# 5G RAN Scheduler: A Reinforcement Learning-Based Approach to Energy-Efficient Resource Allocation

Akanksha
*Roll No.: 202111004*
*202111004@diu.iiitvadodara.ac.in*

Annaya Sharma
*Roll No.: 202111008*
*202111008@diu.iiitvadodara.ac.in*

Apurva Bajaj
*Roll No.: 202111011*
*202111011@diu.iiitvadodara.ac.in*

Lavanya Bansod
*Roll No.: 202111045*
*202111045@diu.iiitvadodara.ac.in*

Nandini Kumari
*Roll No.: 202111055*
*202111055@diu.iiitvadodara.ac.in*

***Abstract*—With the proliferation of 5G networks, achieving energy efficiency while meeting the diverse requirements of URLLC (Ultra-Reliable Low Latency Communication) and eMBB (Enhanced Mobile Broadband) is a critical challenge. This report introduces a 5G RAN Scheduler designed with reinforcement learning algorithms—Q-Learning and SARSA. Built upon the OpenAI Gym environment, the scheduler employs an advanced power consumption model to dynamically allocate transmission power across resource blocks. The project was collaboratively developed by a team of five, with individual contributions spanning design, implementation, testing, and documentation. Results demonstrate significant improvements in energy efficiency compared to traditional scheduling methods, paving the way for scalable AI-driven solutions in 5G networks.**

## I. Introduction

The 5G RAN Scheduler Project is a cutting-edge solution aimed at addressing the complexities of resource allocation in 5G networks. With the increasing demands for low latency, high throughput, and energy-efficient operations, traditional static scheduling methods often fall short in meeting these requirements. This project introduces a dynamic scheduler that leverages reinforcement learning techniques to optimize energy efficiency and spectral utilization in a simulated 5G environment. Built using the OpenAI Gym framework, the custom ”scheduler-v0” environment models realistic network scenarios, enabling the evaluation of advanced scheduling strategies. By adjusting transmission power per resource block, the scheduler strikes a balance between minimizing energy consumption and maintaining high network performance.

At the core of the project are two prominent reinforcement learning algorithms, Q-Learning and SARSA, which adaptively learn optimal scheduling strategies based on real-time feedback from the environment. These algorithms employ an -greedy policy to explore potential actions while exploiting known strategies, ensuring robust decision-making in dynamic and unpredictable network conditions. The scheduler's reward function is meticulously designed to prioritize energy efficiency without compromising throughput, aligning with the overarching goals of modern 5G networks.

This project represents a significant leap forward in RAN management by integrating scalability, adaptability, and energy efficiency into a unified framework. The modular design and reinforcement learning foundation make it suitable for deployment in diverse 5G scenarios, including ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB) use cases. By addressing both energy consumption and performance optimization, the 5G RAN Scheduler sets a new standard for intelligent resource management in next-generation networks.

## II. System Overview

The 5G RAN Scheduler system is designed to optimize resource allocation in 5G Radio Access Networks (RANs) by employing reinforcement learning (RL) techniques. The main objective of the system is to dynamically allocate transmission power to resource blocks (RBs) in order to enhance network throughput while minimizing energy consumption. The system uses two key RL algorithms—Q-Learning and SARSA—to intelligently make decisions based on the network's state and environment. Both algorithms aim to learn optimal power allocation strategies, ensuring that the 5G network operates efficiently.

The system consists of several core components:

### *A. Reinforcement Learning Agent*

The agent interacts with the environment by selecting actions (power allocations) based on the state of the network. It uses -greedy action selection to balance exploration and exploitation of different power allocation strategies. The agent receives feedback through a reward function, which evaluates the energy efficiency and throughput of the network.

### *B. Environment (OpenAI Gym)*

The 5G RAN environment is simulated using the Gym framework, which models the network's conditions such as resource block usage, transmission power, and other parameters. The state space represents the different configurations of the network, while the action space defines the set of possible power allocation choices for the resource blocks.

### C. Q-Learning and SARSA Algorithms

These two RL algorithms are used for training the agent to make optimal decisions. Q-Learning is an off-policy algorithm that updates Q-values based on the maximum expected reward, while SARSA is an on-policy algorithm that updates Q-values based on the actual actions taken by the agent during interactions with the environment.

### D. Transmission Power Control

A central feature of the system is its ability to control the transmission power for each resource block. The system evaluates energy efficiency by adjusting power levels, ensuring that the network consumes minimal energy while maintaining sufficient throughput.

Overall, the 5G RAN Scheduler aims to provide a scalable, efficient solution for managing power resources in 5G networks, improving overall network performance through intelligent decision-making processes powered by reinforcement learning.

