A Review of Deep Reinforcement Learning-Based Energy Saving for 6G Network

Tao Weng*

Nanjing University Of Posts And Telecommunications, Nanjing, 210023, China

* Corresponding author email: p23000226@njupt.edu.cn

Abstract

With the exponential enhancement of 6G network performance, its energy consumption has become a core challenge constraining sustainable development. This paper systematically surveys the research progress of Deep Reinforcement Learning (DRL) in the field of 6G energy saving, focusing on three key technologies: resource scheduling, power control, and sleep strategies. It analyzes the application effects in typical scenarios such as ultra-dense urban networks, Space Air Ground Sea Integrated Networks (SAGIN), and Industrial Internet of Things (IIoT). Our analysis indicates that DRL, utilizing its autonomous decision-making and optimization capabilities, can effectively address the high dynamics and resource coupling inherent in 6G environments. However, bottlenecks still exist, including policy lag and long-term credit assignment. Future research directions need to make breakthroughs in dynamic coordination of renewable energy sources, lightweight engineering deployment, and cross-scenario generalization through meta-learning. This survey an attempt to construct a DRL driven 6G energy-saving technology system framework, providing a theoretical foundation for academia and industry to collaboratively advance high-performance, low-energy-consumption 6G networks.

Keywords

6G Network, energy saving, deep reinforcement learning.

1. Introduction

Facing exponentially growing mobile data traffic, communication networks confront a pivotal challenge in intelligent transformation. Although significant 5G progress in data rates and connection density, its automation and intelligence capabilities still not have the capability to meet up the traffic volume of 5016 EB/month by 2030 [1]. To address this challenge, the sixth-generation (6G) communication system has arisen, expected to achieve 1Tbps peak rates and sub millisecond latency through innovative technologies such as terahertz transmission and SAGIN [2]. Yet, the performance leap brought by 6G is accompanied by severe energy consumption challenges. It is predicted that by 2030, the energy consumption of the ICT industry will account for more than 20% of the global total energy supply. This situation stands in sharp contrast to the global pursuit of sustainable development and net-zero emissions, highlighting the imperative for 6G energy-saving research.

DRL combines the powerful perceptual ability of deep learning with the sequential decision optimization advantage of reinforcement learning. This intrinsic characteristic renders DRL inherently aligned with the energy-saving optimization imperatives of 6G networks. The 6G network environment exhibits high dynamism characterized by ultra-high-speed user mobility, sudden service demands, and rapidly time-varying channels. It also faces the complexity brought by ultra dense heterogeneous node deployment and tightly coupled multidimensional resources. The optimization objectives include multidimensional

performance indicators including energy efficiency, latency, throughput, reliability, and fairness. Traditional optimization methods based on analytical models or heuristic rules often encounter significant limitations in addressing such challenges, including modeling distortions, computational complexity explosions, and poor environmental adaptability [3].

While existing research partially addresses Artificial Intelligence (AI) in empowering communication networks, current surveys either broadly on AI communications or lack in-depth analysis of DRL's core challenges within specific 6G energy-saving application categories. Aiming to bridge this gap, this paper reviews DRL-based energy-saving advancements for 6G, analyzes its applications across resource scheduling, power control, and sleep strategies, discusses practical efficacy in typical scenarios, identifies prevailing challenges, and outlines future research directions. The overall goal is to establish a comprehensive cognitive framework and reference foundation for further development in this field. The article structure is shown in Figure1.

