frontiers | Frontiers in Communications and Networks

TYPE Original Research
PUBLISHED 26 January 2026
DOI 10.3389/frcmn.2025.1764320

Check for updates

OPEN ACCESS

EDITED BY
Mehran Behjati,
Sunway University, Malaysia

REVIEWED BY
Mohammed Sani Adam,
National University of Malaysia, Malaysia
Javad Haghighat,
TED University, Türkiye

*CORRESPONDENCE
Mohsen Eslami,
meslami1@ualberta.ca

RECEIVED 09 December 2025
REVISED 20 December 2025
ACCEPTED 29 December 2025
PUBLISHED 26 January 2026

CITATION
Saleh V, Eslami M and Kazemi K (2026) DDPG-based energy efficiency optimization for ABS-assisted beyond-5G cellular networks with sleep mode management.
Front. Commun. Netw. 6:1764320.
doi: 10.3389/frcmn.2025.1764320

DDPG-based energy efficiency optimization for ABS-assisted beyond-5G cellular networks with sleep mode management

Vala Saleh[1], Mohsen Eslami[1,2]* and Kamran Kazemi[1]

[1]Department of Electrical Engineering, Shiraz University of Technology, Shiraz, Iran, [2]Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB, Canada

Introduction: Energy efficiency is a critical challenge in Beyond-5G (B5G) cellular networks, where ground base stations (GBSs) are responsible for a substantial portion of network energy consumption. Reducing this consumption while maintaining minimum user data rate requirements remains a key research problem.

Methods: This paper proposes an Aerial Base Station (ABS)-assisted energy optimization framework that integrates ABS deployment with low-power sleep states of GBSs. Traffic is selectively offloaded from lightly loaded GBSs to ABSs, enabling energy savings without violating user quality-of-service constraints. A Deep Deterministic Policy Gradient (DDPG) algorithm is employed to jointly optimize ABS positioning, GBS sleep mode scheduling, and resource allocation under dynamic traffic conditions.

Results: Simulation results demonstrate that the proposed DDPG-based framework significantly reduces network energy consumption while improving achievable user data rates compared to baseline schemes without ABS assistance or learning-based optimization.

Discussion: The results highlight the effectiveness of integrating ABSs with GBS low-power sleep states using reinforcement learning. By enforcing minimum data rate constraints and dynamically adapting to traffic variations, the proposed approach provides a scalable and energy-efficient solution for sustainable operation.

KEYWORDS

deep-double Q-learning (DDQL), deterministic-policy gradient (DDPG), energy efficiency, sleeping ground BS, ABS-assisted beyond-5G network

1 Introduction

The evolution of *Beyond 5G (B5G)* technology builds upon the worldwide deployment of *fifth-generation (5G)* networks, aiming to meet growing connectivity demands through more advanced standards and innovative communication solutions. B5G is expected to deliver enhanced mobility, superior reliability, ultra-high data rates, intelligent network management, and improved energy efficiency. With increasing dependency on high-speed and ubiquitous communication, B5G networks are anticipated to progress toward greater resilience and maturity. To enable this transition, key technologies such as *Artificial Intelligence (AI)*, *Edge Computing*, *Reconfigurable Intelligent Surfaces (RIS)*, *Terahertz (THz)* communication, *Quantum Computing*, and *Unmanned Aerial Vehicles (UAVs)* are being actively explored (Sufyan et al., 2023; Puspitasari et al., 2023; Dogra et al., 2020; Alsharif et al., 2018).

The potential of UAVs in cellular and wireless networks has gained substantial attention in recent literature. *UAV-assisted communication systems* have shown promise in extending network coverage beyond the limitations of terrestrial infrastructure, enhancing link reliability, and offering flexible, resilient, and sustainable connectivity in diverse deployment scenarios (Gryech et al., 2024). A prominent outcome of this integration is the development of *Aerial Access Networks (AANs)*, where UAVs are utilized to provide communication services from the air. Unlike *Terrestrial Access Networks (TANs)*, AANs overcome geographic and infrastructural constraints, offering wide-area coverage, improved communication quality, and high mobility support—particularly in remote or hard-to-reach environments where TANs may be infeasible (Behjati et al., 2025). In such systems, UAVs often operate as *Aerial Base Stations (ABSs)* or airborne relays, enabling direct access to users and thereby improving service availability and network adaptability.

Significant research efforts have been directed toward the integration of UAVs into B5G networks, addressing challenges such as trajectory optimization, user association mechanisms, transmission power management, and the cooperative deployment of UAVs with Intelligent Reflecting Surfaces (IRS) for improved signal propagation (Banafaa et al., 2024; Qazzaz et al., 2024; Shahzadi et al., 2021; Geraci et al., 2022; Gu and Zhang, 2023; Sarkar and Gul, 2023; Jangsher et al., 2022). These advancements improve performance, reliability, and energy efficiency in future wireless communication systems (Amponis et al., 2022).

A diverse set of techniques has been proposed to effectively reduce energy consumption in drone networks (Abubakar et al., 2023). These approaches encompass resource management (Masroor et al., 2021; Basharat et al., 2022), flight and transmission scheduling (Wu et al., 2022), path planning (Azadur et al., 2024), and optimal placement and trajectory design (Elnabty et al., 2022; Won et al., 2023; Azarhava et al., 2024; Tung et al., 2022).

Beyond UAV-assisted solutions, a key energy challenge in B5G networks is the high power consumption of *Ground Base Stations (GBSs)*. To mitigate this issue, researchers are actively developing optimization strategies to improve GBS energy efficiency. One widely adopted approach, applicable to both terrestrial networks and ABS-assisted frameworks, is the *GBS sleep strategy*, which dynamically deactivates underutilized GBSs with a small number of associated users while ensuring that user data rate requirements are satisfied.

GBS sleep strategies are designed to identify optimal opportunities for base stations to enter sleep mode without compromising network coverage or service quality (López-Pérez et al., 2022). These strategies can be broadly classified into *binary on/off schemes* and *multi-level sleep modes*, each offering distinct mechanisms for reducing energy consumption while maintaining network performance.

The binary scheme conserves energy by deactivating underutilized GBSs; however, this approach may negatively impact data transmission rates (Kim et al., 2015; Kooshki et al., 2023). To optimize energy efficiency in ultra-dense networks, researchers in (Amine et al., 2022) and (Ju et al., 2022) proposed *reinforcement learning (RL)*-based cell switching algorithms for managing small cells. Specifically, the work in (Ju et al., 2022) introduces a *Decision Selection Network (DSN)* to streamline the action space within a *Deep Reinforcement Learning (DRL)* framework, demonstrating effective management of active and sleep modes while maintaining essential data rate requirements.