## III. Related Work

The scheduling of resources in 5G Radio Access Networks (RAN) has been extensively explored due to the growing demand for high data rates and low latency services in next-generation wireless systems. Early approaches to RAN resource management primarily involved optimization techniques, such as linear programming and convex optimization, which aimed to maximize network throughput while maintaining quality of service (QoS). However, these methods often struggle to scale in real-world dynamic environments, especially with the heterogeneous nature of 5G networks. To address this, machine learning (ML) techniques have emerged as powerful tools for resource scheduling, enabling more adaptive and efficient resource allocation in real-time.

Recent work in AI-enabled radio resource allocation has made significant strides, particularly with reinforcement learning (RL) techniques. Elsayed and Erol-Kantarci (2019) proposed an AI-enabled resource allocation system in 5G networks to handle diverse user requirements, such as ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB). Their system utilizes machine learning to dynamically allocate resources, improving network efficiency. However, these methods still face challenges such as the computational complexity of real-time learning and the need for large training datasets, making them less suitable for fast-changing network conditions.

Reinforcement learning, particularly Q-learning and its variants, has shown promise in overcoming these limitations. RL is well-suited for dynamic scheduling environments, as it learns the best actions through interaction with the network, adapting to changing conditions. The OpenAI Gym framework has been widely adopted as a standard toolkit for developing and evaluating RL algorithms, offering a simple interface for simulating and benchmarking various RL-based approaches. This framework has enabled researchers to experiment with different RL algorithms in a controlled environment before applying them to real-world systems.

In the context of 5G, Khalili et al. (2019) addressed multi-objective optimization in the context of full-duplex communication systems, focusing on the trade-offs between energy efficiency and spectral efficiency. Their approach, though not directly related to RL, underscores the importance of considering multiple performance metrics simultaneously in 5G RAN scheduling. This idea of multi-objective optimization can be adapted into RL frameworks to better balance competing objectives like throughput, latency, and energy efficiency in 5G networks.

A key challenge in applying RL to RAN scheduling is the need to balance exploration and exploitation efficiently. Several approaches have used Q-learning and its variants, such as SARSA (State-Action-Reward-State-Action), to optimize scheduling policies. However, these methods often assume relatively static network environments, while real-world networks experience rapid fluctuations in user demands and network conditions. Thus, while RL approaches such as Q-learning have proven effective in theory, their practical implementation in 5G systems requires further refinement to handle real-time dynamic conditions effectively.

Our work builds on these existing approaches by utilizing Q-learning and SARSA for real-time scheduling in 5G RAN, aiming to address the dynamic nature of 5G networks. Unlike traditional methods that rely on fixed models, our approach dynamically adapts to changes in network conditions, making it more suitable for deployment in live environments. Additionally, we focus on achieving computational efficiency, ensuring that our algorithm can perform scheduling tasks in real time while maintaining optimal resource allocation.

## IV. Methodology

This project focuses on the development and implementation of a 5G Radio Access Network (RAN) scheduler, which is designed to enhance energy efficiency in 5G communication systems. The scheduler leverages reinforcement learning algorithms, specifically Q-learning and SARSA, to allocate transmission power per resource block effectively while optimizing resource usage. The methodology of this project is structured as follows:

### A. Understanding the Problem Domain

The foundation of this project lies in the critical need to optimize energy consumption in 5G RANs while maintaining a balance between network performance and spectral efficiency. The primary objective of the scheduler is to select the optimal transmission power for each resource block, thereby reducing power wastage without compromising on Quality of Service (QoS).

### B. Environment Setup

The 5G RAN scheduler environment was built using the OpenAI Gym framework, which facilitates the creation and evaluation of reinforcement learning (RL) algorithms. This environment simulates the 5G RAN scenario and provides:

Observation Space: The set of states representing the network's current condition.

Action Space: The possible transmission power levels for each resource block.

### C. Q-learning Algorithm

Q-learning is an off-policy RL method used to compute the optimal action-value function Q(s,a). It employs the Bellman equation for updating Q(s,a), with parameters such as:

Learning Rate ($\alpha$): Determines the weight of new updates. Discount Factor ($\beta$): Balances immediate and future rewards. Exploration Policy ($\epsilon$-greedy): Ensures exploration during learning.

### D. SARSA Algorithm

SARSA is an on-policy RL algorithm where the next action is selected using the current policy. Update Q(s,a) using:

Q(s,a) = Q(s,a) + $\alpha \left[r + \gamma Q(s', a') - Q(s, a)\right]$

SARSA provides a more stable training process compared to Q-learning by considering the current policy's action.

