission and reflection coefficients. Nevertheless, a significant gap exists in the research on modeling and performance analysis of ASRISs. Furthermore, the strategy for utilizing ASRIS to improve Quality of Service (QoS) in communications within an operational IoT and cellular network is currently ambiguous.

2) Symbiotic Radio: The paper [25] provides a comprehensive review of backscatter communication, addressing key challenges such as interference and double-path fading, while exploring advancements and future trends to enhance its role in the evolving IoT ecosystem. In [26], the advanced version of backscatter communication, known as symbiotic radio, has been implemented in the CSR form. This system is specifically designed for scenarios involving multiple backscatter devices (BDs). The primary goal within this framework is to reduce energy consumption, and a method for optimal resource allocation, named Timing SR, has also been introduced. Additionally, the paper [27] presents a system model similar to the one proposed in [26], considering a multi-user scenario at the destination. The aim of this paper is to design transmit beamforming that minimizes the transmit power at the BS while adhering to a cellular transmission outage probability constraint. The study in [28] introduces a symbiotic backscatter NOMA (SBN) system, where a backscatter device transmits information to two users by adjusting its reflection coefficient, ensuring successful decoding of both NOMA and backscatter signals. Similarly, [29] proposes a segmented RIS-based symbiotic ambient backscatter system that improves signal reception and backscatter performance, optimizing the coexistence outage probability and ergodic capacity. Building on these concepts, the work in [30] presents a STAR-RIS segmented SBN system, which optimizes reflection coefficients to maximize ergodic capacity while balancing the decoding performance of primary and backscatter signals. Further, [31] investigates the outage performance of the STAR-RIS segmented SBN system, analyzing the effects of interference and power allocation. Additionally, the letter [32] proposes two mechanisms for backscatter devices to enhance concurrent transmissions in primary and backscatter systems.

Also, the article [15], proposes a novel SR technique for passive IoT devices. This approach involves integrating a BD with a primary communication system, and designing a primary transmitter and receiver to optimize both the primary and BD transmissions. The decoding strategy used in the receiver is based on SIC. So far, articles on SR have primarily focused on IoT device communications. However, the main goal of the SR system is to integrate it into a cellular network. This integration presents various challenges and requires further examination. Furthermore, to efficiently accommodate a significant number of devices in SR networks, it is crucial to design the network to support multiple SBDs. However, challenges, including potential user interference with improperly designed multiple access schemes, may arise. To mitigate these issues, various methods, such as NOMA, are employed to prevent interference [28], [33], [34].

It should be noted that, SR systems typically suffer from the poor system performance owing to the limited communication efficiency of backscattering devices. To overcome this bottleneck, one promising method involves using RIS to enhance the received signal strength in SR transmission. Due to the abundant advantages of these two subjects and their combination, numerous articles, such as [35]–[37], have been presented in this field. Also, in the article [38], a passive STAR-RIS empowered transmission scheme for SR systems. The authors focused on minimizing the transmit power of the BS by designing the active beamforming and simultaneous reflection and transmission coefficients under the practical phase correlation constraint.

Another method involves utilizing the power of a active antennas on BS to relay information from IoT devices. In this scheme, the BS receives data from IoT devices and forwards it to the intended destination. This method has been demonstrated for a single user in [39]. In this paper, to enhance the BS signals for transmission to the destination, a backscatter relay scheme is employed, where multiple passive devices relay the BS information. This approach introduces challenges due to increased interference at the destination and the inherent limitations of passive devices in signal amplification and transmission, particularly in double fading channels.

C. Our Contributions

In this work, we propose a 6G cellular network featuring a CSR subsystem. The BS initially transmits signals to the SBDs on one side using active massive MIMO antennas. The SBDs modulate their information onto the received signal carrier. Due to the significant distance between the passive IoT devices and their intended recipients (SUEs), we employ a relay system. The BS relays the information to the SUEs located on the opposite side of the BS. To enhance the performance of user data transmission in this network, we incorporate an ASRIS device. This device not only amplifies the received signals but also transmits them to users within its coverage area. These users typically lack a line-of-sight (LoS) connection with the BS due to intervening obstacles. In all transmission schemes between nodes, the method for multiple access is NOMA.

The major contributions of this paper are summarized as follows:

- First, We implement an innovative, comprehensive, and practical system model where multiple users, both actively (SUEs) and passively (SBDs), engage in the exchange of diverse information within the 6G cellular network. To facilitate communication between IoT users and other cellular devices, we employ a CSR setup. In this setup, the BS dispatches ambient signals to SBDs using a NOMA approach. The SBDs then harvest wireless energy from these signals, modulate their own information onto the carrier, and transmit backscatter signals back to the BS. Assisted by an ASRIS, the BS efficiently relays this information to the intended SUEs located in both the reflection and transmission regions of the ASRIS, ensuring effective communication for users with LoS and non-LoS links to the BS, respectively.
- Second, To achieve optimal resource allocation, we formulate an optimization problem aimed at maximizing the information throughput among SBDs, the BS, and SUEs. To this end, we define several key variables within the problem, including the active beamforming vectors for two phases: transmitting signals to the SBDs and delivering the desired information to the SUEs. We also consider a timing schedule for these phases and determine the amplifier gain and phase variables for the ASRIS. To ensure QoS, we impose constraints to guarantee a minimum rate for each SBD and SUE. Additionally, we incorporate constraints related to SIC for all transmitted signals from the users.
- Third, SBDs can harvest the required energy from ambient waves whenever they need to transmit information. Due to the considerable distance between SBDs and SUEs, direct transmission to SUEs may not be feasible. To address this issue, we leverage the power of active massive MIMO antennas at the BS. After eliminating interference using SIC, the BS directs each of the signals from the SBDs to their respective SUEs. Additionally, the signal modeling in ASRIS is thoroughly examined and incorporated into the overall design.
- Finally, to solve this complex problem, we employ DRL methods. Three novel approaches, namely PPO, TD3, and A3C, each exhibiting distinctive features, were explored. These algorithms are employed to model the problem, and a comprehensive comparison of all simulated methods is conducted under varying conditions.

This paper is structured as follows. In Section II, we present the proposed system model for the ASRIS assisted by the CSR system. Section III focuses on the throghput maximization problem. In Section IV, we investigate some DRL methods, namely PPO, TD3 and A3C. In Section V, the mentioned methods are modeled and simulated, followed by a comparison with each other. Finally, in Section VI, we summarize our conclusions and discuss future work.

Notations: $\mathbf{A}^H$, $\mathbf{A}^T$, $\|\mathbf{A}\|$ denote the trace, conjugate transpose, transpose, and norm of the matrix $\mathbf{A}$, respectively. Also, the gradient operator denoted as ∇.

TABLE I: List of abbreviations.

AmBC	Ambient backscatter communication
ASRIS	Active simultaneously transmitting and reflecting RIS
A3C	Asynchronous advantage actor critic
BS	Base station
CSR	Commensal symbiotic radio
DDPG	Deep deterministic policy gradient
DRL	Deep reinforcement learning
IoT	Internet of things
MIMO	Multiple input multiple output
NOMA	Non orthogonal multiple access
PPO	Proximal policy optimization
QoS	Quality of Service
RIS	Reconfigurable intelligent surfaces
SIC	Successive interference cancellation
STAR-RIS	Simultaneously transmitting and reflecting RIS
SR	Symbiotic radio
SBD	Symbiotic backscatter device
SUE	Symbiotic user equipment
TD3	Twin delayed DDPG

II. SYSTEM MODEL AND PROBLEM FORMULATION

As illustrated in the system model in Fig. 1, the cellular network with the SR subsystem comprises a BS equipped with N massive MIMO antennas, an ASRIS with M active elements, I SUEs in the transmission and reflection spaces of the ASRIS, and I SBDs with a single antenna.

In this structure, the SBDs are positioned at a long distance from the SUEs (relative to the wavelength of the network's operating frequency). Moreover, each SBD is a passive device and lacks the capability to transmit information over long distances, it cannot establish a direct backscatter communication link with the SUEs. Therefore, to establish communication, we need to use the cooperative relay structure. An appropriate idea for this is to use the BS capabilities, which is located in the center of the cell and can receive and amplify the SBD's signal. This amplified signal reaches the SUEs through a direct link and with the help of ASRIS. As a result, the communication between SBDs and SUEs is established through the cooperation of BS and ASRIS. This method significantly increases the range and quality of communication.

In this section, we analyze the signal model for communication between the SBDs, BS, ASRIS, and SUEs. The BS establishes a direct communication link with SUE$_j$, j $\in \{1, 2, \ldots, i, \ldots, I\}$ in reflection space of ASRIS, while there is no direct link between BS and the SUE$_j$ in transmission space of ASRIS.

Since the primary objective of this paper is to explore the implementation of IoT users communications leveraging the infrastructure and spectrum of a cellular network, and to examine how components of this infrastructure, including the BS and ASRIS, can enhance the reliability and stability of SBD and SUE communications, we have chosen to exclude the study of intra-network communications within the cellular network itself (i.e., communications between the BS and SUEs). It is also worth noting that, given the minimal data volume generated by SBD users, they are unlikely to have any adverse impact on the cellular network under practical conditions.

A. System Model of the Active STAR-RIS

The STAR-RISs are advanced devices that enable independent control of the transmitted and reflected signals. Specifically, the signal incident upon the mth element of the ASRIS is denoted by s_m, where $m \in \mathcal{M} \triangleq \{1, 2, ..., M\}$. The mth element can adjust the amplitude and phase of the incident signal during transmission and reflection, resulting in transmitted and reflected signals given by $\left(\sqrt{\beta_m^t} e^{j\theta_m^t}\right) s_m$ and $\left(\sqrt{\beta_m^r} e^{j\theta_m^r}\right) s_m$, respectively. Here, β_m^t and θ_m^t represent the amplitude and phase shift adjustments made by the mth element during transmission, while β_m^r and θ_m^r represent the corresponding adjustments during reflection. The signals transmitted and reflected by each element of the ASRIS can be accurately modeled with this approach [40]:

$$t_m = \left(\sqrt{\beta_m^t} e^{j\theta_m^t}\right) s_m, \quad \forall m \in \mathcal{M} \tag{1}$$

$$r_m = \left(\sqrt{\beta_m^r} e^{j\theta_m^r}\right) s_m, \quad \forall m \in \mathcal{M}. \tag{2}$$

In the ASRIS structure each element can operates in simultaneous transmission and reflection mode, which is more general than either full transmission mode or full reflection mode [18], [41]. As a result, we adopt the operating protocol for the ASRIS known as energy splitting, where all elements of the ASRIS are assumed to operate in transmission and reflection mode. The energy splitting mode provides more degrees of freedom for optimizing the network, but it also increases the communication overhead between the BS and ASRIS [41]. In order to reduce complexity and provide greater clarity, we propose a simplified equal energy-splitting protocol, similar to that in [42]. This protocol involves setting $\beta_m^t = \beta_m^r = \frac{p_{\text{ASRIS}}}{2}$ in both transmission and reflection modes, where p_{ASRIS} is the power ASRIS amplifier, intended for the elements, remains constant over time.

