UAV-Assisted Resilience in 6G and Beyond Network Energy Saving: A Multi-Agent DRL Approach

Dao Lan Vy Dinh[1], Anh Nguyen Thi Mai[1], Hung Tran[1], Giang Quynh Le Vu[1,6], Tu Dac Ho[2], Zhenni Pan[3], Vo Nhan Van[4], Symeon Chatzinotas[5], Dinh-Hieu Tran[5]

[1]DATCOM Lab, Faculty of DS&AI, College of Technology, National Economics University, Vietnam.
[2]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway.
[3]Waseda University, Japan.
[4]Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam.
[5]Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg.
[1,6]Faculty of Information Technology, National Academy of Education Management, Hanoi, Vietnam.
Emails: hung.tran@neu.edu.vn, tu.d.ho@ntnu.no, zhenni.pan@aoni.waseda.jp
vonhanvan@dtu.edu.vn, symeon.chatzinotas@uni.lu, hieu.tran-dinh@uni.lu

***Abstract*—This paper investigates the unmanned aerial vehicle (UAV)-assisted resilience perspective in the 6G network energy saving (NES) scenario. More specifically, we consider multiple ground base stations (GBSs) and each GBS has three different sectors/cells in the terrestrial networks, and multiple cells are turned off due to NES or incidents, e.g., disasters, hardware failures, or outages. To address this, we propose a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework to enable UAV-assisted communication by jointly optimizing UAV trajectories, transmission power, and user-UAV association under a sleeping ground base station (GBS) strategy. This framework aims to ensure the resilience of active users in the network and the long-term operability of UAVs. Specifically, it maximizes service coverage for users during power outages or NES zones, while minimizing the energy consumption of UAVs. Simulation results demonstrate that the proposed MADDPG policy consistently achieves high coverage ratio across different testing episodes, outperforming other baselines. Moreover, the MADDPG framework attains the lowest total energy consumption, with a reduction of approximately 24% compared to the conventional all GBS ON configuration, while maintaining a comparable user service rate. These results confirm the effectiveness of the proposed approach in achieving a superior trade-off between energy efficiency and service performance, supporting the development of sustainable and resilient UAV-assisted cellular networks.**

***Index Terms*—6G, centralized training and decentralized execution (CTDE), Multi-Agent Deep Reinforcement Learning, Network Energy Saving, Self-Organizing Networks (SONs), Unmanned Aerial Vehicle (UAV).**

I. Introduction

The rapid proliferation of mobile devices and and the shift toward 6G and beyond networks have driven an unprecedented surge in energy consumption. This escalating energy demand not only increases operational expenditure for mobile network operators but also amplifies the environmental impact through higher carbon emissions. Consequently, improving the energy efficiency (EE) and sustainability of wireless infrastructures has become a central design goal for next-generation networks. Beyond energy concerns, modern communication systems must remain resilient to failures caused by natural disasters, backhaul congestion, or equipment malfunctions, necessitating rapid and reliable service recovery under dynamic conditions [2].

Unmanned Aerial Vehicles (UAVs) have recently emerged as key enablers for green and resilient communications due to their high mobility, adaptive deployment, and strong line-of-sight (LoS) connectivity. When functioning as Aerial Base Stations (ABSs), UAVs can promptly compensate for service disruptions in Cell Outage Compensation (COC) scenarios or temporarily replace low-traffic Ground Base Stations (GBSs) placed in sleep mode to reduce network energy consumption [3]. However, jointly determining UAV trajectories, transmit power levels, and the on/off states of GBSs constitutes a complex, non-convex, and temporally coupled optimization problem that challenges traditional model-based methods.

Several works have studied UAV-assisted energy-efficient (EE) architectures via classical optimization. Chang et al. [12] propose a UAV-GBS scheme for small cells, using a heuristic integer linear program (ILP) to jointly schedule GBS sleep and place UAVs, improving EE while preserving coverage. Alsharoa et al. [13] optimize a three-tier heterogeneous network (HetNet) with solar-powered drone cells via a per-block binary program, reducing total energy through joint drone deployment and dynamic GBS activation under minimum received-power (QoS) constraints. Li et al. [3] formulate a unified problem over GBS sleep states, UAV trajectories, and transmit powers; their block-coordinate descent with successive convex approximation (SCA) and branch-and-bound balances flight, transmit, and GBS operational energy. Gaddam et al. [14] analyze ground radio station infrastructures for UAV links, showing throughput–coverage–EE trade-offs under coordinated scheduling and sleep control.

