Deep Reinforcement Learning for Energy-Efficient Networking with Reconfigurable Intelligent Surfaces

Gilsoo Lee[†], Minchae Jung[†], Ali Taleb Zadeh Kasgari[†], Walid Saad[†], and Mehdi Bennis[‡]

[†] Wireless@VT, Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, Emails: {gilsoolee,hosaly,alitk,walids}@vt.edu.

[‡] Centre for Wireless Communications, University of Oulu, Finland, Email: mehdi.bennis@oulu.fi.

***Abstract*—When deployed as reflectors for existing wireless base stations (BSs), reconfigurable intelligent surfaces (RISs) can be a promising approach to achieve high spectrum and energy efficiency. However, due to the large number of RIS elements, the joint optimization of the BS and reflector RIS configuration is challenging. In essence, the BS transmit power and RIS's reflecting configuration must be optimized so as to improve users' data rates and reduce the BS power consumption. In this paper, the problem of energy efficiency optimization is studied in an RIS-assisted cellular network endowed with an RIS reflector powered via energy harvesting technologies. The goal of this proposed framework is to maximize the average energy efficiency by enabling a BS to determine the transmit power and RIS configuration, under uncertainty on the wireless channel and harvested energy of the RIS system. To solve this problem, a novel approach based on deep reinforcement learning is proposed, in which the BS receives the state information, consisting of the users' channel state information feedback and the available energy reported by the RIS. Then, the BS optimizes its action composed of the BS transmit power allocation and RIS phase shift configuration using a neural network. Due to the intractability of the formulated problem under uncertainty, a case study is conducted to analyze the performance of the studied RIS-assisted downlink system by asymptotically deriving the upper bound of the energy efficiency. Simulation results show that the proposed framework improves energy efficiency up to 77.3% when the number of RIS elements increases from 9 to 25.**

I. Introduction

Reconfigurable intelligent surfaces (RISs), mounted on walls and buildings, have emerged as an effective approach to enhance the ever increasing need for capacity [1]. The advantage using an RIS as a reflector that assists existing cellular base stations (BSs) stems from the ability of RISs to provide near-field communications while having a very low carbon footprint relative to conventional BSs [2]. If properly deployed in an urban environment, RISs can provide nearly line-of-sight (LOS) communication channels [3]. Indeed, the users in RIS environment will be able to maintain reliable wireless connections and low-latency data transmission compared to conventional antenna-array systems [4]. However, to reap the benefits of RISs, architectural and operational challenges must be addressed [5]–[10].

To develop RISs, the authors in [5] study the use of a field programmable gate array (FPGA) made with tunable metasurfaces electrically controllable via software. When using an RIS in wireless communications, radio resource allocation to optimize the network performance is a prime concern. For instance, the work in [6] studies an RIS-assisted downlink system design that minimizes the BS transmit power by optimizing the continuous transmit beamforming of the BS and the discrete phase shifter of the RIS. Moreover, the authors in [7] develop a joint active and passive beamforming design to maximize the received signal power of RIS users. Also, the authors in [8] investigate the problem of maximizing the downlink capacity to design the optimal RIS phase shift by exploiting statistical channel state information. In [9], a passive beamformer is proposed to achieve an asymptotic optimal performance by controlling the incident wave properties while considering a limited RIS control link and practical reflection coefficients. Moreover, the work in [10] studies the energy efficiency maximization problem in an RIS environment when all involved channels are perfectly known at BS to use zero forcing transmission.

In all of these existing RIS works [5]–[10], it is generally assumed that information on the environment such as wireless channels and power consumption is completely known. However, in practice, the wireless channel gains change in a fading environment, and the wireless channel can be uncertain if the RIS configuration is dynamically updated. Hence, the BS cannot know the exact channel gain. Indeed, it is challenging for a cellular BS to perform precoding with incomplete channel information. Thus, there exists an inherent uncertainty stemming from the unknown RIS configuration and the spatio-temporal dynamics of the channel. Further, most of the existing works [5]–[10] typically assume that an RIS system is operated by using a power grid. However, in practice, the use of energy harvesting at an RIS can be necessary to reduce the reliance on the conventional power grid and enable the vision of green communications. Moreover, since no amplifier is used in an RIS, it will consume very low energy, and, therefore, it can be suitable for an RIS to adopt energy harvesting technologies. Consequently, unlike the existing literature [5]–[10] which assumes full knowledge about the network environment, uses power grid as energy source, and relies on model-based optimization techniques, our goal is to design a *deep reinforcement learning (RL) approach* to make a decision using on-the-fly information on the cellular networks, under channel uncertainty and energy harvesting, while maximizing the average energy efficiency.