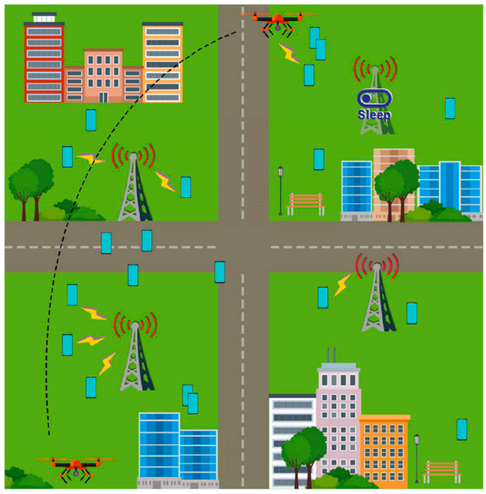

FIGURE 1
A wireless network consisting of multiple Ground Base Stations (GBSs) and Aerial Base Stations (ABSs), where the GBSs are equipped with sleep mode capability to optimize energy efficiency.

In contrast, multi-level sleep modes leverage mobile traffic prediction to dynamically transition idle small cells into different sleep states, further optimizing energy efficiency while ensuring network performance (Kim et al., 2023).

An often-overlooked application of drones is their use as *Aerial Base Stations (ABSs)* to facilitate GBS sleep strategies. In Chowdary et al. (2021), the authors propose a resource allocation algorithm that leverages active GBSs to serve users in areas where some GBSs have entered sleep mode. While this approach ensures user connectivity in sleeping areas, it introduces challenges such as service instability for users in non-sleeping areas and increased algorithmic complexity due to resource reallocation after GBS deactivation. Moreover, this work does not explore the potential of drones operating explicitly as ABSs to enable GBS sleep modes.

Motivated by these limitations, we propose an *ABS-assisted GBS sleep strategy* that selectively allows lightly loaded GBSs to enter sleep mode during periods of reduced traffic demand. This approach minimizes overall network energy consumption while ensuring that each ABS satisfies the minimum data rate required to maintain *Quality of Service (QoS)* during GBS downtime. Our analysis demonstrates that effective ABS deployment not only enhances overall network transmission rates but also significantly reduces GBS power consumption.

To further amplify the energy-saving benefits, we introduce a *joint optimization framework* that integrates GBS sleep scheduling, resource allocation, and ABS position optimization to minimize network-wide energy consumption. The resulting decision-making process forms a complex *binary integer programming* problem, motivating the need for efficient learning-based optimization techniques.

To address this challenge, we propose a *hybrid Deep Reinforcement Learning (DRL)* framework that combines *Deep Deterministic Policy Gradient (DDPG)* and *Deep Double Q-Learning (DDQL)* algorithms. The rationale for this integration is that the considered optimization includes both *continuous* variables (e.g., ABS horizontal positioning and transmit power allocation) and *discrete* decisions (e.g., GBS sleep mode control and association selection). DDPG is well-suited to continuous control, whereas DDQL is effective in discrete action spaces and reduces Q-value overestimation. By combining them, the proposed framework can efficiently handle the mixed discrete–continuous action space in a unified learning process.

Unless otherwise stated, users are assumed *quasi-static* during each optimization interval, i.e., user locations remain fixed while decisions on association, sleep scheduling, and resource allocation are optimized.

The remainder of this paper is organized as follows: Section 2 introduces the system model, including the sleep model, terrestrial and aerial channel models, and the power consumption model, followed by the optimization problem formulation. Section 3 presents the theoretical preliminaries of the DRL algorithms and details the proposed hybrid framework. Section 4 discusses the simulation results and provides an in-depth analysis. Finally, Section 5 concludes the paper with key findings.

2 System model

This section provides a structured overview of the key components considered in this study, including the network architecture, user/ABS location assumptions, channel models, energy consumption model, and problem formulation.

As illustrated in Figure 1, this study focuses on an ABS-assisted downlink wireless network. The analysis is conducted within a $L_x \times L_y$ km area, where K users are served by M strategically positioned Ground Base Stations (GBSs). User distribution follows a Poisson point process, modeling realistic spatial randomness. In line with many related works and to focus on network-level optimization, users are assumed static during each optimization interval (i.e., user locations do not change during a decision epoch).

Additionally, the network includes U Aerial Base Stations (ABSs), each hovering at an altitude h_u, where $1 \le u \le U$. The x-y coordinates of the ABSs will be optimized as part of the network design to enhance coverage, ensure seamless connectivity, and improve overall network performance.

The network employs Orthogonal Frequency Division Multiplexing (OFDM) to serve users across L subcarriers, where $L \gg K$. Each GBS m transmits a signal with power $p_{m,l}$ on the lth subcarrier, subject to the total power constraint (Equation 1):

$$\sum_{l=1}^{L} p_{m,l} \le P_m, \tag{1}$$

where P_m represents the maximum transmit power of GBS m. Similarly, the uth ABS transmits with power $p_{u,l}$ on the lth subcarrier, constrained by (Equation 2):

$$\sum_{l=1}^{L} p_{u,l} \le P_u, \tag{2}$$

where P_u is the maximum transmit power of the ABS.

To model user association, we define a binary variable $\beta_{m,k,l} \in \{0,1\}$ that indicates whether the kth user is associated with GBS m on subcarrier l. Specifically, $\beta_{m,k,l} = 1$ implies that $p_{m,l} > 0$, and the kth user is receiving a signal from GBS m. The set of users served by GBS m is denoted as (Equation 3):

$$\mathcal{D}_m = \{1, 2, \ldots, N_m\}, \tag{3}$$

where the total number of users in the network satisfies (Equation 4):

$$\sum_{m=1}^{M} N_m = K. \tag{4}$$

To represent the operational status of each GBS, we introduce a binary sleep indicator α_m, defined as (Equation 5):

$$\alpha_m = \begin{cases} 1, & \text{if GBS } m \text{ is active,} \\ 0, & \text{if GBS } m \text{ is in sleep mode.} \end{cases} \tag{5}$$

The uth UAV's location is denoted by (Equation 6):

$$\bar{d}_u = (\bar{x}_u, \bar{y}_u, \bar{z}_u), \tag{6}$$

while each GBS is positioned at the center of its respective cell. The distance between the uth UAV and the kth user at time slot n is given by (Equation 7)

$$d_{u,k}[n] = \sqrt{\bar{z}_u^2 + \|\mathbf{w}_u[n] - \mathbf{w}_k\|^2}, \tag{7}$$

where $\mathbf{w}_u[n] = [\bar{x}_u[n], \bar{y}_u[n]]$ and $\mathbf{w}_k = [x_k, y_k]$ denote the horizontal locations of the uth UAV and the kth user, respectively. Furthermore, $\bar{z}_u$ represents the altitude of the UAV, which, without loss of generality, is assumed to be constant in this work.

2.1 Air-to-ground (A2G) channel model

The Line-of-Sight (LoS) channel model is commonly employed in UAV-assisted networks to facilitate communication between Aerial Base Stations (ABSs) and Cellular Users (CUs) (Kim et al., 2015; Khawaja et al., 2019). The expected channel power gain from the kth user to the UAV on the lth subchannel can be expressed as

$$G_{u,k,l}[n] = \left[P_{\text{LoS}}^{u,k}[n] + \left(1 - P_{\text{LoS}}^{u,k}[n]\right)\hat{\kappa}\right] h_0 \left(d_{u,k}[n]\right)^{-\hat{\alpha}}, \tag{8}$$

where $h_0 = \left(\frac{\lambda}{4\pi}\right)^2$ represents the channel power gain under LoS conditions at a reference distance of 1 m, λ denotes the carrier wavelength, $\hat{\alpha}$ is the path-loss exponent constrained within $2 < \hat{\alpha} < 6$, and $\hat{\kappa} > 1$ is the additional attenuation factor due to Non-Line-of-Sight (NLoS) propagation.