Collectively, these studies have demonstrated the potential of integrating UAV deployment and GBS sleep mechanisms to achieve greener communication. However, their reliance on deterministic or heuristic optimization techniques, such as ILP, SCA, or block-coordinate methods, limits their scalability and adaptability to highly dynamic network environments. Moreover, none of the above frameworks incorporate learning-based mechanisms that allow agents to autonomously adapt to temporal traffic variations, user mobility, and stochastic wireless conditions.

In contrast, deep reinforcement learning (DRL) has emerged as a powerful approach for complex, non-convex control in wireless networks. Liu et al. [4] use DRL for energy-efficient UAV coverage control, and Wang et al. [5] apply Deep Deterministic Policy Gradient (DDPG) to UAV-assisted computation offloading, illustrating DRL's ability to learn adaptive policies. Likewise, [1] shows that Deep Q-Learning (DQN) can tune GBS configurations for energy savings while preserving QoS. However, few works jointly leverage UAV-assisted architectures and DRL to co-optimize UAV mobility, transmit power, and GBS sleep scheduling within a unified green-communication framework.

To address these limitations, this paper makes four contributions in a single unified framework. First, to the best of our knowledge, it is the first to study UAV-assisted resilience for 6G-and-beyond NES using multi-agent DRL. Second, we model a multi-cell radio access network with scheduled inactive cells, distributed users, and user rate demands with occasional sudden surges. To ensure these surges reflect realistic usage, the traffic patterns are synthesized based on the characteristics of the open 5G dataset from Choi et al. [15], which contains real-world traffic from diverse applications such as video conferencing (e.g., Google Meet, MS Teams) and video streaming. Under this model, we aim to maximize user coverage while minimizing UAVs' energy consumption to ensure both network resilience and UAV sustainability. Third, we employ a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) approach with centralized training and decentralized execution to jointly optimize UAV trajectories, transmit power, and user–UAV association. Finally, extensive simulations show that, compared with benchmark schemes, our learning-driven method offers superior adaptation, scalability across heterogeneous environments, and robustness to unpredictable network dynamics.

II. System Model

We considers a downlink communication network in which UAVs provide service to ground users when nearby cells are temporarily turned off to save energy or due to power outages or failures. The goal is to simultaneously optimize the UAV trajectory and transmit power to maximize user coverage while minimizing the UAV's energy consumption.

As shown in Fig. 1, each GBS consists of three sectors or cells. Each cell serves the associated mobile users within its coverage area. When the cells encounter problems such as hardware failures, natural disasters, or energy-saving problems, but the traffic demand from mobile users in cells suddenly increases unexpectedly, turning on the cells when they are in turn-off/deep sleep modes will take many minutes. Even when turned on, these cells are still overloaded and cannot serve all users in time. Therefore, we use UAVs acting as ABSs to serve users in a timely manner. We design the scheduling so that users are allocated to the UAV with the best SINR. This ensures that each user is connected to the most favorable serving UAV based on the instantaneous channel quality and the UAV's location.

1) Cells: There are K fixed cells, each is a hexagonal cell. The cell operational state is given by $s_k(t) \in \{0, 1\}$, where 1 denotes ON and 0 denotes OFF. The cell ON/OFF schedule is predetermined during data synthesizing process and provided as input. The set of inactive cells requiring UAV assistance at time t is: $A(t) = \{k | s_k(t) = 0\}$.

2) UAVs: There are N UAVs in the network with each UAV i flies at a fixed altitude H with position $q_i(t) = [x_i(t), y_i(t)]$. UAVs can move within a bounded area, with movement constrained by:

$$||\Delta q_i(t)|| \leq v_{max}. \quad (1)$$

Each UAV has limited transmit power:

$$0 \leq P_i(t) \leq P_{max}. \quad (2)$$

The total system transmit power throughout flying time must satisfy:

$$\sum_i P_i(t) \leq P_{max}. \quad (3)$$

3) Users: M users are distributed randomly within the coverage area. Each user j has a fixed location (x_j, y_j) and a heterogeneous, dynamic rate requirement $R_{req,j}$.