The main contribution of this paper is a *novel framework for RIS-assisted cellular communications using energy harvesting technologies at the RIS*. This framework allows a BS to dynamically adapt to wireless environment in the presence of uncertainty on the wireless channel gains and future available

This research was supported by the U.S. National Science Foundation under Grants CNS-1814477 and IIS-1633363.

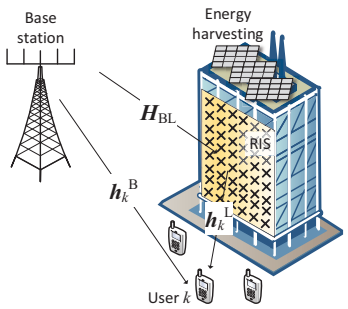

Fig. 1: System model of the RIS-assisted system powered by energy harvesting.

energy of the RIS (acting as a *reflector*). Therefore, this downlink system can jointly use both the direct and indirect wireless paths between the BS and users. We formulate an optimization problem whose objective is to maximize the average energy efficiency of the BS by properly allocating the downlink transmit power while also properly determining the phase of the RIS elements. To solve this optimization problem, we propose a novel approach based on deep RL by defining the state, action, reward, and policy. In the proposed framework, the BS decides the *action* about the power allocation, phase shift, and RIS ON/OFF states. Then, the environment nodes including the RIS and users send the feedback information consisting of the *state* about the wireless channel and energy and *reward* about energy efficiency. Throughout the proposed learning process, the *policy* of the BS can select the best possible actions depending on different states. Finally, a case study is conducted to analyze the performance of the studied RIS-assisted downlink system by asymptotically deriving an upper bound of the energy efficiency. Simulation results show that the energy efficiency can increase by up to 24.6% by increasing RIS phase shifter resolution from 3 to 5 bits.

The rest of this paper is organized as follows. In Section II, the system model is presented. In Section III-A, we formulate the proposed problem. Section III-B presents our approach to solve the problem, and Section III-C includes a case study for a performance analysis. Simulation results are analyzed in Section IV while conclusions are drawn in Section V.

II. System Model

A. Wireless Network Model

We consider the downlink of a wireless network with a single BS with M antennas. The BS is assisted by an RIS reflecting surface that serves a set of K single-antenna user devices, as shown in Fig. II. In this system model, the RIS is a reflecting surface that includes an antenna array with N passive phase shifters used to reflect the received signal while changing the phase of the signal. Fig. 1 illustrates our model in which one of the sides of a building is equipped with an RIS having a large number of antennas, i.e., $N > M$.

A user can receive signal from the BS via direct and reflected wireless links, respectively. The direct path between the BS and the user is a non-LOS (NLOS) channel. The channel of the direct link between the BS and user k is given by

$$\boldsymbol{h}_k^{\mathrm{B}} = g_{\mathrm{B}k}[h_{1k}^{\mathrm{B}}, \ldots, h_{Mk}^{\mathrm{B}}] \in \mathbb{C}^{1\times M},$$

where $g_{\mathrm{B}k} = d_{\mathrm{B}k}^{-\alpha_{\mathrm{NL}}/2}$ is the path loss between the BS and user k at a distance $d_{\mathrm{B}k}$ and path loss exponent α_{NL}, and h_{mk}^{B} is the small-scale fading between BS antenna m and user k. Therefore, the channel between the BS and the users will be:

$$\boldsymbol{H}_{\mathrm{BU}} = [(\boldsymbol{h}_1^{\mathrm{B}})^T, \ldots, (\boldsymbol{h}_K^{\mathrm{B}})^T]^T \in \mathbb{C}^{K\times M}.$$

The RIS-reflected path includes two links: an NLOS link between the BS and the RIS and an LOS link between the RIS and the users. The NLOS channel between the BS and the RIS system is given by

$$\boldsymbol{H}_{\mathrm{BL}} = g_{\mathrm{BL}} \{h_{nm}\}_{n\in[1,N], m\in[1,M]} \in \mathbb{C}^{N\times M},$$

where g_{BL} is the path loss between the BS and the RIS system, and h_{mn} is the complex Gaussian random variable, $\mathcal{CN}(0,1)$. Therefore, $\boldsymbol{H}_{\mathrm{BL}}$ is a matrix where $g_{\mathrm{BL}}h_{nm}$ is the element at row n and column m. When the BS transmits a signal to the RIS through $\boldsymbol{H}_{\mathrm{BL}}$, the RIS is used to reflect the signal to the users with the following vector of phase shifts:

$$\boldsymbol{\Phi} = [\phi_1, \ldots, \phi_N] \in \mathbb{C}^{1\times N}.$$

Each RIS element n can select a phase shift value ϕ_n from the feasible set of phase shifting values:

$$\phi_n \in \mathcal{F}_{\mathrm{RIS}} \triangleq \left\{ \exp\left(\frac{j2\pi m}{2^b}\right) \right\}_{m=0}^{2^b-1},$$

where b is the resolution of the RIS element's phase shifter [10]. Also, the ON or OFF state of RIS element n is denoted by σ_n and defined as follows:

$$\sigma_n = \begin{cases} 1, & \text{if RIS element } n \text{ is turned ON,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Then, the vector of the ON/OFF states of the RIS elements is defined as $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_N] \in \mathbb{R}^{1\times N}$ If an RIS element n is turned OFF, the signal from the BS will not be reflected by this RIS element n. Therefore, considering the ON/OFF states of the RIS elements, the phase shifting matrix can be defined as:

$$\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\Phi} \odot \boldsymbol{\sigma}) \in \mathbb{C}^{N\times N},$$

where $\mathrm{diag}(\cdot)$ is a block-diagonal matrix, and notation $\odot$ is an element-wise vector multiplication.

When the BS's signals are reflected to the users, the channel between the RIS and user k becomes

$$\boldsymbol{h}_k^{\mathrm{L}} = [g_{1k}h_{1k}^{\mathrm{L}}, \ldots, g_{Nk}h_{Nk}^{\mathrm{L}}] \in \mathbb{C}^{1\times N},$$

where $g_{nk} = 1/\sqrt{4\pi d_{nk}^2}$ is a free space path loss attenuation between RIS element n and user k, and h_{nk}^{L} is the LOS channel state between RIS element n and user k. In Fig. 1, the users are located in front of the RIS within a two-dimensional space in the xy-plane at $z = 0$ in Cartesian coordinates. We define the location of user $k \in [1, K]$ as (x_k, y_k, z_k). The RIS element $n \in [1, N]$ is located at $(x_n, y_n, 0)$. Then, the channels between the RIS system and user k are assumed to be LOS links as in [4], and, thus, channel h_{nk}^{L} becomes:

$$h_{nk}^{\mathrm{L}} = \exp\left(\frac{-j2\pi d_{nk}}{\lambda}\right), \forall n \in [1, N],$$

where λ is the signal wavelength, and d_{nk} is the distance between user k and RIS element n given by $d_{nk} = \sqrt{(x_k - x_n)^2 + (y_k - y_n)^2 + z_k^2}$. Therefore, the channel between the RIS and its K users is given by

$$\boldsymbol{H}_{\mathrm{LU}} = [(\boldsymbol{h}_1^{\mathrm{L}})^T, \ldots, (\boldsymbol{h}_K^{\mathrm{L}})^T]^T \in \mathbb{C}^{K \times N}.$$

Thus, through the channels $\boldsymbol{H}_{\mathrm{BL}}$ and $\boldsymbol{H}_{\mathrm{LU}}$, the reflected path is used to transmit a signal from the BS to the users.

Given the wireless channel model, the downlink signal received by user k is expressed by:

$$y_k = \left(\boldsymbol{h}_k^{\mathrm{L}} \boldsymbol{\Lambda} \boldsymbol{H}_{\mathrm{BL}} + \boldsymbol{h}_k^{\mathrm{B}}\right) \boldsymbol{x} + n_k,$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is the zero-mean complex white Gaussian noise with variance σ_n^2. The BS transmits the signal $\boldsymbol{x} = \sum_{k=1}^{K} \sqrt{p_k} \boldsymbol{f}_k^H s_k \in \mathbb{C}^{M \times 1}$, where p_k is the transmit power of user k's signal, $\boldsymbol{f}_k^H$ is the precoding vector for user k, and s_k is the unit power information symbol. The sum of the transmit power of all users is smaller than the BS maximum transmit power $P_{\max}$. The transmit power for the users are captured by the vector $\boldsymbol{p} = [p_1, \ldots, p_K]$. At the BS, precoding $\boldsymbol{f}_k$ is applied to obtain the transmitted signal $\boldsymbol{x}_k$. Particularly, each user selects a precoding vector from a pre-defined codebook that is known to both the users and the BS [11]. Then, the user sends the precoding matrix indicator (PMI) of the selected precoding vector to the BS. When selecting a precoding vector, the users decide the precoding vector that is most suitable to maximize their data rate after measuring their channel state [11]. Since $\boldsymbol{h}_k \triangleq \boldsymbol{h}_k^{\mathrm{L}} \boldsymbol{\Lambda} \boldsymbol{H}_{\mathrm{BL}} + \boldsymbol{h}_k^{\mathrm{B}} \in \mathbb{C}^{1 \times M}$ is the effective channel of user k, user k selects its channel direction $\hat{\boldsymbol{h}}_k$ to the closest $\boldsymbol{w}_i$ according to

