
DEVELOPMENT OF A Q-LEARNING-DRIVEN INTERFERENCE MANAGEMENT MODEL WITHIN THE ADAPTIVE RRM FRAMEWORK FOR ADAPTIVE TRANSMIT POWER OPTIMIZATION

**Abubakar Abdulkadir^{1,*}, H. A. Abdulkareem¹, ZM Abdullahi¹, Elvis Obi¹, Aliyu Umar
Abubakar¹**

¹Ahmadu Bello University Zaria

Email: aabubakar@fptb.edu.ng, ha2zx@yahoo.com, amzanna@gmail.com, elvisobi72@gmail.com,
auabubakarphysics@gmail.com

***Corresponding author:** Abubakar Abdulkadir

DOI: <https://doi.org/10.67203/sp.ICESID.2026>

ABSTRACT

The rapid densification of fifth-generation (5G) heterogeneous networks has improved network capacity and coverage but also increased co-tier and cross-tier interference. In ultra-dense multi-tier Het-HetNets, traditional static and semi-static radio resource management techniques are insufficient to address rapidly changing interference, traffic demands, and user distribution. This paper presents a Q-learning-driven interference management model within an adaptive Radio Resource Management (RRM) framework to dynamically optimize transmit power for PeNBs. The hybrid Q-learning power allocation framework models each PeNB as an autonomous reinforcement-learning agent that observes network conditions, selects transmit power actions, receives performance-based rewards, and updates its Q-table using the Bellman learning mechanism. Key factors such as network capacity, interference, transmit power consumption, and Quality of Service (QoS) are incorporated into the learning process. A 5G Het-HetNet simulation in MATLAB, featuring one MeNB, one gNB, sixteen PeNBs, and diverse user equipment, evaluates three communication scenarios with different interference and spectrum-sharing conditions. The Q-learning process demonstrates stable convergence within 80,000 iterations. Results show an average improvement of 3.37% in 5G user equipment (5G-UE) capacity and 14.81% in the minimum capacity of small-cell users compared to the benchmark model by Iqbal et al. Additionally, the proposed approach maintains 5G-UE capacity above the QoS threshold under the tested conditions, highlighting its effectiveness in interference management and RRM in ultra-dense 5G networks.

Keywords: 5G and B5G; Het-HetNets; Q-learning; interference management; power allocation; adaptive RRM

1. INTRODUCTION

The evolution toward fifth-generation (5G) and beyond-5G (B5G) networks demands higher capacity, spectral efficiency, and low latency, necessitating network densification through small-cell deployments. Heterogeneous Networks (HetNets) and Heterogeneous-Heterogeneous Networks (Het-HetNets) integrate multiple network tiers, such as Macro evolved NodeBs (MeNBs), Pico evolved NodeBs (PeNBs), and 5G gNBs, to significantly enhance coverage and capacity (Luong et al., 2019; Yadav et al., 2021). However, this dense architecture generates severe co-tier and cross-tier interference, degrading the Signal-to-Interference-plus-Noise Ratio (SINR), throughput, and Quality of Service (QoS) (Abbass et al., 2022). This interference is further exacerbated by unequal user association, where users preferentially connect to high-power MeNBs, causing macrocell congestion while leaving neighboring small-cell resources underutilized. Such imbalances severely limit the overall effectiveness of the infrastructure (Iqbal et al., 2023).

Conventional interference-management techniques, such as fixed power allocation and frequency reuse, rely on predetermined policies and struggle to adapt to dynamic network conditions, traffic demands, and user distributions (Iqbal et al., 2021). Thus, machine learning, particularly reinforcement learning (RL), has emerged as a powerful tool for adaptive Radio Resource Management (RRM), allowing network entities to learn optimal resource-allocation decisions through environmental interaction rather than predefined rules (Shabbir et al., 2024).

Existing literature highlights the potential of learning-based approaches for interference management in these complex environments. Recent studies by researchers demonstrate that RL and optimization techniques can effectively improve QoS, energy efficiency, and throughput across various network configurations. However, optimization based on predefined system conditions often lacks the necessary adaptability required for dynamically changing network environments and fluctuating user distributions.