A. Communication Model

The formulation for the communication model, including the distance, Rician fading channel, SINR, and achievable data rate (Eqs. (4) – (9)), is adopted from the model presented in [10], [11].

Given the location $(x_j,\ y_j,\ 0)$ of each user $j \in M$ at ground level and the current UAV location $(x_i,\ y_i,\ H)$ in time slot n, the distance d_{ij} between the UAV and the user is:

$$d_{ij}(t) = \sqrt{H^2 + (x_i - x_j)^2 + (y_i - y_j)^2}. \quad (4)$$

The channel between the user j and the UAV i follows a Rician fading channel model with a factor K, where the channel coefficient h_{ij} can be written as:

$$h_{ij}(t) = \sqrt{w_{ij}(t)}.\tilde{h}_{ij}(t), \quad (5)$$

where $w_{ij}(t)$ represents for the large-scale fading effects and $h_{ij}(t)$ accounts for Rician small-scale fading coefficient. Specifically, $w_{ij}(t)$ can be modeled as:

$$w_{ij}(t) = w_0.d_{ij}^{-\alpha}(t), \quad (6)$$

where w_0 is the average channel power gain at the reference distance $d = 1m$, and $\alpha \geq 2$ is the path loss exponent for the Rician fading channel. The small scale fading $\tilde{h}_{ij}$ with an expected value $\mathbb{E}[|\tilde{h}_{ij}^2|] = 1$, is given by:

$$\tilde{h}_{ij}(t) = \sqrt{\frac{G}{1+G}}\bar{h}_{ij}(t) + \sqrt{\frac{1}{1+G}}\hat{h}_{ij}(t), \tag{7}$$

where $\bar{h}_{ij}(t)$ and $\hat{h}_{ij}(t) \sim \mathcal{CN}(0, 1)$ denote the deterministic LoS and non-line-of-sight (NLoS) component (Rayleigh fading) during time slot t, respectively; G is the Rician factor. The SINR of each user j is:

$$\gamma_{ij}(t) = \frac{P_i(t).|h_{ij}(t)|^2}{\sigma^2 + \sum_{k\neq i} P_k(t).|h_{kj}(t)|^2}, \tag{8}$$

where $P_i(t)$ is the power of the UAV i at the timeslot t, σ^2 is the thermal noise power which is linearly proportional to the allocated bandwidth, interference includes other UAVs and active GBSs. Thus, the data rate achievable for user j when served by UAV i at time t is:

$$R_{ij}(t) = B\log_2(1+\gamma_{ij}(t)), \tag{9}$$

where B is the frequency range allocated to this communication link.

Each user is associated with the UAV that provides the highest SINR:

$$a_{ij}(t) = \begin{cases} 1, & \text{if } i = \arg\max_k \gamma_{kj}(t), \\ 0, & \text{otherwise.} \end{cases}$$

Each user can be served by at most one UAV:

$$\sum_i a_{ij}(t) \leq 1, \quad \forall j. \tag{10}$$

To ensure satisfactory communication quality, each user must achieve a minimum data rate requirement

$$R_{ij}(t) \geq R_{\text{req},j}, \quad \forall j \tag{11}$$

where $R_{\text{req},j}$ denotes the required data rate threshold of user j.

The total network throughput at time t is expressed as

$$R_{\text{total}}(t) = \sum_i \sum_j a_{ij}(t)R_{ij}(t). \tag{12}$$

The objective of the UAV communication system is typically to maximize $R_{\text{total}}(t)$ under the above user assignment and QoS constraints.