For ease of expression, we define the transmission (when $l = t$) or reflection (when $l = r$) beamforming vector as $\mathbf{R}_l = [\sqrt{\beta_1^l} e^{j\theta_1^l}, \sqrt{\beta_2^l} e^{j\theta_2^l}, \ldots, \sqrt{\beta_M^l} e^{j\theta_M^l}]^H \in \mathbb{C}^{M\times 1}$, and let $\mathbf{\Theta}_l = \text{diag}(\mathbf{R}_l^H) \in \mathbb{C}^{M\times M}$ be the corresponding diagonal beamforming matrix of ASRIS.

In the context of the passive structure of STAR-RIS, it is important to note that according to the energy conservation law, the energy of the incident signal must equal the sum of the energies of the transmitted and reflected signals, expressed as $|s_m|^2 = |t_m|^2 + |r_m|^2$ for all $m \in \mathcal{M}$. This implies a coupling between the amplitudes of the transmission and reflection coefficients, necessitating that $\beta_m^t + \beta_m^r = 1$.

B. Problem Formulation

We propose the implementation of a CSR as the considered structure for the SR network. In this configuration, the BS transmits K symbols (where $k = 1, 2, ..., K$ and $K \gg 1$) for every SBD's data symbol. This means that the duration of each symbol transmitted by SBD (T_{SBD}) is equal to K times of the duration of symbol transmission by BS (T_{BS}). With this CSR design, we ensures that the SBDs and BS signals do not interfere with each other at the destination.

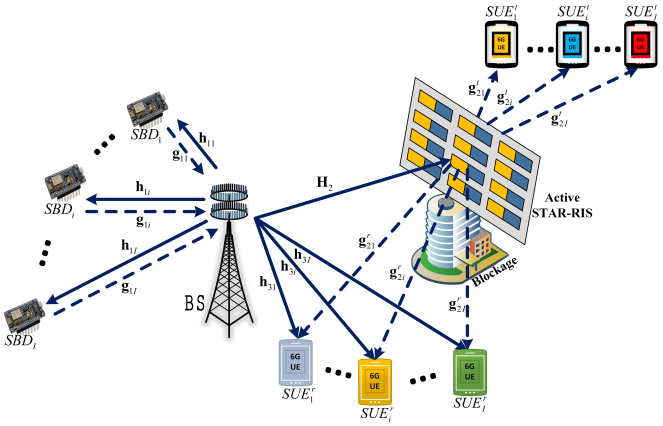

Fig. 1: Symbiotic radio system model with SBDs, STAR-RIS and SUEs.

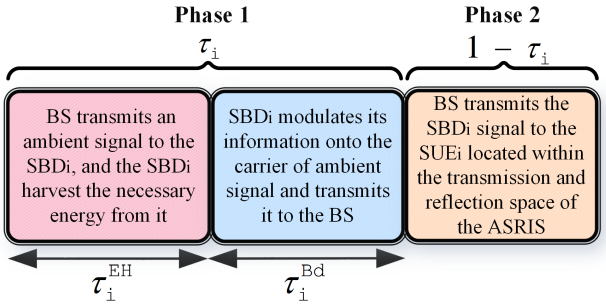

Fig. 2: The time division duplexing in the 6G Cellular-CSR network.

In the proposed cellular-CSR network, time division duplexing for sending and receiving information from the BS to the SBDs and then from SBDs to BS and SUEs is considered in two phases, which occurs within a time unit of 1. This is illustrated in Fig. 2. In the first phase, the CSR network is implemented, and the information from the SBDs reaches the BS. In the second phase, this information is sequentially delivered to its intended destination within a cellular network. The signal analysis for each phase is as follows.

In the first phase, the BS transmits the known signal $s_k(t)$ to SBD$_i$ through the complex channel $\mathbf{h}_{\text{li}}^H \in \mathbb{C}^{1\times N}$. This transmission is achieved using the transmit beamforming vector $\mathbf{w}_{\text{li}} \in \mathbb{C}^{N\times 1}$ and with power P_i. The received signal at SBD$_i$ can be obtained as follows:

$$y_k^{\text{SBD}_i}(t) = \sqrt{P_i} \mathbf{h}_{\text{li}}^H \mathbf{w}_{\text{li}} s_k(t) + n_k^{\text{SBD}_i}(t), \forall i \in \psi, \tag{3}$$

where $\psi \triangleq \{1, 2, ...i, ..., I\}$ represents the collection of all SBDs or all SUEs. Upon receiving the ambient signal $y_k^{SBD_i}(t)$, the SBDs harvests energy for their electronic circuits. These steps are carried out within the time duration τ_i^{EH}.

In this paper, we utilize NOMA to facilitate multiple access for SBDs and SUEs, optimize resource allocation among them, and mitigate excessive interference at the destination. Therefore, the BS transmits signals to each of the SBDs with different power levels based on their distances and channels gain. As shown in Fig. 1, the allocated power to the SBDs is

in the form of $P_I > ... > P_i > ... > P_1$. Here, we assume that the transmitted symbol from the ambient signal sent to all SBDs is the same and equal to $s_k(t)$.

After the SBDs are charged with the ambient signal $y_k^{\mathrm{SBD}_i}(t)$, they subsequently modulate their information, denoted as $c_i(t)$, $i \in \psi$, onto the ambient carrier signal. It then backscatters the modulated signal back to the BS using the SBD's reflection coefficient η_i, $0 \leq \eta_i \leq 1$. These steps are carried out within the time duration τ_i^{Bd}. It should be noted that since SBDs does not have any power-consuming active components, $n_k^{\mathrm{SBD}_i}(t)$, which is the complex gaussian noise at the SBD's antenna, is very small and can be disregarded [43]. Assuming that the beamforming vector used during signal reception at the BS is denoted as $\mathbf{w}_{Ri}^H \in \mathbb{C}^{1\times N}$, the received signal at the BS can be expressed as follows:

$$y_k^{\mathrm{BS}}(t) = s_k(t) \sum_{j\in\psi} \sqrt{P_j \eta_j}\mathbf{w}_{Rj}^H \mathbf{g}_{1\mathrm{j}} \mathbf{h}_{1\mathrm{j}}^H \mathbf{w}_{1\mathrm{j}} c_j(t) + n_k^{\mathrm{BS}}(t), \tag{4}$$

where $n_k^{BS} \sim \mathcal{CN}\left(0, \sigma_{BS}^2\right)$ is the circularly symmetric complex gaussian (CSCG) noise at the BS and $\mathbf{g}_{1\mathrm{i}} \in \mathbb{C}^{N\times 1}$ is the complex channel vector from SBD_i to BS.

In this paper, we assume that all channels experience are flat fading and remain constant within a given time frames. Channel estimation for active users is typically performed using the pilot signal method. For passive users in SR networks, several channel estimation techniques can be applied, some of which are explored in the papers [44], [45], [46]. Therefore, because it is beyond the scope of this paper, we assume that the channel state information (CSI) for all channels is readily available.

The BS is designed with a robust infrastructure and powerful processors, which enable it to perform complex signal processing tasks. Moreover, since the signal $s_k(t)$ is specific to the BS and well known, and given the availability of CSI for channels $\mathbf{g}_{1\mathrm{i}}$ and $\mathbf{h}_{1\mathrm{i}}^H$, the BS is capable of efficiently extracting the information $c_i(t)$ for SBD_i in its baseband and remove any other noise from it. Furthermore, to eliminate interference caused by other SBDs, which have a stronger signal to interference plus noise ratio (SINR) compared to SBD_i and simultaneously transmit their signals to the BS, the SIC method is employed.

So far, the desired signal for SBD_i has been successfully received and decoded by the BS. To enhance the signal strength and improve reception quality, the maximal ratio combining (MRC) technique is applied at the BS receiver. The corresponding beamforming vector is given by $\mathbf{w}_{Ri} = \frac{\mathbf{g}_{1\mathrm{i}}}{\|\mathbf{g}_{1\mathrm{i}}\|}$.

In the CSR setup, a single SBD symbol is transmitted over K consecutive BS symbol intervals, enabling the primary signal $s_k(t)$ to be interpreted as a spread-spectrum code of length K for the SBD symbols. Consequently, the SINR for decoding the SBD symbol $c_i(t)$ increases by a factor of K, though this comes at the cost of reducing the symbol rate by a factor of $1/K$. Therefore, by considering Eq. 4 and assuming $\mathbb{E}\left[|s_k(t)|^2\right] = 1$, the achievable rate, denoted as $R_{\mathrm{SBD}_i}^1$, after applying the SIC technique at the BS during the first phase, can be determined as follows:

$$R_{\mathrm{SBD}_i}^1 = \frac{B\tau_i^{Bd}}{K}\log_2\left(1 + \frac{KP_i\eta_i\left|\mathbf{g}_{1\mathrm{i}}\mathbf{h}_{1\mathrm{i}}^H\mathbf{w}_{1\mathrm{i}}\right|^2}{\sum\limits_{j\in\upsilon} P_j\eta_j\left|\mathbf{g}_{1\mathrm{j}}\mathbf{h}_{1\mathrm{j}}^H\mathbf{w}_{1\mathrm{j}}\right|^2 + B\sigma_{\mathrm{BS}}^2}\right), \tag{5}$$

where B represents the bandwidth of the receiver filter in the BS. Without loss of generality, for the sake of simplicity, we can consider the bandwidth is $B = 1$ Hz in all scenarios. Furthermore, $\tau_i = \tau_i^{\mathrm{EH}} + \tau_i^{\mathrm{Bd}}$ denotes the allocated time for completing the first phase. Also, $\upsilon \triangleq \{1, ..., i-1\}$ represents the set of all SBDs whose distance to the BS is less than that of SBD_i, and consequently, these users receive lower power allocation. These SBDs cause interference on SBD_i, whose total interference is shown in the denominator of the SINR fraction. It is worth noting that, since in this design we aim to examine the data exchange rate of users, we conduct this analysis in the worst-case scenario, which involves the simultaneous reception of signals from all SBDs at the BS. It is obvious that in a real and operational scenario, where signals are not received simultaneously, there will be less interference, and as a result, the output of the design will be more favorable.

