Outage Minimization in RIS Assisted Self-Powered Sensor Network using DDPG

Sayantika Banerjee[1], Avik Banerjee[2], and Santi Prasad Maity[3]

[1]Department of Information Technology,Indian Institute of Engineering Science and Technology, Shibpur, India
[2]Department of Electronics and Communication Engineering, RV College of Engineering, Bengaluru, India
[3]Department of Information Technology,Indian Institute of Engineering Science and Technology, Shibpur, India

[1]sayantika160@gmail.com, [2]avikbanerjee@rvce.edu.in, [3]santi.maity@gmail.com

***Abstract*—6G wireless communication faces challenges of energy requirement to support billions of sensor devices in the Internet of Things (IoT). Challenges can be overcome if sensor nodes harvest energy from the RF signal of primary user in the network. Duration for non overlapping time slot for energy harvesting (EH) and data transmission in a single time frame is a trade off in challenging wireless network, that can be addressed by Reconfigurable Intelligent Surfaces (RIS). Insufficient energy at the sensor node causes failure of data transmission which leads to increased outage probability. This work aims to minimize outage probability using reinforcement learning (RL) approaches. A Deep Deterministic Policy Gradient (DDPG) RL framework is proposed to minimize outage probability while optimizing the energy harvesting time fraction and RIS phase-shift control. Performance of DDPG algorithm is compared with a baseline Q-learning approach. Simulation results demonstrate that the proposed DDPG method significantly outperforms Q-learning, reducing outage probability by 28% in the optimal energy harvesting region and achieving more than 56% improvement as the number of RIS elements increases with reduced learning latency by nearly 40%.**

I. Introduction

Emergence of Internet of Things (IoT) has shifted the human centric mobile communication to the connection between things and smart devices to human, leading to several application specific services namely healthcare, weather and gas monitoring, intelligent transport system. Internet of Consumer Electronics (IoCE) devices exchange data over the Internet for monitoring [1], automation, remote control, consumer-centric electric vehicle charging in industry 5.0 environment [2]. Most IoCE devices require embedded sensors for their operation [3].Increased functionality of these devices demands higher energy consumption [4], which challenges the environment sustainability. Energy harvesting by the sensor for its operation from primary user (PU) available in the network can reduce the demand of conventional energy source and become sustainable for the environment[5]. To realize this each time frame for data transmission by the sensor is divided into two parts: one for energy harvesting and the other for data transmission [6].

6G aims to deliver ultra-high data rates, ultra-low latency, massive device connectivity and support for extreme applications [7]. It is difficult to deliver these services through the conventional communication link where signal scattering and attenuation are the major challenges [8]. Reconfigurable Intelligent Surfaces (RIS) can improve the performance of wireless communications by enlarging the coverage of the cell, reducing the inter-user interference, and improving the security of networks [9]. RIS consumes low energy and may harness tools from Artificial Intelligence (AI) and Machine Learning (ML) to enable systems operation and optimization [10], potentially deployed for both indoor and outdoor operations.

Insufficient energy at the sensor node causes data transmission failure. Failure of meeting the target data rate of transmission, described as outage probability, depends on the combined factors of the energy–harvesting time fraction, channel gain, RIS phase configuration, and PU activity. The work in [11] explores the application of deep reinforcement learning (DRL) and RIS to improve edge computing performance, specifically focusing on maximizing the computation rate in wireless networks for consumer applications.

The authors in [12] used a model to integrate Internet of Things (IoT) based self-powered gas sensors with software-defined networking (SDN)-driven control, forming an SDN-IoCE architecture for reliable transmission of sensing data through RIS-assisted underlay cognitive radio networks (CRNs). In that work, deep Q-networks (DQNs) were used for time-critical analysis and energy-harvesting optimization through analytical methods. This paper works on IoT based sensor model that harvests energy from PU or ambient RF signal and transmit data to the edge node via RIS and uses Deep Deterministic Policy Gradient algorithm (DDPG) to optimize the EH time fraction in relation to outage probability.

The major contributions of this paper are summarized as follows:

- A realistic system model with mathematical expression of outage probability is derived and is optimize for the the energy-harvesting (EH) time fraction incorporating RIS reflection, channel randomness, and PU activity.
- To overcome the limitations of static parameter tuning, DDPG algorithm are developed to optimize the EH time fraction including RIS configuration and compare the result with Q-learning approaches.
- The proposed DDPG method reduces outage probability up to 28% in the optimal energy harvesting region and achieves improvement of 56% when the number of RIS elements increases compared to the Q learning method.