B. UAV Energy Model

The energy consumption of UAV i at time t consists of both propulsion and communication components:

$$E_i(t) = E_{\text{prop},i}(t) + E_{\text{comm},i}(t), \tag{13}$$

where $E_{\text{prop},i}(t)$ and $E_{\text{comm},i}(t)$ are formulated, respectively, as follows:

$$E_{\text{prop},i}(t) = \alpha_1||\Delta q_i(t)||^2 + \alpha_2||\Delta q_i(t)||, \tag{14}$$
$$E_{\text{comm},i}(t) = P_i(t)\Delta t, \tag{15}$$

with Δt is a timestep duration. Accordingly, the total energy of all UAVs is defined as:

$$E_{\text{UAV}}(t) = \sum_{i=1}^{n} E_i(t). \tag{16}$$

C. Cell ON/OFF Energy Model

For each cell c, instantaneous cell power is modeled as: $P_{c,t} = z_{c,t}(P_{\text{static}} + \Delta_p^{\text{cell}}\phi_{c,t})$ with cell ON or OFF at timestep t: $z_{c,t} \in \{0, 1\}$, load fraction: $\phi_{c,t} \in [0, 1]$ and load dependent slope: Δ_p^{cell} as referred to [8] [9]. Episode energy per cell is calculated as: $E_c = \sum_{t=1}^{T} P_{c,t}\Delta t$. Network energy will include the static power drawn when any cell linked to the GBS is ON: $E_{\text{total}} = \sum E_c + \sum E_s$ with $E_s = \sum_t P_{s,t}\Delta t$.

D. Problem Formulation

The optimization objective of maximizing the users coverage while minimizing the total UAVs' energy consumption by optimizing the UAVs' trajectory, UAVs' transmit power, and user-UAV association. Our problem is formulated as:

$$\mathcal{P}_1: \tag{17}$$
$$\max_{\{q_i(t),P_i(t),a_{ij}(t)\}} \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{M} a_{ij}(t)R_{ij}(t) - \lambda \sum_{t=1}^{T}\sum_{i=1}^{N} E_i(t) \tag{18}$$

$$\textbf{s.t. } R_{ij}(t) \geq R_{\text{req},j}, \quad \forall i,j,t, \tag{19}$$
$$\sum_i a_{ij}(t) \leq 1, \quad \forall j,t, \tag{20}$$
$$a_{ij}(t) \in \{0,1\}, \quad \forall i,j,t, \tag{21}$$
$$||\Delta q_i(t)|| \leq v_{max}, \quad \forall i,t, \tag{22}$$
$$0 \leq P_i(t) \leq P_{max}, \quad \forall i,t, \tag{23}$$
$$\sum_i P_i(t) \leq P_{max}, \quad \forall t, \tag{24}$$

where λ is a trade-off coefficient balancing coverage maximization and energy minimization. Constraint (19) ensures that each served user meets its minimum QoS requirement, while (20) guarantees that each user is associated with at most one UAV. Constraints (22)-(24) enforce UAV mobility and power limitations.

III. POMDP Formulation for UAV-Assisted Resilience in 6G NES

A. POMDP Formulation

We model the control problem as a partially observable Markov decision process (POMDP) [6]. Each UAV i observes a local state $s_i(t)$ while a global state $s_{global}(t)$ is available for centralized training.

1) Local State of UAV i:

$$s_i(t) = [q_i(t), P_i(t), q_{\text{other}_i}(t), P_{\text{other}_i}(t), (x_j, y_j, R_{req,j})_{j\in U_i(t)}, s_k(t)_{k=1}^{K}], \tag{25}$$

where $q_i(t)$ is position of UAV i at time t; $P_i(t)$ is transmit power of UAV i at timestep t; $q_{\text{other}_i}(t), P_{\text{other}_i}(t)$ are positions

