

Environment-Aware Power Beacon Placement for RF-WET IoT Networks Using DDPG

João Gabriel Ferreira Ribeiro, Gilberto Cezar Silveira Júnior e Victoria Dala Pegorara Souto

Abstract—This paper proposes a Deep Deterministic Policy Gradient (DDPG)-based method for optimizing the horizontal placement of power beacons (PBs) in radio-frequency wireless energy transfer (RF-WET) IoT networks. The model considers Rician fading, distance-dependent path loss, nonlinear energy harvesting, and environmental effects from temperature and wind speed. The goal is to maximize the number of IoT devices whose harvested energy exceeds a minimum threshold. To handle the resulting nonconvex continuous optimization problem, DDPG is used to learn a continuous PB positioning policy. Results show that the proposed method learns stable policies and improves the number of served devices, highlighting the potential of deep reinforcement learning for RF-WET optimization in IoT scenarios.

Keywords—IoT, Wireless Energy Transfer, Reinforcement Learning.

I. Introduction

According to Statista projections [1], the number of Internet of Things (IoT) devices is expected to reach 40.6 billion by 2034, driven by applications in health-care, home automation, manufacturing, and critical infrastructure. This growth intensifies the challenge of energy supply, since battery-powered devices are limited by finite lifetime, degradation, and maintenance costs, while fixed power supply increases deployment complexity, particularly in large-scale or remote scenarios.

RF-WET is a promising solution to reduce manual battery replacement in IoT networks. However, its efficiency is constrained by severe path loss, transmit power limitations, and design aspects related to coverage, scalability, and energy consumption [2]. In this context, beamforming and power beacon (PB) placement are key factors. Although beamforming can enhance received energy, it typically requires channel state information, increasing signaling overhead and system complexity [2]. Moreover, inadequate PB placement may intensify the doubly near-far problem, create energy coverage gaps, and increase infrastructure power consumption [2].

PB placement has been addressed under different assumptions and objectives, including omnidirectional

PB deployment [3], wavelet-based near-optimal positioning [4], directional PB orientation [5], cost-aware and task-oriented formulations, reinforcement learning approaches, and online placement models [6], [7]. Nevertheless, efficient PB positioning in continuous and dynamic environments remains challenging, particularly when multiple performance criteria must be considered. Therefore, this work investigates the use of Deep Deterministic Policy Gradient (DDPG), a continuous reinforcement learning method, to optimize PB placement and improve energy transfer efficiency in wireless-powered IoT networks without relying on exhaustive search or restrictive analytical models.

II. System Model and Problem Formulation

This paper considers a wireless-powered IoT network composed of L PBs and K single-antenna IoT devices. Each PB is equipped with a uniform linear array of N antennas and wirelessly transfers energy to the IoT devices. The RF power received by the k -th IoT device is given by

$$P_k^{\text{rx}} = \sum_{l=1}^L g_{k,l} |\mathbf{w}_{k,l} \mathbf{h}_{k,l}^H|^2, \quad (1)$$

The harvested energy at the k -th IoT device during one coherence interval τ_c is modeled by the nonlinear energy-harvesting function

$$E_k = \frac{\tau_c(\mu - \mu\Omega) [1 + e^{-a(P_k^{\text{rx}} - b)}]}{(1 - \Omega) [1 - e^{-a(P_k^{\text{rx}} - b)}]}, \quad (2)$$

where $\Omega = 1/(1 + e^{ab})$, μ is the maximum harvested power, and a and b are circuit-dependent parameters [8]. All IoT devices are assumed to have identical harvesting circuit parameters.