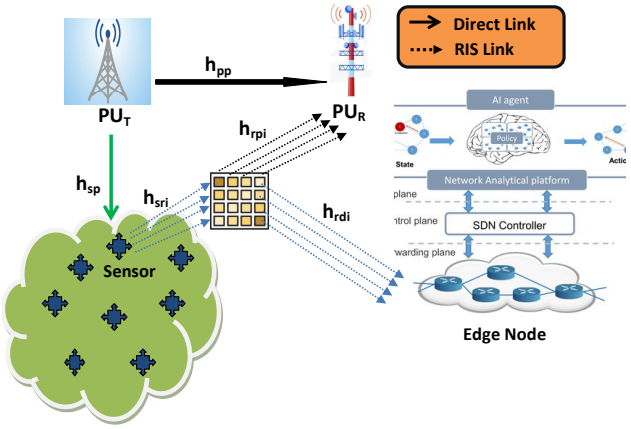

Fig. 1: system model with SDN architecture (adapted from [12])

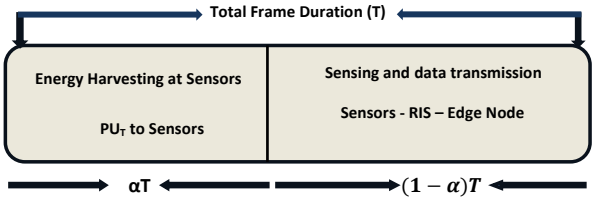

Fig. 2: Frame structure (adapted from [12])

Furthermore, learning latency is nearly 40% lower for the proposed DDPG method than for the Q learning method.

The remainder of the paper is organized as follows: Section II illustrates the system model and frame structure, while problem formulation is given in Section III. Section IV presents the solution methodology. Results and conclusion are provided in Sections V and VI respectively.

II. System Model

The system model comprises two modules: 1. IoT Sensor–EH–RIS–Cognitive Radio(CR) involved in sensing and data transmission, and 2. SDN-IoT architecture for time–critical analysis for energy harvesting, sensing, and transmitting sensor data.

Fig. 1 presents the proposed SDN-IoT-sensor-CR-RIS-EH system model involving a self-powered sensor, followed by sensor data transmission in CR underlay mode and an optimization policy for time-critical analysis using RL at the edge node. Fig. 2 shows the periodic time frame structure that includes operations of energy harvesting (EH) and sensor data transmission. The frame consists of two slots, where over some fraction of time αT, the sensor harvests energy from RF signal of primary user (PU), i.e., existing WiFi RF signals. The remaining time $(1-\alpha)T$ is used for transmitting sensor data from the Sensor to the RIS and to edge devices for time-critical analysis. The IoT sensor communicates with the access point (AP) via the RIS in a wireless communication link, where the sensor harvests energy from the RF signal of the PU. A software-defined networking (SDN) controller is used to manage the entire network operation, including reconfiguration of RIS phases. Each transmission frame of unit duration is divided into two phases: 1.Energy Harvesting (EH) phase 2.Information Transmission phase. In the Energy Harvesting Phase a fraction of time $\alpha \in [0,1]$ in the time frame, the sensor harvests energy from the RF signal transmitted by the PU. The harvested energy is expressed as:

$$E_H = \eta \alpha P_{\mathrm{PU}} |h_{\mathrm{eff}}|^2, \tag{1}$$

where η is the RF-to-DC conversion efficiency, P_{PU} is the transmit power of the PU, and h_{eff} is the effective channel gain of the RIS-assisted link. In Information Transmission Phase the sensor transmits its data to the AP via the RIS during the remaining fraction of time $(1-\alpha)T$. The baseband equivalent of fading channel for Sensor-RIS, RIS-Edge node and RIS-PU_R link vectors are denoted as $h_{sr_i} = [h_{sr_1}, h_{sr_2}, ...h_{sr_M}] \in \mathbb{C}^{1\times M}$, $h_{rd_i} = [h_{rd_1}, h_{rd_2}, ...h_{rd_M}]^T \in \mathbb{C}^{M\times 1}$ and $h_{rp_i} = [h_{rp_1}, h_{rp_2}, ...h_{rp_M}]^T \in \mathbb{C}^{M\times 1}$, respectively. Here h^T represents transpose of h matrix. The fading channel coefficient of the direct link for PU_T-PU_R and PU_T-GS are denoted as $h_{pp} \in \mathbb{C}^1$ and $h_{sp} \in \mathbb{C}^1$, respectively. It is assumed that L number of reflecting elements are present per side of RIS such that $M = L \times L$ and the inter element spacing is given as $d_{space} = \frac{c}{2Lf_o}$, where c and f_0 denote the speed of light and RIS operating frequency, respectively. The phase shift (PS) matrix at RIS is given as