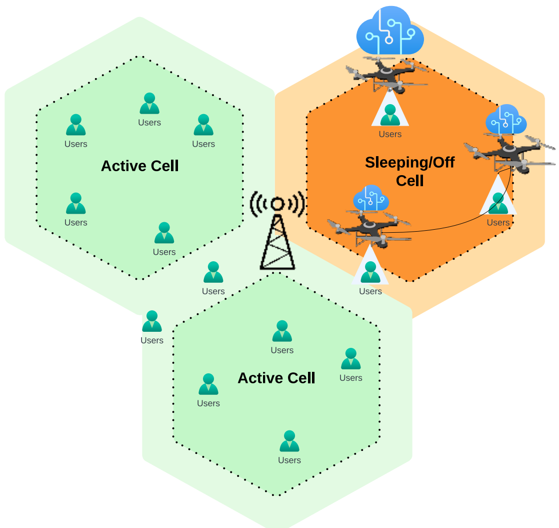

Fig. 1. System Model Illustration

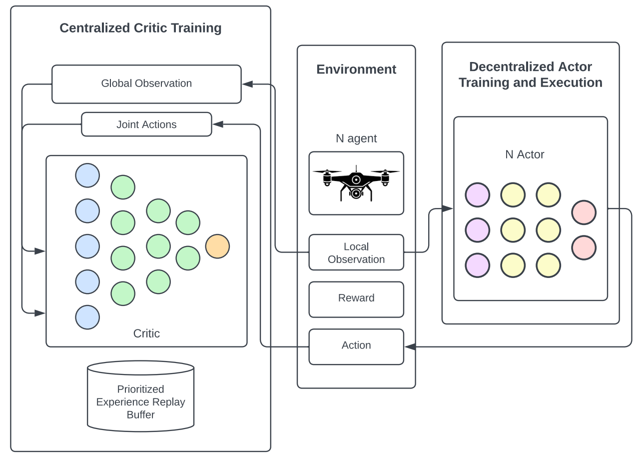

Fig. 2. CTDE-MADDPG Framework

and powers of other UAVs; $(x_j, y_j, R_{req,j})_{j \in U_i(t)}$ is information of 6 nearest users within UAV i coverage diameter; $s_k(t)_{k=1}^K$ is states of all cells.

2) Global State:

$$s_{\text{global}}(t) = [\{q_i(t), P_i(t)\}_{i=1}^N, \{(x_j, y_j, R_{req,j})\}_{j=1}^M, s_k(t)_{k=1}^K]. \tag{26}$$

3) Action of UAV i:

$$a_i(t) = [\Delta x_i(t), \Delta y_i(t), P_i(t)]. \tag{27}$$

B. Reward Function

The reward emphasizes coverage maximization and energy minimization, with coverage rate is measured based on successfully served UAV-needed users:

$$R_i(t) = \omega_1 \cdot \mathcal{R}_i^C(t) + \omega_2 \cdot (1 - \mathcal{P}_i^E(t)), \tag{28}$$

where $\mathcal{R}_i^C(t)$ and $\mathcal{P}_i^E(t)$ are calculated, respectively, as

$$\mathcal{R}_i^C(t) = \frac{|\{j : R_{ij}(t) \geq R_{\text{req},j},\ s_{\text{cell},j}(t) = 0\}|}{|\{j : s_{\text{cell},j}(t) = 0\}|}, \tag{29}$$

$$\mathcal{P}_i^E(t) = \frac{E_i(t)}{E_{max}}, \tag{30}$$

with $E_{max} = P_{max}\Delta t + \alpha_1 v_{max}^2 + \alpha_2 v_{max}$.

C. Objective

Each UAV learns a policy $\pi_i(a_i|s_i)$ that maximizes its expected cumulative discounted reward:

$$\max_{\pi_1, \pi_2} E\left[\sum_{t=0}^{T} \gamma^t R_i(t)\right]. \tag{31}$$

IV. The Proposed CTDE-MADDPG Framework

A. CTDE-MADDPG Framework

We adopt a Centralized Training with Decentralized Execution (CTDE) variant of MADDPG for cooperative UAVs control with better memory and training efficiency as shown 2. Each UAV is modeled as an agent with its own actor–critic pair. During training, each agent's critic has access to the global state $s_{\text{global}}(t)$ and actions of all agents, enabling coordinated learning. However, during execution, each UAV only uses its local observations $s_i(t)$ through the actor network, making the system scalable and robust to partial observability.

B. Network Architecture

Both actor and critic networks are implemented as fully connected multilayer perceptrons (MLPs) with ReLU activations. Actor outputs continuous actions $[v_x, v_y, P]$ normalized to valid ranges via tanh activation, with $v_x, v_y \in [-v_{max}, v_{max}]$ and $P \in [0, P_{max}]$. Critic receives concatenated joint observations and actions. Optimization is performed with the Adam optimizer and gradient norms are clipped to prevent instability. Target networks are softly updated using Polyak averaging.