Despite these advancements, a significant research gap remains regarding lightweight Q-learning frameworks that dynamically optimize PeNB transmit power while simultaneously mitigating co-tier and cross-tier interference in multi-tier Het-HetNet architectures. Methodologically, hybrid Q-learning power allocation differs from Iqbal et al. and Amiri et al. in its learning formulation and network design. hybrid Q-learning power allocation models each PeNB as an autonomous Q-learning agent with a state space representing interference and QoS conditions, while its action space consists of 32 discrete transmit-power levels from -15 to 15 dBm. Its reward function jointly considers capacity, interference, and energy consumption, with additional QoS rewards or penalties. In contrast, Iqbal et al. primarily focus on power-control/resource-allocation optimization within its considered heterogeneous network, whereas Amiri et al. employs a different learning and network-management formulation. hybrid Q-learning power allocation further incorporates a multi-tier topology comprising a MeNB, gNB, multiple PeNBs, and heterogeneous UEs, enabling adaptive power decisions under varying interference and network-density conditions. To address this critical gap, we propose the novel Hybrid Q-Learning Power Allocation framework. hybrid Q-learning power allocation integrates Q-learning into an adaptive RRM process, enabling PeNBs to dynamically select optimal transmit-power levels based on real-time network states, interference levels, and QoS requirements. By jointly considering network capacity, interference reduction, transmit-power consumption, and QoS, the proposed hybrid Q-learning power allocation framework provides a robust, adaptive mechanism to significantly improve resource utilization and overall network performance in ultra-dense 5G and B5G Het-HetNets. Ultimately, this intelligent approach ensures sustainable and highly efficient connectivity for future wireless communication systems.

2. METHODOLOGY

A 3GPP-compliant MATLAB simulation models a multi-tier Het-HetNet to execute iterative Q-learning loops and collect performance metrics. Custom reinforcement learning scripts then analyze this data to evaluate the convergence and effectiveness of the proposed hybrid Q-learning power allocation framework.

2.1 Het-HetNet Architecture

The proposed system is based on a multi-tier 5G Het-HetNet architecture incorporating both LTE and 5G-NR technologies as depicted in Figure 1.

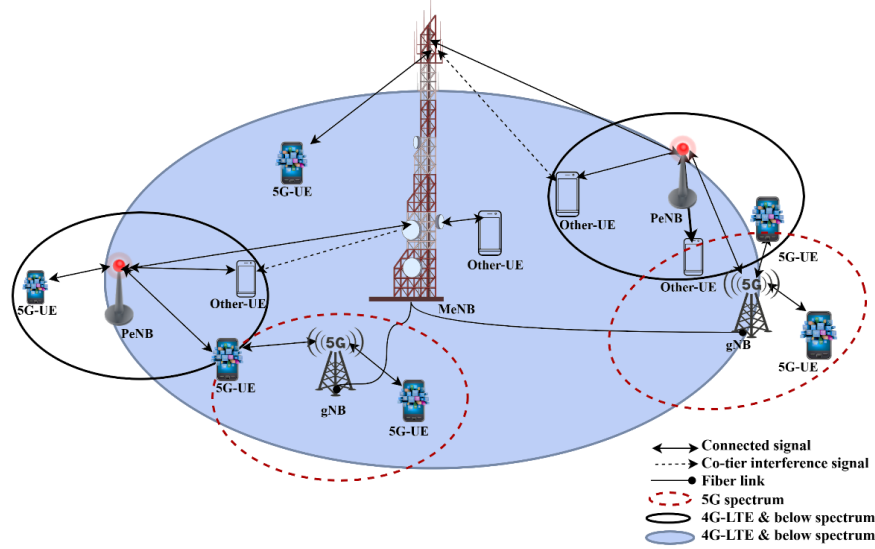


Figure 1: Proposed Multi-tier 5G Het-HetNet architecture

The network simulation includes the parameters listed in Table 1.

Table 1: Simulation parameters

Network Component	Quantity
Macro evolved NodeB (MeNB)	1
Pico evolved NodeBs (PeNBs)	16
5G Next Generation NodeB (gNB)	1
MeNB User Equipment	1
Other/small-cell UEs	16

The MeNB provides wide-area coverage and operates at 2 GHz, while the small-cell layer operates at 3.5 GHz. The PeNBs are uniformly deployed within the MeNB coverage region to improve localized capacity and offload traffic. The model assumes shared spectrum operation and universal frequency reuse. Perfect Channel State Information is assumed to be available, and users remain stationary during each Q-learning episode.