The objective of this paper is to maximize the number of served IoT devices, i.e., devices whose harvested energy is not lower than the threshold $E_{\min}$. Thus, the beamforming vectors and the horizontal PB positions $(x_{\text{PB}_l}, y_{\text{PB}_l})$ are jointly optimized as

$$\begin{aligned} & \underset{\{\mathbf{w}_{k,l}\}, \{x_{\text{PB}_l}, y_{\text{PB}_l}\}}{\text{maximize}} && \sum_{k=1}^K \mathbb{I}(E_k \geq E_{\min}) \\ & \text{subject to} && x_{\text{PB}_l}, y_{\text{PB}_l} \in [0, R], \quad \forall l \in \{1, \dots, L\}, \\ & && |w_{k,l,m}|^2 = \frac{P_T}{N}, \quad \forall m \in \{1, \dots, N\}, \end{aligned} \quad (3)$$

João Gabriel Ferreira Ribeiro, Gilberto Cezar Silveira Júnior e Victoria Dala Pegorara Souto, National Institute of Telecommunication, e-mail: {joao.g@get.inatel.br; gilberto.junior@gec.inatel.br; victoria.souto@inatel.br}.

where $\mathbb{I}(\cdot)$ is the indicator function defined as

$$\mathbb{I}(E_k \geq E_{\min}) = \begin{cases} 1, & E_k \geq E_{\min}, \\ 0, & E_k < E_{\min}. \end{cases}$$

The problem in (3) is non-convex due to the nonlinear energy-harvesting model, the dependence of the received power on PB positions and beamforming vectors, and the stochastic wireless channel. Moreover, the continuous decision variables make exhaustive search impractical, motivating the deep reinforcement learning solution proposed in the next section.

III. Proposed DDPG-Based Solution

The proposed solution formulates the PB placement problem as a sequential decision-making task and employs the DDPG algorithm to learn a continuous control policy. DDPG is an actor-critic deep reinforcement learning method suitable for continuous action spaces. In the proposed framework, the wireless-powered IoT network is modeled as the environment, and the agent observes the system state, selects continuous PB displacement actions, receives a reward based on network performance, and updates its policy to maximize the long-term accumulated reward. The agent-environment interaction is defined as follows:

- **State:** The state is defined by the normalized horizontal coordinates of the L PBs, i.e.,

$$\mathbf{s}_t = \left[\frac{x_{\text{PB}_1}}{R}, \frac{y_{\text{PB}_1}}{R}, \dots, \frac{x_{\text{PB}_L}}{R}, \frac{y_{\text{PB}_L}}{R} \right] \in [0, 1]^{2L},$$

where the physical PB coordinates belong to $[0, R]$.

- **Action:** The action represents the horizontal PB displacements, $\mathbf{a}_t = [\Delta x_{\text{PB}_1}, \Delta y_{\text{PB}_1}, \dots, \Delta x_{\text{PB}_L}, \Delta y_{\text{PB}_L}]$, with each component bounded by $[-\delta, \delta]$. The updated PB positions are clipped to remain within the $R \times R$ m² deployment area.
- **Reward:** The reward is the number of IoT devices satisfying the minimum harvested-energy requirement, $r_t = \sum_{k=1}^K \mathbb{I}(E_k(t) \geq E_{\min})$, encouraging PB placements that maximize the number of served devices.

The DDPG agent is composed of an actor and a critic network. The actor maps the system state into continuous PB displacement actions, whereas the critic estimates the corresponding state-action value to guide policy improvement. Training stability is enhanced through experience replay, in which transitions $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$ are stored and randomly sampled in minibatches, and through slowly updated target networks. Exploration is promoted by adding stochastic noise to the actor output during training. Thus, the proposed DDPG-based approach learns adaptive PB placement policies that consider network geometry and propagation conditions while avoiding exhaustive search over the continuous deployment region.

A. Environmental-Aware Effective Channel Gain

For the received RF power computation, phase-aligned beamforming is assumed for each PB-IoT link. To account for practical outdoor deployments, especially with drone-mounted PBs, temperature- and wind-speed-dependent degradation factors are incorporated into the channel gain. Thus, the effective gain between the l -th PB and the k -th IoT device is defined as $g_{k,l}^{\text{eff}} = g_{k,l}/t_{k,l}v_{k,l}$, where $t_{k,l}$ and $v_{k,l}$ represent the temperature and wind-speed degradation factors, respectively.

