Deep Reinforcement Learning for Downlink Power Control in 5G/6G Networks

1st Alberto López Casanova
R&D Dept.
Future Connections
Linares, Spain
alberto.lopez@futureconnections.com

2nd Rocío Pérez de Prado
Telecommunications Dept.
University of Jaén
Linares, Spain
rperez@ujaen.es

3rd Miguel A. Regueira Caumel
R&D Dept.
Future Connections
Málaga, Spain
miguelangel.regueira@futureconnections.com

4th Carlos Hidalgo-Luque
Telecommunications Dept.
University of Jaén
Linares, Spain
chluque@ujaen.es

5th Chuan Heng Foh
Communication Systems Dept.
University of Surrey
Guildford, United Kingdom
c.foh@surrey.ac.uk

***Abstract*—From the creation of early mobile network standards to current 5G and beyond, interference management has been extensively researched as it is crucial for improving network performance and service quality. Among the different techniques, downlink power control, which consists of dynamically adjusting the power with which base stations transmit to users, has demonstrated to be critical to mitigate inter-cell interference. In this context, through the years, traditional optimization methods have been applied, achieving good performance but facing challenges in current standards characterized by high complexity in terms of service requirements. In this paper, we propose a deep reinforcement learning based approach for downlink power control, motivated by the results achieved in recent years by data-driven approaches to radio resource management optimization. We train an agent based on the double dueling deep Q-network algorithm, relying on a centralized single-agent reinforcement learning architecture. The OMNeT++ discrete-event simulator with the Simu5G library is chosen to simulate a 5G Standalone scenario where the agent is trained with the goal of maximizing the network-wide signal-to-interference-plus-noise ratio. We tested the trained agent in this scenario against the baseline approach native to the simulator, based on fixed power transmission. Validation results show that the proposed approach improves the signal-to-interference-plus-noise ratio while achieving a reduction in physical resource block usage compared with the baseline approach. While the work presented in this paper is based on scalar network key performance indicators, it establishes a baseline for future improvements to incorporate multimodal data, enabling the use of multimodal architectures to achieve a deeper understanding of the network state.**

***Index Terms*—Deep Reinforcement Learning, Deep Q-Learning, Multimodal Neural Networks, 5G/6G Networks, OMNeT++, Simu5G, Downlink Power Control, O-RAN**

I. Introduction

In the last two decades, mobile networks have experienced exponential growth in terms of the number and diversity of connected devices, a trend that is expected to continue [1]. Mobile network standards have evolved from early generations to meet the demands of growing traffic driven by smartphones and the Internet of Things (IoT). These technologies have enabled a wide variety of connected devices and the development of new services that introduce new types of traffic in recent years, driving the need for highly dynamic networks and ultra-dense deployments. Current 5G networks are designed to support three types of strict Quality of Service (QoS) requirements with the aim of adapting to this landscape: Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC) [2]. To meet these demands, network architectures are evolving toward flexible topologies with advanced Radio Resource Management (RRM) mechanisms, significantly increasing system complexity.

Several research areas have been extensively explored to address the challenges posed by dense and heterogeneous networks (HetNets), including handover management [3], energy efficiency optimization, and interference management. In the context of interference management, downlink power control (DPC) has proven to be an effective approach to reduce inter-cell interference (ICI) while maintaining or improving spectral efficiency, which is essential to minimize power consumption per transmitted bit and reduce physical resource block (PRB) usage. As ultra-dense deployments become more common, ineffective transmit power management of the base stations may lead to high ICI, affecting coverage and thus degrading the QoS and user experience. DPC is not a straightforward optimization problem as it needs to tackle a multi-objective problem where a balance between coverage, throughput, and interference is required, despite being opposing objectives. This, coupled with the variability of traffic in HetNets, makes this optimization problem even more complex.

Traditional power control techniques are characterized by the use of heuristics or statistical models that, although achieving good performance in simple networks, are prone to fail when adapting to the dynamic conditions of modern networks. In recent years, as the use of data-driven solutions and Artificial Intelligence (AI) has expanded into all areas of society, the research community has extensively studied the use of deep reinforcement learning (DRL) for RRM optimization problems. Its exploratory nature in training matches the variability of the wireless environment, thus achieving remarkable performance. While DRL solutions in this context typically rely solely on scalar inputs, Multimodal Neural Networks are gaining attention as they enable the integration of heterogeneous data sources [4]—such as combining network key performance indicators (KPIs) with visual embeddings from simulation scenario images or text embeddings compressing contextual information, helping the DRL agent achieve a deeper understanding of the network state.