As stated, SBDs must first harvest the necessary amount of energy wirelessly from ambient waves in order to establish communication. According to Eq. 3, the energy that can be harvested by $\mathrm{SBD_i}$ ($\varepsilon_{\mathrm{SBD}_i}$) during the time slot τ_i^{EH}, can be expressed as follow:

$$\varepsilon_{\mathrm{SBD}_i} \leq \Gamma_i P_i \tau_i^{\mathrm{EH}}(1-\eta_i)|\mathbf{h}_{1\mathrm{i}}^H\mathbf{w}_{1\mathrm{i}}|^2, \quad i \in \psi \tag{6}$$

where, $0 \leq \Gamma_i \leq 1$ is the energy conversion efficiency by SBD_i. The maximum energy harvested by $\mathrm{SBD_i}$, occurs when the power reflection coefficient η_i is equal to zero.

Although ASRIS could indeed be used to support and improve SBD communications, we refrained from using it in this phase to reduce the complexity of the mathematical relationships. Additionally, since we are investigating the effects of using ASRIS in the cellular network (second phase) and the same effects can occur in the assumed CSR network, we chose not to apply ASRIS here.

In the second phase, the BS transmits the signal $c_i(t)$, $i \in \psi$ to the destinations ASRIS and SUE_i, $i \in \psi$. Specifically, SUE_i is located in the reflection space of ASRIS, denoted as SUE_i^r. Both ASRIS and SUE_i^r have direct links with the BS. The transmission is carried out with a power P_i to SUE_i^r using the beamforming vector $\mathbf{w}_{2\mathrm{i}} \in \mathbb{C}^{N\times 1}$. In this section, we also utilize NOMA for accessing multiple SUEs present in the reflecting space of ASRIS. In this scenario, SUEs are located at different distances from BS and ASRIS, and therefore, the power of the transmitted signal is allocated based on their distance and channels gain. In this scenario, the received signal that reaches the ASRIS through the channel $\mathbf{H}_2^H \in \mathbb{C}^{M\times N}$ is given by:

$$\mathbf{y}_{\mathrm{ASRIS}}(t) = \mathbf{H}_2^H \sum_{j\in\psi} \sqrt{P_j}\mathbf{w}_{2\mathrm{j}} c_j(t) + \mathbf{n}_{\mathrm{ASRIS}}(t). \tag{7}$$

As mentioned above, since the ASRIS has active components, the presence of CSCG noise, denoted as $\mathbf{n}_{\mathrm{ASRIS}} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{\mathrm{ASRIS}}^2\mathbf{I_M}\right)$, should be taken into account.

On the other hand, considering Fig. 1, the received signal at SUE_i^r, which arrives through both the direct link from the BS and the reflected signal from ASRIS, can be obtained as follows:

$$\begin{aligned} y_{\text{SUE}_i}^r(t) = \left(\mathbf{g}_{2i}^r \boldsymbol{\Theta}_r \mathbf{H}_2^H + \mathbf{h}_{3i}^H\right) \sum_{j\in\psi} \sqrt{P_j}\mathbf{w}_{2j} c_j(t) \\ + \mathbf{g}_{2i}^r \boldsymbol{\Theta}_r \mathbf{n}_{\text{ASRIS}}(t) + n_{\text{SUE}_i}^r(t), \end{aligned} \tag{8}$$

where $\mathbf{g}_{2\text{i}}^r \in \mathbb{C}^{1\times M}$ is the channel vector between BS and SUE_i^r in the reflection space of the ASRIS, $\mathbf{h}_{3\text{i}}^H \in \mathbb{C}^{1\times N}$ is the channel vector for the direct link between BS and SUE_i^r, and $n_{\text{SUE}_i}^r \sim \mathcal{CN}\left(0, \sigma_{\text{SUE}_i^r}^2\right)$ is the CSCG noise at the SUE_i^r.

According to (8) the maximum achievable rate after using the SIC technique in the SUE_i^r, denoted as $R_{\text{SUE}_i}^{2,r}$ is

$$R_{\text{SUE}_i}^{2,r} = (1-\tau_i)\log_2\left(1+\gamma_{\text{SUE}_i}^{2,r}\right), \tag{9}$$

where

$$\gamma_{\text{SUE}_i}^{2,r} = \frac{P_i \mathbf{w}_{2i}^2 \left|\overline{\mathbf{h}_i^r}\right|^2}{\left|\overline{\mathbf{h}_i^r}\right|^2 \sum\limits_{j\in\upsilon} P_j \mathbf{w}_{2j}^2 + \left\|\mathbf{g}_{2i}^r \boldsymbol{\Theta}_r\right\|^2 \sigma_{\text{ASRIS}}^2 + \sigma_{\text{SUE}_i^r}^2}, \tag{10}$$

where $1-\tau_i$ denotes the allocated time for completing the second phase and $\overline{\mathbf{h}_i^r} \triangleq \mathbf{g}_{2\text{i}}^r \boldsymbol{\Theta}_r \mathbf{H}_2^H + \mathbf{h}_{3\text{i}}^H$. In this relation, υ represents the set of all SUEs in reflection space, whose channels gain (BS-SUE_i^r and BS-ASRIS-SUE_i^r) is higher than SUE_i^r and as a result have lower power allocation.

Also, due to obstacle, a direct link between the BS and SUE_i in the transmission space (SUE_i^t) of ASRIS is not available. Consequently, only ASRIS has the capability to transmit the signal to SUE_i^t using its own transmission space. It's important to note that if any other structure, such as an active or passive RIS, were used instead of ASRIS, SUE_i^t would be located in a cellular blind spot. In this scenario, the received signal at SUE_i^t , which arrives through the transmission link by ASRIS, can be obtained as follows:

$$\begin{aligned} y_{\text{SUE}_i}^t(t) = \mathbf{g}_{2\text{i}}^t \boldsymbol{\Theta}_t \mathbf{H}_2^H \sum_{j\in\psi} \sqrt{P_j}\mathbf{w}_{2\text{j}} c_j(t) \\ + \mathbf{g}_{2\text{i}}^t \boldsymbol{\Theta}_t \mathbf{n}_{\text{ASRIS}}(t) + n_{\text{SUE}_i}^t(t), \end{aligned} \tag{11}$$

where $\mathbf{g}_{2\text{i}}^t \in \mathbb{C}^{1\times M}$ is the channel vector between BS and SUE_i^t in the transmission space of the ASRIS, and $n_{\text{SUE}_i}^t \sim \mathcal{CN}\left(0, \sigma_{\text{SUE}_i^t}^2\right)$ is the CSCG noise at the SUE_i^t.

According to (11) the maximum achievable rate after using the SIC technique in the SUE_i^t, denoted as $R_{\text{SUE}_i}^{2,t}$ is

$$R_{\text{SUE}_i}^{2,t} = (1-\tau_i)\log_2\left(1+\gamma_{\text{SUE}_i}^{2,t}\right), \tag{12}$$

where

$$\gamma_{\text{SUE}_i}^{2,t} = \frac{P_i \mathbf{w}_{2i}^2 \left|\overline{\mathbf{h}_i^t}\right|^2}{\left|\overline{\mathbf{h}_i^t}\right|^2 \sum\limits_{j\in\upsilon} P_j \mathbf{w}_{2j}^2 + \left\|\mathbf{g}_{2i}^t \boldsymbol{\Theta}_t\right\|^2 \sigma_{\text{ASRIS}}^2 + \sigma_{\text{SUE}_i^t}^2}, \tag{13}$$

where $\upsilon \triangleq \{1, ..., i-1\}$ represents the set of all SUEs whose distance to ASRIS is less than SUE_i^t and as a result have higher channel gain. Also, $\overline{\mathbf{h}_i^t} \triangleq \mathbf{g}_{2\text{i}}^t \boldsymbol{\Theta}_t \mathbf{H}_2^H$.

It should be noted that in this article, all communication channels with ASRIS are considered to be Rician, and all communication channels with BS are considered to be Rayleigh. Additionally, due to the direct LoS between BS and ASRIS, we also consider this channel to be Rician.

Furthermore, the assumption of an equal number of SUEs and SBDs is made purely for clarity and simplicity in formulating the equations. Any variation in these numbers would impact the interference levels for each user but would not alter the overall concept presented in the paper.

III. THROUGHPUT MAXIMIZATION

Based on the explanations provided in the preceding section and in accordance with equations (5), (9), and (12), the main objective of this article is to maximize the throughput in the network. This objective is achieved when the minimum information rate between users in Phase 1 and Phase 2 is maximized. To accomplish this, we define the optimization problem as follows:

$$\max_{\eta_i,\tau_i,\tau_i^{\text{EH}},\tau_i^{\text{Bd}},P_i,\beta_m^l,\theta_m^l} \min\left(\text{R}_{\text{SBD}_i}^1, R_{\text{SUE}_i}^{2,l}\right) \tag{14}$$

$s.t.$

$$\left(|\beta_m^t|^2, |\beta_m^r|^2\right) \le \frac{p_{\text{ASRIS}}}{2}, \quad p_{\text{ASRIS}} = \text{cte} \tag{14a}$$

$$0 \le \theta_m^l \le 2\pi, \qquad \forall m \in \mathcal{M} \tag{14b}$$

$$P_i \le p_{\text{BS}}, \qquad i \in \psi, p_{\text{BS}} = \text{cte} \tag{14c}$$

$$0 \le \eta_i \le 1, \quad i \in \psi \tag{14d}$$

$$0 \le \tau_i \le 1, \quad i \in \psi \tag{14e}$$

$$\tau_i^{\text{EH}} + \tau_i^{\text{Bd}} \le \tau_i \tag{14f}$$

$$\varepsilon_{\text{SBD}_i} \le \Gamma P_i \tau_i^{\text{EH}} |\mathbf{h}_{1\text{i}}^H \mathbf{w}_{1\text{i}}|^2, \quad i \in \psi \tag{14g}$$

$$R_{\text{SUE}_i}^{2,l} > R_{\text{SUE}_{i+1}}^{2,l}, \quad i \in \psi, \tag{14h}$$

where l denotes the SUEs located in the transmission and reflection areas ($l = t, r$). The (14a) constraint limits the power of the ASRIS to its maximum value in both transmission and reflection modes, while the constraint (14b) pertains to the phase of the ASRIS elements. Additionally, constraint (14f) specifies the maximum power of the BS for transmitting signals to the SBD_i, SUE_i^r and ASRIS. To allocate resources more effectively, this BS power is treated as a variable in the problem, which cannot exceed the maximum specified value p_{BS}. Furtheremore, constraint (14d) is considered to limit the rate (or power reflection) of modulated the information on the ambient carrier by SBD_i. Constraint (14e) sets the maximum duration for completing phase 1. Relation 14f is intended to set constraints on the energy harvesting and data transmission times for each SBD. Equation (14g) restricts the energy harvested by the SBD_i to its maximum specified value. Since we assume that all SBDs are of approximately the same type in this design, we can consider the Γ value to be the same for all of them. Also, to guarantee the SIC performed successfully, the conditions (14h) should be satisfied.