$$\hat{\boldsymbol{h}}_k = \arg\max_{\boldsymbol{w}_n \in \mathcal{C}_{\mathrm{BS}}} \left| \boldsymbol{h}_k \boldsymbol{w}_n^H \right|,$$

where $\boldsymbol{w}_n$ is a precoding vector. The BS uses a discrete Fourier transform (DFT)-based codebook [12] given as:

$$\mathcal{C}_{\mathrm{BS}} = \left\{ \exp\left(\frac{j 2\pi n}{2^c} \right) \right\}_{n=0}^{2^c - 1},$$

where c is the number of feedback bits for PMI. Based on the feedback information from user k, the BS selects the precoding vector $\boldsymbol{f}_k = \hat{\boldsymbol{h}}_k$, and, therefore, the received SINR at user k becomes:

$$\gamma_k = \frac{p_k |\boldsymbol{h}_k \hat{\boldsymbol{h}}_k^H|^2}{\sum_{\substack{j=1 \\ j \neq k}}^{K} p_j |\boldsymbol{h}_k \hat{\boldsymbol{h}}_j^H|^2 + \sigma_n^2}. \quad (2)$$

The achievable sum rate of the RIS-assisted multi-user MISO system is given by $R = \sum_{k=1}^{K} \log_2(1 + \gamma_k)$.

B. Energy Consumption Model

In our considered system, the RIS reflector is self-powered and relies exclusively on energy harvesting sources. For example, the RIS reflector can be equipped with solar panels to procure energy for its operation. Since the characteristics of the harvested energy can be highly dynamic, we do not make any specific assumption on the energy harvesting process. Thus, our model can accommodate any type of energy harvesting mechanism. To enhance the overall energy efficiency of the system, the RIS elements can be turned ON or OFF, depending on the network performance and harvested energy status. Therefore, the RIS system is equipped with a battery or energy storage systems (ESS) to manage the intermittent and uncertain energy harvesting process. Also, the RIS system can harvest energy irrespective of the ON or OFF status of its RIS elements. Moreover, when the RIS elements are turned ON, the RIS system can still store the excess of harvested energy if the instantaneous harvested energy is enough to operate the RIS elements.

When the transmission power of the BS for user k at time t is $p_k(t)$, the power consumption for the wireless communication link between the BS and K users becomes $P(t) = \sum_{k=1}^{K} \mu p_k(t)$ where $1/\mu$ is the efficiency of the transmit power amplifier [10]. When $\boldsymbol{\sigma}(t)$ is the vector of the ON/OFF states of the RIS elements at time t, the power consumption of the RIS system at time t is modeled as $P_{\mathrm{RIS}}(t) = \sum_{n=1}^{N} \boldsymbol{\sigma}(t) P_b$ where P_b is the power consumption of a phase shifter with b-bit resolution. When all RIS elements are configured to use b-bit resolution, the power consumption to turn an RIS element ON will be identical.

We consider a time-slotted system with a time slot duration Δ. The energy consumption of the RIS system during time slot t becomes $S(t) = P_{\mathrm{RIS}}(t)\Delta$. When the RIS elements consume the harvested energy, the available amount of energy stored in the ESS at time t is given by

$$E(t) = \min\left(E(t-1) - S(t-1) + \Omega(t-1), E_{\max}\right), \quad (3)$$

where $E(t-1) \geq 0$ is the stored energy of the RIS system at time $t-1$, $S(t-1)$ is the energy used by the RIS elements, $\Omega(t)$ is the amount of harvested energy at the RIS system, and $E_{\max}$ is the maximum energy storage capacity of the ESS [13]. Since $\Omega(t)$ is randomly generated because the RIS system is unable to know the harvested energy in the future, and, therefore, randomness captures the uncertainty of the energy harvesting process over time. When the RIS reflectors are turned ON by only using the harvested energy, all RIS elements can be turned OFF at time t if $E(t)$ becomes zero at a certain time t. In this case, the users will only receive the signal from the BS without any reflected signal from the RIS. Also, if the RIS system does not have enough energy to all of its RIS elements ON at time t, i.e., $S(t) > E(t)$, then, a partial set of RIS elements must be turned OFF, and the users' data rate can change over the ON/OFF configuration of the RIS elements, i.e., $\boldsymbol{\sigma}(t)$.