$$\mathbf{\Theta}_k = \mathcal{A}_k \mathrm{diag}(e^{j\phi_{k1}}, e^{j\phi_{k2}}, ...e^{j\phi_{kM}}) \in \mathbb{C}^{M\times M}, \tag{2}$$

where, $\mathcal{A}_k$ and ϕ_k $(k \in \{sr_i, sp_i\},\ i \in 1,2,...M)$ denote the amplitude reflection coefficient and phase shift of all RIS elements, respectively. Here, $\mathcal{A}_k \in [0,1]$ and $\phi_k \in (0, 2\pi]$. Furthermore, $|\mathbf{\Theta}_k|^2 \approx |\mathcal{A}_k|^2 = 1$ denote 100% reflection with zero attenuation from RIS, where $|e^{j\phi_{kM}}|^2 = 1$. The present work models the gain of the instantaneous channel fading coefficient h_k as $|h_k|^2 \sim \mathcal{CN}(0, d_k^{-\psi_k})$, where $k \in \{pp, sp, sr_i, rd_i, rp_i\},\ i \in 1,2,...M$. Symbols '$d$' and '$\psi$' denote the respective distance and path loss exponent between any two nodes, respectively.

A. RIS and NL-EH model

1) RIS model: The overall channel gain from Sensor to Edge node and PU_R, during data transmission slot can be expressed as

$$h_{sd}^C = \underbrace{\sum_{i=1}^{M} h_{sr_i} \mathbf{\Theta_{sr_i}} h_{rd_i}}_{\textbf{RIS reflected link}}, \quad h_{sp}^C = \underbrace{\sum_{i=1}^{M} h_{sr_i} \mathbf{\Theta_{sp_i}} h_{rp_i}}_{\textbf{RIS reflected link}}. \tag{3}$$

2) NL-EH model: Sensor harvests energy from the received RF signal transmitted from PU_T during αT duration. Such signal can be expressed as mentioned below

$$y_{sp}(n) = h_{sp} X_p(n) + n_s(n), \tag{4}$$

where, n = 1 to $\alpha T f_s$. The symbols f_s, X_p and n_s denote the sampling frequency, the transmitted signal from PU_T and the additive noise at the receiver of GS, respectively. X_p and n_s

both have zero mean and variances P_p and σ_s^2, respectively, where, $E[|X_p(n)|^2] = P_p$ and $E[|n_s(n)|^2] = \sigma_s^2$. Here P_p denotes PU_T transmitted power.

Power harvested at Sensor, adopting NL-EH model, can be expressed similarly as given in [12]

$$\Phi_{\text{EH}}^{\text{NL}} = \frac{\frac{\mathcal{X}}{1+\exp(-a(P_\Gamma - b))} - \frac{\mathcal{X}}{1+\exp(ab)}}{1 - \frac{1}{1+\exp(ab)}} \tag{5}$$

where $\mathcal{X}$ denotes the constant maximum harvested power at Sensor when the EH circuit gets saturated. The symbols 'a' and 'b' denote the constant parameters utilized as part of the circuit characteristics.

Here the received RF power (P_Γ) at sensor is given by

$$P_\Gamma = d_{sp}^{-\psi_{sp}} P_p + \sigma_s^2, \tag{6}$$

where, $|h_{sp}|^2 \sim \mathcal{CN}(0, d_{sp}^{-\psi_{sp}})$. Harvested energy at Sensor during αT is given by $E_H = \Phi_{\text{EH}}^{\text{NL}} \alpha T$.

III. Problem Formulation

Our objective is to minimize the outage probability while maintaining the constraints of residual energy, PU and SU throughput.

$$\begin{aligned} &\min_{\alpha} P_{\text{out}} \\ &\text{s.t. } E_{\text{res}} \geq 0,\ R_{\text{pr}} \geq R_{\text{th}}^p,\ R_{\text{sr}} \geq R_{\text{th}}^s, \\ &\quad 0 < \phi_k < 2\pi,\ k \in \{sr, sp\}. \end{aligned} \tag{7}$$

Outage probability is defined as the probability that the transmission rate falls below threshold:

$$P_{\text{out}} = \Pr(R < R_{\text{th}}). \tag{8}$$

The goal is to minimize P_{out} of IoCE Sensor–EH–RIS–CR data transmission by determining the optimal α using reinforcement learning.

