International IT Journal of Research (IITJR), ISSN: 3007-6706
Volume 3, Issue 3, July-September, 2025
Available online at: https://itjournal.org/index.php

Dynamic Resource Allocation in 6G Non-Terrestrial Networks via Multi-Agent Deep Reinforcement Learning

Dr. Seung-Min Cho

Associate Professor, School of Electrical Engineering and Computing, Pohang University of Science and Technology, South Korea

ABSTRACT

The evolution of sixth-generation (6G) wireless communication networks is expected to transform global connectivity through the integration of terrestrial and non-terrestrial networks (NTNs), including low Earth orbit (LEO) satellites, high-altitude platform stations (HAPS), unmanned aerial vehicles (UAVs), and other aerial communication infrastructures. These heterogeneous network architectures offer significant opportunities for providing ubiquitous connectivity, improving communication coverage, and supporting emerging applications such as autonomous transportation, Internet of Things (IoT), remote healthcare, and real-time industrial automation. However, dynamic resource allocation in 6G non-terrestrial networks presents several challenges due to highly variable network conditions, satellite mobility, fluctuating traffic demands, limited spectrum availability, energy constraints, and frequent changes in network topology. Traditional resource allocation techniques often struggle to maintain optimal performance in such complex and dynamic environments.

This study proposes a Multi-Agent Deep Reinforcement Learning (MADRL)-based framework for intelligent and adaptive resource allocation in 6G non-terrestrial networks. The proposed approach enables multiple autonomous agents to collaboratively optimize communication resources, including bandwidth allocation, transmission power, spectrum utilization, channel selection, and computational resources. By incorporating deep reinforcement learning algorithms, the framework enables agents to learn optimal allocation strategies through continuous interaction with dynamic network environments. A cooperative decision-making mechanism is introduced to improve coordination among heterogeneous network nodes while minimizing communication interference, transmission latency, and energy consumption. Furthermore, the proposed framework considers quality of service (QoS), network reliability, throughput maximization, and load balancing as essential optimization objectives.

The proposed methodology can be evaluated through simulation-based experiments under different network traffic conditions, mobility patterns, and satellite connectivity scenarios. Performance evaluation focuses on throughput, resource utilization efficiency, end-to-end latency, energy consumption, and convergence stability. Comparative analysis with conventional optimization methods and single-agent reinforcement learning approaches can provide insights into the effectiveness of collaborative multi-agent learning. The study highlights the potential of MADRL to support scalable, autonomous, and intelligent resource management in future 6G non-terrestrial communication infrastructures. Overall, this research establishes an AI-driven resource allocation framework that can contribute to reliable, energy-efficient, and adaptive connectivity for next-generation integrated terrestrial and non-terrestrial networks.

Keywords: 6G Networks, Non-Terrestrial Networks (NTN), Multi-Agent Deep Reinforcement Learning, Dynamic Resource Allocation, Satellite Communications.

INTRODUCTION

The rapid evolution of wireless communication technologies has created increasing demand for high-speed connectivity, ultra-low latency, massive device connectivity, and reliable communication services across diverse geographical environments. While fifth-generation (5G) networks have significantly improved mobile communication capabilities, emerging applications such as autonomous vehicles, intelligent transportation systems, smart cities, industrial automation, extended reality (XR), and large-scale Internet of Things (IoT) infrastructures require more advanced communication capabilities. In this context, sixth-generation (6G) wireless networks are expected to introduce a new communication paradigm characterized by intelligent connectivity, integrated sensing and communication, artificial intelligence (AI)-driven network management, and seamless integration of terrestrial and non-terrestrial communication infrastructures.

Non-Terrestrial Networks (NTNs) are expected to play a fundamental role in the development of 6G communication systems by extending network coverage beyond conventional terrestrial infrastructure. These networks integrate multiple aerial and space-based communication platforms, including Low Earth Orbit (LEO) satellites, Medium Earth Orbit (MEO) satellites, Geostationary Earth Orbit (GEO) satellites, High-Altitude Platform Stations (HAPS), and Unmanned Aerial Vehicles (UAVs). Such heterogeneous infrastructures can provide connectivity in remote, rural, maritime, mountainous, and disaster-affected regions where traditional cellular networks are difficult or economically challenging to deploy. Furthermore, the integration of terrestrial and non-terrestrial communication systems can facilitate ubiquitous connectivity and support emerging applications requiring high reliability and continuous network availability.