C. Training Mechanism

1) Prioritized Experience Replay (PER): Standard experience replay samples transitions uniformly, which can be inefficient when most experiences are not informative. To improve sample efficiency and stability, we integrate a prioritized replay buffer based on the proportional prioritization scheme. Each transition's sampling probability is determined by its temporal-difference (TD) error magnitude, stored in a sum-tree structure for efficient $O(logN)$ sampling and priority updates.

2) Exploration Strategy: Besides dynamically adjusting exploration through PER, we also add temporally correlated Ornstein-Uhlenbeck (OU) Noise to actions during training to encourage stable exploration: $N_t = N_{t-1} + \theta(\mu - Nt - 1)\Delta t + \sigma W_t$ with noise parameter σ is decayed by 0.9999 per step with a minium of 0.01 [7]. After substantial number of

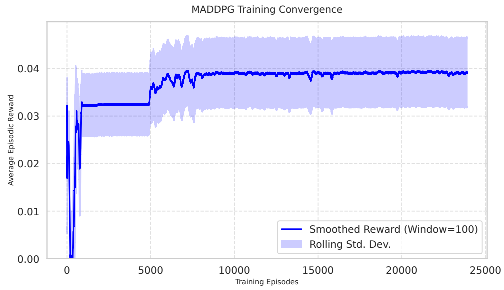

Fig. 3. Training convergence curve, showing the average reward per step over the training process.

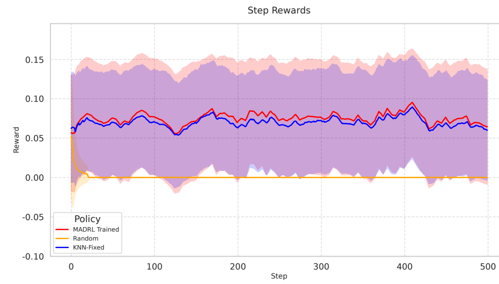

Fig. 4. omparison of instantaneous reward per step, averaged over 15 testing episodes.

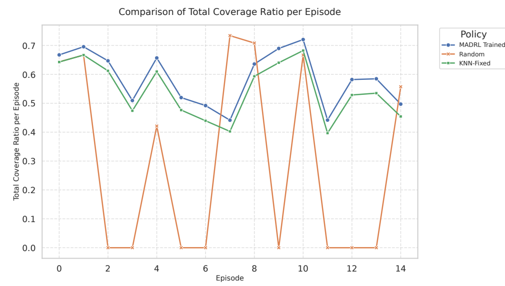

Fig. 5. Comparison of Total Coverage Ratio per Episode

episodes, exploration noise is disabled to exploit the learned policy. This could prevents erratic movements that could violate physical constraints, and produces smooth, correlated actions suitable for UAV dynamics.

3) Interference Management: The system accounts for inter-UAVs interference when computing SINR values. Each UAV's transmission contributes to the interference perceived by users connected to other UAVs. This coupling is reflected in the environment dynamics and achievable rate calculation. The mechanism encourages UAV agents to learn cooperative spatial and power control strategies that balance coverage maximization and energy minimmization.

V. Results and Discussions

A. Reward-based Analysis

Fig. 3 illustrates the convergence of the proposed MADDPG framework. The x-axis represents the Training Episodes (up to 25,000), and the y-axis is the normalized average episodic reward (per step). The average reward rises sharply during the initial exploration phase and stabilizes around 0.032. It exhibits a clear stepped improvement around the 5,000-episode mark, which coincides with the learning-rate decay. This jump is followed by a brief period of high-variance fine-tuning. The policy then settles into its final converged state of approximately 0.039, and the stable plateau from roughly 8,000 episodes onward confirms robust convergence.

As shown in Fig. 4, the learned MADDPG policy maintains the highest average reward across all timesteps, outperforming both the k-nearest neighbors (KNN)-Fixed and Random baselines. The Random policy quickly collapses toward zero reward, while the KNN-Fixed method remains myopic and less adaptive. In contrast, the MADDPG agents, through cooperative learning, appears to find a more refined balance in spatial positioning and power control. This ability to make better sequential decisions yields a consistent, albeit small, performance gain throughout the episode.