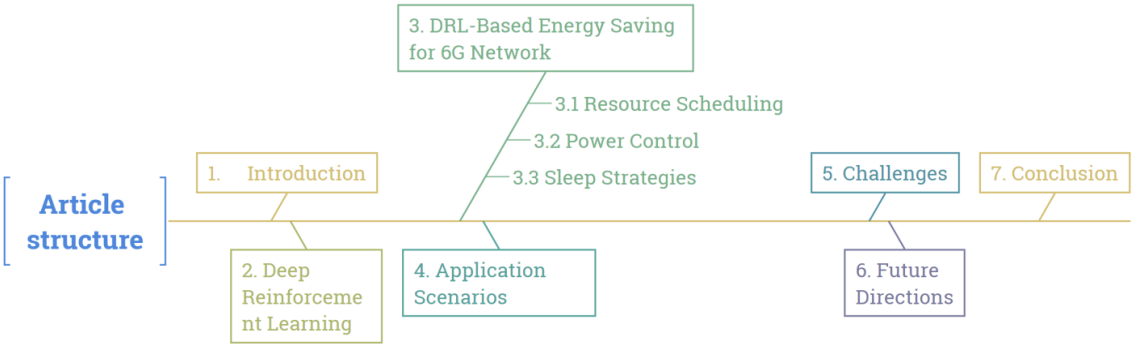

Figure 1. article structure

2. Deep Reinforcement Learning

DRL has become a key enabling technology for addressing the formidable energy-saving challenges of 6G networks, mainly due to its unique core strengths [3]. Its essence lies in the ability to autonomously learn and optimize sequential decision-making policies through interactive learning within complex, dynamic, and uncertain environments. The high dynamism and non-stationarity of 6G environments require optimization strategies with real-time adaptability. DRL, through continuous interaction with the environment, achieves autonomous learning and dynamic policy adjustments without relying on precise analytical environmental models, thereby facilitating robust decision-making under uncertainty. Consequently, DRL provides a powerful tool for addressing the complex energy-saving challenges in 6G networks, which involve high dimensionality, strong dynamics, multi-objective requirements, and distributed characteristics. In practical applications, key decisions such as resource allocation, task offloading, base station sleep scheduling, and parameter configuration are typically modeled as Markov Decision Processes (MDPs) or Partially Observable MDPs. The DRL agents then autonomously learn approximately optimal energy-saving strategies through environmental interactions, driving the network towards a more environmentally friendly, intelligent, and sustainable direction [4].

3. DRL-Based Energy Saving for 6G Network

This chapter reviews the energy-saving technology of 6G network based on DRL, mainly including three aspects: Resource Scheduling, Power Control, and Sleep Strategies. As shown in Table 1.

Table 1. DRL-based energy conservation research work.

Aspect	Approach	Method Proposed
Resource Scheduling	Saeed et al. [5]	Constructs an end-to-end DRL model that formulates energy consumption as a Markov process for dynamic computing offloading and resource allocation
	She et al. [6]	Develops a multi-agent DDPG framework enabling distributed task offloading through local observation sharing among base stations.
	Qian et al. [7]	Coordinates UAV trajectories and communication links via DRL, encoding channel-energy coupling into a multi-dimensional reward function.
Power Control	Bassoy et al. [8]	Integrates power amplifier nonlinearity into DRL state space and employs delayed TD3 learning for predistortion compensation.
	Diamanti et al. [9]	Applies PPO algorithm to jointly optimize massive MIMO beamforming and transmit power while suppressing multi-user interference.
Sleep Strategies	Ju et al. [10]	Models traffic tidal patterns using Double Q-learning, transforming micro-BS cluster activation/deactivation into spatiotemporal sequence prediction.
	Lee et al. [11]	Employs multi-agent DRL for collaborative LEO satellite state adjustment via inter-satellite coordination in non-stationary networks.

3.1. Resource Scheduling

In the field of spectrum and computational resource scheduling, the Deep Reinforcement Learning-based Computing Offloading and Resource Allocation algorithm proposed by Saeed et al. [5] constructs an end-to-end decision-making model. Its innovation lies in modeling energy consumption as a Markov process, enabling the agent to dynamically finish the computing offloading, energy consumption, and resource allocation based on real-time service load. A key advantage of this approach is its independence from historical traffic patterns required by traditional optimization methods, making it particularly suitable for edge node scenarios with limited computational resources. It reduces transmission energy consumption through joint optimization of communication and computational resource mapping. However, the method has inherent delays in responding to cross domain state changes. When high-speed movement of Low Earth Orbit (LEO) satellite causes drastic changes in topology, delayed resource reconfiguration can lead to notable Quality of Service (QoS) degradation. Additionally, its static action space design is difficult to adapt to sudden traffic surges triggered by large-scale events. Future research needs to integrate spatiotemporal graph convolutional networks to predict service distributions and build proactive resource reservation mechanisms. At the same time, integrating federated learning frameworks can enable privacy-preserving collaborative decision-making between base stations and reduce the risk of sensitive data leakage.