Most existing DRL-based approaches for RRM target user-level or distributed multi-agent reinforcement learning (MARL) architectures [5], [6]. This paper proposes a centralized DRL-based approach for cell-level DPC. The main contributions are:

- Development of a single-agent reinforcement learning (SARL) agent based on the double dueling deep Q-network (D3QN) algorithm for cell-level DPC, aiming to maximize network-wide signal-to-interference-plus-noise ratio (SINR).
- Integration of OMNeT++ [7] with the Simu5G [8] library to generate simulated network scenarios, extract KPIs, and train DRL agents in a controlled environment.
- Demonstration of the effectiveness of DRL-based solutions through simulation results, showing improved SINR and PRB utilization, thereby highlighting the potential of neural network-based approaches for optimizing DPC in 5G and beyond.
- Establishment of an extensible DRL framework designed to support future multimodal data integration alongside standard network KPIs for richer state representations.

The remainder of this paper is organized as follows: Section II reviews related works on downlink power control and reinforcement learning in 5G/6G networks. In Section III, IV and V we formulate the downlink power control problem and present our practical DRL proposal. Section VI presents evaluation results and discussion. Finally, Section VII concludes the paper and outlines potential future research directions.

II. Related Work

Interference management in mobile cellular networks has been an active research topic since the early generations of cellular systems. The aggressive reuse of radio frequencies, essential to maximize system capacity, inevitably makes ICI a dominant performance-limiting factor, rendering cellular networks interference-constrained. In this context, power control is an RRM optimization that has proven to be key in modern mobile networks to regulate interference levels and balance signal quality across user equipment (UEs), with the multi-objective of improving capacity, coverage, energy efficiency, and QoS [9].

The first research efforts in the line of interference management with DPC are presented in [10], [11], where the objective is to suppress ICI to improve cell sum-rate by iteratively adjusting transmission power based on the current Carrier-to-Interference Ratio (CIR) to minimize outage probability. Further research based on this foundational work focused on model-based optimizations, with the Weighted Minimum Mean Square Error (WMMSE) algorithm [12] and Fractional Programming (FP) [13] regarded as standard reference methods against which to evaluate modern approaches. Among traditional approaches, convex approximation [14], heuristics [15], and game theory [16] have also been implemented to tackle this power allocation optimization problem. Even though these traditional methods provide a strong theoretical foundation and improvements in small networks based on early standards, their application in large and modern HetNets is not feasible due to their limited scalability and high computational complexity. Furthermore, these algorithms typically rely on the assumption of available instantaneous global Channel State Information (CSI), which is often unrealistic in practice due to signaling overhead and estimation errors.

In the current era of AI, DRL has emerged as a promising alternative to traditional methods for RRM optimization. An early use of DRL for power control, integrating RL with an Artificial Neural Network to approximate a Q-function—typically known nowadays as DQN—is presented in [17]. Similarly, in [18], the authors present the first MARL solution applying DQN for dynamic power allocation, where the objective is to maximize a weighted sum-rate utility function. DQN has been widely used since then in 5G networks [19] and applied to systems such as Non-Orthogonal Multiple Access (NOMA) to efficiently maximize spectrum sharing [20]. In order to map wireless environment fluctuations to power control actions, an auxiliary Deep Neural Network (DNN) is added to the DQN architecture in [21]. Although value-based methods are the preferred type within the DRL landscape for their application in this optimization problem, policy-based methods are also explored, with both types being compared in multi-user scenarios in [22]. The MARL architecture is predominant in the research often utilizing the Centralized Training with Distributed Execution (CTDE) framework introduced in [17], [18] using DQN and further extended using Proximal Policy Optimization (PPO) in [23]. Although usually preferred due to the scalability issues that centralized solutions present—the observation space grows drastically with network size—, MARL also presents important limitations. As highlighted in [24], distributed agents suffer from partial observability, which significantly increases training complexity.