B. Coverage Performance Evaluation

Fig. 5 illustrates the total coverage ratio per episode for the proposed MADDPG policy compared against two baseline methods. It can be seen from Fig. 5 that the MADDPG policy (blue line) achieves the highest total coverage ratio in all 15 testing episodes. This superior performance suggests that the agents have successfully learned a complex, cooperative strategy, optimizing their trajectories and power control to serve the maximum number of users in the sleeping cells.

The KNN-Fixed policy (green line), which serves as a stable heuristic, is consistently outperformed by MADDPG. This performance gap can be explained by the policy's myopic and non-cooperative nature. Lacking coordination, multiple agents can be independently drawn to the same dense user cluster. This fact results in redundant coverage in one area while neglecting users in other areas. The MADDPG agents, in contrast, appear to learn to divide this workload efficiently via the centralized training mechanism.

Conversely, the Random policy (orange line) exhibits extremely poor and highly volatile performance, with its coverage ratio frequently dropping to zero. This result is expected, as its actions have no correlation with the system's objective. The UAVs are just as likely to fly away from users as they are to fly towards them.

C. Energy Performance Evaluation

Table I compares the overall energy consumption and user service performance of the proposed MADDPG framework against 3 baseline methods. The total energy consumption is decomposed into the UAV operational energy E_{UAV} and the energy consumed by active cell E_{cell}.

The All Cell ON baseline represents the scenario in which all GBSs remain fully active at fixed transmission power, consuming the highest amount of energy (121.06 Wh). In contrast, other methods that involve UAV assistance or selective GBS activation demonstrate energy savings by dynamically adjusting coverage and transmission behavior. The KNN +

TABLE I
ENERGY CONSUMPTION COMPARISON

Method	$\mathbf{E}_{\mathrm{UAV}}$ **(Wh)**	$\mathbf{E}_{\mathrm{cell}}$ **(Wh)**	$\mathbf{E}_{\mathrm{total}}$ **(Wh)**	**Served Users (%)**	**Saving vs. MADDPG (%)**
KNN + Fixed Power	34.29	86.37	120.66	76.20	23.78
MADDPG (Ours)	**5.60**	86.37	**91.97**	**78.42**	–
Random	34.32	86.37	120.69	63.13	23.80
All Cell ON	–	93.28	121.06	77.66	**24.03**

Fixed Power and Random baselines consume over 120 Wh due to their static or non-optimized decision-making, leading to redundant UAV movement or inefficient transmission scheduling. Our MADDPG-based approach achieves the lowest total energy consumption (91.97 Wh) while maintaining a comparable user service rate (78.42%). This indicates that the trained agents effectively balance between activating UAVs and GBSs to minimize energy usage without sacrificing network service quality. Specifically, MADDPG reduces the total energy consumption by approximately 24% compared to the All Cell ON configuration, and by around 23.8% compared to the random policy, while serving a higher percentage of users.

It is worth noting that the total served user percentage slightly differs from the coverage rate in the previous section. Here, the metric accounts for all users successfully served by either UAVs or active GBSs, whereas the coverage rate previously measured only the UAV-needed coverage proportion. Overall, the MADDPG-based coordination strategy achieves the best trade-off between energy efficiency and service performance, confirming its capability to dynamically adapt to traffic variations and power constraints in green UAV-assisted cellular networks.

VI. CONCLUSION

This paper presents a UAV-assisted approach to green, resilient wireless networks using a MADDPG framework with CTDE. The method jointly optimizes UAV trajectories, transmit power, and user-UAV association under a sleeping-cell strategy to maximize user coverage while minimizing UAV energy; in simulations it outperforms conventional baselines, achieving the highest coverage and reducing total energy consumption by 24% versus an All-Cell-ON configuration. Future work includes lightweight policy distillation and transfer learning to speed convergence; integrating adaptive rate control, user prioritization, and dynamic bandwidth allocation to better satisfy per-user QoS; and jointly optimizing the GBS sleep schedule with UAV trajectories and transmit power. Overall, integrating MADRL with UAV-assisted architectures is a promising path toward sustainable, autonomous, and resilient 6G-era networks.

REFERENCES

[1] 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," 2025 IEEE Wireless Communications and Networking Conference (WCNC), Milan, Italy, 2025, pp. 1-6, doi: 10.1109/WCNC61545.2025.10978758.