It should be noted that, since SBDs merely place their own information on the carrier frequency of ambient waves, no decoding of these signals is performed. Therefore, there is no need to consider the SIC constraint for the first phase of this

system model. In scenarios where decoding the BS signal is necessary, or when information exchange between SBDs in the network is required, information decoding in the SBD would also be necessary, and therefore, the SIC technique would be needed. In this case, we must consider a constraint such as $R^1_{\mathrm{SBD}_i} > R^1_{\mathrm{SBD}_{i+1}}, i \in \psi$ for it. However, in this case, the SBD hardware would become more complex and might not be implementable with simple passive circuits.

To make the optimization problem (14) implementable, certain modifications are required. Given the requirement of uninterrupted and collision-free transmission of information from SBD_i to the SUE^l_i, the maximum achievable throughput is attained when $R^1_{\mathrm{SBD}_i} = R^{2,l}_{\mathrm{SUE}_i} = R$. Therefore, the final optimization problem for maximizing the achievable rate can be stated as follows:

$$\max_{R,\eta_i,\mathbf{w}_{1\mathrm{i}},\mathbf{w}_{2\mathrm{i}},\tau_i,\tau_i^{\mathrm{EH}},\tau_i^{\mathrm{Bd}},P_i,\beta^l_m,\theta^l_m} R \tag{15}$$

$s.t.$

$$2^{\left(\frac{KR}{\tau_i^{Bd}}\right)} - 1 \leq \frac{KP_i\eta_i\left|\mathbf{g}_{1\mathrm{i}}\mathbf{h}^H_{1\mathrm{i}}\mathbf{w}_{1\mathrm{i}}\right|^2}{\sum\limits_{j\in\upsilon} P_j\eta_j\left|\mathbf{g}_{1\mathrm{j}}\mathbf{h}^H_{1\mathrm{j}}\mathbf{w}_{1\mathrm{j}}\right|^2 + \sigma^2_{\mathrm{BS}}} \tag{15a}$$

$$2^{\left(\frac{R}{1-\tau_i}\right)} - 1 \leq \gamma^{2,r}_{\mathrm{SUE}_i} \tag{15b}$$

$$2^{\left(\frac{R}{1-\tau_i}\right)} - 1 \leq \gamma^{2,t}_{\mathrm{SUE}_i} \tag{15c}$$

$$(14a), (14b), (14f), (14d), (14e), (14f), (14g), (14h), \tag{15d}$$

in problem (15), the constraints (15a), (15b), and (15c) correspond to the limitations imposed by the maximum information rates in a communication channel, which are defined by the relations (5), (9), and (12), respectively. Additionally, in the simulations for this problem, (15b) is used when $l = r$, and (15c) is used when $l = t$.

Under the aforementioned conditions, depending on factors such as the data size transmitted from the SBDs to the BS, the parameter K, CSI between nodes and the BS, and ASRIS/BS beamforming coefficients, the execution times of Phase 1 and Phase 2 may be equal or different.

Problem (15) is a non-convex problem, and it can be implemented using convex optimization methods or various machine learning techniques. Considering that resource allocation problems in various articles have mostly been simulated using convex optimization methods and the CVX toolbox in Matlab, in this paper, we will model the problem using advanced deep reinforcement learning (DRL) methods and simulate it using a python program.

It is important to note that convex optimization methods are effective for well-posed, static problems with convex objectives. However, they are less suitable for the highly complex, non-convex problem formulated in Eq.(15), which involves dynamic, high-dimensional decision variables such as power allocations, beamforming vectors, and reflection coefficients, along with intricate constraints. DRL, on the other hand, offers several significant advantages in this context. It is capable of handling non-convexity and large-scale action spaces, adapting to dynamic environments with uncertainty, and optimizing long-term rewards through sequential decision-making. Moreover, DRL does not require an exact closed-form model of the system, making it more robust to uncertainties and enabling effective learning in situations where convex optimization methods may fail due to the problem's inherent complexity or non-stationarity. Consequently, DRL provides a flexible and scalable solution to this optimization problem, overcoming challenges that traditional convex optimization methods would struggle to address.

IV. Deep Reinforcement Learning

In this section, we initially transform the non-convex problem (15) into a model-free Markov decision process (MDP). Subsequently, we develop DRL algorithms utilizing PPO, TD3 and A3C to address and resolve problem (15).

A. MDP

The MDP formulation constructs a 4-tuple, denoted as $(\mathbf{s}_t, \mathbf{a}_t, \mathbf{r}_t, \mathbf{s}_{t+1})$, where the current state, action, reward function, and next state are represented by $\mathbf{s}_t$, $\mathbf{a}_t$, r_t, and $\mathbf{s}_{t+1}$, respectively. The agent interacts with the environment, observing the current state $\mathbf{s}_t$ from the state space $\mathcal{S}$ with $\mathbf{s}_t$ and selecting action $\mathbf{a}_t$ from the action space $\mathcal{A}$ with $\mathbf{a}_t$ according to its policy. Additionally, the formulation of problem (15) further details the state, action, and reward function.

1) State: The current state $\mathbf{s}_t \in \mathcal{S}$ at time step t encompasses crucial environmental information associated with problem (15). This configuration enables the policy to improve and adjust in response to the dynamic environment. To elaborate, the state $\mathbf{s}_t$ for the analyzed system includes the all channels of environment expressed as follows:

$$\mathbf{s}_t = \{h_{1\mathrm{i}}, h_2, h_{3\mathrm{i}}, g_{1\mathrm{i}}, g^l_{2\mathrm{i}}\}, \quad \forall i \in \psi, \forall t \in T, l = r, t. \tag{16}$$

2) Action: The term action denoted by $\mathbf{a}_t \in \mathcal{A}$ at time step t encompasses the decisions and choices undertaken by an agent as it interacts with the relevant states. These actions represent the agent's responses and strategies employed to navigate and influence the dynamics of the considered system during the specified time instance [47]. Every variable within problem (15) serves as an action, implying that each element or parameter in the problem formulation represents a distinct decision or manoeuvre in addressing and resolving the challenges posed by the problem. In essence, these variables act as the modifiable components through which the agent, system, or solver can influence and effect changes in pursuit of optimal solutions or outcomes for problem (15). As a result, the set of actions can be described as follows:

$$\begin{gathered}\mathbf{a}_t = \{R, \eta_i, \mathbf{w}_{1\mathrm{i}}, \mathbf{w}_{2\mathrm{i}}, \tau_i, P_i, \beta^l_m, \theta^l_m\}, \\ \forall i \in \psi, \forall m \in M, \forall t \in T, l = r, t.\end{gathered} \tag{17}$$

3) Reward: The DRL methodologies instruct agents to make appropriate decisions to maximize the reward. In the optimization problem (15), the reward function aligns with the objective function, guiding the training process towards

actions that contribute to the overall goal of optimizing the specified objective

$$\mathbf{r}_t = R_t + \sum_{j=1}^{11} l_{C_j}, \quad \forall t \in T, \tag{18}$$

where $l_{C_j} = \alpha_j R_t$, and the index j corresponds to all constraints, i.e., $\forall j \in \{1, 2, ..., 11\}$. Besides, $\alpha_j = 1$, if the C_j-th constraint is satisfied and $\alpha_j = 0$, otherwise.

B. PPO Algorithm

The PPO algorithm, functioning as an actor-critic on-policy gradient method, simplifies the intricate computations involved in earlier policy gradient methods like trust region policy optimization (TRPO) [48].

In the context of reinforcement learning, the main objective is to maximize the expected cumulative reward, taking into account a protracted temporal process. Consequently, the cumulative reward at time step t is represented as $\mathcal{R}_t = \sum_{t=0}^{\infty} \lambda^t r_t$, where $\lambda \in [0, 1)$ signifies the discount factor. To delve into specifics, both actor and critic networks are employed to portray the parameterized stochastic policy of action selection, denoted as $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)$, and the state-value function $V_\phi(\mathbf{s}_t)$, respectively. Here, θ and ϕ stand for the parameters of the actor and critic networks. Subsequently, a surrogate objective function, constructed based on the PPO approach, can be articulated as follows:

$$\mathcal{L}(\theta, \mathbf{s}_t, \mathbf{a}_t) = \mathbb{E}\left[\beta_t(\theta)\Omega(\mathbf{s}_t, \mathbf{a}_t)\right]. \tag{19}$$