Despite these advantages, efficient resource allocation remains one of the most challenging problems in 6G non-terrestrial networks. Unlike conventional terrestrial networks, NTNs operate under highly dynamic environmental and operational conditions. Satellite mobility, changing coverage areas, variable propagation delays, Doppler frequency shifts, limited onboard energy resources, and fluctuating traffic demands significantly affect network performance. Additionally, the coexistence of multiple communication platforms introduces challenges related to spectrum sharing, interference management, transmission power optimization, bandwidth allocation, and computational resource distribution. Consequently, static resource allocation techniques and conventional optimization algorithms may not provide sufficient adaptability for complex and rapidly changing 6G network environments.

Dynamic resource allocation is therefore essential for maintaining efficient network operations and satisfying diverse Quality of Service (QoS) requirements. Different applications have different communication requirements. For example, autonomous transportation and emergency communication systems require extremely low latency and high reliability, whereas high-definition video streaming requires substantial bandwidth and stable throughput. Similarly, massive IoT applications prioritize energy efficiency and reliable connectivity for numerous low-power devices. An intelligent allocation mechanism must continuously analyze network conditions, predict resource requirements, and distribute available resources according to application-specific priorities.

PROPOSED MODELS AND METHODOLOGIES

The proposed research develops a Multi-Agent Deep Reinforcement Learning (MADRL)-based framework for dynamic resource allocation in sixth-generation (6G) Non-Terrestrial Networks (NTNs). The primary objective is to establish an intelligent, distributed, and adaptive resource management mechanism capable of optimizing communication resources across heterogeneous satellite and aerial network infrastructures. The proposed framework integrates artificial intelligence, cooperative decision-making, deep reinforcement learning, and multi-objective optimization to address challenges associated with bandwidth allocation, transmission power management, spectrum utilization, energy efficiency, interference reduction, and Quality of Service (QoS).

The methodology considers a heterogeneous network environment consisting of Low Earth Orbit (LEO) satellites, Medium Earth Orbit (MEO) satellites, High-Altitude Platform Stations (HAPS), Unmanned Aerial Vehicles (UAVs), terrestrial base stations, and ground user equipment. Multiple intelligent agents interact with the network environment, observe local network conditions, and collaboratively determine resource allocation policies.

2.1 Proposed 6G Non-Terrestrial Network Architecture

The proposed architecture consists of multiple interconnected communication layers designed to provide seamless connectivity and efficient resource utilization.

The architecture is divided into the following components:

1. Space-Based Communication Layer: This layer consists of LEO, MEO, and GEO satellites responsible for providing wide-area communication coverage. LEO satellites offer relatively low propagation delays and high mobility, whereas MEO and GEO satellites provide additional coverage and connectivity capabilities.

2. Aerial Communication Layer: HAPS and UAV platforms provide flexible communication infrastructure. These platforms can dynamically adjust their positions and coverage areas according to network traffic demands.

3. Terrestrial Communication Layer: This layer includes cellular base stations, terrestrial gateways, edge computing facilities, and ground communication terminals. It provides connectivity between non-terrestrial platforms and end users.

4. User Equipment Layer: This layer includes smartphones, IoT sensors, autonomous vehicles, industrial devices, and remote communication terminals requiring reliable 6G connectivity.

5. Intelligent Resource Management Layer: This layer implements the proposed MADRL algorithm to coordinate resource allocation decisions across network components.

2.2 Mathematical Modeling of Dynamic Resource Allocation

The resource allocation problem is formulated as a multi-objective optimization problem involving bandwidth, transmission power, spectrum, and communication resources.

Let the network consist of N intelligent agents and K communication users. Each agent manages a subset of network resources while coordinating with other agents.

The total available bandwidth is represented as:

$$B_{\text{total}}=\sum_{i=1}^{N}B_i$$

Where:

- B_{total} = Total available network bandwidth.
- B_i = Bandwidth allocated to agent i.
- N = Total number of agents.

The transmission rate for user k can be expressed using the Shannon capacity model:

$$R_k=B_k\log_2\left(1+ \frac{P_k h_k}{N_0B_k+I_k}\right)$$

Where:

- R_k = Achievable data transmission rate.
- B_k = Allocated bandwidth.
- P_k = Transmission power.
- h_k = Effective channel gain.
- N_0 = Noise power spectral density.
- I_k = Interference power.