[2] T. R. Pijnappel, J. L. Van Den Berg, S. C. Borst and R. Litjens, "Joint Use of Drone-Mounted Base Stations and Cell Outage Compensation in Emergency Scenarios," 2024 15th IFIP Wireless and Mobile Networking Conference (WMNC), Venice, Italy, 2024, pp. 1-8, doi: 10.52545/3-1.

[3] H. Li et al., "UAV Assisted BS Sleep Strategy for Green Communication," in IEEE Transactions on Network Science and Engineering, vol. 12, no. 5, pp. 3770-3783, Sept.-Oct. 2025, doi: 10.1109/TNSE.2025.3565316.

[4] C. H. Liu, Z. Chen, J. Tang, J. Xu and C. Piao, "Energy-Efficient UAV Control for Effective and Fair Communication Coverage: A Deep Reinforcement Learning Approach," in IEEE Journal on Selected Areas in Communications, vol. 36, no. 9, pp. 2059-2070, Sept. 2018, doi: 10.1109/JSAC.2018.2864373.

[5] Wang, Y., Fang, W., Ding, Y., & Xiong, N. (2021). Computation offloading optimization for UAV-assisted mobile edge computing: a deep deterministic policy gradient approach. Wireless Networks, 27(4), 2991–3006.

[6] Leslie Pack Kaelbling, Michael L. Littman, Anthony R. Cassandra, Planning and acting in partially observable stochastic domains, Artificial Intelligence, Volume 101, Issues 1–2, 1998, Pages 99-134, ISSN 0004-3702.

[7] A. Surriani, O. Wahyunggoro and A. I. Cahyadi, "Noise Parameterization of Continuous Deep Reinforcement Learning for a Class of Non-linear System," 2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2022, pp. 24-29.

[8] M. Oikonomakou, A. Khlass, D. Laselva, M. Lauridsen, M. Deghel and G. Bhatti, "A Power Consumption Model and Energy Saving Techniques for 5G-Advanced Base Stations," 2023 IEEE International Conference on Communications Workshops (ICC Workshops), Rome, Italy, 2023, pp. 605-610.

[9] S. K. G. Peesapati, M. Olsson, M. Masoudi, S. Andersson and C. Cavdar, "An Analytical Energy Performance Evaluation Methodology for 5G Base Stations," 2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Bologna, Italy, 2021, pp. 169-174.

[10] D. -H. Tran, S. Chatzinotas and B. Ottersten, "Satellite- and Cache-Assisted UAV: A Joint Cache Placement, Resource Allocation, and Trajectory Optimization for 6G Aerial Networks," in IEEE Open Journal of Vehicular Technology, vol. 3, pp. 40-54, 2022.

[11] D. -H. Tran, S. Chatzinotas and B. Ottersten, "Throughput Maximization for Backscatter- and Cache-Assisted Wireless Powered UAV Technology," in IEEE Transactions on Vehicular Technology, vol. 71, no. 5, pp. 5187-5202, May 2022.

[12] Chang, W., Meng, Z. T., Liu, K. C., & Wang, L. C. (2021). Energy-Efficient Sleep Strategy for the UBS-Assisted Small-Cell Network. IEEE Transactions on Vehicular Technology, 70(5), 5178-5183. Article 9416880. https://doi.org/10.1109/TVT.2021.3075603.

[13] A. Alsharoa, H. Ghazzai, A. Kadri and A. E. Kamal, "Energy Management in Cellular HetNets Assisted by Solar Powered Drone Small Cells," 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 2017, pp. 1-6, doi: 10.1109/WCNC.2017.7925568.

[14] Gaddam, Akhileswar Chowdary & Ramamoorthi, Yoghitha & Kumar, Abhinav & Cenkeramaddi, Linga Reddy. (2021). Joint Resource Allocation and UAV Scheduling With Ground Radio Station Sleeping. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3111087.

[15] Y. -H. Choi et al., "ML-Based 5G Traffic Generation for Practical Simulations Using Open Datasets," in IEEE Communications Magazine, vol. 61, no. 9, pp. 130-136, September 2023, doi: 10.1109/MCOM.001.2200679.