The probability ratio between the current policy and the previous one is denoted as $\beta_t(\theta) = \pi_\theta(\mathbf{a}_t|\mathbf{s}_t)/\pi_{\theta^{\text{old}}}(\mathbf{a}_t|\mathbf{s}_t)$, where θ^{old} represents the parameter associated with the old policy in the actor network. Additionally, the advantage function is expressed as follows:

$$\Omega(\mathbf{s}_t, \mathbf{a}_t) = r(\mathbf{s}_t, \mathbf{a}_t) + \lambda V_{\phi^{\text{old}}}(\mathbf{s}_{t+1}) - V_{\phi^{\text{old}}}(\mathbf{s}_t), \tag{20}$$

where ϕ^{old} signifies the parameter associated with the critic network for the previous state-value estimation function. Subsequently, a mini-batch stochastic gradient descent (SGD) technique is employed to update the associated θ across a set of Q transitions denoted as $(\mathbf{s}_t^q, \mathbf{a}_t^q, \mathbf{r}_t^q, \mathbf{s}_{t+1}^q)$ sampled from an experience buffer, which is given by

$$\theta = \theta^{\text{old}} - \delta_A \frac{1}{Q} \sum_{q=1}^{Q} \nabla_\theta \tilde{\mathcal{L}}_q\left(\theta, \mathbf{s}_t^q, \mathbf{a}_t^q\right), \tag{21}$$

where δ_A denotes the learning rate (LR), and $\tilde{\mathcal{L}}_q\left(\theta, \mathbf{s}_t^q, \mathbf{a}_t^q\right)$ represents the instantiation of $\mathcal{L}\left(\theta, \mathbf{s}_t, \mathbf{a}_t\right)$ with the q-th transition, respectively. The mini-batch stochastic gradient descent (SGD) utilized for updating ϕ employs the mean squared error (MSE) loss function in the following manner:

$$\phi = \phi^{\text{old}} - \delta_C \frac{1}{Q} \sum_{q=1}^{Q} \nabla_\phi \big(V_\phi(\mathbf{s}_t^q) - \hat{\mathcal{R}}(\mathbf{s}_t^q, \mathbf{a}_t^q)\big)^2, \tag{22}$$

where the learning rate is denoted as δ_C. Additionally, the target state-value function, indicated as $\hat{R}(\mathbf{s}_t, \mathbf{a}_t)$, is expressed as:

$$\hat{R}(\mathbf{s}_t, \mathbf{a}_t) = r(\mathbf{s}_t, \mathbf{a}_t) + \lambda V_{\phi^{\text{old}}}(\mathbf{s}_{t+1}). \tag{23}$$

The PPO-based approach are outlined in Algorithm (1). To elaborate, the action $\mathbf{a}_t$ is generated based on the particular policy within the current state $\mathbf{s}_t$, resulting in the acquisition of the reward $\mathbf{r}_t$. Subsequently, the transition $(\mathbf{s}_t, \mathbf{a}_t, \mathbf{r}_t, \mathbf{s}_{t+1})$ is recorded in the experience buffer, from which Q instances are sampled. In the following, the advantage function $\Omega(\mathbf{s}_t, \mathbf{a}_t)$ in (20) is computed. Finally, the relevant actor and critic parameters undergo updating through mini-batch SGD.

Algorithm 1 The PPO Algorithm

1- Initialize the environment parameters and the parameters of actor and critic networks, i.e., θ, ϕ, ε, δ_A and δ_C
2- Set $\theta^{\text{old}} = \theta$ and $\phi^{\text{old}} = \phi$
3- **For** each episode do
4- Reset environment and initialize position of users randomly
5- Initialize state $\mathbf{s}_0$ accroding to (16)
6- **For** each step do
7- Generate action $\mathbf{a}_t$ according to $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)$ in state $\mathbf{s}_t$
8- Calculate reward $\mathbf{r}_t$
9- Observe the new state $\mathbf{s}_{t+1}$
10- Store $(\mathbf{s}_t, \mathbf{a}_t, \mathbf{r}_t, \mathbf{s}_{t+1})$ in the experience buffer
11- Calculate the advantage function $\Omega(\mathbf{s}_t, \mathbf{a}_t)$ in (20)
12- Calculate $\nabla_\theta \tilde{\mathcal{L}}_q\left(\theta, \mathbf{s}_t^q, \mathbf{a}_t^q\right)$ in (21)
13- Calculate $\nabla_\phi \big(V_\phi(\mathbf{s}_t^q) - \hat{R}(\mathbf{s}_t^q, \mathbf{a}_t^q)\big)^2$ in (22)
14- Calculate $\hat{R}(\mathbf{s}_t, \mathbf{a}_t)$ in (23)
15- Update θ and ϕ in (21) and (22), respectively
16- Update $\theta^{\text{old}} = \theta$ and $\phi^{\text{old}} = \phi$
17- **End FOR**
18- **End FOR**

C. TD3 Algorithm

The TD3 algorithm represents a reinforcement learning approach that is both model-free and off-policy. The state-action value function as follows:

$$q_\mu(\mathbf{s}_t, \mathbf{a}_t) = \mathbb{E}_{\Pr(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)}[\sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mu(\mathbf{s}_t)|\mathbf{s}_t = \mathbf{s}_0, \mathbf{a}_t = \mu(\mathbf{s}_0)], \tag{24}$$

where μ is parameters of the actor network, and $\gamma \in (0, 1]$ denotes the discount factor. The optimal policy is given by

$$\mu^\star(\mathbf{s}_t) = \arg \max_{\mu(\mathbf{s}_t) \in \mathcal{A}} \quad q_\mu(\mathbf{s}_t, \mu(\mathbf{s}_t)). \tag{25}$$

TD3 represents an enhanced iteration of the deep deterministic policy gradient (DDPG), introducing adjustments to mitigate the overestimation of state-action value and avert the generation of sub-optimal policies [49]. The specifics of these modifications are elaborated in the training network description. Similar to the PPO algorithm, the agent chooses its action based on the observed state, and the corresponding experience $(\mathbf{s}_t, \mathbf{a}_t, \mathbf{r}_t, \mathbf{s}_{t+1})$ is stored in the buffer. Consequently, the loss function for critic networks with parameter α_i is calculated as follows:

$$\mathcal{L}(\alpha_i) = \frac{1}{|Q|} \sum_{k=1}^{Q} \Big(q_\mu(\mathbf{s}_t^k, \mathbf{a}_t^k; \alpha_i) - y(r_t^k, \mathbf{s}_{t+1}^k)\Big)^2, \tag{26}$$

where y is calculated by

$$y(r_l^k, \mathbf{s}_{t+1}^k) = r_t^k + \gamma \min_i q_{\bar{\mu}}(\mathbf{s}_{t+1}^k, \tilde{\mathbf{a}}_{t+1}^k; \bar{\alpha}_i). \quad (27)$$

To adjust the parameters of critic networks, α_i, the gradient descent algorithm is employed on the loss function (26) using the following equation:

$$\alpha_i = \alpha_i - \theta_i \nabla_{\alpha_i} \mathcal{L}(\alpha_i), \quad (28)$$

where θ_i represents the learning rate. Also, the loss function and update parameters of the actor network are as follows:

$$\mathcal{L}(\mu) = \frac{1}{|Q|} \sum_{k=1}^{Q} q_\mu(\mathbf{s}_l^k, \mathbf{a}_l^k), \quad (29)$$

$$\mu = \mu - \tilde{\theta} \nabla_\mu \mathcal{L}(\mu). \quad (30)$$

The pseudo-code for the TD3 algorithm is presented in Algorithm (2).

Algorithm 2 The TD3 Algorithm

1- Initialize the environment and networks parameters, i.e., α, and μ
2- Set $\theta^{\text{old}} = \theta$ and $\phi^{\text{old}} = \phi$
3- **For** each episode do
4- Reset environment and initialize position of users randomly
5- Initialize state $\mathbf{s}_0$ accroding to (16)
6- **For** each step do
7- Observe state $\mathbf{s}_t$ and select action $\mathbf{a}_t$
8- Observe next state $\mathbf{s}_{l+1}$ and receive reward r_t
9- Store $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$ in $\mathcal{M}$.
10- Randomly sample a batch set B from reply buffer
11- Update the network parameters α_i, and μ using (28), (30), respectively.
12- **End FOR**
13- **End FOR**

D. A3C Algorithm

The A3C algorithm employs an actor-critic architecture, where an actor network makes policy decisions and a critic network evaluates the value of these decisions. The "Advantage" in A3C refers to the use of an advantage function, which measures the advantage of taking a particular action in a given state over the average action value [50]. the advantage function is used to decrease the variance in estimation, as expressed by the following equation:

$$\mathcal{A}_t(\mathbf{s}_t, \mathbf{r}_t; \mu, \alpha) = \mathcal{R}_t - V(s_t; \alpha), \quad (31)$$

where μ and α are the parameters of actor and critic network, respectively. Furthermore, $\mathcal{R}$ represents cumulative reward as follows:

$$\mathcal{R} = \sum_{i=0}^{k} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}; \alpha), \quad (32)$$

where k is the number of steps which A3C used for parameter updating.

Derived from the advantage function $\mathcal{A}_t$ the actor's loss function is expressed as

$$f_\pi(\mu) = \log \pi(a_t|s_t; \mu)(\mathcal{A}_t) + \beta H(\pi(s_t; \mu)), \quad (33)$$

Here, $H(\pi(s_t; \mu))$ represents an entropy term incorporated to promote exploration during training, preventing potential premature convergence [51]. The parameter β is utilized to regulate the intensity of entropy regularization, facilitating the balance between exploration and exploitation. The loss function for the estimated critic network is specified as:

$$f(\alpha) = (\mathcal{R}_t - V(s_t; \alpha))^2. \quad (34)$$

This is employed to update the value function $V(st; \theta_v)$. The update for the critic is executed using the following accumulated gradient:

$$d\alpha \leftarrow d\alpha + \frac{\partial (\mathcal{R}_t - V(s_t; \alpha))^2}{\partial \alpha'}. \quad (35)$$

The actor undergoes an update through the following process:

$$d\mu \leftarrow d\mu + \nabla_{\mu'} \log \pi(a_t|s_t; \mu')(\mathcal{A}_t) + \beta \nabla_{\mu'} H(\pi(s_t; \mu')). \quad (36)$$

All the steps of the A3C algorithm are outlined in Algorithm (3).

Algorithm 3 The A3C Algorithm

1- Initialize the global actor network and global critic network with parameters, α, and μ
2- Initialize the thread-specific actor and thread-specific critic network parameters α', and μ'
3- **For** each episode do
4- Reset environment and initialize position of users randomly
5- **For** each worker do
6- Initialize the gradients of global agent: $d\alpha = 0$, $d\mu = 0$
7- Synchronous parameters of each worker with global parameters $\alpha' = \alpha$, and $\mu' = \mu$
6- Obtain initial state s_0.
7- **For** each step do
8- Perform at under policy $\pi(s_t; \mu')$
9- Obtain reward r_t and new state s_{t+1}
10- **End FOR**
11- $R = \begin{cases} 0, & \text{for terminal state} \\ V(s_t; \alpha'), & \text{for non-terminal state} \end{cases}$
12- $R = r_t + \gamma R$
13- Obtain accumulate gradient α based on (35)
14- Obtain accumulate gradient μ based on (36)
15- **End FOR**
16- **End FOR**

In the following, each of the PPO, TD3, and A3C methods, modeling and simulations are conducted, and the outputs of each are compared with each other in relation to this modeling system.