III. System Model

A. Network Scenario

We consider a 5G Standalone (SA) cellular network composed of M next-generation Node Bs (gNBs), denoted by the set $\mathcal{M} = \{1, \ldots, M\}$, and K UEs, denoted by the set $\mathcal{K} = \{1, \ldots, K\}$. The network operates under a frequency reuse-1 scheme, implying that all gNBs transmit simultaneously over the entire system bandwidth W.

We define $\mathcal{K}_m(t) \subset \mathcal{K}$ as the dynamic set of UEs served by gNB m. Within each cell, UEs are scheduled using Orthogonal Frequency Division Multiple Access (OFDMA); therefore, we consider intra-cell interference negligible so that the system performance is mainly limited by the ICI generated by neighboring gNBs.

B. Signal and Channel Model

The wireless channel between gNB m and UE k accounts for large-scale fading (path loss and shadowing) and small-scale fading. The channel gain from gNB m to UE k is represented by $h_{m,k}(t)$.

The SINR measured by a UE k served by gNB m at time slot t is given by:

$$SINR_k(t) = \frac{P_m(t) h_{m,k}(t)}{\sum_{j \in \mathcal{M}, j \neq m} P_j(t) h_{j,k}(t) + \sigma^2} \tag{1}$$

where:

- $P_m(t)$ is the transmit power of the serving gNB m.
- The term $\sum_{j \neq m} P_j(t) h_{j,k}(t)$ represents the aggregate ICI.
- σ^2 denotes the additive white Gaussian noise (AWGN) power.

C. Problem Formulation

The primary objective of the proposed power control scheme is to maximize the average network-wide signal quality while strictly adhering to power constraints. We formulate this as a utility maximization problem where the objective function is the mean SINR of the system:

$$\max_{\mathbf{P}} \frac{1}{K} \sum_{k=1}^{K} SINR_k(t) \tag{2}$$

Subject to the following constraints:

Power Budget: The transmit power of each gNB must remain within the hardware operational limits:

$$P_{min} \leq P_m(t) \leq P_{max}, \quad \forall m \in \mathcal{M} \tag{3}$$

Discrete Power Control Dynamics: Following the network standards (e.g., 3GPP), we utilize discrete Transmit Power Control (TPC) commands. We model the power update rule as a step-based adjustment:

$$P_m(t+1) = \text{clip}_{[P_{min}, P_{max}]} \left(P_m(t) + \Delta P_m(t) \right) \tag{4}$$

where the adjustment step is determined by a discrete action set $\mathcal{A}$:

$$\Delta P_m(t) \in \{-\delta, 0, +\delta\} \text{ dB} \tag{5}$$

Here, δ represents the fixed power control step size (e.g., 1 dB).

As the optimization variables appear in both the numerator and the denominator of the SINR expression (1), the utility function (2) is non-convex with respect to the transmit power $\mathbf{P}$.

IV. Proposal

In this section we describe the DRL agent based on a centralized SARL architecture that we propose as a solution to address the combinatorial complexity of the DPC optimization problem formulated in Section III.

A. MDP Formulation

First, we formulate the DPC problem as a Markov Decision Process (MDP) defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$. At each time step t, the agent observes the global state s_t, executes an action a_t, and receives a scalar reward r_t calculated by a reward function.

Domain-Aware State Space ($\mathcal{S}$):

The global state s_t is constructed by concatenating the local observations from all M gNBs into a single flattened feature vector: $s_t = \text{flatten}(\{\mathbf{o}_1(t), \ldots, \mathbf{o}_M(t)\})$. For each gNB m, the observation vector $\mathbf{o}_m(t)$ consists of:

- **SINR Distribution:** The agent is provided with a discretized Probability Density Function (PDF) representing the SINR distribution of all attached UEs. This metric is computed with a 1 dB bin resolution, allowing the agent to perceive the fine-grained distribution of interference across the cell.
- **Power Headroom (PHR):** The margin between the current transmit power and the configured limits ($PHR_m^{up} = P_{max} - P_m(t)$, $PHR_m^{down} = P_m(t) - P_{min}$).
- **Load Metrics:** The Uplink and Downlink PRB utilization ratios (ρ_m^{UL}, ρ_m^{DL}) and the aggregated Packet Error Rate (PER).