III. Machine Learning Framework for Energy-Efficient Networking

Given the defined system model, our goal is to analyze the joint problem for allocating transmit power to the BS antennas and optimizing the operation of the RIS elements by deciding the ON/OFF status and phase shifting. Particularly, since the future energy status of the RIS system is unpredictable, the ON/OFF states of the RIS elements dynamically change. Also, when the ON/OFF status of the RIS elements is determined depending on the harvested energy at the RIS system, the wireless channel can dynamically change over time. Therefore, there is uncertainty on wireless channels when a self-powered RIS is used. Thus, it is highly challenging for the BS to optimize the transmit power each time the ON/OFF status of the RIS elements is updated. In fact, even if turning ON

more RIS elements can improve the data rate, it may have a detrimental effect on future data rates due to the lack of the stored energy at an RIS. To cope with the uncertainty of the harvested energy and channel states, we introduce an artificial intelligence framework that uses machine learning techniques to maximize the energy efficiency of the cellular system while properly managing the harvested energy at the RIS.

A. Problem Formulation

First, we formulate the following optimization problem whose goal is to maximize the energy efficiency of the RIS-assisted communication system:

$$\max_{\boldsymbol{\sigma}(t),\boldsymbol{\Phi}(t),\boldsymbol{p}(t)} \quad \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} \frac{R(t)}{P(t)}, \tag{4}$$

$$\text{s.t.} \quad S(t) \leq E(t), \forall t, \tag{4a}$$

$$\textstyle\sum_{k=1}^{K} p_k(t) \leq P_{\max}, \forall t. \tag{4b}$$

The objective function (4) is the average energy efficiency achieved by the decision of the BS. Constraint (4a) is an energy causality condition. Constraint (4b) means that the sum of transmit power is smaller than or equal to the maximum transmit power. In the formulated problem, the BS is assumed to optimize the transmit power, RIS phase shifts, and ON/OFF states. To determine the transmit power and RIS phase shifts, the BS needs to calculate the objective function (4). However, since the channel information required to calculate the SINR in (2) is unknown to the BS, the value of (4) is not directly known by the BS. In particular, when the users observe the channel through pilot signaling from the BS, the users send channel feedback information to the BS by using the PMI, thus resulting in quantization errors. Therefore, the BS can only know the estimated channel [14]. Due to the uncertainty on the channel between the BS and users, the BS is unable to calculate the data rate function that includes both original channel and estimated channel term. Thus, when the estimation error is unknown to the BS, conventional optimization techniques such as convex and combinatorial programming cannot be applied to solve problem (4) [15]. To this end, a deep RL technique can be used to solve problem (4) by using a deep neural network function to approximate the relationship between the optimization variables and achieved performance. Next, we propose an RL-based framework used to seek a solution to maximize the time-averaged energy efficiency of the system when the performance metric is not directly accessible for the optimization algorithm running on the BS.

B. Deep Reinforcement Learning Approach

As aforementioned, the BS is unable to directly evaluate the objective function with under channel uncertainty. Also, the wireless resource optimization and RIS ON/OFF scheduling problem in (4) involves a large state space, i.e., $\mathcal{O}(2^N)$. Therefore, a closed-form solution does not exist without any knowledge about the state transition probabilities, and an exhaustive search, i.e., a brute-force algorithm, is impractical. Thus, we propose a deep RL framework to seek a solution that maximizes the energy efficiency so that the data rate

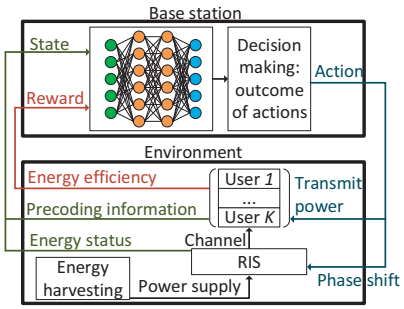

Fig. 2: Machine learning framework for energy-efficient networking of an RIS-assisted cellular system.

of the users are improved while consuming a lower energy. A deep RL framework can handle a control problem with a large state space since deep RL uses a deep neural network for approximation of the RL's action-value function for a RIS system [16]. Beyond being able to maximize the energy efficiency, the key advantage of the deep RL framework is that the BS can learn the performance outcome of the system through a trial-and-error process while updating the weights of a deep neural network [17]. In doing so, the BS can make an immediate decision to allocate the radio resource while knowing only estimated channel without having a knowledge of the exact information on channel.

