A Review on The Current Research Status of Base Station Sleep Strategies Using Deep Reinforcement Learning

Pushen Chen *

Nanjing University of Posts and Telecommunications, Nanjing, 210023, Jiangsu, China

* Corresponding Author Email: 3468088202@qq.com

Abstract. This paper presents a comprehensive review of the current research status and application advancements of deep reinforcement learning (DRL) in base station (BS) sleep control. With the ultra-dense deployment of 5G/6G networks, the energy consumption of BSs has become a critical issue, making dynamic sleep control a key technology for improving energy efficiency (EE). This work systematically analyzes core DRL-based approaches for BS sleep strategies, including DQN, PPO, DDPG, and others, as well as their trade-off mechanisms for multi-objective optimization involving EE, spectral efficiency (SE), and quality of service (QoS). Furthermore, it explores joint optimization schemes integrating emerging technologies such as renewable energy, cell zooming, and reconfigurable intelligent surfaces (RIS). A tripartite taxonomy is proposed, categorizing existing methods into value-based, policy gradient-based, and hybrid algorithms. Representative studies are summarized, highlighting the advantages and challenges of current DRL techniques in BS sleep control.

Keywords: Deep Reinforcement Learning; Base Station Sleep; Energy Efficiency Optimization; Deep Q Network; Ultra-Dense Network.

1. Introduction

In recent years, the rapid advancement of fifth-generation (5G) mobile communication technologies and the widespread adoption of IoT-enabled smart devices have led to exponential growth in global wireless communication demands. Against this backdrop, the high energy consumption of 5G networks has become increasingly prominent. As the core equipment responsible for the conversion between wireless radio signals and optical fiber signals, base stations (BSs) have emerged as a critical factor influencing the overall energy efficiency of communication systems. Functioning as the fundamental unit for physical coverage in mobile networks, BSs serve as physical-layer entities in the radio access network (RAN), where their deployment density and topological configuration directly determine network coverage performance. With the advent of 6G on the horizon, terahertz (THz) communications will necessitate even denser BS deployments. Whether from the perspective of current 5G technology or future 6G advancements, BSs retain irreplaceable significance in wireless infrastructure.

Compared with fourth-generation (4G) mobile communication systems that utilize low-frequency signals, fifth-generation (5G) base stations (BSs) exhibit significantly reduced coverage ranges, necessitating an exponential increase in deployment density to achieve comparable network coverage. To maintain equivalent coverage quality to 4G networks, the deployment scale of 5G BSs must be at least twice that of 4G BSs. From an energy consumption perspective, this leads to operational power costs several times higher than those of 4G systems for equivalent coverage areas, compounded by frequent underutilization of BS resources—a critical challenge that has directly spurred the development of deep sleep techniques for active antenna units (AAUs).The advancement of BS sleep technologies has demonstrated measurable benefits, including effective carbon emission reduction and unexpected operational expenditure (OPEX) savings across entire networks. Traditional heuristic algorithms, which represent a class of approximate optimization methods relying on empirical rules and local search mechanisms, provide suboptimal solutions for NP-hard problems within polynomial time complexity. These approaches can be broadly categorized into: constructive algorithms based on single-point search, improvement algorithms employing neighborhood search, and swarm

intelligence algorithms inspired by natural metaphors. However, such heuristic methods are inherently limited by their strong problem-specific dependency and lack of convergence guarantees, which significantly constrain their effectiveness in dynamic network environments.

In comparison with traditional heuristic algorithms, deep reinforcement learning (DRL) provides an innovative solution for achieving high energy efficiency by virtue of its data-driven characteristics and real-time decision-making advantages. This paper systematically reviews the methodological innovations, application scenarios, and technical challenges of DRL in base station sleep control. The research covers the following dimensions: first constructing an intelligent sleep architecture, then conducting in-depth analysis of DRL modeling methods, verifying algorithm effectiveness through typical cases, and finally discussing current technical difficulties and future development directions. This paper first briefly introduces the background knowledge of base station sleep and deep reinforcement learning (DRL). Subsequently, it proposes a three-category classification method to systematically categorize and review DRL-based base station sleep and energy conservation. The logical structure of this paper is shown in the following Figure 1.

Then it critically presents the challenges and shortcomings encountered in the viewpoints within the literature. Finally, it points out the current challenges and problems in the field of DRL-based base station sleep control, and proposes suggestions for future research directions.Compared with existing review papers on base station sleep strategies, this paper innovatively and comprehensively summarizes DRL-based base station sleep algorithms. Some existing review literature lacks descriptions of practical applications, and the cited references are not recently published.