The Multi-Agent Deep Deterministic Policy Gradient framework developed by She et al. [6] establishes a distributed resource negotiation mechanism within the DRL architecture. Its core value lies in dynamically optimizing task offloading paths through the sharing of local observations between base station agents. This method significantly reduces the energy consumption of cloud edge transmission and demonstrates excellent energy efficiency improvement, especially in computation-intensive scenarios like intelligent manufacturing factories. Nevertheless, the framework faces two fundamental limitations: Firstly, the communication overhead between agents increases exponentially with the number of base stations, resulting in decision delays in ultra dense networks reaching millisecond level, which is inadequate for Ultra-Reliable Low-Latency Communication (URLLC) services. Secondly, it

fails to consider the energy efficiency differences of heterogeneous hardware, resulting in resource allocation strategies deviating from physical layer characteristics. Breakthroughs require the design of lightweight communication protocols to compress observation data and establishment of a hardware energy efficiency fingerprint libraries to guide differentiated scheduling, ensuring policies precisely match the energy efficiency curves of underlying devices.

In addition, regarding the optimization of offshore base stations, high backhaul energy consumption of offshore base stations can be innovatively solved by coordinating Unmanned Aerial Vehicle (UAV) trajectories and communication link selection [7]. Encode the coupling relationship between channel state and energy consumption into a multi-dimensional reward function based on DRL principles, allowing the agent to autonomously balance transmission quality and energy consumption. However, there are some drawbacks: Firstly, it ignores the non-stationary channel characteristics induced by vessel movement, resulting in a sharp decline in the generalization ability of the training strategy in harsh sea conditions. Secondly, it overlooks the impact of thermal control systems on total energy consumption. Future development requires the integration of meteorological prediction models to enhance environmental adaptability and incorporate physical parameters like base station surface temperature and cooling efficiency into the state space.

3.2. Power Control

For the power control, the SEEDRL framework [8] was the first to integrate the nonlinear characteristics of power amplifiers into the DRL state space. Its important breakthrough involves using delayed twin delayed deep deterministic policy gradient learning to derive predistortion compensation strategies. This solution directly addresses the critical issue of PA energy consumption dominating in base stations, reduces hardware-layer losses by suppressing harmonic distortion and achieves doubled energy efficiency in dense urban scenarios. However, the framework exhibits two major shortcomings. It relies on idealized PA models and fails to adapt to the electromagnetic dynamic characteristics of 6G's novel reconfigurable intelligent surfaces. policy stability is insufficient under high-power scenarios, and sudden load changes can easily lead to oscillation instability. There is an urgent need to build Hardware-in-the-Loop simulation platforms injecting real-time electromagnetic field monitoring data into the DRL training environment. Meanwhile, developing action space shearing mechanisms with safety constraints is essential to limit the damage to device lifespan caused by extreme power fluctuations.

The research on joint massive MIMO beamforming and power control by Diamanti et al. [9] employs the Proximal Policy Optimization (PPO) algorithm to coordinate array radiation patterns. This DRL-based scheme jointly optimizes beam patterns and transmit power, while suppressing multi-user interference and reducing energy consumption, making it particularly suitable for high-density access scenarios like stadiums. Nevertheless, the method has theoretical limitations. It assumes perfect Channel State Information, whereas actual measurement errors can significantly degrade the performance of the strategy. And it lacks compatibility with the heterogeneous propagation characteristics of millimeter-wave and Sub-6GHz frequency bands, which hinders its applicability to the trend of 6G multi band fusion. Future breakthroughs require integrating blind beamforming techniques to reduce dependency on channel estimation and designing dual-band collaborative reward functions to guide agents autonomous learning of inter band energy transfer strategies.