The outage probability is defined as the probability that the achievable data rate is lower than a minimum threshold requirement. where R denotes the achievable data rate of the sensor data and R_{th} is the minimum required rate for successful transmission.

For self-powered operation, the IoCE sensor must maintain a positive residual energy level such that:

$$E_{\text{res}} \geq 0, \tag{9}$$

ensuring sufficient energy storage for transmission.

As the proposed model operates in underlay cognitive radio (CR) mode, the existing WiFi communication of the primary user (PU) should not be hampered. This constraint is satisfied by maintaining the PU data rate above a predefined threshold:

$$R_{\text{pu}} \geq R_{\text{pu}}^{th}, \tag{10}$$

which protects the PU from interference.

High-speed connectivity for the sensor data in 6G networks is essential to fulfill the desired throughput requirement. Therefore, the GS data rate must satisfy:

$$R_{\text{sr}} \geq R_{\text{sr}}^{th}, \tag{11}$$

where R_{sr}^{th} is the minimum throughput requirement for sensor data. where E_{res}, R_{pr}, R_{sr}, R_{th}^p and R_{sr}^{th} denote the total residual energy (self-powering level), the primary user (PU) data rate, the secondary/ground station (SU/GS) data rate, the minimum throughput requirement for the PU, and the minimum throughput requirement for the Ground station or sensor data rate, respectively.

The RIS phase-shift constraint ensures proper intelligent reflection in the 5G/6G environment and is given by:

$$0 < \phi_k < 2\pi, \qquad k \in \{sr, sp\}. \tag{12}$$

This paper aims to minimize the outage probability P_{out} for the IoT Sensor–EH–RIS–CR data transmission model by finding the optimal energy harvesting time fraction α using RL. The outage probability (P_{out}) calculated at the edge node for the link Sensor-RIS-Edge node is defined as the probability that the instantaneous data rate ($\mathcal{R}_{sr}$) falls below a predefined threshold $\mathcal{R}_{th}$. P_{out} is given by

$$\begin{aligned} P_{\text{out}} &= \mathbb{P}\left(\mathcal{R}_{sr} < \mathcal{R}_{th}\right) = \mathbb{P}\left(\gamma_{sr}^C < \gamma_{\text{th}}\right) \\ &= \mathbb{P}\left(|h_{sd}^C|^2 < \frac{\gamma_{\text{th}} \sigma_d^2}{P_s}\right), \end{aligned} \tag{13}$$

where, the SNR threshold is defined by $\gamma_{\text{th}} = 2^{\frac{2\mathcal{R}_{th}}{1-\alpha}} - 1$. Here, $|h_{sd}^C|^2$ in Eq. 13 can be approximated as $|h_{sd}^C|^2 \approx \left(\sum_{i=1}^{M} |h_{sri}||h_{rdi}|\right)^2$.

Since each product $|h_{sri}||h_{rdi}|$ is a product of Rayleigh random variables, the same can be approximated as $|h_{sri}||h_{rdi}| \sim$ Gamma and $|h_{Csd}|^2 \sim$ Gamma$(k = M, \theta = \mathcal{D}_s/M)$.

Using the Gamma cumulative distribution function (CDF), the closed-form outage probability approximation is given by

$$P_{\text{out}} \approx \frac{1}{\Gamma(M)} \gamma\left(M, \frac{M\gamma_{\text{th}}\sigma_d^2}{P_s \mathcal{D}_s}\right), \tag{14}$$

where, M and $\gamma(a, x)$ denote the total number of RIS elements and lower incomplete gamma function, respectively. Classical mathematical optimization provide suboptimal performance due to dynamic channel variations and discrete action limitations. Therefore, this work adopts RL-based techniques Deep Deterministic Policy Gradient (DDPG) to analyze and compare performance under random environmental conditions.

IV. Optimal Solution

In this section actor critic DDPG model is proposed to solve the EH time fraction optimization problem with introduction to Q learning components.