Furthermore, the probability of establishing an LoS link, $P_{\text{LoS}}^{u,k}[n]$, is modeled as (Equation 9):

$$P_{\text{LoS}}^{u,k}[n] = \frac{1}{1 + \alpha' \exp\left(-\beta' \theta_{u,k}[n] - \alpha'\right)}, \tag{9}$$

where α' and β' are environment-dependent parameters, and (Equation 10)

$$\theta_{u,k}[n] = \sin^{-1}\left(\frac{\bar{z}_u}{d_{u,k}[n]}\right) \tag{10}$$

denotes the elevation angle between the uth UAV and the kth user at time slot n.

2.2 Ground-to-ground (G2G) channel model

For the terrestrial part of the network, which refers to the channel between Ground Base Stations (GBSs) and users, we adopt a fading channel model. The small-scale fading coefficient, $\hat{H}_{m,k,l} \sim \mathcal{CN}(0,1)$, follows a complex Gaussian distribution, while the large-scale fading coefficient, $g_{m,k}$, is modeled using the Hata-COST231 model (Singh, 2012). The frequency domain downlink channel gain $G_{m,k,l}$ between the mth GBS and the kth Cellular User (CU) on the lth subchannel consists of small-scale fading, path loss and shadowing components. Pathloss is given by (Equation 11):

$$PL_{m,k}^{\star} = \begin{cases} -L - 35\log_{10} d_{m,k}, & d_{m,k} > d_1, \\ -L - 15\log_{10} d_0 - 20\log_{10} d_{m,k}, & d_0 < d_{m,k} \le d_1, \\ -L - 15\log_{10} d_1 - 20\log_{10} d_0, & d_{m,k} \le d_0. \end{cases} \tag{11}$$

where $d_{m,k}$ is the distance between the kth user and the mth GBS. The parameter L is obtained as (Equation 12):

$$L = 46.3 - 33.9\log_{10} f - 13.82\log_{10} h_{\text{GBS}} \quad -(1.1\log_{10} f - 0.7)h_{\text{user}} - (1.56\log_{10} f - 0.8), \tag{12}$$

where h_{GBS} and h_{user} denote the heights of the GBSs and users, respectively, and f represents the carrier frequency.

Adding shadowing to the pasthloss, we have (Equation 13):

$$PL_{m,k} = PL_{m,k}^{\star} \times 10^{\frac{\chi_{m,k}\sigma_{\text{th}}}{10}}, \tag{13}$$

where $10^{\frac{\chi_{m,k}\sigma_{\text{th}}}{10}}$ represents the shadowing fading. Here, $\chi_{m,k} \sim \mathcal{N}(0,1)$ is a Gaussian random variable modeling the shadowing effect, and σ_{th} is the standard deviation that determines the scale of the shadowing effect.

Hence, the channel gain is expressed as (Equation 14):

$$G_{m,k,l} = \sqrt{PL_{m,k}} \times \hat{H}_{m,k,l}. \tag{14}$$

The Signal-to-Interference-plus-Noise Ratio (SINR) of the kth user served by the mth GBS is given by (Equation 15):

$$\gamma_{m,k,l} = \frac{p_{m,l}|G_{m,k,l}|^2}{\sum_{\tilde{m}=1:\,M,\,\tilde{m}\neq m} p_{\tilde{m},l}|G_{\tilde{m},k,l}|^2 + \sigma_{m,l}^2}. \tag{15}$$

Similarly, for the kth user served by the uth ABS, the SINR is (Equation 16):

$$\gamma_{u,k,l} = \frac{p_{u,l}|G_{u,k,l}|^2}{\sum_{\tilde{u}=1:\,U,\,\tilde{u}\neq u} p_{\tilde{u},l}|G_{\tilde{u},k,l}|^2 + \sigma_{u,l}^2}. \tag{16}$$

The achievable data rate for user k is then expressed as (Equation 17):

$$R_k = \log_2\left(1 + \gamma_{q,k,l}\right), \quad q \in \{m, u\}. \tag{17}$$

If no user is served on a given subcarrier $\hat{l}$, then the corresponding transmit power is set to zero, i.e., $p_{q,\hat{l}} = 0$. To represent the operational state of each GBS, we introduce a binary parameter $\alpha_m \in \{0,1\}$, which determines whether the mth GBS is active or in sleep mode. If a GBS is forced into sleep mode, all users associated with it will be disconnected and must connect to neighboring GBSs. Consequently, the transmit power of a sleeping GBS is forced to zero, i.e., $\alpha_m P_m = 0$.

2.3 Power consumption model

The total power consumption in UAV-assisted networks comprises two primary components: the power consumption of the ground network and the power consumption of UAVs.

2.3.1 Power consumption of the ground network

The power consumption of the ground network includes the following components:

1. Transmission Power: $P_{m,k,l}^{t}$, which represents the transmission power of the mth GBS.
2. Circuit Power: P_m^{cte}, the power consumed by the circuits and other electronic components of the mth GBS, which is considered constant.
3. Mode-dependent Power Consumption: The power consumed by a GBS due to its active or sleep mode operation, including power supply and air conditioning, given by (Equation 18):

$$P_m^{mod} = \alpha_m P_m^{on} + (1 - \alpha_m) P_m^{off}, \tag{18}$$

where P_m^{on} and P_m^{off} represent the power consumption of the GBS in active mode and sleep mode, respectively.

1. Mode Transition Power: The power associated with transitioning the base station m between operational modes is given by (Equation 19):

$$P_m^{tran} = |\alpha_m - \alpha_m^{prev}| P_m^{tran} \tag{19}$$

where α_m^{prev} denotes the operational state (active or sleep) of the mth GBS during the previous time slot, and P_m^{tran} represents the energy consumed for transitioning between modes.

2.3.2 Power consumption of UAVs

The power consumption of a UAV consists of three main components (Equation 20):

$$P_u = P_u^{com} + P_u^{hov} + P_u^{cte} \tag{20}$$

where P_u^{com} is the power consumed by the UAV's communication systems, including the radio transmitter, receiver, antennas, and other related components, expressed as (Equation 21):

$$P_u^{com} = \sum_{l=1}^{L} P_{u,l}. \tag{21}$$

P_u^{cte} is the power consumed by the UAV's internal circuits and other onboard electronic devices, and P_u^{hov} represents the power consumed when the UAV is hovering or stationary in the air. According to Ghorbel et al. (2019), the hovering power is given by (Equation 22):

$$P^{hov} = \sqrt{\frac{(m^{tot}\, g)^3}{2\pi r_p^2 n_p \rho}} \tag{22}$$

where m^{tot} is the total mass of the UAV, g is the gravitational acceleration, ρ is the air density, and r_p and n_p are the radius and number of propellers, respectively.