Figure 1. The logical structure of this paper

2. Background Knowledge

2.1. Base Stations and Base Station Sleep

The system architecture of base stations mainly includes macro base stations, small cells, user equipment, and reconfigurable intelligent surfaces (RIS). These components work together through a hierarchical heterogeneous network architecture where macro base stations provide wide-area coverage, small cells achieve capacity enhancement and flexible deployment functions, and RIS dynamically regulates the propagation environment. This architecture is particularly suitable for high-

frequency band communications, where RIS can effectively compensate for high-frequency propagation losses while densely deployed small cells can solve high-frequency signal penetration loss problems.Base station sleep is a network energy efficiency optimization paradigm based on dynamic resource modulation. Its operational logic lies in realizing the switching of base station functional modules through programmable radio frequency unit state transition mechanisms, specifically manifested as intelligent switching of power amplifier operating modes, dynamic offloading of baseband processing unit computational resources, and adaptive reconstruction of antenna array radiation patterns. Base station sleep constitutes a constrained Markov decision process whose state space contains key parameters such as the dynamic equivalent noise coefficient of RF links, the buffer state matrix of baseband processors, and the spatiotemporal distribution characteristics of network load.

2.2. Deep Reinforcement Learning Model

The Deep Reinforcement Learning (DRL) model is a trial-and-error learning process that progressively achieves objectives through continuous reward feedback, focusing on long-term cumulative rewards rather than immediate gains [1]. By integrating deep learning with reinforcement learning, this approach enables decision-making in high-dimensional perceptual input spaces, significantly expanding the applicability of reinforcement learning. In DRL implementations, environmental dynamics manifest as partially observable complex systems where state transitions incorporate time-varying channel characteristics, traffic burst patterns, and user mobility uncertainties. The policy function adopts parametric representations, with policy gradient optimization algorithms continuously adjusting decision parameters to enhance energy efficiency while maintaining quality-of-service constraints. The reward mechanism integrates multi-dimensional optimization objectives including power consumption metrics, service quality parameters, and handover success rates, achieving Pareto optimality through dynamic weight adjustment. State-value functions employ deep neural networks for high-dimensional feature extraction and function approximation, providing agents with accurate long-term reward predictions. However, despite significant theoretical and practical advancements in deep reinforcement learning, several limitations persist, including challenges with stochastic policy formulation, continuous action space handling, and large state space applications. In the specific context of base station sleep strategy optimization, DRL agents utilize deep neural networks to process real-time network state information while strategically evaluating long-term rewards. When multiple agents operate concurrently, the operational efficiency of base station sleep can be precisely improved.

3. DRL-Based Optimization Methods for Base Station Sleep Strategies

3.1 Classification Overview

Through comprehensive literature review, this paper categorizes DRL-based optimization methods for base station sleep strategies into three classes: value function-based algorithms, policy gradient-based algorithms, and hybrid algorithms. This classification is established according to different optimization objectives and technical characteristics.Value function-based algorithms achieve sleep optimization by constructing accurate estimations of state-action value functions, with their core lying in establishing parametric representations of Q-functions. Policy gradient algorithms directly optimize parameterized policy functions through gradient ascent methods to realize timely sleep control, with Proximal Policy Optimization (PPO) as a typical representative that employs importance sampling and policy change magnitude constraints.

Hybrid algorithms adopting the Actor-Critic framework attempt to integrate the dual advantages of value function estimation and policy optimization. In this architecture, the Critic component is responsible for accurate state value evaluation, while the Actor component focuses on policy improvement.This section innovatively provides a detailed summary of DRL-based optimization

algorithms for base station sleep strategies through this tripartite classification, accompanied by critical commentary.