A. Q-Learning Component

Q-Learning is a model free RL algorithm used when the action space is discrete and the system dynamics are unknown or too complex for analytical modeling. It is used to determine the optimal energy-harvesting time fraction α to minimizes the outage probability. The optimization problem is modeled as an MDP with: **State space:**$\mathcal{S} = \{(PU, E)\}$,where $PU \in \{0, 1\}$ denotes PU transmit or non transmit state and $E \in \{0, 1, \ldots, E_{\max}\}$ is the battery level; **Action space:**$\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_K\}$; **Reward function:**$r = 1 - P_{\text{out}}(\alpha, M, \theta^2)$; **Q-learning update:**$Q(s,a) \leftarrow Q(s,a) + \nu\big[r + \gamma \max_{a'} Q(s', a') - Q(s,a)\big]$, where ν is the learning rate and γ is the discount factor.

B. DDPG Component

The Deep Deterministic Policy Gradient (DDPG) represents an off-policy actor-critic algorithm [13] that combines the advantages of deep Q-learning and policy gradient [14]. The proposed system (i) follows an actor-critic architecture, where the actor network deterministically outputs a continuous value of α based on the current system observation and the critic network evaluates the performance of current action.(ii) uses two target networks - one for the actor and another for the critic to copy the time delayed version of original network for stabilizing training. DDPG optimizes continuous action α using actor–critic architecture. Reward:$r = 1-P_{\text{out}}(\alpha, M, \theta^2)$ where Pout is the outage probability dependent on frame–split fraction α, RIS dimension M, and channel fading variance θ^2. A supervised training approach is integrated into the critic to improve system stability and convergence. Here, the critic is trained using synthetic data generated from the analytical outage model rather than random exploration alone. This allows the critic to accurately approximate the outage characteristics and consequently direct the actor to better solutions. During training, the critic parameters are optimized via mean squared error minimization. *Critic loss:*

$$L_c = \big(Q_c(\alpha, M, \theta^2) - (1 - P_{\text{out}}(\alpha, M, \theta^2))\big)^2 \tag{15}$$

Deterministic policy gradient:

$$\nabla_\theta J \approx \mathbb{E}\big[\nabla_\alpha Q_c(\alpha, M, \theta^2)\nabla_\theta \pi(\alpha)\big] \tag{16}$$

Algorithm 1 Outage Approximation vs α using DDPG

Initialize actor network $\pi(\alpha|M_0, \Theta_2)$ with parameters θ^π
Initialize critic network $Q_c(\alpha, M_0, \Theta_2)$ with parameters θ^Q
Initialize target actor π' with $\theta^{\pi'} \leftarrow \theta^\pi$
Initialize target critic Q'_c with $\theta^{Q'} \leftarrow \theta^Q$
Initialize replay buffer $\mathcal{D}$
for epoch $= 1$ to E **do**
  Sample $\alpha \sim U(0, 1)$
  Execute action and observe reward $r = 1 - P_{out}$
  Store transition (α, r) in replay buffer $\mathcal{D}$
  Sample mini-batch from $\mathcal{D}$
  Compute target value $y = r + \gamma Q'_c(\alpha', M_0, \Theta_2)$
  Update critic by minimizing loss $L_c = (Q_c(\alpha, M_0, \Theta_2) - y)^2$
  Update actor using policy gradient

$$\nabla_{\theta^\pi} J \approx \nabla_\alpha Q_c(\alpha, M_0, \Theta_2)\nabla_{\theta^\pi} \pi(\alpha|M_0, \Theta_2)$$

  Update target networks

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$

$$\theta^{\pi'} \leftarrow \tau\theta^\pi + (1-\tau)\theta^{\pi'}$$

Output: Predicted reward $\hat{r}$
Compute outage approximation $\hat{P}(\alpha) = 1 - \hat{r}$
Apply Gaussian smoothing

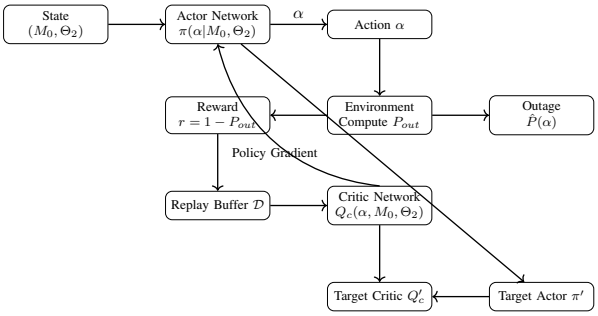

Fig. 3: DDPG learning framework for outage approximation including actor, critic, replay buffer, and target networks.

Fig. 3 illustrates the general block diagram of the DDPG learning framework used for outage approximation, comprising the actor network, critic network, replay buffer, and target networks.