2.3.3 Total weighted power consumption

In UAV-based networks, the energy consumed by UAVs for hovering and flying is typically much higher than the power required for communication. To address this imbalance, a weighting factor w is introduced to balance the trade-off between hovering/flying power and communication-related power consumption (Zhu et al., 2021). This approach allows for an optimized allocation of energy resources across the network.

The total weighted power consumption of the UAV-assisted network is given by (Equation 23):

$$\begin{aligned} P^{tot} = w\Bigg(&\sum_{m=1}^{M}\sum_{l=1}^{L}\alpha_m P^{t}_{m,k,l} + \sum_{m=1}^{M}\left(P_m^{mod} + P_m^{tran} + P_m^{cte}\right) \\ &+ \sum_{u=1}^{U}\left(P_u^{com} + P_u^{cte}\right)\Bigg) + (1-w)\sum_{u=1}^{U} P_u^{hov}. \end{aligned} \tag{23}$$

In the simulations, the weighting factor is set to $w = 0.5$, providing a balanced trade-off between GBS-related power consumption and ABS hovering energy, consistent with Fährmann et al. (2022).

2.4 Energy efficiency

The Energy Efficiency (EE) criteria serve as a critical framework for evaluating the effectiveness of resource allocation within the network, particularly when GBSs are in sleep mode. By computing the Energy Efficiency metric, we ensure that the power consumption of GBSs does not compromise users' quality of service. The Energy Efficiency (EE) criterion is defined as (Equation 24):

$$\eta_{EE} = \frac{\tau_K}{P^{tot}}, \tag{24}$$

where $\tau_K = \sum_{k=1}^{K} R_k$ represents the total achievable data rate of all K users, and P^{tot} denotes the total power consumption of the network.

2.5 Optimization problem formulation

The optimization problem for maximizing the energy efficiency of the cellular network is formulated as (Equation 25):

$$\max_{\{\alpha_m\},\{\beta_{k,m,l}\},\{\beta_{k,u,l}\},\{p_{m,l}\},\{p_{u,l}\},\{(x_u,y_u)\}} \eta_{EE} \tag{25}$$

subject to the following constraints:

$$\textbf{(C1)}\quad 0 \le p_{m,l} \le P_{G,L}^{\max}, \quad 0 \le p_{u,l} \le P_{U,L}^{\max}, \quad \forall m,u,l \tag{26a}$$

$$\textbf{(C2)}\quad \alpha_m \in \{0,1\}, \quad \forall m \tag{26b}$$

$$\textbf{(C3)}\quad \sum_{m=1}^{M}\left(\alpha_m \beta_{k,m,l}\right) + \sum_{u=1}^{U}\beta_{k,u,l} = 1, \quad \forall k,l \tag{26c}$$

$$\textbf{(C4)}\quad \sum_{l=1}^{L} p_{m,l} \le \alpha_m P_m, \quad \forall m \tag{26d}$$

$$\textbf{(C5)}\quad \sum_{l=1}^{L} p_{u,l} \le P_u^{\max}, \quad \forall u \tag{26e}$$

$$\textbf{(C6)}\quad 0 \le x_u \le L_x, \quad \forall u \tag{26f}$$

$$\textbf{(C7)}\quad 0 \le y_u \le L_y, \quad \forall u \tag{26g}$$

$$\textbf{(C8)}\quad R_k \ge R_{\min}, \quad \forall k \tag{26h}$$

Constraint **(C1)** ensures that the transmission power per subcarrier remains within the maximum allowable thresholds for both GBSs and ABSs, as defined in (Equation 26a).

Constraint **(C2)** determines the operational status of each GBS, where $\alpha_m = 1$ indicates that the GBS is active and $\alpha_m = 0$ signifies that it is switched to sleep mode, as given in (Equation 26b).

Constraint **(C3)** guarantees exclusive user association by ensuring that each user is connected to exactly one serving node—either a single GBS or a single ABS—on each subcarrier. The inclusion of the activity indicator α_m prevents user association with sleeping GBSs, thereby avoiding multiple or invalid connections, as stated in (Equation 26c).

Constraint **(C4)** enforces that the total transmit power of each GBS does not exceed its maximum permissible value, accounting for both active and sleep states, as given in (Equation 26d). Constraint **(C5)** ensures that ABSs operate within their defined power limitations, as specified in (Equation 26e).

Constraints **(C6)** and **(C7)** confine ABSs within the designated region of interest, ensuring that they remain within operational limits, as enforced in (Equation 26f) and (Equation 26g), respectively. Finally, Constraint **(C8)** guarantees that each user achieves the minimum required data rate, thereby maintaining the network's quality of service (QoS), as defined in (Equation 26h).

The optimization problem formulated in (Equation 25) is a mixed-integer non-convex problem involving both discrete and continuous variables, rendering it intractable for conventional mathematical optimization techniques. To address this complexity, a learning-based approach is introduced in the following section.

3 DRL-based framework for optimizing complex problems

This section presents the hybrid DDPG–DDQL framework developed to address the energy efficiency optimization problem in ABS-assisted B5G networks incorporating a sleep strategy, as formulated in (Equation 25).

3.1 Basics of deep reinforcement learning (DRL)

Reinforcement Learning (RL) has significantly advanced Artificial Intelligence (AI) by enabling agents to make decisions, observe outcomes, and iteratively refine their strategies to determine an optimal policy (Morocho-Cayamcela et al., 2019; Huang et al., 2019). However, due to its reliance on extensive exploration, traditional RL can be slow and computationally expensive, limiting its applicability in large-scale networks.

Deep Reinforcement Learning (DRL) integrates Deep Neural Networks (DNNs) into RL, significantly enhancing learning speed and efficiency. In applications such as IoT and UAV-assisted networks, devices often need to make independent decisions to optimize network performance. These scenarios are frequently modeled as Markov Decision Processes (MDPs), which are formally defined as a quintuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \zeta)$:

- $\mathcal{S}$ represents a K-dimensional state space, with each state at time t denoted as s_t.
- $\mathcal{A}$ defines the finite action space available to the agent.
- $\mathcal{P}$: $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the transition probability function, specifying the likelihood of transitioning from state s to state s' after taking action a, expressed as $P(s, a, s')$.
- $\mathcal{R}$: $\mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the expected reward function, quantifying the anticipated reward upon executing action a in state s.
- $\zeta \in [0, 1)$ is the discount factor that determines the importance of future rewards.

Although traditional methods such as dynamic programming and value iteration can solve MDPs, they become computationally impractical for large-scale and complex networks. DRL techniques, particularly Deep Q-Learning (DQL), provide scalable solutions by approximating value functions using deep neural networks.