3.2 Value Function-Based Algorithms

Value function-based algorithms are particularly suitable for addressing discrete sleep depth selection problems. By incorporating experience replay mechanisms and fixed target networks, Deep Q-Networks (DQN) can effectively mitigate instability issues in high-dimensional state spaces compared to traditional Q-learning. The derived Double DQN algorithm significantly reduces value estimation bias by decoupling action selection from value evaluation. Reference [1] proposed a dual deep Q-network approach for solving base station sleep and power allocation problems, where the base station sleep DQN optimizes traditional system energy efficiency while the power allocation DQN jointly optimizes both EE and system spectral efficiency, thereby avoiding the limitations of single-objective optimization. Each small base station (SBS) is treated as an intelligent agent that interacts with the environment to obtain reward values. According to the requirements of base station sleep and power allocation, the state, action, and reward functions are designed as two schemes: For the base station sleep DQN, the state includes user load, base station status, total rate, and total energy consumption, with actions being base station on/off switching and rewards representing long-term accumulation; for the power allocation DQN, the state includes power level, channel gain, task load, and remaining time, with actions being power level adjustments and rewards representing immediate reward accumulation. Reference [2] proposed improved DQN algorithms including double-buffered DQN and T-DQN, which classify user demand according to base station bandwidth to reduce state dimensionality. Compared with classical DQN, these improvements achieve faster processing speeds and more stable training processes. The enhanced DQN algorithm designs conventional memory banks and burst memory banks to separately store transition samples under different load states, dynamically adjusting the memory sampling ratio during training. Early stages emphasize burst memory to quickly adapt to sudden scenarios, while later stages gradually approach the true distribution to improve stability. Reference [3] proposed the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm, which employs double DQN to suppress Q-value overestimation and implements delayed policy updates with action noise regularization to enhance system stability. Reference [4] presented Deep Nap, a dynamic sleep control algorithm using deep reinforcement learning, where deep Q-networks learn energy-efficient sleep strategies from high-dimensional observations. Although value function-based algorithms are effective for solving discrete base station sleep problems, they face inherent limitations in handling constrained action space dimensionality. Since base station sleep fundamentally constitutes a continuous control problem involving precise adjustment of transmission power, the forced discretization of continuous action spaces in value function-based methods inevitably leads to reduced control accuracy.

3.3 Policy Gradient-Based Algorithms

Policy gradient-based algorithms achieve base station sleep operation objectives by directly optimizing parameterized policy functions, with parameter updates following the policy gradient theorem. Compared to value function-based approaches, these methods can directly output continuous sleep depth values, enabling joint optimization of beamforming parameters while automatically learning trade-off relationships between objectives to produce optimized sleep strategies. The policy gradient algorithm family includes Reinforce, Proximal Policy Optimization (PPO), Trust Region Policy Optimization (TRPO), and other improved variants. Reference [5] models RIS-assisted multi-cell networks as Markov decision processes, employing PPO to optimize sleep modes, cell zooming, and user association while utilizing dual cascaded correlation networks to optimize RIS reflection coefficients. Compared to baseline DQN, PPO achieves better energy-delay trade-offs, resulting in significant energy reduction. Reference [6] addresses dynamic renewable energy source (RES) harvesting and traffic load variations by designing a multi-discrete PPO algorithm to combat action space dimensionality explosion caused by increasing SBS numbers. The

authors demonstrate that PPO maintains robust performance during coordinated multi-SBS operation, whereas traditional methods like DQN and Deep Deterministic Policy Gradient fail due to dimensionality explosion. Reference [7] proposes a Multi-Agent PPO with DRl constraints (MAPPO-DRL) algorithm, constructing a communication network architecture considering resource constraints and optimal trajectory planning. Through joint optimization of user association, resource allocation, and UAV-BS trajectories using block coordinate descent - balancing K-means clustering for user association subproblems while applying resource-constrained MAPPO-DRL for joint resource-trajectory optimization - the approach maximizes total system throughput. Reference [8] develops a PPO-based low-altitude network base station planning method, where the PPO reinforcement learning model avoids unnecessary base station operation computations to reduce computational burden, while discretizing critical base station configuration parameters overcomes high-dimensional action space challenges. In multi-agent scenarios, the exponential growth of joint action space dimensionality creates computational bottlenecks for value function estimation methods. In contrast, policy gradient algorithms demonstrate superior scalability through their policy gradient framework combined with clipped objective functions, enabling decoupled optimization of discrete action dimensions. Incorporating offline reinforcement learning with prioritized experience replay reduces environmental interaction samples. However, the temporal non-stationarity of base station traffic poses adaptation challenges for policy-based methods, suggesting future research directions incorporating meta-reinforcement learning for rapid policy adaptation to traffic pattern variations.