TABLE II: Parameters set in the learning simulation.

Parameters	PPO	TD3	A3C
Number/Size of Actor,Critic	2/128,128	2/400,300	2/128,128
Min Batch Size	32	64	64
Actor Learning Rate	0.0001	0.0001	0.0001
Critic Learning Rate	0.001	0.001	0.001
Target Network Update	0.0005	0.0005	0.0005
Discount Factor	0.99	0.99	0.99
Policy Entropy Coefficient	0.01	-	-
Number of Workers	-	-	3
Number of Episodes	30000	30000	30000
Number of Steps	200	200	200

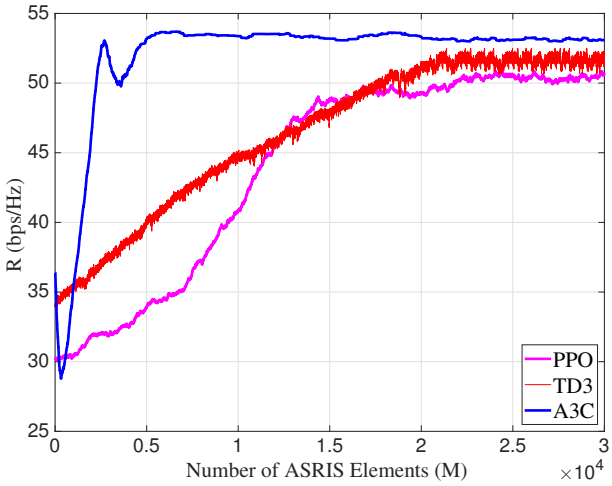

Fig. 3: Convergence plot of the three learning methods, PPO, TD3, and A3C, in the case of $N = 8, M = 16, I = 3, \varepsilon_{\text{SBD}_i} = 1$ μW.

V. Simulation Results

According to the proposed model, SBDs and SUEs are randomly distributed across the network space, as depicted in Fig. 1. In all simulations, constant values of $K = 100$, Rician factor = 10, $\Gamma = 0.8$, and $\sigma^2_{\text{ASRIS}} = \sigma^2_{\text{BS}} = \sigma^2_{\text{SUE}} = -120$ dBm have been considered. Additionally, we assume the carrier frequency of the ambient signal is 28 GHz, the path loss exponent is 3, and the transmit power of the BS and ASRIS are 16 watts and 8 watts, respectively. All simulation results are generated by averaging over 100 random channels. The maximum distance between the BS and SBDs is approximately 200 meters, while the distance between the BS and ASRIS extends up to about 400 meters. Within the ASRIS transmission area, the distance from the ASRIS to the SUEs is 300 meters. SUEs in the Reflection area are randomly distributed within the designated space between the BS and ASRIS.

The simulations were carried out on a laptop featuring an Intel Core i7-6500U 8 GB DDR3-RAM. It should be noted that the parameters related to learning simulation for each method are given in Table II.

A. Convergence of the Proposed Methods

As described, in this paper, we utilized three learning methods, namely PPO, TD3, and A3C, to tackle the non convex optimization problem. In this scenario, the convergence plot for each of these methods is depicted in Fig. 3.

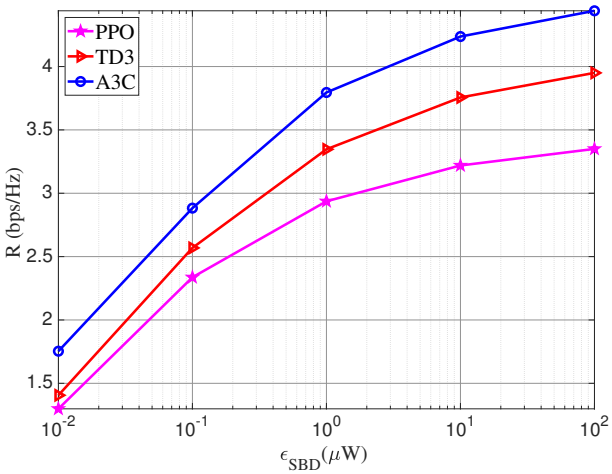

Fig. 4: Throughput maximization based on the energy harvested by SBDs in all three methods, PPO, TD3, and A3C, under the conditions $N = 8, M = 16, I = 3$.

Based on the information presented in Fig. 3, it is clear that the A3C method exhibits markedly superior convergence compared to the other two methods. Specifically, the convergence points for PPO, TD3, and A3C occur at approximately 17000, 22000, and 5000 episodes, respectively. These values indicate when network learning completes for each method. Furthermore, their convergence demonstrates the validity of their performance on this network.

B. Throughput Maximization Based on Energy Harvesting by SBDs

In the design of passive symbiotic radio networks, a pivotal factor is the quantity of energy harvested by the SBD devices. In this section, adhering to the constraint outlined in (14g), we depict the rate of information in the network as it correlates with the fluctuations in the energy harvesting capacity for each SBD.

As illustrated in Fig. 4, the information throughput in the network increases with the energy harvesting capability of the devices. This rise in throughput is due to the enhanced ability of each SBD to store energy from ambient waves, which in turn accelerates the speed of information modulation. Consequently, the first phase rate, linked to the CSR structure, also experiences an increase. It is worth noting that the average energy required for transmitting a pulse by passive IoT devices is about 1-10 μW [52]–[55].

Moreover, in this approach, it is apparent that utilizing the A3C algorithm leads to an increase in the total information exchange rate compared to the other two methods.

C. Throughput Maximization versus Transmit Power of BS and ASRIS

Since both BS and STAR-RIS are considered active in the proposed system model, in this section, we investigate the impact of the transmit power of each on the network's performance.

In Fig. 5, we investigate the influence of augmenting the transmit power of the BS from 4 to 32 watts across all three

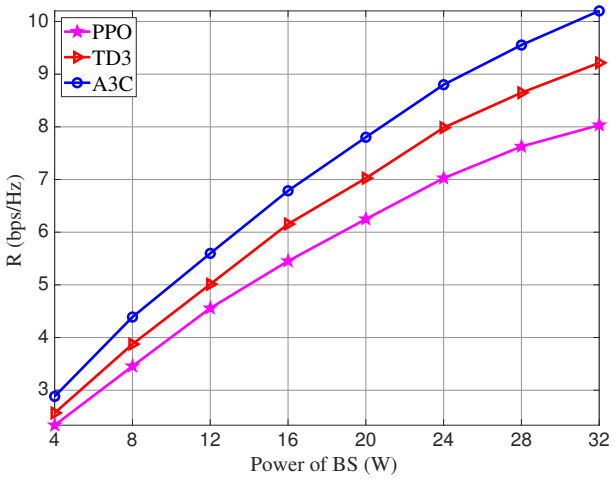

Fig. 5: Impact of increasing the transmitted signal power at the BS in relation to the user data rate in the case of $N = 8, M = 16, I = 3, \varepsilon_{\text{SBD}_i} = 1$ µW, $P_{\text{ASRIS}} = 10$ W.

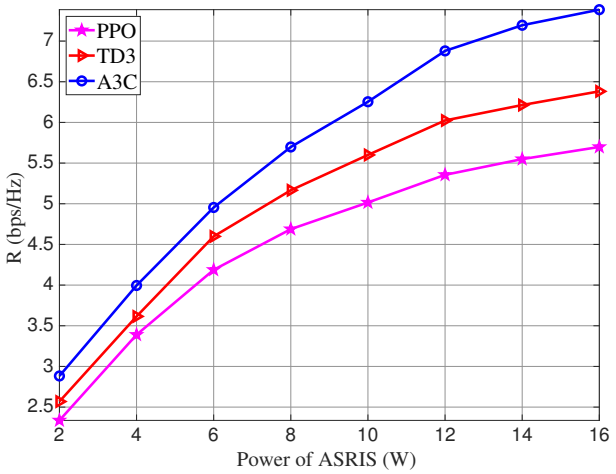

Fig. 6: Impact of increasing the transmitted signal power in ASRIS in relation to the user data rate in the case of $N = 8, M = 16, I = 3, \varepsilon_{\text{SBD}_i} = 1$ µW, $P_{\text{BS}} = 20$ W.

methods. Additionally, Fig. 6 illustrates the consequences of amplifying the transmit power of the ASRIS from 2 to 16 watts, which is caused by the reflection and transmission of the signals by it. The analysis encompasses all three methods, namely PPO, TD3, and A3C.

As expected, with the increase in the transmit power of signals in each BS or ASRIS, the data rate also increases. According to Figs. 5 and 6 , the A3C, TD3, and PPO methods allocate the highest processing efficiency for optimal resource allocation among users, consequently increasing the users' information rate. Also, considering the convergence depicted in Fig. 3, indicating better convergence of the A3C method compared to the other two methods, we conclude that this method is more suitable for modeling and implementing the proposed system in this research.

D. Throughput Maximization versus the Number of Elements in BS and ASRIS

As mentioned in the first part of this chapter, our system model incorporates a BS equipped with Massive MIMO antennas and an active STAR-RIS within the network. In this section, our objective is to examine the influence on the data rate of all users in the network by manipulating the number of elements in both the BS and ASRIS.

As illustrated in Fig. 7-(a), which belongs to the PPO method, increasing the number of elements in ASRIS from 16 to 128 exhibits a notable upward slope in the total data exchange rate. This phenomenon is attributed to more precise beamforming allocation and higher power for each of the SUEs. On the other hand, by augmenting the number of elements in BS from 8 to 32, the network's performance sees a significant improvement of approximately 133%. Consequently, users can engage in information exchange at a much higher rate compared to the previous rate In this scenario, the cumulative data exchange rate can achieve 10 bps/Hz.

An noteworthy consideration in this scenario is the increase in the volume of information processing in the network, which amplifies the computational complexity and implementation challenges as the number of elements grows. Therefore, careful consideration must be given to determining an optimal number of elements.