3.3. Sleep Strategies

The impact of sleep strategies on energy consumption in 6G networks is significant. The deep sleep mechanism [10], based on Double Q-learning to understand traffic tidal patterns, achieves its technical breakthrough by modeling the activation and deactivation decisions of

micro-base station clusters as spatiotemporal sequence prediction problem, significantly reducing energy consumption during low-load periods. This DRL solution has shown substantial energy savings in residential nighttime scenarios, effectively reducing the idle loss of base stations. But the strategy suffers from fundamental drawbacks. Excessive reliance on historical traffic statistics leads to delays of up to minutes in transitioning from sleep to active mode, while sudden disasters can trigger emergency communication needs. It ignores the energy overhead associated with state transitions, and frequent switching actually increases the total power consumption. Innovation needs to introduce causal reinforcement learning to distinguish features of normal operations and emergency events, and construct transitional energy penalty functions to enable agents to achieve Pareto optimality between energy savings and switching costs.

The emergency Random Access Channel protocol framework proposed by Lee et al. [11] employs multi-agent deep reinforcement learning interacting with non-stationary network environments to adjust the activation states of LEO satellites through inter-satellite collaboration. It solves the problem of resource idle caused by dynamic topology changes. This solution extends network lifespan and offers a novel approach for energy savings in space information networks. But it fails to quantify the energy consumption of inter-satellite signaling transmission, and excessive coordination may reduce net energy savings. Breakthrough pathways involve designing distributed decision mechanisms driven by local observations to reduce signaling overhead and developing onboard incremental learning chips for online policy evolution to meet rapid networking demands.

4. Application Scenarios

4.1. Ultra-Dense Urban Networks

As a typical 6G scenario, this leverages square-kilometer-scale micro-base station deployment and DRL-based multi-agent collaboration to solve dynamic energy savings problems in fluctuating tidal traffic patterns. The sharing of local observation data by base stations enables the MADDPG algorithm to coordinate sleep decisions and resource allocation [10], [12]. For example, in the O-RAN architecture, the Asynchronous Advantage Actor-Critic framework reduces peak energy consumption by disabling redundant processing units while maintaining QoS. However, communication overhead becomes a important performance bottleneck at the scale of thousands of nodes .

4.2. Space-Air-Ground-Sea Integrated Network (SAGIN)

As a core 6G architecture, SAGIN integrates satellite communications, High-Altitude Platforms (HAPs), terrestrial base stations, and maritime communication nodes to achieve seamless global coverage. Its main challenge lies in latency heterogeneity, requiring a hierarchical optimization structure. In this case, DRL adopts global and local control, the satellite layer utilizes PPO to plan macro level resource allocation strategies, while the UAV layer optimizes local trajectory and transmission power based on Deep Deterministic Policy Gradients (DDPG) [13], [14]. Validation proves that the architecture can coordinate energy consumption across LEO satellites, HAPs, and terrestrial base stations[7]. Nevertheless, asynchronous cross-layer state updates induce policy drift, which is an unresolved challenge.

4.3. Industrial Internet of Things (IIoT)

The strict reliability requirements and harsh electromagnetic interference environments in IIoT pose significant challenges. DRL addresses this by constructing joint device optimization models that deeply couples production rhythm with communication scheduling [6], [15]. A representative scheme involves real-time perception of the operating status of machine tools to dynamically adjust base station power levels and multiple access parameters, thereby

optimizing energy consumption while ensuring reliable transmission of control commands [9]. However, the rapid channel fading induced by metallic workshop structures necessitates the development of robust decision-making models incorporating adversarial training mechanisms.

5. Challenges

5.1. Policy Lag in Dynamic Environments

This directly limits the real-time applicability of DRL in 6G energy-saving. During inter-cell handovers triggered by high-speed user mobility, the strategy update period of traditional DRL is significantly lower than the channel coherence time, causing delayed power control commands [10] . Consequently, the performance of DRL based energy efficiency strategies significantly decreases at high speeds [12] , and there is an urgent need to develop lightweight online learning architectures.

5.2. Long-Term Credit Assignment Dilemma

This stems from the delayed nature of energy consumption feedback. Annual energy savings from macro-base station deep sleep require quarterly-level data for validation, making it difficult for DRL agents to establish causal links between sleep actions and energy efficiency gains [9]. While existing studies employ intermediate rewards, there lacks theoretical proof of their mathematical correlation with total energy consumption[16], [17]. The risks strategies improving short-term metrics while inadvertently increasing overall power usage.