3.2 Deep Q-learning (DQL) and its limitations

Deep Q-Learning (DQL) is a fundamental DRL algorithm that estimates Q-values for state-action pairs using neural networks (Braga et al., 2020). For an agent parameterized by θ^Q at time t, the DQL update equation after taking action a_t in state s_t, receiving immediate reward r_{t+1}, and transitioning to the next state s_{t+1} is (Equation 27):

$$Q\left(s, a|\theta^Q_{t+1}\right) = Q\left(s, a|\theta^Q_t\right) + \nu\Big[r_{t+1} + \zeta \max_{a'} Q\left(s_{t+1}, a'|\theta^Q_t\right) - Q\left(s, a|\theta^Q_t\right)\Big], \tag{27}$$

Here, ν represents the learning rate. However, the Q-learning update consistently overestimates Q-values due to the use of bootstrapping, where estimates are derived from other estimates. This bias is exacerbated by using the same Q-network for both action selection and evaluation.

3.3 Double deep Q-learning (DDQL)

To mitigate the overestimation bias in DQL, the authors in (Fährmann et al., 2022; Shokrnezhad et al., 2024) introduced Double Deep Q-Learning (DDQL), which decouples action selection and evaluation employing two distinct Q networks:

- The primary Q-network: Q selects actions.
- The target network: Q' evaluates actions, using separate parameters θ^Q and $\theta^{Q'}$.

The DDQL update equation is (Equation 28):

$$Q\left(s, a|\theta^Q_{t+1}\right) = Q\left(s, a|\theta^Q_t\right) + \nu\Big[r_{t+1} + \zeta Q\left(s_{t+1}, \arg\max_{a'} Q'\left(s_{t+1}, a'|\theta^{Q'}_t\right)|\theta^Q_t\right) - Q\left(s_t, a_t|\theta^Q_t\right)\Big]. \tag{28}$$

The target network Q' is periodically updated using Polyak avenging (Equation 29):

$$\theta^{Q'}_{t+t_0} = (1 - \tau)\theta^{Q'}_t + \tau\theta^Q_t, \tag{29}$$

where $\tau \in [0, 1]$ controls the update rate.

Although DDQL effectively reduces overestimation and improves convergence in discrete action spaces (such as GBS sleep mode decisions), it struggles with continuous action spaces, such as power allocation and ABS positioning.

3.4 Deep deterministic policy gradient (DDPG)

For continuous action spaces, Deep Deterministic Policy Gradient (DDPG) algorithm (Yu et al., 2021) is a more suitable approach. DDPG is an actor-critic algorithm that efficiently handles sequential decision making. It optimizes a policy function π, mapping states to actions, by maximizing the objective function (Equation 30):

$$J(\theta) = \mathbb{E}[Q(s, a)|S = s_t, a = \pi(a|s_t)]. \tag{30}$$

Unlike DQL, where policies output a probability distribution over discrete actions, DDPG directly maps states to actions through a policy network $\pi(a|s_t)$. The policy network parameters θ^μ are updated using the gradient (Equation 31):

$$\nabla J_{\theta_\mu}(\theta) \approx \nabla_a Q(s, a) \nabla\mu(s|\theta^\mu), \tag{31}$$

where $\mu(s|\theta^\mu)$ is the policy network (the actor), and $\nabla_a Q(s, a)$ represents the gradient derived from a Q-network (the critic).

In large-scale environments with numerous actions, the actor-critic framework efficiently approximates Q-values using (Equation 32):

$$\max_a Q(s, a) \approx Q\big(s, a|\theta^Q\big)|_{a=\mu(s|\theta^\mu)}. \tag{32}$$

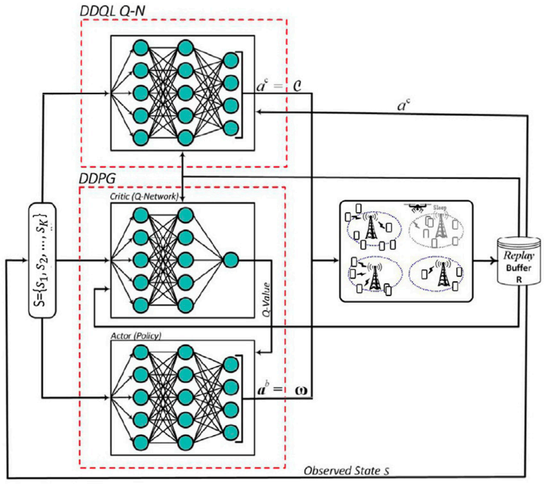

FIGURE 2
Hybrid DDPG–DDQL framework for optimizing ABS-assisted networks.

Similarly to DDQL, DDPG enhances stability by using:

- Experience replay to train the critic network.
- Target networks for both actors and critics, updated using Polyak averaging.

3.5 Hybrid DDPG-DDQL for UAV-assisted B5G networks

Given the nature of our optimization problem, which involves both discrete decisions (e.g., GBS sleep mode and discrete association choices) and continuous variables (e.g., power allocation and ABS positioning), we propose a Hybrid DDPG–DDQL framework. In this hybrid design, DDQL handles the discrete decision component and mitigates overestimation bias, while DDPG learns a deterministic continuous-control policy through actor–critic training. This explicit separation helps stabilize learning and reduces the overall search complexity compared to using a single algorithm to handle a mixed action space.

3.6 Proposed hybrid DDPG-DDQL framework with ABS-assisted sleep strategy

The objective of this research is to develop a DRL-based framework that optimizes the sleep scheduling of GBSs, the power allocation vector, and the ABS positioning vector, based on a given Channel State Information (CSI) matrix, defined as (Equation 33):

$$\mathbf{H} = [\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k], \tag{33}$$

where without loss of generality, a single UAV in the network has been assumed and hence, the UAV index of the channel matrix discarded. Furthermore, $\mathbf{G}_k = [G_{k,1} \ldots G_{k,L}]$. To achieve this, we propose a hybrid DRL system that integrates DDPG and DDQL to jointly learn optimal sleeping configurations and continuous control actions in relation to the CSI matrix $\mathbf{H}$ (or any related network metric). A schematic representation of the proposed hybrid DDPG–DDQL architecture is illustrated in Figure 2.

For a given sleep configuration, the actor–critic DDPG algorithm (Zhou et al., 2022) is used to optimize the power allocation and ABS horizontal positioning. DDPG continuously outputs the continuous control vector, denoted by $\mathbf{p}_{LK\times 1}$, and the ABS location vector $\mathbf{w} = (x, y)$.

To determine the optimal sleep configuration, the DDQL algorithm (Van Hasselt et al., 2016) is used, since the number of possible sleep configurations is finite and each configuration index is discrete.

3.7 Optimization problem formulation

As illustrated in Figure 2, both algorithms interact with a simulated ABS-assisted network environment to address the optimization problem formulated in Equation 24.