Now, in light of the aforementioned details, for better comparison between these methods, we plot the graphs for the TD3 and A3C methods.

As evident in Figs. 7-(b), and 7-(c), the data rate in the A3C method is, on average under similar conditions, higher at 1 bps/Hz compared to the TD3 method. This scale holds approximately true for the TD3 graphs in comparison to the PPO ones as well.

E. Comparison of Throughput Maximization versus the Number of Elements in Active and Passive STAR-RIS

In this section, we aim to assess the efficacy of activating a STAR-RIS as opposed to its passive counterpart. In the optimization problem (15) considered in the system model, we have constraint (14a), corresponding to the ASRIS mode in the network. To examine and compare the ASRIS structure with the passive STAR-RIS mode, we need to replace constraint (14a) with the relationship $\beta_m^t + \beta_m^r = 1$.

As depicted in Fig. 8, the information throughput is almost 2 bps/Hz higher when actively employing STAR-RIS compared to the passive utilization of STAR-RIS. This value further increases with the addition of more elements to the STAR-RIS configuration. Furthermore, it is evident that an ASRIS structure with 64 elements achieves the same efficiency as a passive structure with 128 elements. Put differently, the desired performance can be achieved with half the number of elements in this configuration. Therefore, reducing the number of elements in the active RIS within the network can significantly decrease implementation complexity and facilitate the estimation of communication channels in complex urban environments.

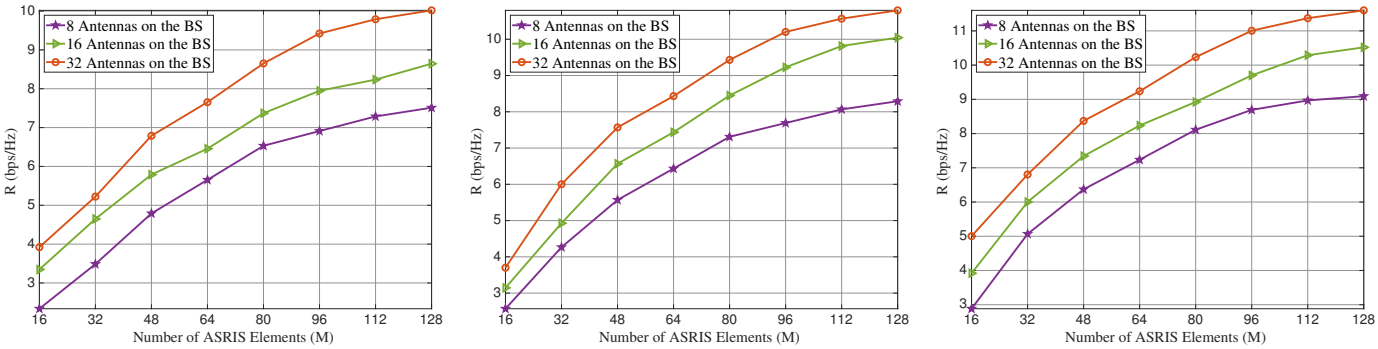

Fig. 7: Throughput maximization versus varying the number of elements in BS and ASRIS in **(a)** PPO, **(b)** TD3, and **(c)** A3C methods, with $I = 3, \varepsilon_{\text{SBD}_i} = 1$ μW.

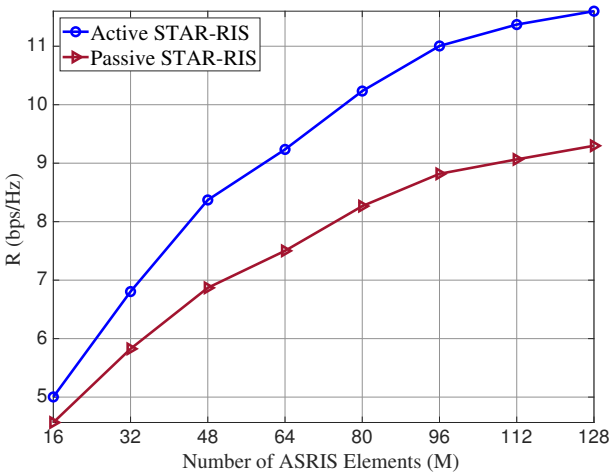

Fig. 8: Throughput maximization versus the number of elements in active and passive STAR-RIS in A3C method.

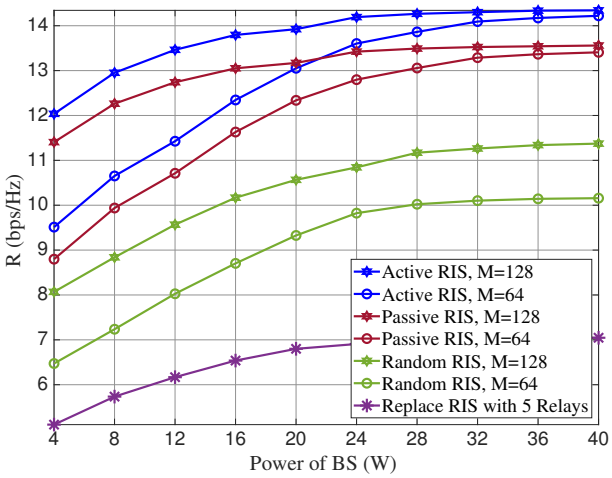

Fig. 9: Comparison of active, passive, random phase RIS states, and their replacement with 5 passive relays, $N = 8, I = 3, P_{\text{ASRIS}} = 10\ W$.

F. Comparison of Proposed Model with Other Baselines

To comprehensively evaluate the proposed scheme, we compare four different structures in this section:

1) ASRIS with 64 and 128 elements
2) Passive STAR-RIS with 64 and 128 elements
3) Random phase selection for RIS (without optimization)
4) Replacing RIS with several passive relays

For a fair comparison of the above schemes, we consider users on only one side since a relay cannot simultaneously serve users on both its front and back. Therefore, in this section, the STAR-RIS is converted to the RIS structure.

The baseline 3) indicates that the RIS parameters have not been optimized by the agent. Instead, we use randomly generated θ_m values during the simulation. By adopting this approach, we can assess the performance of the system without the influence of optimized parameters and compare it with more sophisticated strategies. In the fourth baseline, we assume that instead of using RIS in the network, several passive relays are employed to transmit information and enhance diversity to the destination. This method is discussed in paper [39]. Each of these relays has a structure similar to the SBDs used in the first phase, thereby creating a backscatter relay configuration within a cellular network.

We mathematically model all four of the aforementioned cases and simulate them using the A3C method, resulting in Fig. 9. As depicted in the figure, the proposed method in this paper demonstrates a significant advantage over other methods. In this configuration, using five relays not only fails to achieve optimal efficiency but also introduces considerable complexity to the network. It is noteworthy that increasing the number of relays makes it impossible to eliminate the interference they cause at the destination.

Another significant observation is the convergence of the 128 and 64 element diagrams in both active and passive modes of the RIS. This suggests that as the transmission power from the BS increases, it reaches a saturation point beyond which further increments do not substantially enhance the total network rate. Additionally, augmenting the number of RIS elements also ceases to significantly boost the overall rate beyond a specific threshold. On the other hand, comparing the Random phase diagrams with other diagrams where RIS phase optimization has been performed clearly shows that phase optimization in the network is essential for implementing telecommunication networks.

G. Power Consumption Comparison: passive STAR-RIS vs. ASRIS

We compare the power consumption of passive STAR-RIS and ASRIS with a 16-element structure. For passive STAR-RIS, each element employs a passive phase-shifting control circuit with a power consumption of approximately 6 mW (The power consumption of each phase shifter, which are 1.5, 4.5, 6, and 7.8 mW for resolutions of 3, 4, 5, and 6 bits, respectively [56]). Additionally, a switching circuit required for dual-mode operation (reflection and transmission) consumes about 1 mW per element [23]. Therefore, the total energy consumption for passive STAR-RIS is calculated as $P_{\text{total, passive STAR-RIS}} = M \times (P_{\text{control}} + P_{\text{switching}})$, where $M = 16$. Substituting the values yields $P_{\text{total, passive STAR-RIS}} = 16 \times (6\,\text{mW} + 1\,\text{mW}) = 112\,\text{mW}$.

For ASRIS, each element has a control circuit and switching circuit with the same power consumption as STAR-RIS, but also includes an active amplifier to boost the transmitted or reflected signal. The power consumption of the amplifier ranges from $50\,\text{mW}$ to $100\,\text{mW}$, with an average of $75\,\text{mW}$ considered in this analysis [23]. Thus, the total energy consumption for ASRIS is given by $P_{\text{total, ASRIS}} = M \times (P_{\text{control}} + P_{\text{switching}} + P_{\text{amplifier}})$. Substituting the values gives $P_{\text{total, ASRIS}} = 16 \times (6\,\text{mW} + 1\,\text{mW} + 75\,\text{mW}) = 1312\,\text{mW}$.

As can be observed, the energy consumption of ASRIS is significantly higher than that of STAR-RIS. Specifically, ASRIS consumes approximately 11.7 times more energy than passive STAR-RIS for a 16-element structure. Factors influencing ASRIS energy consumption include amplifier gain, number of elements, and material efficiency.

VI. Conclusion

In this article, we have explored a comprehensive system model incorporating various cutting-edge technologies, including massive MIMO, ASRIS, and both passive and active users. The objective of this initiative is to achieve optimal resource allocation among users, aiming to maximize the throughput across the entire network. In this system, for a practical modeling approach, we have incorporated constraints to ensure the minimum required QoS, impose limits on the maximum power for both BS and ASRIS, account for the constraints on the amount of energy harvesting for SBDs, and adhere to the requirements of satisfying SIC in the NOMA multiple access scheme. After mathematically modeling the target system, we encounter a non-convex and complex problem. To address the objectives inherent in this challenge, we leverage advanced DRL methods, including PPO, TD3, and A3C. The conducted simulations reveal that the A3C method, apart from achieving faster convergence, exhibits the capability to enhance the total throughput rate in the network when compared to the other two methods, TD3 and PPO. Additionally, the TD3 method significantly outperforms the PPO method. In the final segment of the simulation section, we draw the conclusion that the adoption of an active structure significantly influences the network throughput rate in comparison to the passive STAR-RIS mode. Additionally, in this section, we observed that using RIS in the network is significantly more efficient than using the backscatter relay structure. Moreover, with the help of the provided diagrams, we can determine the optimal network performance point for implementation by selecting the optimal number of RIS elements and the transmission power from the BS.