2.2 Communication Scenarios

Two interference scenarios are evaluated: (1) a high-interference scenario characterized by strong cross-tier and co-tier interference, and (2) a low-interference scenario with relatively reduced interference levels. In addition, a separate dense-deployment experiment is conducted to assess scalability under very high network density. No third or intermediate-interference scenario is evaluated or presented as a separate result.

2.3 Q-Learning Formulation

Each PeNB is treated as an autonomous reinforcement-learning agent. The agent observes the network environment, selects a transmit-power action, receives a reward, and updates its Q-value. The Q-learning model can be represented by:

$$\mathcal{M} = (S, A, R, Q) \quad (1)$$

where (S) is the state space, (A) is the action space, (R) is the reward function, and (Q) is the Q-value table.

A. STATE SPACE

The network state is represented using the prevailing interference level and QoS conditions of the served users:

$$S = (I, C_{5G-UE}, C_{other-UE}) \quad (2)$$

where I denotes the observed interference level, C_{5G-UE} denotes the QoS status of the 5G-UE, and $C_{other-UE}$ denotes the QoS status of other UEs. The 5G-UE QoS requirement is satisfied when normalized capacity reaches at least **1 bit/s/Hz**, while the corresponding requirement for other UEs is **0.5 bit/s/Hz**.

B. ACTION SPACE

The agent selects a discrete transmit-power level for the corresponding PeNB. The proposed framework uses **32 discrete power actions**, uniformly distributed between **-15 dBm and +15 dBm**, giving an approximate power step of **0.968 dBm**. Thus,

$$A = \{P_1, P_2, \dots, P_{32}\} \quad (3)$$

This discretization provides a practical balance between power-control granularity and Q-table complexity.

C. HYBRID REWARD FUNCTION

The proposed reward function combines four objectives:

$$r_h = w_1 C_{tot} - w_2 I - w_3 P_i + w_4 Q \quad (4)$$

where C_{tot} is aggregate normalized network capacity, I is aggregate interference, P_i is PeNB transmit power, and (Q) represents the QoS reward. The w_i ($i = 1, 2, 3, \dots$) parameters control the relative importance of the objectives. The reward weights used in the simulations are $w_1=0.50$, $w_2=0.30$, $w_3=0.20$, and $w_4=0.10$ for capacity, interference, and energy, respectively. QoS is incorporated separately through a ± 0.10 bonus/penalty rather than as $w_4 Q$. These weights are fixed empirically, while α , γ , and ε are optimized through sequential grid-search tuning. The reward therefore encourages the agent to maximize network capacity; reduce co-tier interference; reduce cross-tier interference; minimize unnecessary transmit power; maintain QoS; and improve overall network performance. The QoS component provides a positive reward when both the 5G-UE and other-UE requirements are satisfied and a penalty when either requirement is violated.

D. Q-VALUE UPDATE

The learning process is based on the Bellman principle. The standard Q-learning update is expressed as:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (5)$$

Where the learning rate α determines the extent to which newly received reward information modifies the current Q-value. In the proposed hybrid Q-learning power allocation model, α is explicitly incorporated into the Bellman Q-learning update, ensuring consistent learning and gradual adaptation of the power-allocation policy.

E. E-GREEDY POLICY

An ε -greedy policy is adopted to balance exploration and exploitation. The agent selects a random action during exploration and selects the action with the highest Q-value during exploitation. The study uses an exploration probability of approximately 0.1, corresponding to 10% exploration and 90% exploitation. The proposed hybrid Q-learning power allocation framework employs an ε -greedy action selection policy. Under this strategy, the agent selects a random action with probability ε and selects the action with the highest Q-value with probability $(1 - \varepsilon)$. The ε -greedy action selection policy is expressed as

$$a = \begin{cases} \text{Random Action}, & \text{if } rand < \varepsilon \\ \arg \max_{a \in A} Q(s, a), & \text{otherwise} \end{cases}$$

3. RESULTS AND DISCUSSION

In this section, we will provide a detailed presentation of the findings, using a series of visually engaging figures and an informative table that summarize the overall improvements observed. This display aims to clearly illustrate the progress made, allowing for a deeper understanding of the results.