The network environment state is represented as (Equation 34):

$$\mathbf{S} = (s_1, s_2, \ldots, s_K), \tag{34}$$

where each state s_k is defined by the Signal-to-Interference-plus-Noise Ratio (SINR) values of users, represented as (Equation 35):

$$s_k = \left(\gamma_k, \alpha_m, \beta_{m,k,l}, \beta_{k,u,l}, p_{m,l}, p_{u,l}\right). \tag{35}$$

3.8 Reward function and action space

The immediate reward function r plays a crucial role in estimating optimal policies and Q-values. While energy efficiency is the primary objective, the reward must also discourage QoS violations. Therefore, we define the reward as the energy efficiency when all users satisfy the minimum rate requirement, and apply a penalty otherwise (Equation 36):

$$r_t = \begin{cases} \eta_{EE}(t), & \text{if } R_k(t) \geq R_{\min},\ \forall k, \\ \eta_{EE}(t) - \lambda \sum_{k=1}^{K} \max(0, R_{\min} - R_k(t)), & \text{otherwise}, \end{cases} \tag{36}$$

where $\lambda > 0$ controls the penalty severity. This reward structure encourages the agent to improve energy efficiency while maintaining feasibility with respect to QoS constraints. (If the original Equation 27 corresponds to the reward, Equation 36 replaces it and is referenced accordingly in the revised manuscript).

In this framework, as shown in Figure 2, the action space $\mathcal{A}$ consists of an action pair (Equation 37):

$$a = \left(a^b, a^c\right) = (\boldsymbol{\omega}, \mathbf{c}). \tag{37}$$

Here,

- The continuous action vector $\boldsymbol{\omega}$, which includes power allocation and ABS position, is generated using the DDPG algorithm.

TABLE 1 Simulation parameters.

Parameter	Value
f_c	2 GHz
BW	20 MHz
r_p	0.2
n_p	4
ρ	102 kg/s^3
m_{ToT}	750 gr
$\hat{\alpha}$	3
h_{GBS}, h_{user}	15 m, 1.65 m
d_o, d_1, D	10 m, 50 m
α'	9.53
β'	0.41
p^{max}	5 W
Number of GBSs	12
Number of users	20
ζ	0.99
ν	5×10^{-5}
Polyak avaraging parameter τ	10^{-3}
R	2000
w	0.5
β_1 & β_2	0.9 & 0.999

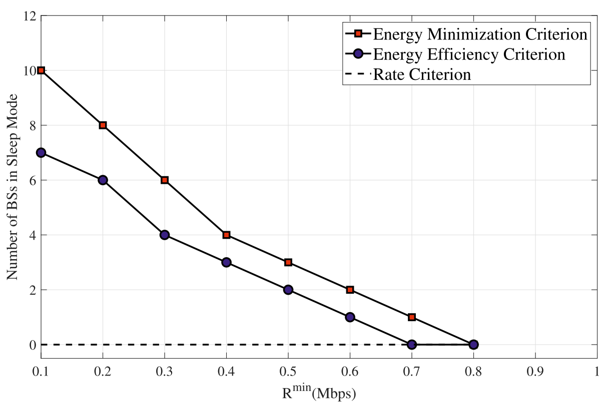

FIGURE 3
The number of sleeping GBSs as a function of rate requirement $R_{\min}$.

- tThe discrete action vector **c**, which includes the GBS sleep configuration (and any discrete association selection if applicable), is produced by the DDQL algorithm.

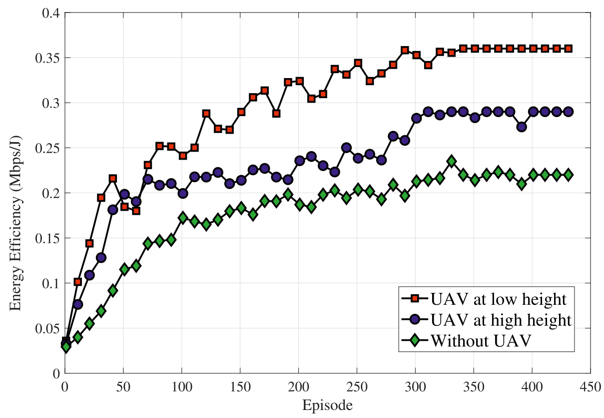

FIGURE 4
Energy efficiency criterion as a function of the number of training episodes for $R_{\min} = 0.2$.

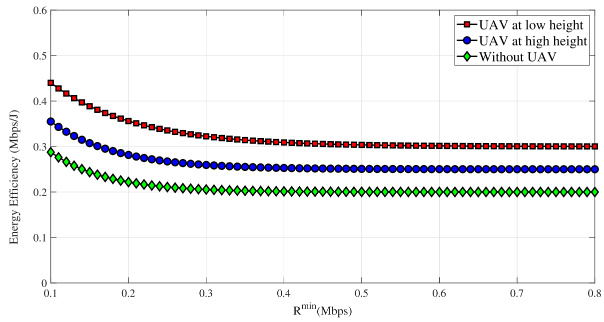

FIGURE 5
Energy efficiency criterion as a function of rate requirement $R_{\min}$.

3.9 Computational complexity analysis

The computational complexity of the proposed method mainly arises during the offline training stage due to iterative neural network updates in both DDPG and DDQL. For a fully connected network layer i, let m_i and n_i denote the input and output dimensions, and let b be the batch size. The dominant cost per training update is due to matrix multiplications, leading to (Equation 38):

$$\sum_i \mathcal{O}(m_i n_i b). \tag{38}$$

After training, the online inference stage requires only forward passes through the trained networks. With $b = 1$ (as in online RL decision-making), the complexity becomes (Equation 39):

$$\sum_i \mathcal{O}(m_i n_i). \tag{39}$$

In addition, the DDQL component evaluates a discrete action among a finite set of sleep configurations; thus, the per-step selection overhead scales with the number of discrete actions (sleep configurations) considered by the DDQL output layer. Overall, the proposed approach is computationally intensive during offline training but has low online computational overhead, making it suitable for real-time operation once trained.

4 Results and discussions

In this section, we present the results of the proposed hybrid-DRL algorithm to optimize energy efficiency in ABS-assisted cellular networks. The simulation parameters are summarized in Table 1.

The actor and critic networks, along with their respective target networks, are designed with two hidden layers comprising 256 and 128 neurons, respectively. In contrast, the DDQN architecture includes two fully connected layers with 64 neurons each, followed by ReLU activation functions, and terminates with a linear output layer. To enhance convergence and ensure training stability, the Adam optimizer is employed for the critic network, using its default hyperparameters β_1 and β_2, which control the exponential moving averages of the gradients and their squared values, respectively (Van Hasselt et al., 2016).

Figure 3 illustrates the relationship between the number of sleeping GBSs and the rate requirement across three optimization scenarios: (i) optimizing achievable rate without considering EE, (ii) prioritizing EE while ignoring rate constraints, and (iii) jointly optimizing EE under the minimum-rate constraint. The results show that $R_{\min}$ strongly limits the number of GBSs that can enter sleep mode. In the first scenario, no GBS enters sleep mode due to strict rate-driven operation. In the second scenario, more GBSs can be deactivated, improving EE but violating user rate requirements. The third scenario balances both objectives and is effective in the range $R_{\min} \in [0.1, 0.6]$ bps/Hz. When $R_{\min} = 0.6$ bps/Hz, up to seven GBSs can be switched off, while at $R_{\min} = 0.1$ bps/Hz, only one GBS is switched off. For $R_{\min} > 0.6$ bps/Hz, the method cannot place GBSs into sleep mode without violating the QoS constraint.