The proposed framework aims to maximize aggregate network throughput while minimizing latency and energy consumption.

The optimization objective is defined as:

$$\max_{\mathbf{x},\mathbf{p}} \left[\alpha R_{\text{total}} -\beta D_{\text{total}} -\gamma E_{\text{total}} - \delta I_{\text{total}} \right]$$

Where:

- R_{total} = Aggregate network throughput.
- D_{total} = Total communication latency.
- E_{total} = Energy consumption.
- I_{total} = Network interference.
- $\alpha,\beta,\gamma,\delta$ = Optimization weights.
- $\mathbf{x}$ = Resource allocation decisions.
- $\mathbf{p}$ = Transmission power decisions.

The optimization is subject to bandwidth, transmission power, connectivity, and QoS constraints.

2.3 Multi-Agent Deep Reinforcement Learning Framework

The central component of the proposed methodology is the Multi-Agent Deep Reinforcement Learning framework. It enables multiple autonomous agents to learn resource allocation policies through interactions with the network environment.

Multi-UAV Autonomous Path Planning in Reconnaissance Missions Considering Incomplete Information: A Reinforcement Learning Method

Each agent independently observes its local network state and selects appropriate resource allocation actions. The agents collaborate through shared information and coordinated learning strategies.

The proposed MADRL framework incorporates the following elements:

1. Agent: Each satellite, UAV, HAPS platform, or network controller is represented as an intelligent agent.

2. Environment: The environment represents the dynamic 6G NTN infrastructure, including communication channels, network traffic, mobility, interference, and available resources.

3. State Space: The state space includes bandwidth availability, transmission power, channel quality, traffic demand, network congestion, mobility information, and QoS requirements.

4. Action Space: The action space includes bandwidth allocation, channel selection, transmission power adjustment, user association, and resource scheduling.

5. Reward Function: The reward function evaluates the effectiveness of resource allocation decisions based on throughput, latency, energy efficiency, interference, and QoS satisfaction.

2.4 State, Action, and Reward Formulation

The multi-agent learning problem is formulated as a decentralized partially observable Markov decision process (Dec-POMDP).

A. State Space

The state of agent i at time step t is represented as:

$$ s_i(t)= [B_i(t),P_i(t),Q_i(t),D_i(t),H_i(t),L_i(t)] $$

Where:

- $B_i(t)$ = Available bandwidth.
- $P_i(t)$ = Current transmission power.
- $Q_i(t)$ = Channel quality.
- $D_i(t)$ = Network latency.
- $H_i(t)$ = Channel gain.
- $L_i(t)$ = Network traffic load.

B. Action Space

The action selected by agent i is represented as:

$$ a_i(t)= [b_i(t),p_i(t),c_i(t),u_i(t)] $$

Where:

- $b_i(t)$ = Bandwidth allocation.
- $p_i(t)$ = Transmission power.
- $c_i(t)$ = Channel selection.
- $u_i(t)$ = User association decision.

C. Reward Function

The reward function is designed to encourage efficient resource allocation while maintaining network reliability.

$$ r_i(t)= \alpha \frac{R_i(t)}{R_{\max}} -\beta \frac{D_i(t)}{D_{\max}} -\gamma \frac{E_i(t)}{E_{\max}} -\delta \frac{I_i(t)}{I_{\max}} -\lambda V_i(t) $$

Where:

- $R_i(t)$ = Achieved throughput.
- $D_i(t)$ = Communication latency.
- $E_i(t)$ = Energy consumption.
- $I_i(t)$ = Interference.
- $V_i(t)$ = QoS violation penalty.
- λ = QoS penalty coefficient.

Normalization ensures that different performance indicators contribute appropriately to the reward function.

RESULTS & ANALYSIS

The performance evaluation of the proposed Multi-Agent Deep Reinforcement Learning (MADRL)-based dynamic resource allocation framework demonstrates how intelligent cooperative learning can improve resource management in sixth-generation (6G) Non-Terrestrial Networks (NTNs). The evaluation focuses on network throughput, end-to-end latency, energy efficiency, spectrum utilization, QoS satisfaction, and resource allocation stability.