C. Algorithm

The DDPG algorithm uses the following settings: state space $\mathcal{S} = \{(PU, E)\}$, action set $\mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$, outage model $P_{\text{out}}(\alpha, M, \theta^2)$, learning rate η, discount factor γ, exploration parameters $\varepsilon_{\text{start}}$ and ε_{end}, and training parameters N_{ep} episodes with N_{step} steps per episode. The outage probability approximation vs α and M using DDPG are summarized in Algorithm 1 and 2, respectively.

Algorithm 2 Outage Approximation vs. M using DDPG

Input: Training RIS sizes $\mathcal{M}_{train}$, Evaluation RIS sizes $\mathcal{M}_{eval}$
Initialize: Actor network μ_θ, Critic network Q_ϕ, Target networks
Replay buffer $\mathcal{D} = \emptyset$

— Training Phase —
for episode $= 1$ to N **do**
  Randomly select $M \in \mathcal{M}_{train}$
  Initialize state s_0
  for each step t **do**
    Select action $a_t = \mu_\theta(s_t, M) + \text{noise}$ *(energy harvesting ratio α)*
    Apply action and observe next state s_{t+1} and outage $P_{out}(s_t, a_t, M)$
    Reward $r_t = 1 - P_{out}(s_t, a_t, M)$
    Store $(s_t, a_t, r_t, s_{t+1}, M)$ in buffer $\mathcal{D}$
    Sample batch from $\mathcal{D}$ and update critic and actor networks
    Soft-update target networks

— Outage Approximation over M —
for each $M \in \mathcal{M}_{eval}$ **do**
  for each α in search grid **do**
    Predict reward $\hat{r}(\alpha, M) = Q_\phi(s, \alpha)$
    Compute outage estimate $\hat{P}_{out}(\alpha, M) = 1 - \hat{r}(\alpha, M)$
  Best outage $\hat{P}^{DDPG}_{out}(M) = \min_\alpha \big(\hat{P}_{out}(\alpha, M)\big)$
Output: Approximated outage curve $\hat{P}^{DDPG}_{out}(M)$

V. Results and Discussion

Fig. 4 illustrates the outage probability versus the energy harvesting time fraction α comparing Q-learning and DDPG (critic). Initially, the outage probability decreases as α increases due to improved harvested energy. However, further

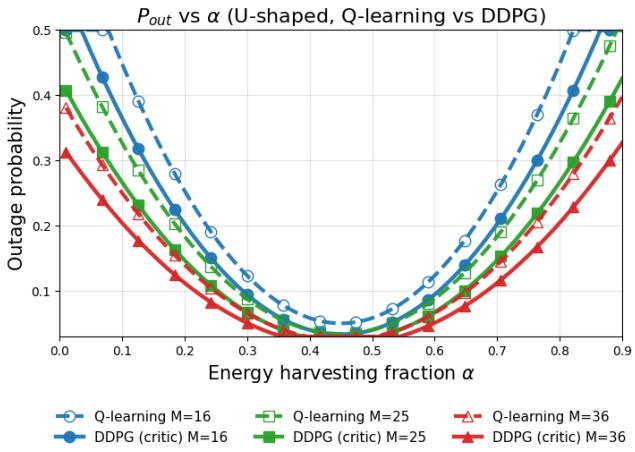

Fig. 4: Outage probability vs α.

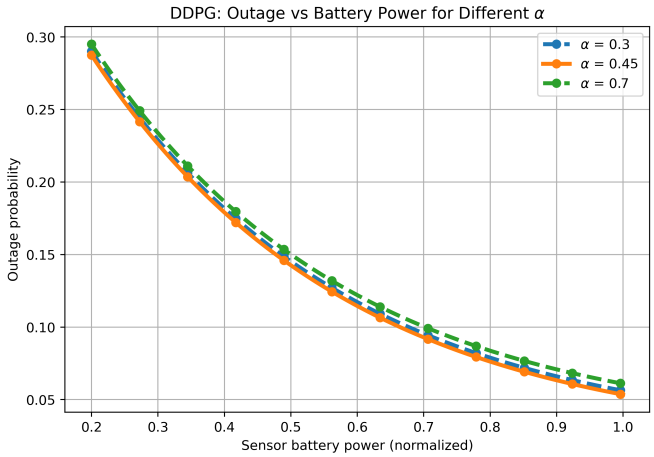

Fig. 6: Outage probability vs battery power.