3.1 Q-Learning Convergence

The learned Q-values demonstrate that the Q-learning agent successfully distinguishes between favourable and unfavourable transmit-power actions. High Q-values correspond to state-action combinations that produce improved network performance, while low Q-values correspond to less desirable power decisions.

The learned policy heatmap further illustrates that the algorithm does not rely on a single fixed transmit-power level across all network conditions. Instead, varying network states lead to distinct preferred power levels, highlighting the adaptive nature of the proposed approach.

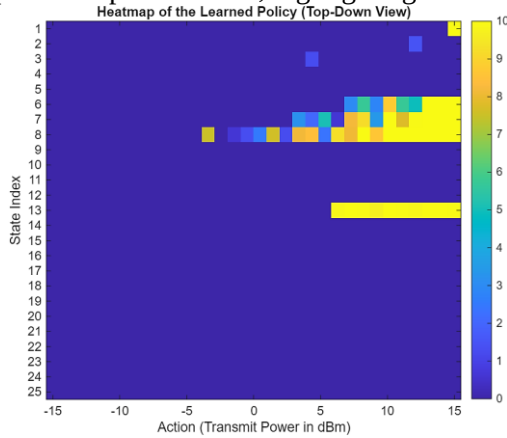


Figure 2: Heatmap of the transmit-power policy based on the trained Q-table.

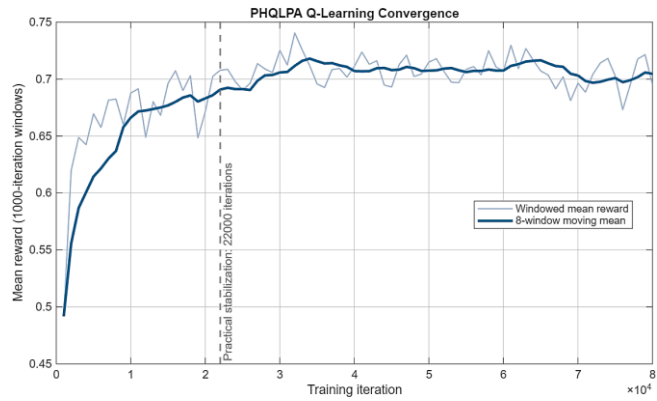


Figure 3: Q-learning Convergence

The convergence analysis reveals a swift increase in average reward during the initial learning phase, followed by stabilization with relatively minor fluctuations. The Q-learning process converges within 80,000 iterations, suggesting that the chosen learning configuration yields a stable policy for the simulated Het-HetNet environment.

3.2 5G-UE Capacity under High and Low Interference Conditions

Under Scenario 1, increasing the number of PeNBs increases the level of interference and consequently reduces user capacity for all evaluated approaches. However, the proposed hybrid Q-learning power allocation framework exhibits a more gradual reduction in capacity. The hybrid Q-learning power allocation approach maintains 5G-UE capacity above the specified minimum QoS threshold throughout the evaluated deployment range. By dynamically adjusting PeNB transmit power, the learning agent avoids unnecessarily high transmission power that would otherwise increase interference.

Compared with the benchmark approaches of Iqbal et al. (2021) and Amiri et al. (2019) the proposed framework demonstrates improved stability as network density increases. Under Scenario 2, the lower interference level allows all evaluated approaches to achieve higher user capacity. The proposed hybrid Q-learning power allocation framework consistently produces the highest capacity across the evaluated PeNB deployment range.

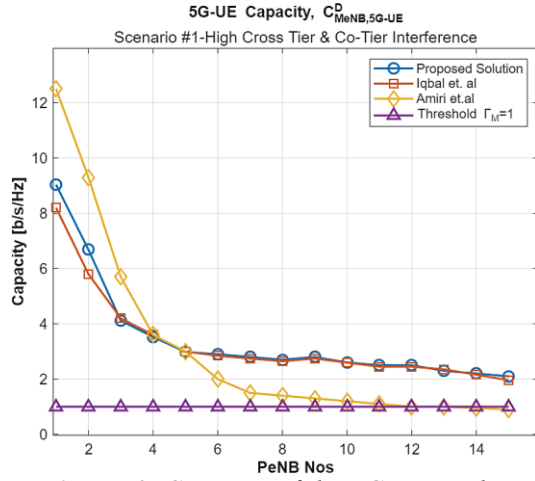


Figure 4. Capacity of the 5G-UE under Scenario 1

(High Cross-Tier and Co-Tier Interference).