The proposed framework considers heterogeneous communication environments involving Low Earth Orbit (LEO) satellites, High-Altitude Platform Stations (HAPS), Unmanned Aerial Vehicles (UAVs), terrestrial gateways, and ground-based users. The performance of MADRL is compared with conventional resource allocation techniques, single-agent Deep Reinforcement Learning (DRL), and alternative multi-agent learning algorithms.

Note: The numerical values presented in the following tables are illustrative simulation results designed to demonstrate the proposed evaluation methodology. They should not be interpreted as experimentally validated findings.

3.1 Overall Network Performance Evaluation

The first experiment evaluates the overall performance of different resource allocation strategies under identical network conditions.

The evaluation considers six approaches: Static Allocation, Round Robin, Greedy Optimization, Single-Agent DRL, MADDPG, and the proposed MADRL framework.

Algorithm	Throughput (Gbps)	Latency (ms)	Energy Efficiency (Mbps/W)
Static Allocation	4.2	48.5	72
Round Robin	4.8	42.3	78
Greedy Optimization	5.4	37.8	85
Single-Agent DRL	6.1	31.5	96
MADDPG	6.8	26.4	108
Proposed MADRL	**7.5**	**20.8**	**125**

The illustrative results indicate that the proposed MADRL framework achieves the highest throughput and energy efficiency while maintaining the lowest communication latency.

The improvement is attributed to cooperative resource allocation, adaptive transmission power management, and intelligent bandwidth distribution among heterogeneous network agents.

3.2 Network Throughput Analysis

Network throughput represents the total amount of successfully transmitted information over a given period. It is an important performance indicator for evaluating the effectiveness of resource allocation in 6G NTNs.

Network Throughput Comparison
Illustrative throughput performance of different allocation algorithms.
0 Gbps2 Gbps4 Gbps6 Gbps8 GbpsStaticRound RobinGreedySingle DRLMADDPGProposed
The proposed framework achieves an illustrative throughput of 7.5 Gbps, compared with 4.2 Gbps for static allocation. This represents approximately a 78.6% improvement over the static approach.
The improvement can be explained by the following factors:

- Dynamic bandwidth distribution according to real-time traffic demand.
- Cooperative resource management among multiple network agents.
- Improved channel selection and interference management.
- Adaptive transmission power optimization.
- Efficient allocation of resources to high-priority communication services.

These findings suggest that cooperative learning can improve the utilization of available communication resources in heterogeneous NTN environments.

3.3 End-to-End Latency Analysis

Latency is a critical performance indicator for 6G applications such as autonomous transportation, remote healthcare, industrial automation, and real-time satellite communication.

End-to-End Latency Comparison
Illustrative latency values; lower values indicate better performance.
0 ms15 ms30 ms45 ms60 msStaticRound RobinGreedySingle DRLMADDPGProposed
The proposed MADRL framework achieves an illustrative average latency of 20.8 ms, compared with 48.5 ms under static allocation.
This represents an approximate 57.1% reduction in latency.

The reduction can be attributed to:

1. Intelligent traffic-aware resource scheduling.
2. Improved user association and network load balancing.
3. Dynamic bandwidth allocation based on application requirements.
4. Reduced queuing delays through adaptive resource distribution.
5. Cooperative decisions among satellite and aerial network agents.

However, propagation delay remains an important limitation, particularly for satellite links involving long communication distances. Therefore, the proposed framework primarily targets controllable scheduling, queuing, and processing delays rather than eliminating physical propagation delays.

3.4 Energy Efficiency Analysis

Energy efficiency is particularly important for non-terrestrial networks because satellites and UAVs operate under limited onboard power and energy constraints.
Energy Efficiency Comparison
Illustrative energy efficiency across resource allocation approaches.

0 Mbps/W35 Mbps/W70 Mbps/W105 Mbps/W140 Mbps/WStaticRound RobinGreedySingle DRLMADDPGProposed
The proposed MADRL framework achieves an illustrative energy efficiency of 125 Mbps/W, compared with 72 Mbps/W for static allocation.

This represents approximately a 73.6% improvement.
The improvement is associated with adaptive transmission power control, reduced unnecessary resource allocation, and better coordination between network agents.
Energy-aware learning policies can also help prevent excessive power consumption during periods of low network traffic.