The wind-speed degradation factor is modeled as

$$v_{k,l} = \begin{cases} (e^{0.01 \times 31.7} - 1) \frac{V_{k,l}}{31.7} + 1, & V_{k,l} \leq 31.7, \\ e^{0.01 V_{k,l}}, & V_{k,l} > 31.7, \end{cases} \quad (4)$$

where $V_{k,l} \in [0, 90]$ km/h denotes the wind speed. This model captures mild degradation under moderate winds and stronger degradation at high wind speeds. The temperature-dependent factor is given by $t_{k,l} = 0.0002(T_{k,l} - 25)^2 + 1$, where $T_{k,l} \in [-29, 62.3]^\circ\text{C}$ is the temperature. This expression models the degradation caused by deviations from the reference temperature of 25°C .

By including these factors, the channel model captures the impact of environmental conditions on wireless energy transfer. Since the harvested energy is sensitive to variations in the effective channel gain, the proposed DDPG-based policy learns PB placement strategies that account for both network geometry and environmental impairments.

IV. Results and Discussion

This section evaluates the proposed DDPG-based PB placement strategy. The simulations considered $R = 30$ m, $K = [50, 100, 200, 300]$, $L = [1, 2, 3, 4]$, $N = 4$, $P_T = 2$ W, $f = 915$ MHz, $\mu = 10.73$ mW, $\alpha = 1.5$, $a = 5.365$, and $b = 0.2308$. The DDPG agent used actor and critic learning rates of 10^{-3} and 2×10^{-3} , respectively, with two hidden layers of 64 and 128 neurons. The discount factor and soft-update rate were set to $\gamma = 0.95$ and $\tau = 0.001$, respectively. The minibatch size was $B = 32$, and the replay-buffer capacity was $C_{\text{buffer}} = 256109$ transitions. The standard deviation of the exploration noise was $\sigma_n = 0.1$, with a clipping limit of $n_{\text{clip}} = 0.2$. The parameter γ controls the contribution of future rewards, whereas τ determines the soft-update rate of the target networks. Training was conducted for 500 episodes, with each episode limited to 200 steps. The minimum harvested-energy requirement was set to $E_{\min} = 1$ μJ .

The reward was normalized by $K \times$ steps to enable a fair comparison among scenarios with different network densities. All curves remain close to unity, indicating high normalized performance for all evaluated values of K . The observed fluctuations are mainly

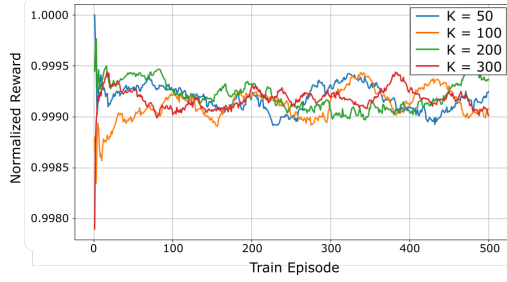


Fig. 1: DDPG convergence for different values of K , with $L = 3$, $N = 4$, and 200 steps per episode.

caused by exploration noise and random device distributions. In addition, the absence of a clear increasing trend suggests that the agent rapidly reaches a high-performance operating region or that the reward becomes saturated under the adopted simulation conditions.

The percentage of served IoT devices was evaluated for $K = \{50, 100, 200, 300\}$ and $L = \{1, 2, 3, 4\}$. In all cases, the proposed method achieved 100% service, serving all devices for each value of K . This indicates that, under the adopted parameters, even a single PB was sufficient to meet the minimum harvested-energy requirement, and additional PBs did not improve this already saturated metric. Due to space limitations, the corresponding figure is omitted.

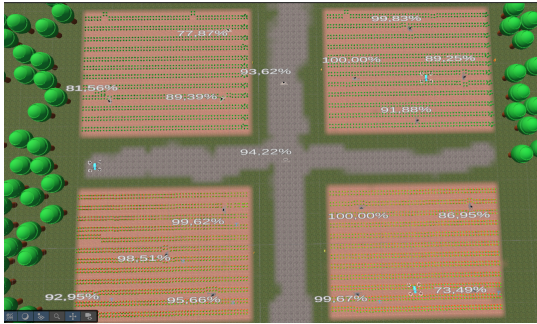


Fig. 2: Unity-based visualization of the proposed PB placement strategy in a wireless-powered IoT scenario.