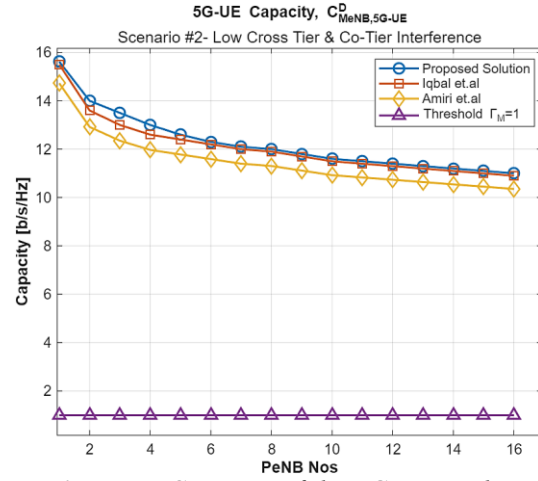


Figure 5. Capacity of the 5G-UE under Scenario 2

(Low Cross-Tier and Co-Tier Interference).

While capacity gradually decreases as PeNB density increases, the decline remains relatively moderate for hybrid Q-learning power allocation. This indicates that the adaptive power-control mechanism is able to compensate for additional interference generated by network densification. The results also show that all evaluated methods remain above the minimum QoS threshold under the relatively favourable interference conditions. Amiri et al. also provides a comparative benchmark for the evaluated scenarios. Under low-interference conditions, HQLPA maintains the highest capacity as network density increases, while Amiri et al. shows a declining trend. Under intermediate-interference conditions, Amiri et al. is initially competitive and slightly outperforms HQLPA, but its performance decreases with increasing PeNB density, whereas HQLPA maintains a more stable capacity trend.

3.3 Performance under Very High Network Density

The dense-deployment experiment evaluates the proposed method under very high co-tier and cross-tier interference. Increasing PeNB density naturally reduces capacity because of accumulated interference. But hybrid Q-learning power allocation maintains substantially higher capacity than the Iqbal et al. benchmark throughout the deployment range.

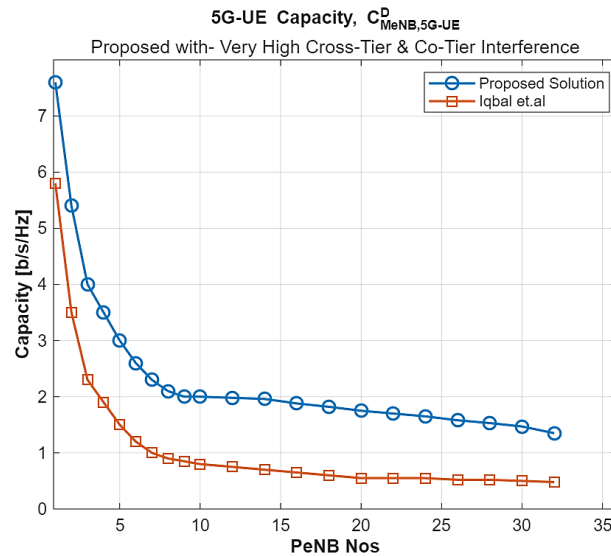


Figure 6: Capacity Performance of the Proposed Hybrid Q-Learning Power Allocation Framework Under Very High Cross-Tier and Co-Tier Interference Conditions.

The benchmark capacity declines more rapidly with increasing network density, whereas hybrid Q-learning power allocation exhibits relatively stable communication performance. This observation

indicates that adaptive power control can improve the scalability of interference management in ultra-dense Het-HetNets.

3.4 Quantitative Improvement

The improvement in the overall simulation results is shown in Table 2.