COMPARATIVE ANALYSIS IN TABULAR FORM

The comparative analysis evaluates the performance of the proposed Multi-Agent Deep Reinforcement Learning (MADRL)-based dynamic resource allocation framework against conventional and intelligent resource allocation techniques in 6G Non-Terrestrial Networks (NTNs). The comparison considers network throughput, end-to-end latency, energy efficiency, spectrum utilization, Quality of Service (QoS), adaptability, scalability, and coordination capabilities.

The objective is to identify the advantages of cooperative multi-agent learning over traditional resource allocation mechanisms and single-agent reinforcement learning approaches.

4.1 Comparative Performance Analysis

Algorithm	Throughput (Gbps)	Latency (ms)	Energy Efficiency (Mbps/W)	Spectrum Utilization (%)
Static Allocation	4.2	48.5	72	58
Round Robin	4.8	42.3	78	64
Greedy Optimization	5.4	37.8	85	70
Single-Agent DRL	6.1	31.5	96	77
MADDPG	6.8	26.4	108	84
Proposed MADRL	**7.5**	**20.8**	**125**	**91**

Note: *All numerical values are illustrative simulation results, not experimentally validated measurements.*

The comparison indicates that the proposed MADRL framework achieves the highest throughput, energy efficiency, and spectrum utilization while maintaining the lowest latency among the evaluated approaches.

4.2 Comparative Analysis Based on QoS Performance

Algorithm	QoS Satisfaction (%)	Interference Management	Traffic Adaptability
Static Allocation	72	Low	Low
Round Robin	76	Limited	Moderate
Greedy Optimization	81	Moderate	Moderate
Single-Agent DRL	86	Good	Good
MADDPG	91	Very Good	Very Good
Proposed MADRL	**96**	**Excellent**	**Excellent**

The proposed approach demonstrates improved adaptability to changing traffic demands and heterogeneous communication requirements. Cooperative agents can prioritize latency-sensitive services and dynamically distribute resources according to network conditions.

4.3 Comparison of Algorithmic Characteristics

Parameters	Static Allocation	Conventional DRL	MADDPG	MAPPO	Proposed MADRL
Learning capability	No	Yes	Yes	Yes	Yes
Multiple agents	No	No	Yes	Yes	Yes
Cooperative learning	No	No	Yes	Yes	Yes
Dynamic bandwidth allocation	Limited	Yes	Yes	Yes	Yes
Power optimization	Fixed/limited	Yes	Yes	Yes	Yes
Spectrum management	Limited	Adaptive	Adaptive	Adaptive	Adaptive
Mobility adaptability	Low	Moderate	High	High	High
Distributed execution	Limited	Limited	Yes	Yes	Yes
Multi-objective optimization	Limited	Possible	Yes	Yes	Yes
Scalability potential	Low	Moderate	High	High	High

The proposed framework integrates cooperative decision-making, adaptive resource management, and multi-objective optimization to address the complex requirements of heterogeneous 6G NTNs.

4.4 Comparative Analysis of Resource Allocation Strategies

Resource Parameter	Traditional Approach	Proposed MADRL Approach	Expected Benefit
Bandwidth allocation	Fixed or predefined	Dynamic and adaptive	Improved bandwidth utilization
Transmission power	Static power configuration	Learning-based power control	Reduced energy consumption
Channel selection	Predetermined or heuristic	Intelligent channel selection	Better interference management
User association	Conventional selection	Adaptive agent-based association	Improved load balancing
Network scheduling	Fixed scheduling policies	Cooperative scheduling	Reduced queuing delays
Energy management	Limited optimization	Energy-aware learning	Improved energy efficiency
QoS management	Predefined priorities	Reward-based prioritization	Better service reliability
Mobility management	Conventional handover mechanisms	Mobility-aware resource decisions	Improved network adaptability

4.5 Comparative Analysis under Different Network Conditions

The following table presents an illustrative comparison of the proposed MADRL framework under different network traffic conditions.

Network Scenario	Static Allocation Throughput (Gbps)	Proposed MADRL Throughput (Gbps)	Performance Improvement
Low Traffic	2.8	3.9	39.3%
Moderate Traffic	3.5	5.1	45.7%
High Traffic	4.0	6.2	55.0%
Very High Traffic	4.3	7.0	62.8%
Peak Traffic	4.5	7.5	66.7%

The results suggest that the proposed framework can maintain better resource utilization as network traffic increases. However, performance under extreme traffic conditions must also be evaluated against physical network capacity and available spectrum constraints.