It can be observed from Figure 3 that when the minimum rate requirement exceeds $R_{\min} = 0.6$ bps/Hz, the number of GBSs entering sleep mode drops to zero. This outcome highlights the trade-off between stringent QoS constraints and achievable energy savings. Under high QoS requirements, maintaining user data rates necessitates keeping all GBSs active, which limits the applicability of sleep strategies. Consequently, the proposed ABS-assisted framework is most effective in low-to-medium QoS regimes, such as off-peak or lightly loaded traffic conditions, where energy efficiency gains can be achieved without compromising user performance.

As illustrated in Figure 4, the energy efficiency criterion is expressed in terms of training episodes. To evaluate the role of the ABS in the proposed system, three scenarios were considered: one with the ABS at a higher altitude, another at a lower altitude, and a scenario without an ABS. In this context, $R_{\min}$ was set to 0.2 bps/Hz based on the results in Figure 3. The results indicate that the learning process improves energy efficiency over training as the agent refines sleep scheduling, power allocation, and ABS positioning decisions. Figure 5 presents the energy efficiency criterion as a function of the minimum rate requirement. The observations indicate that, with an average $R_{\min}$ of 0.4, the proposed methods employing low-altitude and high-altitude ABSs can improve the energy efficiency criterion by 30% and 25%, respectively, compared to the scenario without an ABS.

5 Conclusion

This paper presents an ABS-assisted energy optimization framework for beyond-5G (B5G) cellular networks, utilizing selective ground base station (GBS) sleep modes and traffic offloading through aerial base stations (ABSs). To address the dynamic and non-convex nature of the problem, a hybrid reinforcement learning algorithm combining Deep Deterministic Policy Gradient (DDPG) and Double Deep Q-Learning (DDQL) is developed. This algorithm jointly optimizes ABS positioning, GBS sleep scheduling, and resource allocation. Simulation results demonstrate that the proposed framework significantly reduces overall network energy consumption while maintaining service quality.

These findings underscore the potential of hybrid deep reinforcement learning techniques in enabling intelligent and energy-efficient wireless communication systems. Future research directions include incorporating renewable-powered ABSs, implementing cooperative multi-ABS coordination, and developing real-time adaptive mechanisms to further enhance system scalability and performance.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

VS: Writing – review and editing, Writing – original draft. ME: Writing – original draft, Writing – review and editing. KK: Writing – review and editing, Writing – original draft.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI was used exclusively to edit and improve the clarity of the manuscript's text.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Abubakar, A. I., Ahmad, I., Omeke, K., Ozturk, M., Ozturk, C., Abdel-Salam, A., et al. (2023). A survey on energy optimization techniques in UAV-based cellular networks: from conventional to machine learning approaches. *Drones* 7 (3), 214. doi:10.3390/drones7030214

Alsharif, M., Nordin, R., Abdullah, N. F., and Kelechi, A. H. (2018). How to make key 5G wireless technologies environmental friendly: a review. *Trans. Emerg. Telecommun. Technol.* 29 (1), e3254. doi:10.1002/ett.3254

Amine, A. E., Chaiban, J. P., Hassan, H. A. H., Dini, P., Nuaymi, L., and Achkar, R. (2022). Energy optimization with multi-sleeping control in 5G heterogeneous networks using reinforcement learning. *IEEE Trans. Netw. Serv. Manag.* 19 (4), 4310–4322. doi:10.1109/tnsm.2022.3157650

Amponis, G., Lagkas, T., Zevgara, M., Katsikas, G., Xirofotos, T., Moscholios, I., et al. (2022). Drones in B5G/6G networks as flying base stations. *Drones* 6 (2), 39. doi:10.3390/drones6020039

Azadur, R.Md., Pawase, C. J., and Chang, K. (2024). Multi-UAV path planning utilizing the PGA algorithm for terrestrial IoT sensor network under ISAC framework. *Trans. Emerg. Telecommun. Technol.* 35 (1), e4916. doi:10.1002/ett.4916

Azarhava, H., Abdollahi, M. P., and Musevi Niya, J. (2024). Placement and power assignment for hierarchical UAV networks under hovering fluctuations in mmWave communications. *Trans. Emerg. Telecommun. Technol.* 35 (11), e70002. doi:10.1002/ett.70002

Banafaa, M. K., Pepeoğlu, Ö., Shayea, I., Alhammadi, A., Shamsan, Z. A., Razaz, M. A., et al. (2024). A comprehensive survey on 5G-and-Beyond networks with UAVs: applications, emerging technologies, regulatory aspects, research trends and challenges. *IEEE Access* 12, 7786–7826. doi:10.1109/access.2023.3349208

Basharat, M., Naeem, M., Qadir, Z., and Anpalagan, A. (2022). Resource optimization in UAV-Assisted wireless networks—A comprehensive survey. *Trans. Emerg. Telecommun. Technol.* 33 (7), e4464. doi:10.1002/ett.4464

Behjati, M., Alobaidy, H. A. H., Nordin, R., and Abdullah, N. F. (2025). UAV-assisted federated learning with hybrid LoRa P2P/LoRaWAN for sustainable biosphere. *Front. Commun. Netw.* 6, 1529453. doi:10.3389/frcmn.2025.1529453

Braga, I. M., Cavalcante, E. d. O., Fodor, G., Silva, Y. C. B., e Silva, C. F. M., and Freitas, W. C. (2020). User scheduling based on multi-agent deep Q-learning for robust beamforming in multicell MISO systems. *IEEE Commun. Lett.* 24 (12), 2809–2813. doi:10.1109/lcomm.2020.3015462

Chowdary, A., Ramamoorthi, Y., Kumar, A., and Cenkeramaddi, L. R. (2021). Joint resource allocation and UAV scheduling with ground radio station sleeping. *IEEE Access* 9, 124505–124518. doi:10.1109/access.2021.3111087

Dogra, A., Jha, R. K., and Jain, S. (2020). A survey on beyond 5G network with the advent of 6G: architecture and emerging technologies. *IEEE Access* 9, 67512–67547. doi:10.1109/access.2020.3031234

Elnabty, I. A., Fahmy, Y., and Kafafy, M. (2022). A survey on UAV placement optimization for UAV-Assisted communication in 5G and beyond networks. *Phys. Commun.* 51, 101564. doi:10.1016/j.phycom.2021.101564

Fährmann, D., Jorek, N., Damer, N., Kirchbuchner, F., and Kuijper, A. (2022). Double deep Q-Learning with prioritized experience replay for anomaly detection in smart environments. *IEEE Access* 10, 60836–60848. doi:10.1109/access.2022.3179720

Geraci, G., Garcia-Rodriguez, A., Azari, M. M., Lozano, A., Mezzavilla, M., Chatzinotas, S., et al. (2022). What will the future of UAV cellular communications be? A flight from 5G to 6G. *IEEE Commun. Surv. Tutor.* 24 (3), 1304–1335. doi:10.1109/comst.2022.3171135