Discrete Action Space ($\mathcal{A}$):

Consistent with the system model, we employ a discrete action space to ensure stability and standard compliance. For each gNB m, the agent selects a command $a_m(t) \in \{0, 1, 2\}$ mapping to the TPC adjustments:

$$\begin{aligned} a_m = 0 &\implies \Delta P = 0 \text{ (No Change)} \\ a_m = 1 &\implies \Delta P = +1\,\text{dB (Increase Power)} \\ a_m = 2 &\implies \Delta P = -1\,\text{dB (Decrease Power)} \end{aligned}$$

This approach mitigates the oscillatory behavior often associated with continuous power control in high-variance channels.

Reward Function ($\mathcal{R}$):

The objective is to maximize the network-wide signal quality. The system calculates the average linear SINR across all active users. To ensure numerical stability and facilitate neural network convergence, this value is normalized via Min-Max scaling to the range $[-1, 1]$:

$$r_t = 2 \cdot \frac{\overline{\text{SINR}}_t - \text{SINR}_{min}}{\text{SINR}_{max} - \text{SINR}_{min}} - 1 \tag{6}$$

where $\overline{\text{SINR}}_t$ is the instantaneous average network SINR, and $[\text{SINR}_{min}, \text{SINR}_{max}]$ represents the expected operational range of the system.

B. Proposed D3QN Agent

Once the problem is formulated as an MDP defined by $\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$, the objective is to learn an optimal policy π^* through training that maximizes the expected cumulative reward. To achieve this, we implement and train a D3QN agent.

DQN approximates the optimal action-value function $Q^*(s,a)$ by minimizing the loss between the current prediction and the Temporal-Difference (TD) target, calculated as:

$$y = r + \gamma \max_{a'} Q(s', a'; \theta^-) \tag{7}$$

The Double DQN extension [30] helps to avoid overestimation, while the Dueling architecture [31] aims to decouple the state-value function from the advantage function, making the training more efficient in complex scenarios. Prioritized Experience Replay (PER) is also selected for the replay buffer to improve sample efficiency. The specific parameters, hyperparameters, and training settings are listed in Table I.

TABLE I
D3QN PARAMETERS, HYPERPARAMETERS AND TRAINING SETTINGS

Parameter	**Value**
Optimizer	Adam
Learning Rate (α)	5×10^{-5}
Discount Factor (γ)	0.99
Batch Size	96
Hidden Layers	1×256 neurons
Activation Function	ReLU
Loss Function	Huber
Replay Buffer Size	75000 transitions
Steps Before Learning	5000
Target Network Update	Every 2000 steps
N-step Returns	3
Prioritized Replay	$\alpha = 0.6$, $\beta = 0.4$
Exploration (ϵ)	Linear ($1.0 \to 0.05$)
ϵ Decay Steps	60000
Ray Seed	0,500,1500,2000,2500
OMNeT++ Base Seed	0,500,1500,2000,2500 (incremental)
Observation Period (T)	1 s (simulation time)

V. EXPERIMENTAL SETUP

A. Network Scenario and Parameters

We designed a 5G SA network scenario composed of 4 gNBs and 20 UEs in OMNeT++ with Simu5G for the training and validation of our D3QN agent, the characteristics of which are summarized in Table II. This network scenario allows us to study the DPC problem and present preliminary results in a controlled environment. Training in more complex and varied scenarios will be carried out in future work.

TABLE II
SIMULATED 5G SA NETWORK PARAMETERS

Parameter	**Value**
Layout	2×2 Grid (4 gNBs)
Coverage Area / ISD	4600×4600 m
Distance between gNBs	1000 m
Number of UEs (K)	20 (Uniformly placed)
Mobility Model	Gauss-Markov ($1 - 1.5$ m/s)
Traffic Type	CBR (100 bytes, 20 ms interval)
Carrier Frequency / BW	3.5 GHz / 50 PRBs
Frequency Reuse	Reuse-1
Power Range $[P_{min}, P_{max}]$	$[10, 50]$ dBm ($P_{init} = 46$ dBm)
UE Transmit Power	26 dBm
Channel Model	3GPP TR 38.901 [26]
Fading Type	JAKES

B. DRL Framework

Regarding the architecture of the solution (Fig. 2), we extended the framework developed by Hidalgo-Luque [27] to enable DPC within the simulator and the training of the D3QN agent for this optimization problem. A custom OMNeT++ submodule embedded in each gNB extracts cell-level KPIs and applies power changes in runtime. A separate custom module aggregates these metrics every observation period T, forwarding the raw network state to the Python environment via the Veins-Gym framework [28]. Within a custom Gymnasium wrapper, these raw KPIs are processed to construct the observation space and compute the reward function. The D3QN agent is implemented using Ray RLlib [29].