To complement the numerical analysis, a Unity-based scenario was developed to visualize the proposed PB placement strategy in a three-dimensional environment. Unity provides a more realistic representation of the spatial distribution of PBs and IoT devices, coverage behavior, and positioning dynamics, serving as an intermediate validation step between numerical simulations and future experimental tests.

As shown in Fig. 2, the proposed method places the PBs in suitable regions to provide effective energy coverage to many IoT devices. The results also show that the DDPG-based solution remains robust under environmental degradation caused by temperature and wind-speed losses, which may affect beam alignment

and propagation quality. This confirms the ability of the proposed approach to adapt to both network geometry and realistic outdoor impairments, especially in drone-mounted PB scenarios.

V. Conclusion

This paper proposed a DDPG-based method for PB placement in RF-WET IoT networks, considering Rician fading, nonlinear energy harvesting, and environmental losses due to temperature and wind speed. Results showed that the proposed approach learns effective and robust placement policies, achieving high service levels under realistic impairments. The Unity-based visualization further supported its practical applicability, while future work will address more challenging scenarios and experimental validation.

Acknowledgments

The following research projects have funded this work: Brasil 6G Project with support from RNP/MCTI (01245.020548/2021-07), xGMobile Projects code XGM-AFCCT-2024-4-1-1 and XGM-AFCCT-2026-4-1-1 with resources from EMBRAPII/MCTI (Grant 052/2023 PPI IoT/Manufatura 4.0 and FAPEMIG Grant PPE-00124-23), by FAPEMIG (Grant APQ-04523-23, APQ-05305-23, APQ-01558-24, RED-00194-23, BIS-00529-25), by CNPq (305021/2021-4, 443974/2024-1, INCT STREAM 409179/2024-8), and by Centro de Competência em Redes 5G e 6G Ref. Finep nº 1060/2 contract nº 0-1-25-0883-00.

References

- [1] Statista, "Number of Internet of Things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2034," 2024, statista.
- [2] S. Zhou, H. Dai, H. Sun, G. Tan, and B. Ye, "On the deployment of clustered power beacons in random wireless powered communication," *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 2424–2438, 2023.
- [3] O. M. Rosabal, O. L. A. Lopez, H. Alves, S. Montejo-Sanchez, and M. Latva-Aho, "On the Optimal Deployment of Power Beacons for Massive Wireless Energy Transfer," *IEEE Internet of Things Journal*, vol. 8, no. 13, p. 10531–10542, Jul. 2021.
- [4] D. Arivudainambi and S. Balaji, "Optimal Placement of Wireless Chargers in Rechargeable Sensor Networks," *IEEE Sensors Journal*, vol. 18, no. 10, pp. 4212–4222, 2018.
- [5] H. Dai, X. Wang, A. X. Liu, H. Ma, G. Chen, and W. Dou, "Wireless Charger Placement for Directional Charging," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1865–1878, 2018.
- [6] X. Ding, Y. Wang, G. Sun, C. Luo, D. Li, W. Chen, and Q. Hu, "Optimal charger placement for wireless power transfer," *Computer Networks*, vol. 170, p. 107123, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1389128619307947>
- [7] D. Zorbas, P. Raveneau, Y. Ghamri-Doudane, and C. Douligeris, "The charger positioning problem in clustered RF-power harvesting wireless sensor networks," *Ad Hoc Networks*, vol. 78, pp. 42–53, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1570870518302440>
- [8] O. L. A. López, S. Montejo-Sanchez, R. D. Souza, C. B. Papadiaz, and H. Alves, "On csi-free multi-antenna schemes for massive rf wireless energy transfer," *IEEE Internet of Things Journal*, vol. 8, pp. 278–296, 2021.