3.4 Hybrid Algorithms

Hybrid algorithms employ the Actor-Critic dual-network architecture to achieve coordinated policy optimization and value estimation, combining the dual utilities of value functions and policy functions. These approaches can simultaneously handle discrete sleep mode selection and continuous parameter adjustment while reducing policy gradient variance through advantage functions to enable multi-timescale optimization. Representative algorithms include Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C), Deep Deterministic Policy Gradient (DDPG), Twin Delayed DDPG (TD3), and other innovative variants. Reference [9] proposes a geo-semantic spatiotemporal network modeling framework that concurrently considers geographical proximity and semantic similarity of traffic flows, utilizing DDPG algorithm within an Actor-Critic (AC) architecture. The method introduces baseline transformation (BT) to reduce cost estimation variance caused by dynamic traffic and employs an exploration network (EN) to enhance exploration efficiency in high-dimensional action spaces. Reference [10] presents a DDPG algorithm that combines deep Q-learning with PPO, where DPG extends DQN to continuous action spaces for learning deterministic policies (unlike original DQN limited to discrete spaces) through deterministic decision model design. Reference [11] develops a DDPG-PSO algorithm that improves particle swarm optimization via deep reinforcement learning, maintaining PSO's convergence speed while obtaining superior deployment solutions by applying customized DDPG during each PSO iteration to determine optimal particle velocities. Reference [12] addresses the highly coupled nature of base station beamforming and RIS phase control by employing DRL for joint optimization, incorporating RIS element gains and antenna radiation patterns into traditional geometric channel models. As a model-free off-policy algorithm, DDPG integrates DPG and DQN principles for continuous action space learning, enabling joint optimization of BS beamforming matrices and RIS reflection coefficient matrices to maximize total system rate. Reference [13] focuses on UAV base stations, proposing a continuous action space Actor-Critic strategy for UAV-BS positioning in mobile terminal environments. This strategy provides greater flexibility than traditional discrete action spaces by allowing agents to select actions within continuous ranges. While hybrid algorithms demonstrate unique advantages in concurrent discrete-continuous optimization and multi-dimensional reward fusion through Critic networks (maintaining sleep operation stability even under burst traffic), they face critical challenges including stringent Critic network inference latency requirements, high training energy consumption, potential coverage hole durations exceeding 2 seconds during

exploration, increased wake-up failure probabilities, and core network signaling overload. The summaries of each specific algorithm are shown in Table 1.

Table 1. Summary of existing work

References	Category	algorithm	advantage
[1][2][3][4]	Value-function-based algorithm	DQN Double DQN	solve the instability problem in the high-dimensional state space.
[5][6][7][8]	Strategy-function based algorithm	TRPO, PPO, MAPPO	directly output continuous values of sleep depth, and achieve joint optimization of beamforming parameters
[9][10][11][12][13]	Mixed algorithm	DDPG-PSO, TD3, AC3	handle the selection of discrete sleep modes and the continuous parameter adjustment meanwhile

4. Key Challenges

4.1 Cross-Domain Trade-off Optimization Between Energy Efficiency and Performance

The implementation of base station sleep strategies, while reducing energy consumption, may induce multi-dimensional performance degradation. When adopting sleep strategies, potential side effects such as increased system response latency or degraded quality of service (QoS) necessitate careful balancing between energy savings and performance metrics. Significant energy efficiency differences exist between deep sleep modes and light sleep schemes. For latency-sensitive services like short video streaming, base station wake-up delays may cause playback interruptions, requiring predictive wake-up mechanisms such as service forecast-based pre-activation. Furthermore, deep sleep modes may lead to handover failures for edge users, mandating consideration of dynamic user distribution patterns.

4.2 Scalability Challenges for Distributed DRL Cooperative Sleep in Ultra-Dense Heterogeneous Networks

Ultra-dense network scenarios present critical scalability challenges for DRL-based base station sleep strategies. As base station density increases, traditional centralized DRL architectures suffer exponential growth in state space dimensionality, resulting in expanded policy network parameters and increased decision latency that compromises real-time performance. Current distributed architecture research faces novel theoretical obstacles: in fully distributed decision-making, individual base station agents relying solely on local observations may converge to suboptimal solutions, while inter-agent policy coordination requiring frequent signaling exchange could trigger network congestion under constrained backhaul capacity.