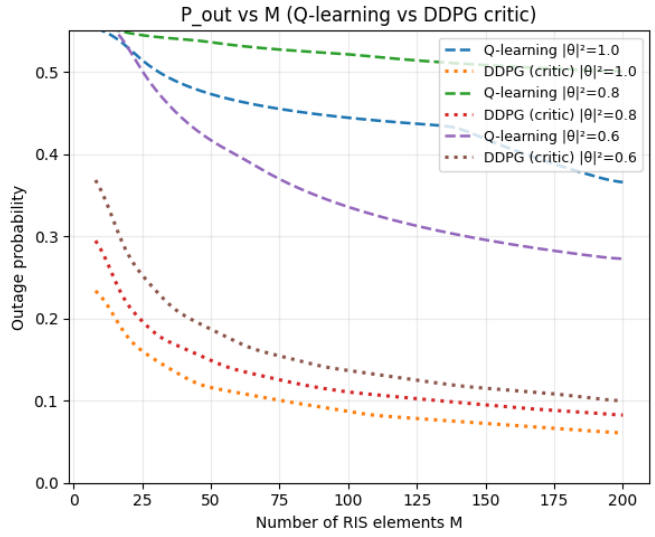

Fig. 5: Outage probability vs M.

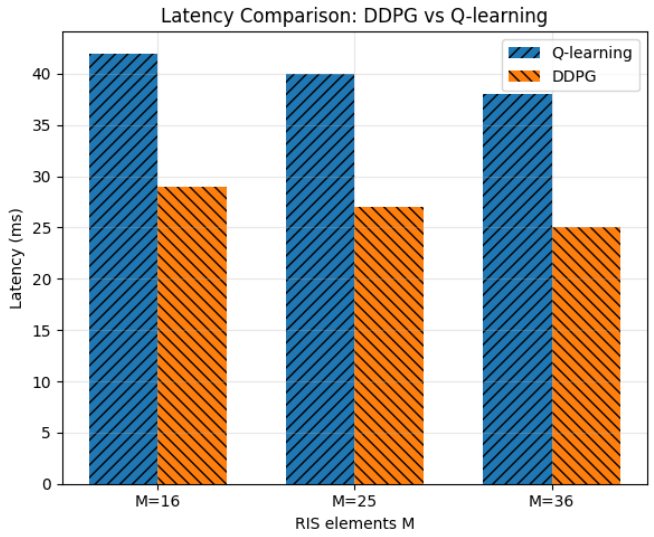

Fig. 7: Latency comparison.

increase in α reduces the transmission duration, resulting in higher outage probability. It is observed that $P_{out}[M = 36] < P_{out}[M = 25] < P_{out}[M = 16]$ for both Q-learning and DDPG approaches, indicating that increasing the number of RIS elements gives better result. The curves achieve minimum outage near $\alpha \approx 0.45$ and exhibit U shaped curve just like analytic mode.

At the optimal point, Q-learning results in outage values of approximately 0.055, 0.048 and 0.044 for $M = 16, 25$, and 36, respectively. In comparison, the proposed DDPG-critic model significantly reduces the outage to about 0.040, 0.036 and 0.032 for the same values of $M = 16, 25$, and 36, respectively. This corresponds to an average outage reduction of 25–28%, demonstrating that DDPG achieves better optimization of energy harvesting and phase-shift control than Q-learning.

Fig. 5 presents the variation on outage probability with the number of RIS elements M for different channel gains $|\theta|^2 = \{1.0, 0.8, 0.6\}$. The outage probability decreases monotonically as M increases due to enhanced passive beamforming gain provided by the RIS. For $|\theta|^2 = 1.0$, Q-learning achieves outages of approximately 0.56, 0.42 and 0.36 at $M = 16, 100$, and 200, respectively, whereas DDPG significantly reduces outage to about 0.48, 0.28 and 0.22. Similarly, for $|\theta|^2 = 0.8$, the DDPG model achieves outage probabilities of approximately 0.18 at $M = 50$ and 0.11 at $M = 200$, compared with 0.32 and 0.22 for Q-learning. For the weakest channel with $|\theta|^2 = 0.6$, DDPG outperforms Q-learning by more than 56% at $M = 200$, demonstrating its ability to learn energy-efficient reflection coefficients.