4.6 Comparative Analysis of Advantages and Limitations

Method	Major Advantages	Major Limitations
Static Allocation	Simple implementation and low computational overhead	Poor adaptability to dynamic traffic
Round Robin	Fair resource distribution and straightforward scheduling	Does not optimize multiple network objectives
Greedy Optimization	Low computational complexity and quick decisions	May produce locally optimal solutions
Single-Agent DRL	Adaptive learning and intelligent decisions	Limited coordination across distributed nodes
MADDPG	Cooperative learning and continuous action optimization	Training complexity and sensitivity to hyperparameters
MAPPO	Stable policy updates and effective multi-agent training	Computational overhead and reward-design dependency
Proposed MADRL	Cooperative, adaptive, and multi-objective resource management	Requires training data, coordination, and computational resources

4.7 Overall Comparative Findings

Based on the illustrative comparative evaluation, the following findings are identified:

1. **Higher Network Throughput:** The proposed MADRL framework achieves 7.5 Gbps, outperforming the conventional resource allocation approaches considered.

2. **Reduced Communication Latency:** The framework reduces illustrative average latency to 20.8 ms through adaptive scheduling and resource distribution.

3. **Improved Energy Efficiency:** Intelligent transmission power management contributes to improved energy utilization.

4. **Enhanced Spectrum Utilization:** Cooperative channel selection and dynamic bandwidth allocation improve the utilization of available spectrum.

5. **Better QoS Satisfaction:** Multi-agent coordination supports application-specific resource prioritization and improved service quality.

6. **Improved Network Adaptability:** The learning framework responds to changing traffic demand, network mobility, and channel conditions.

7. **Scalable Resource Management Potential:** Decentralized execution provides opportunities for extending resource management across heterogeneous satellite and aerial communication platforms, although scalability must be validated experimentally.

CONCLUSION

The research titled **"Dynamic Resource Allocation in 6G Non-Terrestrial Networks via Multi-Agent Deep Reinforcement Learning"** investigates an intelligent and adaptive approach to resource management in emerging sixth-generation (6G) communication infrastructures. The integration of Non-Terrestrial Networks (NTNs), including Low Earth Orbit (LEO) satellites, Medium Earth Orbit (MEO) satellites, High-Altitude Platform Stations (HAPS), and Unmanned Aerial Vehicles (UAVs), is expected to play a significant role in achieving ubiquitous connectivity and supporting advanced communication applications. However, the dynamic nature of these heterogeneous networks introduces several challenges related to bandwidth allocation, spectrum utilization, transmission power management, network interference, energy consumption, and Quality of Service (QoS).

To address these challenges, the study proposes a Multi-Agent Deep Reinforcement Learning (MADRL)-based framework that enables multiple intelligent agents to collaboratively optimize communication resources. The proposed methodology incorporates cooperative learning, dynamic decision-making, centralized training, and decentralized execution to improve resource allocation across distributed network environments. The framework considers multiple optimization objectives, including maximizing network throughput, minimizing end-to-end latency, reducing energy consumption, improving spectrum efficiency, and maintaining reliable communication services.

The illustrative simulation-based evaluation demonstrates the potential advantages of the proposed approach over conventional resource allocation methods and single-agent reinforcement learning techniques. The proposed framework achieves an illustrative network throughput of 7.5 Gbps, compared with 4.2 Gbps for static allocation. Similarly, average latency decreases from 48.5 ms to 20.8 ms, while energy efficiency improves from 72 Mbps/W to 125 Mbps/W. Spectrum utilization reaches 91%, and QoS satisfaction achieves 96% under the assumed simulation configuration. These numerical results are illustrative and require validation through reproducible simulations and practical experimental investigations.