4.3 Distributed Wake-up Control Plane Signaling Storm Mitigation for Sleeping Base Station Clusters

The concurrent wake-up of large-scale nodes triggers random access channel overload effects, where the spatiotemporally coupled preamble allocation mechanism limitations cause instantaneous wake-up procedures to exponentially increase preamble collision probabilities, driving contention-based channel access mechanisms into avalanche-like congestion states. Furthermore, the dynamic reconfiguration of system information generates broadcast signaling storms. State transitions between sleep modes require not only updates to the Master Information Block (MIB) but also cascade-triggered modifications of System Information Blocks (SIBs), presenting dual challenges of version control conflicts and consistency maintenance for these globally synchronized parameters.

4.4 Cross-Domain Coordinated Dynamic Optimization Dilemma

The optimization of base station sleep strategies fundamentally constitutes a cross-domain coupled dynamic system control problem, with its complexity primarily manifested through spatiotemporal coupling effects of multi-physical domain parameters. The multidimensional dynamics confronting sleep decision-making can be categorized into temporal and spatial dimensional variations. In temporal dynamics, fast-fading channel fluctuations induce undesired received signal strength deviations while service arrival processes exhibit non-stationary Poisson characteristics. Regarding spatial dynamics, the narrow-beam properties of MIMO systems render coverage patterns highly sensitive to action variations, and urban canyon effects produce nonlinear signal attenuation characteristics in the vertical dimension.

5. Future Research Directions

Although extensive literature and research exist in the field of deep reinforcement learning (DRL)-based base station sleep control, significant research opportunities remain regarding the integration of large-scale models and balancing the inherent conflict between energy savings and quality of service (QoS) in base station sleep operations.

5.1 Adaptive Lightweight DRL Models (Edge Deployment)

Conventional DRL models such as PPO demand substantial computational resources, while the high-dimensional state-action spaces in multi-cell networks significantly impede convergence speeds. Future research could explore the elimination of redundant neurons or connections within DRL networks to reduce computational overhead, alongside investigating hierarchical reinforcement learning approaches that decompose multidimensional tasks into subtasks to minimize per-decision computation. At the network architecture level, structured pruning techniques based on sensitivity analysis could identify and remove redundant neuronal connections, while incorporating the lottery ticket hypothesis to discover optimal subnetwork architectures. For system implementation, global policy updates could be performed in the cloud while edge nodes deploy compressed lightweight models, maintaining policy consistency through periodic model aggregation.

5.2 6G Terahertz Communication and DRL Cross-Technology Integration

The evolution toward 6G networks presents opportunities to integrate terahertz communications with DRL for advanced sleep control. Terahertz channel characterization faces high measurement costs, while the ultra-dense network deployment and dynamic channel conditions at these frequencies complicate sleep decision-making. Future solutions may employ dual-stream neural network architectures to separately process Sub-6GHz and terahertz channel characteristics, utilizing maximum mean discrepancy minimization for parameter sharing. The DRL design could incorporate multi-scale feature extraction networks, combining 3D sparse convolutions for millimeter-wave frequencies with WaveNet architectures for irregularly sampled terahertz data. For channel modeling, innovative hybrid approaches merging deterministic ray tracing with 3D stochastic geometry could be developed, calibrated through sparse real-world measurements to establish composite channel models.

5.3 Distributed Decision Optimization Under Multi-Dimensional Constraint Coupling

Future research should focus on constrained reinforcement learning frameworks for generating safe sleep policies that simultaneously satisfy coverage continuity, handover success rates, and wake-up latency requirements. This demands novel constraint-handling mechanisms based on dual theory-based constraint embedding methods. At the network coordination level, investigations should prioritize distributed optimization architectures for large-scale base station clusters, developing privacy-preserving training frameworks via federated learning. Hierarchical aggregation mechanisms could be designed within this framework: dynamic clustering at the base station layer, differentially

private aggregation at intermediate layers, and global model distillation at the top layer. Such approaches would address inter-operator data isolation while establishing hierarchical optimization models accounting for base station heterogeneity.

5.4 Intelligent Enhancement for 6G Integrated Sensing and Communication

The DRL co-optimization framework still faces multiple challenges. Theoretically, deeper mathematical analysis is required for joint communication-sensing optimization problems, including developing accurate Pareto frontier analysis models and guaranteeing global convergence for non-convex multi-objective optimization. Architecturally, breakthroughs are needed in distributed agent coordination mechanisms, potentially through game-theoretic Nash equilibrium solutions to achieve policy consistency under partial observability. Quantum machine learning could revolutionize this domain via quantum neural networks for high-dimensional channel state processing, quantum-accelerated policy search algorithms, and entanglement-enhanced sensing signal processing frameworks. For sensing applications, quantum compressed sensing frameworks may be developed, employing quantum Fourier transforms for sparse signal representation and quantum phase estimation for parameter extraction.