Fig. 6 shows outage probability versus sensor battery power for different α. The outage probability decreases as battery power increases. The optimal harvesting fraction $\alpha = 0.45$ provides the minimum outage probability compared to the other values of α. Fig. 7 compares the latency performance where Q-learning achieve convergence in approximately 125 ms, whereas DDPG converges within nearly

75 ms, an improvement of about 40% latency. The reduced convergence time for DDPG results from the continuous action space exploration and efficient gradient-based policy update mechanism, make it more approachable for real-time dynamic wireless environments.

VI. Conclusion

This paper investigated on outage probability minimization of an self sustainable sensors working in RIS-assisted wireless communication system. A DDPG model was developed to jointly optimize the EH time fraction and RIS reflection control under varying channel conditions and battery energy levels. The proposed DDPG approach was compared with Q-learning technique. Simulation results demonstrate that DDPG consistently outperforms Q-learning across all evaluated RIS sizes and channel strengths, achieving faster convergence and significantly lower outage probability. DDPG achieves up to 28% lower outage probability around the optimal energy harvesting point and more than 56% improvement as the number of RIS elements increases. Additionally, DDPG reduces convergence latency by nearly 40% and maintains significantly lower outage probability under low sensor battery conditions, making it useful for dynamic wireless environment.

References

[1] C. Wang *et al.*, "Trustworthy health monitoring based on distributed wearable electronics with edge intelligence," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2333–2341, 2024.

[2] C.-M. Chen *et al.*, "Sustainable secure communication in consumer-centric electric vehicle charging in industry 5.0 environments," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 1544–1555, 2024.

[3] H. N. S. Aldin, M. R. Ghods, F. Nayebipour, and M. N. Torshiz, "A comprehensive review of energy harvesting and routing strategies for IoT sensors sustainability and communication technology," *Sensors Int.*, vol. 5, p. 100258, 2024.

[4] M. H. Alsharif, A. Jahid, A. H. Kelechi, and R. Kannadasan, "Green IoT: A review and future research directions," *Symmetry*, vol. 15, no. 3, Art. 757, 2023, doi: 10.3390/sym15030757.

[5] G. Moloudian, M. Hosseinifard, S. Kumar, R. B. V. B. Simorangkir, J. L. Buckley, C. Song, G. Fantoni, and B. O'Flynn, "RF energy harvesting techniques for battery-less wireless sensing, Industry 4.0, and Internet of Things: A review," *IEEE Sensors J.*, vol. 24, no. 5, pp. 5732–5745, 2024.

[6] F. Al-Azhary and S. Ahmeda, "Radio frequency energy harvesting with power splitting scheme in wireless sensor network," in *Proc. 2024 IEEE 4th Int. Maghreb Meeting Conf. Sci. Tech. Autom. Control Comput. Eng. (MI-STA)*, 2024, pp. 573–578.

[7] M. Banafaa, I. Shayea, J. Din, M. H. Azmi, A. Alashbi, Y. I. Daradkeh, and A. Alhammadi, "6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities," *Alex. Eng. J.*, vol. 64, pp. 245–274, 2023, doi: 10.1016/j.aej.2022.08.017.

[8] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 437–461, 2022.

[9] M. Ahmed, A. A. Soofi, S. Raza, Y. Li, F. Khan, W. U. Khan, M. Asif, and Z. Han, "A comprehensive survey on RIS-enhanced physical layer security in UAV-assisted networks," *IEEE Internet Things J.*, vol. 12, no. 16, pp. 32538–32562, 2025, doi: 10.1109/JIOT.2025.3569716.

[10] H. Zhou, M. Erol-Kantarci, Y. Liu, and H. V. Poor, "A survey on model-based, heuristic, and machine learning optimization approaches in RIS-aided wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 2, pp. 781–823, 2023.

[11] T. Liu, D. Xu, T. Zhang, S. Zhang, J. Chen, K. Yu, and V. C. M. Leung, "Deep reinforcement learning-based computation rate maximization for RIS-aided edge computing in wireless consumer application networks," in *Proc. 2025 IEEE Int. Conf. Consum. Electron. (ICCE)*, 2025, pp. 1–6.

[12] S. P. Maity, A. Banerjee, C. Chakraborty and S. Singh,"SDN-IoCE for Intelligent Self-Powered Gas Sensor Monitoring," *IEEE Consumer Electronics Magazine*, vol. 15, no. 4, pp. 93-98, July 2026, doi: 10.1109/MCE.2025.3628424.

[13] V. R. Konda and J. N. Tsitsiklis, "Actor-Critic Algorithms," in *Advances in Neural Information Processing Systems*, vol. 12, MIT Press, 1999.

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