6. Future Directions

6.1. Dynamic Synergy with Renewable Energy

This represents a breakthrough path for energy efficiency. By constructing coupled models, DRL can achieve precise control on multiple time scales. Adopting PPO for weekly energy storage scheduling plans to cope with weather fluctuations and utilizing DDPG for minute level load matching. A key innovation involves introducing carbon footprint driven reward mechanisms, guiding base stations to prioritize high load tasks during periods of abundant green energy [18], [19].

6.2. Engineering Deployment of Lightweight Models

To solve the limitations of edge computing, it is necessary to establish a cloud pre-training and edge fine-tuning paradigm. Knowledge distillation can compress complex policy networks into edge adapted architectures, while Neural Architecture Search (NAS) can automatically generate hardware optimized models. This approach significantly reduces computational overhead and achieves millisecond-level online decision-making[2], [13], [20].

6.3. Cross-Scenario Meta-Learning Generalization

This is central to handling 6G's heterogeneous environments. Pre-training foundational models are used to absorb prior knowledge from diverse scenarios, enabling them to rapidly adapt to new environments with minimal fine-tuning samples [14]. Innovatively applying adversarial domain adaptation techniques to bridge distribution gaps between satellite links and ground channels is expected to drastically reduce sample requirements in novel scenarios[5], [9].

7. Conclusion

This paper systematically establishes a technical classification for DRL applications in 6G energy conservation, and elucidates the intrinsic relationships and boundaries of the three

major technical branches: resource scheduling, power control, and sleep strategies. By deconstructing the DRL deployment paradigms in scenarios such as ultra dense urban networks, SAGIN, and IIoT, it reveals the causal mechanisms behind core challenges including environmental non-stationarity, reward sparsity, and sample inefficiency. Furthermore, it introduces dynamic coordination strategies for renewable energy sources and a meta learning framework for cross scenario generalization, offering a lightweight deployment path to address the computational constraints of 6G nerworks.

References

[1] J. Sanusi, O. Oshiga, S. Thomas, S. Idris, S. Adeshina, and A. M. Abba, 'A Review on 6G Wireless Communication Systems: Localization and Sensing', in 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), July 2021, pp. 1–5. doi: 10.1109/ICMEAS52683.2021.9692415.

[2] L. Jiao et al., 'Advanced Deep Learning Models for 6G: Overview, Opportunities, and Challenges', IEEE Access, vol. 12, pp. 133245–133314, 2024, doi: 10.1109/ACCESS.2024.3418900.

[3] K. Yu, K. Jin, and X. Deng, 'Review of Deep Reinforcement Learning', in 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Dec. 2022, pp. 41–48. doi: 10.1109/IMCEC55388.2022.10020015.

[4] D.-H. Tran, N. Van Huynh, S. Kaada, V. N. Vo, E. Lagunas, and S. Chatzinotas, 'Network Energy Saving for 6G and Beyond: A Deep Reinforcement Learning Approach', in 2025 IEEE Wireless Communications and Networking Conference (WCNC), Mar. 2025, pp. 1–6. doi: 10.1109/WCNC61545.2025.10978758.

[5] M. M. Saeed, R. A. Saeed, E. S. Ali, R. A. Mokhtar, and O. O. Khalifa, 'Algorithm for Resource Allocation and Computing Offloading in 6G Networks: Deep Reinforcement Learning-based', in 2024 9th International Conference on Mechatronics Engineering (ICOM), Aug. 2024, pp. 188–193. doi: 10.1109/ICOM61675.2024.10652281.

[6] H. She, L. Yan, and Y. Guo, 'Efficient End–Edge–Cloud Task Offloading in 6G Networks Based on Multiagent Deep Reinforcement Learning', IEEE Internet Things J., vol. 11, no. 11, pp. 20260–20270, June 2024, doi: 10.1109/JIOT.2024.3372614.

[7] L. P. Qian, H. Zhang, Q. Wang, Y. Wu, and B. Lin, 'Joint Multi-Domain Resource Allocation and Trajectory Optimization in UAV-Assisted Maritime IoT Networks', IEEE Internet Things J., vol. 10, no. 1, pp. 539–552, Jan. 2023, doi: 10.1109/JIOT.2022.3201017.