C. Training Procedure

The training experiment is composed of 5 independent runs, each using a different Ray RLlib seed and OMNeT++ seeds. Each run lasts 300 episodes of 250 steps. The OMNeT++ base seeds are incremented per episode. Fig. 1 shows the training convergence of the proposed D3QN agent. The solid line represents the mean reward aggregated across runs, while the shaded area denotes the standard deviation.

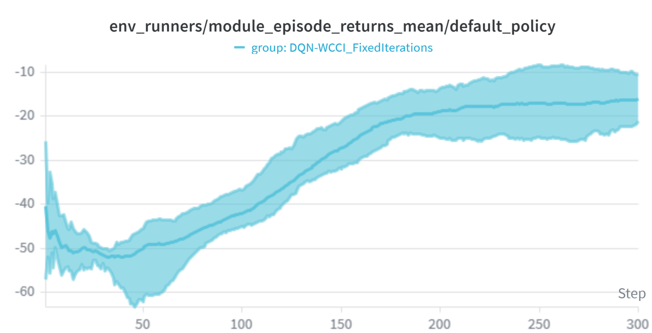

Fig. 1. Training convergence of the proposed D3QN agent.

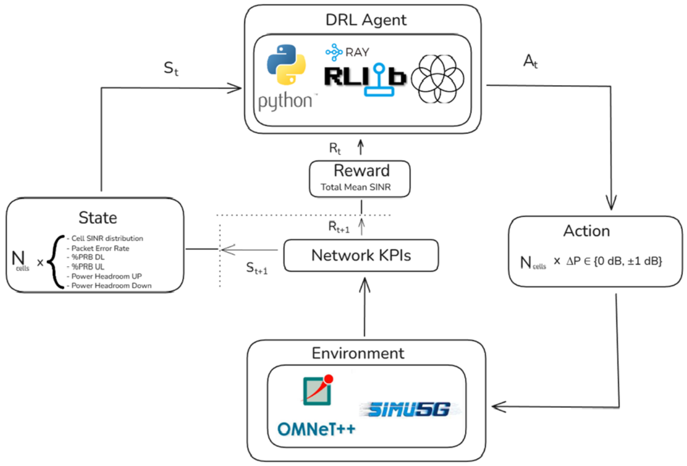

Fig. 2. Architecture of the Proposed DRL-based DPC Solution.

VI. Results

A. Performance Evaluation and Discussion

We compare our D3QN agent against a fixed transmission power baseline, since it is the standard configuration in real deployed networks. FP and WMMSE are not included due to their computational complexity and precise CSI requirements [35]. This would generate higher latency in the RIC; thus, they are not a practical approach and are out of the scope of this work.

For evaluation, the run with highest accumulated reward is selected. The experimental results, obtained over 30 evaluation episodes of 250 steps each and summarized in Table III, show that the D3QN agent achieves a 34.9% improvement in network-wide mean SINR while simultaneously reducing PRB usage by 20.3% and improving the 5th percentile (P5) SINR by 57.1% compared to the fixed-power baseline. The p-values are obtained with the Wilcoxon signed-rank test, demonstrating the statistical significance of our solution in this environment.

VII. Conclusion and Future Work

A. Conclusions

As demonstrated in this work, the proposed D3QN-based DPC approach successfully manages interference, achieving a significant improvement in terms of signal quality and system efficiency. The proposed solution increases the network-wide and P5 SINR while reducing the PRB usage of the transmission, which translates to an increase in system capacity to serve more UEs or save radio resources for the same number of users, while the user experience is not degraded. This solution is validated and benchmarked against a fixed power approach as a baseline, which is an accurate representation of power control in currently deployed networks.

This solution perfectly matches the vision of network operators and the research community towards an AI-native mobile network standard, reflected in the emergence of Open Radio Access Network (O-RAN) architectures, which aim to create interoperable networks across different vendors and enable the integration of data-driven intelligent solutions in the RAN, such as the one presented in this work. The training approach based on a network simulator facilitates the integration of these types of solutions in deployed networks by pre-training the DRL agent with synthetic data, thereby shortening the re-training phase required for real networks.