The proposed framework includes the agent implemented on the BS and the rest of environment nodes including an RIS and the users as shown in Fig. 2. To develop an RL framework, we must define the state, action, policy, and reward. For our model, the *states* consist of the precoding vectors $\hat{\boldsymbol{h}}_k$ from the users and the energy level of our RIS $E(t)$. Meanwhile, the *actions* are the optimization variables in (4) such as transmit power $\boldsymbol{p}(t)$, phase shifting $\boldsymbol{\Phi}(t)$, and ON/OFF status $\boldsymbol{\sigma}(t)$. The *policy* denotes the strategy that maps a state to an action. Therefore, the policy is used to determine the values of the optimization variables. Since the actions are state-dependent, similar actions can yield different outcomes contingent upon the current states.

The goal of RL is to pick the best known action for any given state. To this end, the BS will observe and use feedback. Particularly, in Fig. 2, when the BS decides the *action*, the environment nodes send a feedback to the BS. The feedback message includes the current state and reward. In the proposed framework, the reward returned by the environment is defined as:

$$r(t) = \textstyle\sum_{k=1}^{K} R(t)/P(t). \tag{5}$$

When the users send feedback about the data rate $R(t)$, the BS knows its energy consumption $P(t)$. Therefore, the definition of the reward in (5) can be readily calculated by the BS. The reward value can be affected by various unknown aspects such as uncertainty on exact channel of the users. Therefore, by using the feedback returned from the uncertain environment, the deep neural network uses the difference between the expected reward and the actual reward, and the BS improves the weights of the deep neural network that shows the expected reward of state-action pairs. In Fig. 2, the decision making

block can be implemented by using different deep RL methods such as deep Q-network [17]. Thus, throughout the learning process, the BS is able to learn the policy that is used to select the best possible actions depending on different states. As a result, the proposed framework shown in Fig. 2 has key benefits. First, the agent running on the BS only need to interact with the environment by exchanging a small size of bits. Also, any prior information on the environment is not required to deploy the proposed framework. Moreover, the proposed framework is compatible with the existing cellular standard in that PMI and reward in (5) are calculated by the users and those parameters can be reported to the BS through PUCCH on the control plane.

C. Performance Analysis: Case Study

Due to the uncertainty on the wireless channels and harvested energy of the studied system, the energy efficiency and data rate are not deterministic, and the exact outcomes of an algorithmic solution become mathematically intractable. Therefore, a case study is carried out to examine and analyze the performance of an RIS-assisted downlink system. To this end, we asymptotically analyze the data rate and energy efficiency. This analysis will provide the upper bound of the data rate and energy efficiency that can be asymptotically achieved by using the RL methods.

Theorem 1. *For the studied RIS-assisted downlink system, the upper bound of energy efficiency is given by* $\frac{1}{\ln 2}N^2$ *when* $N, K \to \infty$, *and* $p_k = P_{max}/K, \forall k$.

Proof. We derive the upper bound of the given system model's energy efficiency. The SINR of user k, γ_k, has an upper bound given by $\gamma_k^{\text{SNR}} \triangleq \frac{p_k|\boldsymbol{h}_k\hat{\boldsymbol{h}}_k^H|^2}{\sigma_n^2}$ that is the SNR of the single user case without inter-user interference. Also, γ_k^{SNR} increases with $\mathcal{O}(N^2)$ as $N \to \infty$, as proved in [7]. Therefore, we have the following relationship:

$$\gamma_k \leq \gamma_k^{\text{SNR}} \sim \mathcal{O}(N^2). \tag{6}$$

From (6), an upper bound of the sum data rate can be derived as:

$$\sum_{k=1}^{K} \log_2(1+\gamma_k) \leq \sum_{k=1}^{K} \log_2(1+\gamma_k^{\text{SNR}}),$$

where $\sum_{k=1}^{K} \log_2(1+\gamma_k^{\text{SNR}}) \sim \sum_{k=1}^{K} \mathcal{O}\left(\log_2\left(1+\frac{P_{\max}}{K}N^2\right)\right)$. Now, when $K \to \infty$ and the transmit power is equally allocated to each user, the sum rate can be written as:

$$\lim_{K\to\infty} \sum_{k} \log_2(1+\gamma_k^{\text{SNR}}) \sim \mathcal{O}\left(K \log_2\left(1+\frac{P_{\max}}{K}N^2\right)\right).$$