REFERENCES

1. Akbar, M. S., Hussain, Z., Ikram, M., Sheng, Q. Z., & Mukhopadhyay, S. C. (2025). On challenges of sixth-generation (6G) wireless networks: A comprehensive survey of requirements, applications, and security issues. *Journal of Network and Computer Applications, 233*, 104040. https://doi.org/10.1016/j.jnca.2024.104040
2. Aygul, M. A., Turkmen, H., Cirpan, H. A., & Arslan, H. (2024). Machine learning-driven integration of terrestrial and non-terrestrial networks for enhanced 6G connectivity. *Computer Networks, 255*, 110875. https://doi.org/10.1016/j.comnet.2024.110875
3. Birabwa, D. J., Ramotsoela, D., & Ventura, N. (2023). Multi-agent deep reinforcement learning for user association and resource allocation in integrated terrestrial and non-terrestrial networks. *Computer Networks, 231*, 109827. https://doi.org/10.1016/j.comnet.2023.109827
4. 3rd Generation Partnership Project. (2020). *Study on new radio (NR) to support non-terrestrial networks (3GPP TR 38.811, Release 15).* https://www.3gpp.org/
5. 3rd Generation Partnership Project. (2020). *Solutions for NR to support non-terrestrial networks (3GPP TR 38.821, Release 16).* https://www.3gpp.org/
6. Diarra, D., Absaloms, H. O., & Langat, P. K. (2026). Multi-agent deep reinforcement learning-based power control for energy-efficient user-centric cell-free massive MIMO. *Physical Communication*. https://doi.org/10.1016/j.phycom.2026.103275
7. Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. *Proceedings of the AAAI Conference on Artificial Intelligence, 32*(1). https://doi.org/10.1609/aaai.v32i1.11794
8. Giordani, M., Polese, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2020). Toward 6G networks: Use cases and technologies. *IEEE Communications Magazine, 58*(3), 55–61. https://doi.org/10.1109/MCOM.001.1900411
9. Iqbal, A., Tham, M. L., Wong, Y. J., Al-Habashna, A. A., Wainer, G., Zhu, Y. X., et al. (2023). Empowering non-terrestrial networks with artificial intelligence: A survey. *IEEE Access, 11*, 100986–101006. https://doi.org/10.1109/ACCESS.2023.3314732
10. Jung, S., Jang, H., Park, S., Kim, J., Jang, J., Cho, I.-S., Shin, M., Quek, T. Q. S., Han, Z., & Kim, J. (2026). A survey on reinforcement learning methods for seamless connectivity in 6G non-terrestrial networks: Algorithms and key findings. *IEEE Communications Surveys & Tutorials, 28*, 7472–7511. https://doi.org/10.1109/COMST.2026.3740875
11. Kim, H., Lee, H., & Lee, I. (2020). Multi-agent deep reinforcement learning for distributed resource management in wirelessly powered communication networks. *IEEE Transactions on Vehicular Technology*. https://doi.org/10.1109/TVT.2020.3029609
12. Kodheli, O., Lagunas, E., Maturo, N., Sharma, S. K., Shankar, B., Montoya, J. F. M., Duncan, J. C. M., Spano, D., Chatzinotas, S., Kisseleff, S., Querol, J., Lei, L., Vu, T. X., & Goussetis, G. (2021). Satellite communications in the new space era: A survey and future challenges. *IEEE Communications Surveys & Tutorials, 23*(1), 70–109. https://doi.org/10.1109/COMST.2020.3028247
13. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *International Conference on Learning Representations (ICLR).* https://arxiv.org/abs/1509.02971
14. Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems, 30*. https://arxiv.org/abs/1706.02275
15. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*, 529–533. https://doi.org/10.1038/nature14236
16. Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G., Foerster, J., & Whiteson, S. (2018). QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. *Proceedings of the 35th International Conference on Machine Learning, 80*, 4295–4304. https://proceedings.mlr.press/v80/rashid18a.html
17. Saad, W., Bennis, M., & Chen, M. (2020). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network, 34*(3), 134–142. https://doi.org/10.1109/MNET.001.1900287
18. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv*. https://arxiv.org/abs/1707.06347

19. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book-2nd.html
20. Wang, C. X., You, X., Gao, X., Zhu, X., Li, Z., Zhang, C., Wang, H., Huang, Y., Chen, Y., Haas, H., Thompson, J. S., Larsson, E. G., Di Renzo, M., Tong, W., Zhu, P., Shen, X., Poor, H. V., & Hanzo, L. (2023). On the road to 6G: Visions, requirements, key technologies, and testbeds. *IEEE Communications Surveys & Tutorials*, *25*(2), 905–974. https://doi.org/10.1109/COMST.2023.3249835