6. Conclusion

This paper first provides a systematic overview of base station sleep mechanisms and deep reinforcement learning (DRL) respectively, followed by the proposal of a tripartite classification framework to comprehensively categorize and review DRL-based base station sleep strategies. Graphical representations are employed to deliver more concise and intuitive descriptions and summaries. A critical analysis is presented regarding the challenges and limitations encountered in existing literature. The study concludes by identifying current challenges and open problems in the field of base station sleep control, while proposing concrete suggestions for future research directions. Compared to existing survey papers on DRL-based base station sleep strategies, this work distinguishes itself through the inclusion of more recent references and an innovative, critical synthesis of various DRL-based optimization algorithms, offering substantive evaluative commentary on their respective merits and shortcomings.

References

[1] Q. Chen, X. Bao, S. Chen and J. Zhao, Base station power control strategy in ultra-dense networks via deep reinforcement learning, Physical Communication, Volume 71, Page 10265, August 2025

[2] Zeng Dezhe, Li Yuopeng, Zhao Yuyang, et al. A High-Efficiency Base Station Dynamic Scheduling Method Based on Reinforcement Learning [J]. Computer Science,2021,48(11):363-371.

[3] Yang Fuyu, Zhao Dong. Base Station Sleep Control Algorithm Based on Deep Reinforcement Learning [J]. High-Quality Papers from China Science and Technology Paper Online,2023,16(02):170-178.

[4] J. Liu, B. Krishnamachari, S. Zhou and Z. Niu, DeepNap: Data-Driven Base Station Sleeping Operations Through Deep Reinforcement Learning, IEEE Internet Things J., Volume 5, Issue 6, Page 4273-4282, December 2018

[5] S. Sun, C. Huang, G. Chen, P. Xiaoand R. Tafazolli, Deep Learning-Based Traffic-Aware Base Station Sleep Mode and Cell Zooming Strategy in RIS-Aided Multi-Cell Networks, IEEE Trans. Cogn. Commun. Netw., page1-1, 2024

[6] J. Gan, H. Kou, G. Yang, H. Zhang, Z. Cao andW. Xu, Joint Sleep Control and Energy Sharing Strategy with Deep Reinforcement Learning in Green Ultra-Dense Networks, IEEE Trans. on Green Commun. Netw., page 1-1, 2025

[7] Y. M. Park, S. S. Hassan and C. S. Hong, "Maximizing Throughput of Aerial Base Stations via Resources-based Multi-Agent Proximal Policy Optimization: A Deep Reinforcement Learning Approach," 2022 23rd Asia-Pacific Network Operations and Management Symposium (APNOMS), Takamatsu, Japan, 2022, pp. 1-4

[8] Y. Bo, K. Kang, W. Li, G. Pan and M. Wang, "A Low-Altitude Network Base Station Planning Model Based on PPO Algorithm," 2024 IEEE 10th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE), Guangzhou, China, 2024, pp. 1-4

[9] Q. Wu, X. Chen, Z. Zhou, L. Chen and J. Zhang, Deep Reinforcement Learning With Spatio-Temporal Traffic Forecasting for Data-Driven Base Station Sleep Control, IEEE/ACM Trans. Networking, Volume29, Issue 2, Page 935-948, April 2021

[10] G. Saranya and E. Sasikala, "A Deep Reinforcement Learning based policy gradient for Energy Consumption in Edge Computing," 2022 International Conference on Computer, Power and Communications (ICCPC), Chennai, India, 2022, pp. 71-76

[11] J. Song, B. Zhang and J. Lia, "Deep Reinforcement Learning Empowered Particle Swarm Optimization for Aerial Base Station Deployment," 2023 IEEE Applied Sensing Conference (APSCON), Bengaluru, India, 2023, pp. 1-3

[12] L. Ma, X. Zhang, J. Sun, W. Zhang and C. -X. Wang, "Joint Optimization of Reconfigurable Intelligent Surfaces and Base Station Beamforming in MISO System Based on Deep Reinforcement Learning," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023, pp. 1-5

[13] Lee, H., Eom, C., & Lee, C. (2023). QoS-aware UAV-BS deployment optimization based on reinforcement learning. 2023 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1-4). IEEE.