[8] S. Bassoy, R. Behravesh, and J. Pujol-Roig, 'SEEDRL: Smart Energy Efficiency Using Deep Reinforcement Learning for 6G Networks', in 2023 IEEE Globecom Workshops (GC Wkshps), Dec. 2023, pp. 732–737. doi: 10.1109/GCWkshps58843.2023.10464558.

[9] M. Diamanti, G. Kapsalis, E. E. Tsiropoulou, and S. Papavassiliou, 'Energy-Efficient Rate-Splitting Multiple Access: A Deep Reinforcement Learning-Based Framework', IEEE Open J. Commun. Soc., vol. 4, pp. 2397–2409, 2023, doi: 10.1109/OJCOMS.2023.3322047.

[10] Y. Ju et al., 'Deep Reinforcement Learning Based Joint Beam Allocation and Relay Selection in mmWave Vehicular Networks', IEEE Trans. Commun., vol. 71, no. 4, pp. 1997–2012, Apr. 2023, doi: 10.1109/TCOMM.2023.3240754.

[11] J.-H. Lee, H. Seo, J. Park, M. Bennis, and Y.-C. Ko, 'Learning Emergent Random-Access Protocol for LEO Satellite Networks', IEEE Trans. Wirel. Commun., vol. 22, no. 1, pp. 257–269, Jan. 2023, doi: 10.1109/TWC.2022.3192365.

[12] H. Ju, S. Kim, Y. Kim, and B. Shim, 'Energy-Efficient Ultra-Dense Network with Deep Reinforcement Learning', IEEE Trans. Wirel. Commun., vol. 21, no. 8, pp. 6539–6552, Aug. 2022, doi: 10.1109/TWC.2022.3150425.

[13] H. Zhang, M. Huang, H. Zhou, X. Wang, N. Wang, and K. Long, 'Capacity Maximization in RIS-UAV Networks: A DDQN-Based Trajectory and Phase Shift Optimization Approach', IEEE Trans. Wirel. Commun., vol. 22, no. 4, pp. 2583–2591, Apr. 2023, doi: 10.1109/TWC.2022.3212830.

[14] H. Peng and L.-C. Wang, 'Energy Harvesting Reconfigurable Intelligent Surface for UAV Based on Robust Deep Reinforcement Learning', IEEE Trans. Wirel. Commun., vol. 22, no. 10, pp. 6826–6838, Oct. 2023, doi: 10.1109/TWC.2023.3245820.

[15] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, 'AI Models for Green Communications Towards 6G', IEEE Commun. Surv. Tutor., vol. 24, no. 1, pp. 210–247, 2022, doi: 10.1109/COMST.2021.3130901.

[16] N. Capuano, G. Fenza, V. Loia, and C. Stanzione, 'Explainable Artificial Intelligence in CyberSecurity: A Survey', IEEE Access, vol. 10, pp. 93575–93600, 2022, doi: 10.1109/ACCESS.2022.3204171.

[17] J. J. Alcaraz, F. Losilla, A. Zanella, and M. Zorzi, 'Model-Based Reinforcement Learning with Kernels for Resource Allocation in RAN Slices', IEEE Trans. Wirel. Commun., vol. 22, no. 1, pp. 486–501, Jan. 2023, doi: 10.1109/TWC.2022.3195570.

[18] P. Zhang, Y. Xiao, Y. Li, X. Ge, G. Shi, and Y. Yang, 'Toward Net-Zero Carbon Emissions in Network AI for 6G and Beyond', IEEE Commun. Mag., vol. 62, no. 4, pp. 58–64, Apr. 2024, doi: 10.1109/MCOM.003.2300175.

[19] R. Kamran, S. Kiran, P. Jha, A. Karandikar'T, and P. Chaporkar, 'Green 6G: Energy Awareness in Design', in 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), Jan. 2024, pp. 1122–1125. doi: 10.1109/COMSNETS59351.2024.10427334.

[20] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, 'Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing', ACM Comput. Surv., vol. 55, no. 9, pp. 1–35, Sept. 2023, doi: 10.1145/3560815.