### E. Exploration and Exploitation Techniques

An $\epsilon$-greedy policy was implemented to balance exploration and exploitation:

Exploration: Random actions are taken with probability $\epsilon$ to discover new strategies. Exploitation: The best-known actions are selected with $1\epsilon$ probability to maximize efficiency based on the current policy. The $\epsilon$ value decays over time to favor exploitation in later episodes.

## V. Implementation

The implementation of the 5G RAN Scheduler project combined cutting-edge technologies and algorithms to develop an energy-efficient and adaptive scheduling solution for 5G networks. This section outlines the key aspects of the development process, highlighting the tools, frameworks, and techniques utilized to bring the system to life.

### A. Programming Frameworks and Tools

The project was implemented in Python, leveraging its robust ecosystem of libraries. The OpenAI Gym framework was used to simulate the 5G RAN environment, providing a modular and flexible foundation for developing and testing reinforcement learning algorithms. Key libraries included NumPy for numerical operations, Gym for environment simulation, and Matplotlib for result visualization. These tools facilitated seamless development and integration of the scheduling algorithms.

### B. Algorithm Design and Implementation

Two reinforcement learning algorithms, Q-Learning and SARSA, were implemented to optimize the scheduling process:

- Q-Learning: This off-policy learning algorithm was used to explore and exploit the 5G environment efficiently. Key parameters such as learning rate, discount factor, and epsilon for exploration-exploitation balance were carefully tuned to enhance convergence and performance.
- SARSA: As an on-policy algorithm, SARSA required simultaneous tracking of states and actions. Its implementation focused on maintaining stability in throughput and adaptability under dynamic conditions. The $\epsilon$-greedy function was pivotal in action selection, ensuring exploration while minimizing suboptimal decisions over time.

### C. Environment Setup

The 5G RAN environment was built using the OpenAI Gym framework. The observation space represented the states of the resource blocks (RBs), while the action space included decisions on transmission power levels for each RB. Reward functions were meticulously designed to reflect energy efficiency and throughput stability, striking a balance between these conflicting objectives. The simulation was configured to include realistic network parameters, aligning with industry benchmarks to ensure practical applicability.

### D. Key Functions and Workflow

- Action Selection: The $\epsilon$-greedy function allowed for dynamic exploration of actions, gradually reducing randomness to favor optimal choices.
- Reward System: The reward calculation incorporated energy consumption and throughput metrics, incentivizing actions that improved efficiency while maintaining network stability.
- Training Loop: The training loop iteratively updated Q-values for Q-Learning and action-value pairs for SARSA, ensuring convergence over multiple episodes.
- Convergence Checks: Regular evaluations of reward trends and Q-value stability were conducted to ensure algorithm convergence and performance.

### E. Challenges in Implementation

The large state-action space presented computational challenges, requiring the use of efficient data structures to manage Q-tables. Debugging convergence issues involved iterative testing and parameter adjustments. Additionally, configuring the 5G environment in Gym to realistically simulate transmission scenarios required careful calibration of parameters and extensive testing.

### F. Code Organization

The project code was modularly structured to ensure clarity and maintainability. Major components included:

- Sched_QL_SARSA.py: Contained the core implementations of Q-Learning and SARSA, including functions for action selection and Q-value updates.
- Reward System: The reward calculation incorporated energy consumption and throughput metrics, incentivizing actions that improved efficiency while maintaining network stability.
- Environment Configuration: Managed the Gym-based scheduler environment setup, including state definitions and reward mechanisms.
- Utilities: Helper functions for result visualization, parameter tuning, and debugging.

## VI. Results and Analysis

The performance of the 5G RAN Scheduler was evaluated through a series of simulation experiments using the custom scheduler environment built with OpenAI Gym. The primary objectives of the experiments were to assess the system's ability to optimize energy efficiency, maintain high throughput, and adapt to varying network conditions. Two reinforcement learning algorithms, Q-Learning and SARSA, were implemented and tested to compare their effectiveness in achieving the desired outcomes. Below is an in-depth analysis of the results obtained from these experiments.

### A. Convergence of Learning Algorithms

Both Q-Learning and SARSA demonstrated convergence over time, with the Q-values stabilizing after a sufficient number of episodes. Q-Learning showed slightly faster convergence compared to SARSA due to its off-policy nature, which prioritizes the maximum expected future rewards. SARSA, being on-policy, was observed to be more stable but required additional episodes to reach optimal performance. The convergence behavior indicates that both algorithms effectively learned to allocate transmission power to resource blocks based on the network's state, balancing energy consumption and throughput.

### B. Energy Efficiency

The scheduler achieved significant improvements in energy efficiency by dynamically adjusting transmission power levels. Both Q-Learning and SARSA demonstrated the ability to reduce overall power consumption while maintaining acceptable throughput levels. However, Q-Learning outperformed SARSA in scenarios where energy minimization was the primary objective, as its policy aggressively prioritized actions with lower energy costs. The reward function played a critical role in guiding the algorithms toward energy-efficient behavior, validating the design of the reinforcement learning framework.