Thus, we have the result:

$$\lim_{K\to\infty} K \log_2\left(1+\frac{P_{\max}N^2}{K}\right) \overset{(a)}{=} \frac{1}{\ln 2}P_{\max}N^2,$$

where (a) results from the exponential definition $e^x = \lim_{n\to\infty}(1+x/n)^n$. Hence, by dividing the derived upper bound of the sum rate by the total transmit power $P_{\max}$, the upper bound of EE is given by $\frac{1}{\ln 2}N^2$. □

In the single user case, the user's SNR increases with $\mathcal{O}(N^2)$, and, thus, the data rate will asymptotically follow $\mathcal{O}\left(\log_2(N)\right)$, as $N \to \infty$. However, in a multi-user case, we can observe from Theorem 1 that the derived upper bound of multi-user sum rate increases with $\mathcal{O}(N^2)$, as $N, K \to \infty$.

In a multi-user case, since the upper bound of the sum rate is finite, i.e., $\frac{1}{\ln 2}P_{\max}N^2$, the data rate per user approaches zero as $K \to \infty$. Therefore, it is beneficial to schedule a finite number of users in an RIS-assisted cellular network. From Theorem 1, for a finite K, the upper bound of multi-user sum rate is asymptotically derived as $\mathcal{O}\left(K \log_2\left(1+\frac{P_{\max}}{K}N^2\right)\right) \sim \mathcal{O}\left(\log_2\left(N\right)\right)$. Therefore, we conclude that the data rate of a single user case asymptotically achieves the upper bound of the multi-users' sum rate. Next, we evaluate the performance of the proposed framework throughout simulation experiments in Section IV

IV. Simulation Results

For our simulations, we consider an RIS-assisted downlink system where the distance between the BS and the RIS is 300 m. The width and height of an RIS are 30 m, respectively. When four users are located in front of the RIS, the distance between each user and the RIS surface follows a uniform distribution in range between 0 and 30 m. The BS has 16 antennas while each user has a single antenna. The bandwidth is 10 MHz, and the power spectral density of the noise is -174 dBm/Hz. We use $1/\mu = 22.6$ and $P_{\max} = 43$ dBm. Also, the NLOS path loss exponent α_{NL} is 3.7. We assume that energy arrivals per second follow a Poisson process with an energy arrival rate of 10. The RIS energy consumption is modeled with $\Delta = 1$ and $P_b\Delta/E_{\max} = 100$. Finally, deep Q-network method is implemented as the decision making block in Fig. 2 where the action space includes the discrete transmit power levels quantized with an interval of 1 W. Thus, the performance of the proposed framework is evaluated in the defined environment.

In Fig. 3, we show the energy efficiency and the sum rate of the users for different numbers of RIS elements. Fig. 3 first shows that the sum rate of the users also increases when the number of RIS elements increases. This is due to the fact each user's data rate increases with respect to the number of RIS elements N as shown in Theorem 1. At the same time, in Fig. 3, we can see that the energy efficiency increases when the number of RIS elements increases since the energy efficiency is proportional to the sum rate. From these observations, a larger number of RIS elements can be deployed to maximize the sum rate and energy efficiency of an RIS-assisted downlink system. For instance, if the number of RIS elements N increases from 9 to 25, then the energy efficiency can be improved by 77.3%, and the sum rate can increase up to 2.4 times.

Fig. 4 shows the energy efficiency and sum rate for different energy arrival rates when the resolution of RIS elements changes from 3 to 5 bits. We can first see that the energy efficiency and sum rate increase with respect to the resolution bits. As the RIS uses phase shifters with a higher resolution, it is possible to precisely control the RIS beamformer, thus improving the wireless channel gains. For instance, by increasing the resolution from 3 bits to 5 bits, the energy efficiency

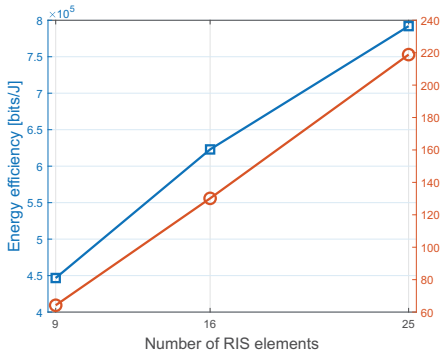

Fig. 3: Energy efficiency and sum rate for different number of RIS elements.