Ghorbel, M. B., Rodriguez-Duarte, D., Ghazzai, H., Hossain, M. J., and Menouar, H. (2019). Joint position and travel path optimization for energy efficient wireless data gathering using unmanned aerial vehicles. *IEEE Trans. Veh. Technol.* 68 (3), 2165–2175. doi:10.1109/tvt.2019.2893374

Gryech, I., Vinogradov, E., Saboor, A., Bithas, P. S., Mathiopoulos, P. T., and Pollin, S. (2024). A systematic literature review on the role of UAV-enabled communications in advancing the UN's sustainable development goals. *Front. Commun. Netw.* 5, 1286073. doi:10.3389/frcmn.2024.1286073

Gu, X., and Zhang, G. (2023). A survey on UAV-assisted wireless communications: recent advances and future trends. *Comput. Commun.* 208, 44–78. doi:10.1016/j.comcom.2023.05.013

Huang, Y., Xu, C., Zhang, C., Hua, M., and Zhang, Z. (2019). An overview of intelligent wireless communications using deep reinforcement learning. *J. Commun. Inf. Netw.* 4 (2), 15–29. doi:10.23919/jcin.2019.8917869

Jangsher, S., Al-Jarrah, M., Al-Dweik, A., Alsusa, E., and Kong, P. Y. (2022). Energy constrained sum-rate maximization in IRS-assisted UAV networks with imperfect channel information. *IEEE Trans. Aerosp. Electron. Syst.* 59 (3), 2898–2908. doi:10.1109/taes.2022.3220493

Ju, H., Kim, S., Kim, Y., and Shim, B. (2022). Energy-efficient ultra-dense network with deep reinforcement learning. *IEEE Trans. Wirel. Commun.* 21 (8), 6539–6552. doi:10.1109/twc.2022.3150425

Khawaja, W., Guvenc, I., Matolak, D. W., Fiebig, U. C., and Schneckenburger, N. (2019). A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles. *IEEE Commun. Surv. Tutor.* 21 (3), 2361–2391. doi:10.1109/comst.2019.2915069

Kim, J., Jeon, W. S., and Jeong, D. G. (2015). Base-station sleep management in open-access femtocell networks. *IEEE Trans. Veh. Technol.* 65 (5), 3786–3791. doi:10.1109/tvt.2015.2445922

Kim, T., Lee, S., Choi, H., Park, H. S., and Choi, J. (2023). An energy-efficient multi-level sleep strategy for periodic uplink transmission in industrial private 5G networks. *Sensors* 23, 9070. doi:10.3390/s23229070

Kooshki, F., Armada, A. G., Mowla, M. M., Flizikowski, A., and Pietrzyk, S. (2023). Energy-efficient sleep mode schemes for cell-less RAN in 5G and beyond 5G networks. *IEEE Access* 11, 1432–1444. doi:10.1109/access.2022.3233430

López-Pérez, D., De Domenico, A., Piovesan, N., Xinli, G., Bao, H., Qitao, S., et al. (2022). A survey on 5G radio access network energy efficiency: massive MIMO, lean carrier design, sleep modes, and machine learning. *IEEE Commun. Surv. Tutor.* 24 (1), 653–697. doi:10.1109/comst.2022.3142532

Masroor, R., Naeem, M., and Ejaz, W. (2021). Resource management in UAV-assisted wireless networks: an optimization perspective. *Ad Hoc Netw.* 121, 102596. doi:10.1016/j.adhoc.2021.102596

Morocho-Cayamcela, M. E., Lee, H., and Lim, W. (2019). Machine learning for 5G/B5G Mobile and wireless communications: potential, limitations, and future directions. *IEEE Access* 7, 137184–137206. doi:10.1109/access.2019.2942390

Puspitasari, A., An, T. T., Alsharif, M. H., and Lee, B. M. (2023). Emerging technologies for 6G communication networks: machine learning approaches. *Sensors* 23, 7709. doi:10.3390/s23187709

Qazzaz, M. M. H., Zaidi, S. A., McLernon, D. C., Hayajneh, A. M., Salama, A., and Aldalahmeh, S. A. (2024). Non-terrestrial UAV clients for beyond 5G networks: a comprehensive survey. *Ad Hoc Netw.* 157, 103440. doi:10.1016/j.adhoc.2024.103440

Sarkar, N. I., and Gul, S. (2023). Artificial intelligence-based autonomous UAV networks: a survey. *Drones* 7 (5), 322. doi:10.3390/drones7050322

Shahzadi, R., Ali, M., Khan, H. Z., and Naeem, M. (2021). UAV assisted 5G and beyond wireless networks: a survey. *J. Netw. Comput. Appl.* 189, 103114. doi:10.1016/j.jnca.2021.103114

Shokrnezhad, M., Taleb, T., and Dazzi, P. (2024). Double deep Q-Learning-Based path selection and service placement for latency-sensitive beyond 5G applications. *IEEE Trans. Mob. Comput.* 23 (5), 5097–5110. doi:10.1109/tmc.2023.3301506

Singh, Y. (2012). Comparison of okumura, hata and COST-231 models on the basis of path loss and signal strength. *Int. J. Comput. Appl.* 59 (11), 37–41. doi:10.5120/9594-4216

Sufyan, A., Khan, K. B., Khashan, O. A., Mir, T., and Mir, U. (2023). From 5G to beyond 5G: a comprehensive survey of wireless network evolution, challenges, and promising technologies. *Electronics* 12, 2200. doi:10.3390/electronics12102200

Tung, T. V., An, T. T., and Lee, B. M. (2022). Joint resource and trajectory optimization for energy efficiency maximization in UAV-based networks. *Mathematics* 10 (20), 3840. doi:10.3390/math10203840

Van Hasselt, H., Guez, A., and Silver, D. (2016). "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*

Won, J., Kim, D. Y., Park, Y. I., and Lee, J. W. (2023). A survey on UAV placement and trajectory optimization in communication networks: from the perspective of air-to-ground channel models. *ICT Express* 9 (3), 385–397. doi:10.1016/j.icte.2022.01.015

Wu, W., Sun, S., Shan, F., Yang, M., and Luo, J. (2022). Energy-constrained UAV flight scheduling for IoT data collection with 60 GHz communication. *IEEE Trans. Veh. Technol.* 71 (10), 10991–11005. doi:10.1109/tvt.2022.3184869

Yu, Y., Tang, J., Huang, J., Zhang, X., So, D. K. C., and Wong, K. K. (2021). Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm. *IEEE Trans. Commun.* 69 (9), 6361–6374. doi:10.1109/tcomm.2021.3089476

Zhou, Q., Guo, C., Wang, C., and Cui, L. (2022). Radio resource management for C-V2X using graph matching and actor–critic learning. *IEEE Wirel. Commun. Lett.* 11 (12), 2645–2649. doi:10.1109/lwc.2022.3213176

Zhu, B., Bedeer, E., Nguyen, H. H., Barton, R., and Henry, J. (2021). UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning. *IEEE Trans. Veh. Technol.* 70 (9), 9540–9554. doi:10.1109/tvt.2021.3102161