Table 2: *Summary of results improvement*

Performance Indicator	Reported Improvement
Average 5G-UE capacity	3.37%
Minimum capacity of small-cell users	14.81%
Q-learning convergence	80,000 iterations
5G-UE QoS	Maintained above minimum threshold
Adaptability	Improved under increasing interference
Scalability	Improved under dense deployment

The reported average 3.37% improvement in 5G-UE capacity and 14.81% improvement in minimum small-cell-user capacity demonstrate the advantage of adaptive power control over the benchmark approach. The larger improvement in minimum small-cell capacity is particularly important because it suggests that the proposed method does not merely maximize aggregate performance but also helps protect users experiencing relatively poor communication conditions.

4. CONCLUSION

This paper presents a Q-learning-based model for interference management within an adaptive RRM framework, focusing on dynamic transmit-power optimization for PeNBs in ultra-dense 5G and beyond-5G Het-HetNets. The hybrid Q-learning power allocation framework formulates transmit-power control as a reinforcement learning challenge, where each PeNB observes network conditions, selects a power level, receives a hybrid reward, and updates its Q-table using the Bellman learning mechanism. The model effectively incorporates factors such as network capacity, interference, power consumption, and QoS into the learning process. A multi-tier Het-HetNet, including one MeNB, one gNB, and sixteen PeNBs, was evaluated under various interference scenarios. The results indicate that the Q-learning agent successfully learns an adaptive transmit-power policy, achieving convergence within 80,000 iterations. Hybrid Q-learning power allocation consistently outperforms benchmark approaches, demonstrating higher and more stable capacity for 5G-UE, with improvements of 3.37% in overall 5G-UE capacity and 14.81% in minimum capacity for small-cell users. This Q-learning approach illustrates a robust solution for interference management and RRM in densely populated wireless networks, providing significant advantages over fixed and semi-static resource allocation techniques. Despite its promising performance, the proposed hybrid Q-learning power allocation model has some limitations. The simulations assume perfect channel state information (CSI) and stationary users, which may not fully represent practical 5G/B5G deployment conditions. Future work will therefore consider imperfect CSI, user mobility, and more dynamic network conditions to further validate the robustness and practical applicability of hybrid Q-learning power allocation.

5. REFERENCES

- Abbass, W., Hussain, R., Frnda, J., Abbas, N., Javed, M. A., & Malik, S. A. (2022). Resource Allocation in Spectrum Access System Using Multi-Objective Optimization Methods. *Sensors*, 22(4), 1–24. <https://doi.org/10.3390/s22041318>
- Amiri, R., Almasi, M. A., Andrews, J. G., & Mehrpouyan, H. (2019). Reinforcement learning for self-organization and power control of two-tier heterogeneous networks. *IEEE Transactions on Wireless Communications*, 18(8), 3933–3947. <https://doi.org/10.1109/TWC.2019.2919611>
- Iqbal, M. U., Ansari, E. A., & Akhtar, S. (2021). Interference mitigation in hetnets to improve the QoS

- using q-learning. *IEEE Access*, 9, 32405–32424. <https://doi.org/10.1109/ACCESS.2021.3060480>
- Iqbal, M. U., Ansari, E. A., Akhtar, S., Rafiq, M. N., Farooq-i-Azam, M., & Hassan, B. (2023). Hybrid Learning based Radio Resource Management in 5G Heterogeneous Networks. *Pakistan Journal of Engineering and Technology*, 6(1), 92–97. <https://doi.org/10.51846/vol6iss1pp92-97>
- Luong, N. C., Wang, P., Niyato, D., Liang, Y. C., Han, Z., & Hou, F. (2019). Applications of Economic and Pricing Models for Resource Management in 5G Wireless Networks: A Survey. *IEEE Communications Surveys and Tutorials*, 21(4), 3298–3339. <https://doi.org/10.1109/COMST.2018.2870996>
- Shabbir, A., Rizvi, S., Shirazi, M. F., Alam, M. M., & Su'ud, M. M. (2024). Maximizing energy efficiency in HetNets through centralized and distributed sleep strategies under QoS constraint. *Scientific Reports*, 14(1), 25839. <https://doi.org/10.1038/s41598-024-70714-x>
- Yadav, I., Chaporkar, P., Jha, P., & Karandikar, A. (2021). Spectrum Allocation in IAB Networks: A Hierarchical Auction-based Approach. *IEEE Vehicular Technology Conference, 2021-Septe.* <https://doi.org/10.1109/VTC2021-Fall52928.2021.9625522>