References

[1] L. Chettri and R. Bera, "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2020.

[2] X. You, C.-X. Wang, J. Huang, X. Gao, Z. Zhang, M. Wang, Y. Huang, C. Zhang, Y. Jiang, J. Wang *et al.*, "Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, pp. 1–74, 2021.

[3] Q. Wu, X. Zhou, W. Chen, J. Li, and X. Zhang, "IRS-aided WPCNs: A new optimization framework for dynamic IRS beamforming," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4725–4739, 2021.

[4] K. Dev, P. K. R. Maddikunta, T. R. Gadekallu, S. Bhattacharya, P. Hegde, and S. Singh, "Energy optimization for green communication in IoT using harris hawks optimization," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 685–694, 2022.

[5] M. N. Mahdi, A. R. Ahmad, Q. S. Qassim, H. Natiq, M. A. Subhi, and M. Mahmoud, "From 5G to 6G technology: meets energy, Internet-of-Things and machine learning: A survey," *Applied Sciences*, vol. 11, no. 17, p. 8117, 2021.

[6] L. Zhang, Y.-C. Liang, and D. Niyato, "6G Visions: Mobile ultra-broadband, super internet-of-things, and artificial intelligence," *China Communications*, vol. 16, no. 8, pp. 1–14, 2019.

[7] L. Zhang, Y.-C. Liang, and M. Xiao, "Spectrum sharing for Internet of Things: A survey," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 132–139, 2018.

[8] Z. Qin, X. Zhou, L. Zhang, Y. Gao, Y.-C. Liang, and G. Y. Li, "20 years of evolution from cognitive to intelligent communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 6–20, 2019.

[9] D. Samanta, C. K. De, and A. Chandra, "Performance analysis of full-duplex multi-relaying energy harvesting scheme in presence of multi-user cognitive radio network," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 2, pp. 626 - 634, 2022.

[10] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 39–50, 2013.

[11] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green Internet-of-Things," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1116–1130, 2018.

[12] R. S. Yeganeh, M. J. Omidi, F. Zeinali, M. Robatmili, and M. Ghavami, "Sum throughput maximization in multi-bd symbiotic radio noma network assisted by active-STAR-RIS," *arXiv preprint arXiv:2401.08301*, 2024.

[13] M. B. Janjua and H. Arslan, "Survey on symbiotic radio: A paradigm shift in spectrum sharing and coexistence," *arXiv preprint arXiv:2111.08948*, 2021.

[14] Z. Chen and B. Ji, "Resource allocation algorithm for IoT communication based on ambient backscatter," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. , pp. 1–5, 2021.

[15] R. Long, Y.-C. Liang, H. Guo, G. Yang, and R. Zhang, "Symbiotic radio: A new communication paradigm for passive Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1350–1363, 2020.

[16] Y.-C. Liang, Q. Zhang, E. G. Larsson, and G. Y. Li, "Symbiotic radio: Cognitive backscattering communications for future wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1242–1255, 2020.

[17] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[18] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360 coverage by intelligent surfaces," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 102–109, 2021.

[19] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3134–3138, 2021.

[20] J. Xu, J. Zuo, J. T. Zhou, and Y. Liu, "Active simultaneously transmitting and reflecting (STAR)-RISs: Modelling and analysis," *IEEE Communications Letters*, vol. 27, no. 9, pp. 2466 - 2470, 2023.

[21] H. Niu, Z. Chu, F. Zhou, P. Xiao, and N. Al-Dhahir, "Weighted sum rate optimization for STAR-RIS-assisted MIMO system," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2122–2127, 2021.

[22] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2684–2688, 2022.

[23] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4962–4975, 2021.

[24] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1707–1725, 2022.

[25] B. Gu, D. Li, H. Ding, G. Wang, and C. Tellambura, "Breaking the interference and fading gridlock in backscatter communications: State-of-the-art, design challenges, and future directions," *IEEE Communications Surveys & Tutorials*, 2024.

[26] R. S. Yeganeh, M. J. Omidi, and M. Ghavami, "Multi-BD symbiotic radio aided 6G IoT network: Energy consumption optimization with QoS constraint approach," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 4, pp. 2067 – 2080, 2023.

[27] J. Wang, Y.-C. Liang, and S. Sun, "Multi-user multi-IoT-device symbiotic radio: A novel massive access scheme for cellular IoT," *IEEE Transactions on Wireless Communications (Early Access)*, 2024.

[28] H. Yang, H. Ding, M. Elkashlan, H. Li, and K. Xin, "A novel symbiotic backscatter-NOMA system," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 8, pp. 11 006–11 011, 2023.

[29] H. Yang, H. Ding, K. Cao, M. Elkashlan, H. Li, and K. Xin, "A RIS-segmented symbiotic ambient backscatter communication system," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 812 – 825, 2023.

[30] J. Liu, H. Ding, M. Elkashlan, W. Chen, S. Wang, and T. Sun, "A STAR-RIS-segmented symbiotic AmBC system," *IEEE Wireless Communications Letters*, vol. 13, no. 5, pp. 1379 – 1383, 2024.

[31] Z. Wen, H. Ding, M. Elkashlan, C. Yuen, J. M. Moualeu, J. Liu, K. Xin, and C. Yang, "Outage analysis for a STAR-RIS-segmented symbiotic backscatter NOMA system," *IEEE Transactions on Communications*, vol. 72, no. 11, pp. 7263 – 7277, 2024.

[32] H. Yang, H. Ding, and M. Elkashlan, "Opportunistic symbiotic backscatter communication systems," *IEEE Communications Letters*, vol. 27, no. 1, pp. 100–104, 2022.

[33] M. Wu, X. Lei, X. Zhou, X. Tang, and O. A. Dobre, "RIS-assisted energy- and spectrum-efficient symbiotic transmission in NOMA systems," *IEEE Transactions on Communications*, vol. 71, no. 5, pp. 2801 - 2815, 2023..

[34] X. Li, Q. Wang, M. Zeng, Y. Liu, S. Dang, T. A. Tsiftsis, and O. A. Dobre, "Physical-layer authentication for ambient backscatter-aided noma symbiotic systems," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2288–2303, 2023.

[35] J. Hu, Y.-C. Liang, Y. Pei, S. Sun, and R. Liu, "Reconfigurable intelligent surface based uplink MU-MIMO symbiotic radio system," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 423–438, 2022.

[36] J. Hu, Y.-C. Liang, and Y. Pei, "Reconfigurable intelligent surface enhanced multi-user miso symbiotic radio system," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2359–2371, 2020.

[37] J. Ye, S. Guo, S. Dang, B. Shihada, and M.-S. Alouini, "On the capacity of reconfigurable intelligent surface assisted mimo symbiotic communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1943–1959, 2021.

[38] C. Zhou, B. Lyu, Y. Feng, and D. T. Hoang, "Transmit power minimization for STAR-RIS empowered symbiotic radio communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 9, no. 6, pp. 1641 - 1656, 2023.

[39] S. Gong, X. Huang, J. Xu, W. Liu, P. Wang, and D. Niyato, "Backscatter relay communications powered by wireless energy beamforming," *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 3187–3200, 2018.

[40] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (star) ris assisted noma systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 611–626, 2022.

[41] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, 2021.

[42] X. Zhai, G. Han, Y. Cai, Y. Liu, and L. Hanzo, "Simultaneously transmitting and reflecting (STAR) RIS assisted over-the-air computation systems," *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1309–1322, 2023.

[43] S. Han, Y.-C. Liang, and G. Sun, "The design and optimization of random code assisted multi-BD symbiotic radio system," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5159–5170, 2021.

[44] Z. Dai, R. Li, J. Xu, Y. Zeng, and S. Jin, "Cell-free symbiotic radio: Channel estimation method and achievable rate analysis," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. IEEE, Conference Proceedings, pp. 25–30.

[45] S. Ma, G. Wang, R. Fan, and C. Tellambura, "Blind channel estimation for ambient backscatter communication systems," *IEEE Communications letters*, vol. 22, no. 6, pp. 1296–1299, 2018.

[46] Z. Wang, H. Xu, L. Zhao, X. Chen, and A. Zhou, "Deep learning for joint pilot design and channel estimation in symbiotic radio communications," *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2056–2060, 2022.

[47] 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," *International Journal of Electronics and Communication Engineering*, vol. 17, no. 9, pp. 217 – 224, 2023.

[48] 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 Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1413–1430, 2023.

[49] L. Zhang, C. She, K. Ying, Y. Li, and B. Vucetic, "Deep reinforcement learning for improving resource utilization efficiency of URLLC with imperfect channel state information," *IEEE Wireless Communications Letters*, vol. 12, no. 10, pp. 1796–1800, 2023.

[50] J. Du, W. Cheng, G. Lu, H. Cao, X. Chu, Z. Zhang, and J. Wang, "Resource pricing and allocation in MEC enabled blockchain systems: An A3C deep reinforcement learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 33–44, 2022.

[51] M. Chen, T. Wang, K. Ota, M. Dong, M. Zhao, and A. Liu, "Intelligent resource allocation management for vehicles network: An A3C learning approach," *Computer Communications*, vol. 151, pp. 485–494, 2020.

[52] R. K. Singh, P. P. Puluckul, R. Berkvens, and M. Weyn, "Energy consumption analysis of LPWAN technologies and lifetime estimation for iot application," *Sensors*, vol. 20, no. 17, p. 4794, 2020.

[53] J. Finnegan and S. Brown, "An analysis of the energy consumption of LPWA-based IoT devices," in *2018 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1–6.

[54] D. Poluektov, M. Polovov, P. Kharin, M. Stusek, K. Zeman, P. Masek, I. Gudkova, J. Hosek, and K. Samouylov, "On the performance of LORAWAN in smart city: End-device design and communication coverage," in *International Conference on Distributed Computer and Communication Networks*. Springer, 2019, pp. 15–29.

[55] M. Lauridsen, R. Krigslund, M. Rohr, and G. Madueno, "An empirical NB-IoT power consumption model for battery lifetime estimation," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pp. 1–5.

[56] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE transactions on wireless communications*, vol. 18, no. 8, pp. 4157–4170, 2019.