B. Future Work

Future work will first focus on improving the current D3QN solution through more complex reward design and the addition of new KPIs tailored for the DPC optimization problem (e.g., Reference Signal Received Power (RSRP) distribution) in the state space. The integration of multimodal data will also be studied, leveraging interference heat-maps images to generate visual embeddings that help the DRL agent better understand spatiotemporal interference patterns.

More algorithms will be tested, ranging from value-based methods like Rainbow DQN [32] to policy-based approaches such as PPO. In order to ensure generalization, scenarios with different complexities and gNB layouts will be developed. Moreover, to ensure scalability, MARL will be investigated.

Furthermore, more advanced topics will be examined, such as adopting Multi-Objective Reinforcement Learning (MORL) to optimize the trade-off between interference minimization and throughput maximization [33], and Federated Learning (FL) [34] to bridge the gap between simulation and reality.

Acknowledgment

This work was supported by UKRI (UK), CDTI (ES), ANI (PI), and TUBITAK (IR) within the EUREKA CELTIC-NEXT project 6G-SMART, C2023/2-20 and project Investigación en Modelos de Autoconfiguración para el Proyecto de Cooperación Tecnológica Internacional CELTICNEXT 6G-SMART. Cooperación de Partners Consorcio Español Future Connections y Universidad de Jaén.

The authors acknowledge the use of Google Gemini AI [36] for grammar and style editing assistance throughout this manuscript.

TABLE III
PERFORMANCE COMPARISON AND STATISTICAL SIGNIFICANCE: D3QN AGENT VS. FIXED-POWER BASELINE

Approach	Performance Metrics (Mean $\pm$ Std)			Comparison & Significance			
	SINR (dB)	**PRB (%)**	**SINR P5 (dB)**	**Δ SINR**	**Δ PRB**	**Δ P5**	**p-value**
Baseline (Fixed)	10.02 ± 2.22	1.58 ± 0.19	-7.00 ± 2.73	–	–	–	–
D3QN Agent	13.52 ± 3.28	1.26 ± 0.27	-3.00 ± 2.73	$+34.9\%$	-20.3%	$+57.1\%$	< 0.001

Specific p-values: SINR (1.51×10^{-5}), PRB (1.10×10^{-5}), SINR P5 (1.10×10^{-5}).

REFERENCES

[1] Ericsson, "Ericsson Mobility Report, November 2025," [Online]. Available: https://www.ericsson.com/4aca6f/assets/local/reports-papers/mobility-report/documents/2025/ericsson-mobility-report-november-2025.pdf.

[2] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *IEEE Access*, vol. 6, pp. 55765–55779, 2018, doi: 10.1109/ACCESS.2018.2872781.

[3] C. Hidalgo-Luque *et al.*, "Digital Twin-Based Deep Q-Learning Strategy for Smart Handover Optimization in 5G/6G Networks," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Rome, Italy, 2025.

[4] Y. Lu, "A multimodal deep reinforcement learning approach for IoT-driven adaptive scheduling and robustness optimization in global logistics networks," *Scientific Reports*, vol. 15, Art. no. 10512, 2025, doi: 10.1038/s41598-025-10512-1.

[5] X. Du *et al.*, "Multi-agent reinforcement learning for dynamic resource management in 6G in-X subnetworks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1900–1914, 2022.

[6] N. Naderializadeh, J. Sydir, M. Simsek, and H. Nikopour, "Resource management in wireless networks via multi-agent deep reinforcement learning," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2020, pp. 1–5.

[7] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. 1st Int. Conf. Simul. Tools Techn. Commun., Netw. Syst. Workshops (Simutools)*, Marseille, France, 2008, Art. no. 60.

[8] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis, "Simu5G—An OMNeT++ library for end-to-end performance evaluation of 5G networks," *IEEE Access*, vol. 8, pp. 181176–181191, 2020, doi: 10.1109/ACCESS.2020.3028550.

[9] S. Khangura, "Power Control Algorithms in Wireless Communication," *Int. J. Comput. Appl.*, vol. 1, pp. 86–91, 2010, doi: 10.5120/253-410.