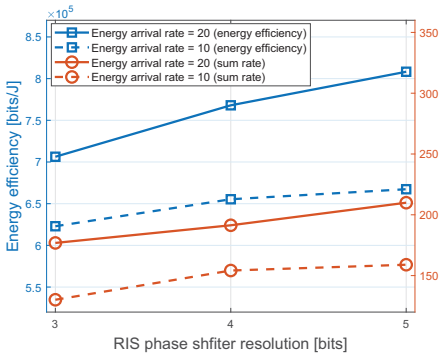

Fig. 4: Energy efficiency and sum rate for different RIS phase shifter's resolution bits.

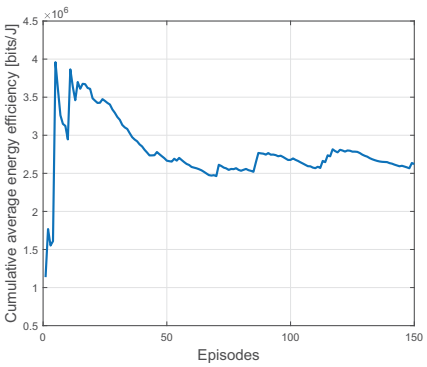

Fig. 5: The cumulative average of energy efficiency over episodes.

can increase up to 24.6% when the energy arrival rate is 20. Also, Fig. 4 shows that the energy efficiency and sum rate increase when more energy is harvested at an RIS. Since the RIS phase shifters are solely operated by harvested energy, harvesting more energy enable the RIS to turn on additional RIS phase shifters. Therefore, additional harvested energy will reduce the impact of the energy harvesting constraint when determining the ON/OFF states of the RIS elements. For example, if the energy arrival rate increases from 10 to 20 with an RIS resolution of 5 bits, then the energy efficiency and sum rate are improved up to 21.1% and 32.1%, respectively.

Fig. 5 shows the cumulative average energy efficiency defined in (4) over episodes. If the energy efficiency reaches or is greater than 10^6, a current episode is assumed to be done. Then, the environment is reset to generate new wireless channels gains and user locations so that a new episode is started. From Fig. 5, we can see that, as the number of episodes increases, the cumulative average of final energy efficiency at each episode tends to converge in a certain range. For instance, when the number of episodes is 150, the value of cumulative average energy efficiency becomes at least 2.5×10^6, achieving a higher value than the preset threshold of 10^6.

V. Conclusion

In this paper, we have proposed a novel framework to optimize the energy efficiency of the BS assisted by an RIS using energy harvesting. We have formulated the problem of maximizing the average energy efficiency which enables the BS to jointly optimize the transmit power allocation, RIS phase shifter, and RIS reflector's ON/OFF state effectively in the presence of uncertainty about wireless channel and available energy of an RIS. We have shown that by using the deep RL approach, the network parameters are suitably determined by the BS without knowing any prior information on wireless environment. The simulation results show that having two times of the harvested energy improves energy efficiency up to 21.1%.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network, to appear*, 2019.

[2] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *arXiv preprint:1911.12296*, 2019.

[3] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects," *IEEE Trans. Wireless Commun.*, to appear, 2020.

[4] M. Jung, W. Saad, and G. Kong, "Performance analysis of large intelligent surfaces (LISs): Uplink spectral efficiency and pilot training," *arXiv preprint arXiv:1904.00453*, 2019.

[5] E. Basar, M. Di Renzo, J. de Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *arXiv preprint arXiv:1906.09490*, 2019.

[6] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoustic, Speech and Sig. Proc. (ICASSP)*, Brighton, United Kingdom, May 2019, pp. 7830–7833.

[7] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[8] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *arXiv preprint arXiv:1812.05429*, 2018.

[9] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *arXiv preprint:1910.00968*, 2019.

[10] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Jun. 2019.

[11] 3rd Generation Partnership Project, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," TS 36.213, May 2016.

[12] J. Suh, C. Kim, W. Sung, J. So, and S. W. Heo, "Construction of a generalized DFT codebook using channel-adaptive parameters," *IEEE Communications Letters*, vol. 21, no. 1, pp. 196–199, Jan. 2017.

[13] G. Lee, W. Saad, M. Bennis, A. Mehbodniya, and F. Adachi, "Online ski rental for on/off scheduling of energy harvesting base stations," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2976–2990, May 2017.

[14] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1739–1755, May 2005.

[15] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials, to appear*, 2019.

[16] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE Int. Wrks. on Sig. Proces. Adv. in Wireless Commun. (SPAWC)*, Cannes, France, 2019, pp. 1–5.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.