[10] S. Grandhi, R. Vijayan, and D. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. Commun.*, vol. 42, no. 2-3-4, pp. 226–228, 1994.

[11] T. Lee, J. Lin, and Y. Su, "Downlink power control algorithms for cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 89–94, 1995.

[12] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[13] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[14] M. Sami and J. Daigle, "User Association and Power Control for UAV-Enabled Cellular Networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 267–270, 2020.

[15] L. Zhou, Y. Wu, and H. Yu, "A Two-Layer, Energy-Efficient Approach for Joint Power Control and Uplink-Downlink Channel Allocation in D2D Communication," *Sensors*, vol. 20, no. 11, 2020.

[16] Z. Fazliu, C. Chiasserini, G. Dell'Aera, and E. Hamiti, "Distributed Downlink Power Control for Dense Networks With Carrier Aggregation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7052–7065, 2017.

[17] E. Ghadimi, F. Calabrese, G. Peters, and P. Soldati, "A reinforcement learning approach to power control and rate adaptation in cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, 2017, pp. 1–7.

[18] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2239–2250, 2019.

[19] S. Saeidian, S. Tayamon, and E. Ghadimi, "Downlink Power Control in Dense 5G Radio Access Networks Through Deep Reinforcement Learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, 2020, pp. 1–6.

[20] M. Youssef, C. Nour, X. Lagrange, and C. Douillard, "A Deep Q-Learning Bisection Approach for Power Allocation in Downlink NOMA Systems," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 316–320, 2022.

[21] R. Liang, H. Lyu, and J. Fan, "A deep reinforcement learning-based power control scheme for the 5G wireless systems," *China Commun.*, vol. 20, no. 3, pp. 109–119, 2023.

[22] F. Meng, P. Chen, L. Wu, and J. Cheng, "Power Allocation in Multi-User Cellular Networks: Deep Reinforcement Learning Approaches," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6255–6267, 2020.

[23] D. Guo, L. Tang, X. Zhang, and Y. Liang, "Joint Optimization of Handover Control and Power Allocation Based on Multi-Agent Deep Reinforcement Learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124–13138, 2020.

[24] S. Kapoor, "Multi-agent reinforcement learning: A report on challenges and approaches," arXiv:1807.09427, 2018.

[25] V. Mnih *et al.*, "Playing Atari with Deep Reinforcement Learning," arXiv:1312.5602, 2013.

[26] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Tech. Rep. TR 38.901, Apr. 2024. [Online]. Available: https://www.3gpp.org/DynaReport/38901.htm.

[27] C. Hidalgo-Luque, "Development of Machine Learning Models to Support the Self-Configuration of 5G/6G Radio Access Networks," M.Sc. thesis, Escuela Politécnica Superior de Linares, Univ. of Jaén, Linares, Spain, Jul. 2025.

[28] M. Schettler, D. S. Buse, A. Zubow, and F. Dressler, "How to train your ITS? Integrating machine learning with vehicular network simulation," in *Proc. 12th IEEE Veh. Netw. Conf. (VNC)*, Virtual Conference, Dec. 2020, doi: 10.1109/VNC51378.2020.9318324.

[29] E. Liang *et al.*, "RLlib: Abstractions for distributed reinforcement learning," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, vol. 80, pp. 3053–3062, Jul. 2018.

[30] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, 2016.

[31] Z. Wang *et al.*, "Dueling network architectures for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 1995–2003, 2016.

[32] M. Hessel *et al.*, "Rainbow: Combining improvements in deep reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, doi: 10.1609/aaai.v32i1.11796.

[33] K. Van Moffaert and A. Nowé, "Multi-objective reinforcement learning using sets of Pareto dominating policies," *J. Mach. Learn. Res.*, vol. 15, pp. 3483–3512, 2014.

[34] H. H. Zhuo, W. Feng, Y. Lin, Q. Xu, and Q. Yang, "Federated deep reinforcement learning," arXiv:1901.08277, 2020.

[35] X. Li, "DQN-Based Resource Allocation for Power Management in Cellular Network," *ITM Web Conf.*, 2025, doi: 10.1051/itm-conf/20257801003.

[36] Google DeepMind, "Gemini AI (Jan 2026 version)," [Large language model], 2023. [Online]. Available: https://deepmind.com/gemini.