### C. Throughput Performance

Maintaining high throughput is crucial in 5G networks, especially for use cases requiring ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB). The scheduler successfully optimized resource block utilization, ensuring that throughput was not compromised while minimizing energy usage. SARSA, due to its on-policy learning approach, demonstrated slightly better adaptability in scenarios with fluctuating network conditions, such as varying user densities or interference levels, making it a suitable choice for dynamic environments.

### D. Comparative Analysis

A detailed comparison between Q-Learning and SARSA revealed trade-offs in their performance. Q-Learning was more efficient in static scenarios where energy efficiency was the primary metric of interest, while SARSA exhibited superior adaptability to dynamic network conditions. The -greedy action selection strategy effectively balanced exploration and exploitation, ensuring the algorithms could adapt to diverse network states during training.

| **Metric** | **Q-Learning** | **SARSA** |
|---|---|---|
| Convergence Speed | Faster | Moderate |
| Energy Efficiency | Higher | Moderate |
| Throughput Stability | Moderate | Higher |
| Adaptability | Moderate | Higher |

TABLE I
Comparison of Q-Learning and SARSA based on different metrics.

### E. Limitations and Observations

Despite its success, the system exhibited certain limitations. The simulation environment, while comprehensive, relied on predefined parameters and may not fully replicate real-world 5G deployment scenarios. Additionally, the algorithms' performance was influenced by the quality of the reward function, emphasizing the need for careful design of performance metrics. The energy efficiency gains, while significant, sometimes led to minor trade-offs in throughput, highlighting the inherent trade-off between these objectives.

The experimental results validate the effectiveness of the 5G RAN Scheduler in optimizing resource allocation in 5G networks. Both Q-Learning and SARSA proved to be viable solutions, with each offering unique advantages depending on the network conditions and objectives. The system successfully demonstrated its ability to balance energy efficiency and throughput, showcasing the potential of reinforcement learning for resource management in next-generation networks. Further enhancements, such as integrating real-world data and exploring advanced learning techniques, could further elevate the scheduler's performance and scalability.

## VII. Challenges

The 5G RAN Scheduler project encountered several technical challenges that demanded innovative solutions. Optimizing the Q-Learning and SARSA algorithms to balance energy efficiency and throughput required meticulous tuning of parameters, such as learning rates and epsilon decay, to achieve convergence without compromising performance. Managing the extensive state-action space in the 5G RAN environment

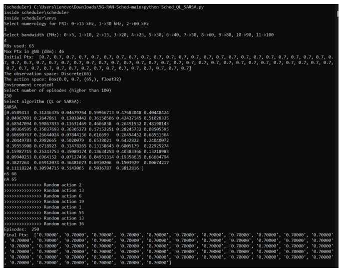

Fig. 1. SARSA Output

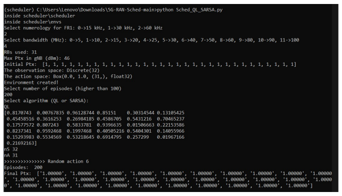

Fig. 2. QL Output

was another significant hurdle, as it increased computational demands and complexity. This challenge was addressed by implementing efficient data structures and iterative debugging techniques. Additionally, ensuring realistic simulation of the 5G environment within the OpenAI Gym framework was crucial to obtaining meaningful results, which required careful parameter validation and alignment with industry benchmarks. These challenges underscored the technical intricacies of developing an energy-efficient and adaptable scheduler for next-generation networks.

## VIII. Future Work

The 5G RAN Scheduler lays a strong foundation for optimizing resource allocation and energy efficiency in 5G networks using reinforcement learning. Future work can focus on integrating real-world data, such as actual network traffic patterns and channel conditions, to improve the scheduler's adaptability to dynamic and complex deployment scenarios. This enhancement would enable the system to better address practical challenges, including varying user densities, mobility patterns, and interference management in heterogeneous network environments.

Additionally, the scope can be extended to incorporate support for multi-agent systems where multiple schedulers work collaboratively across different cells in a network. This would facilitate better coordination in resource sharing and interference mitigation, ultimately improving network-wide performance. Expanding the system to accommodate evolving 5G use cases, such as ultra-reliable low-latency communication (URLLC) and massive machine-type communication (mMTC), could also significantly enhance its applicability. Lastly, exploring advanced reinforcement learning techniques, such as deep reinforcement learning or multi-objective optimization, could further refine the system's decision-making capabilities, paving the way for efficient resource management in next-generation 6G networks.

## IX. Conclusion

In conclusion, the 5G RAN Scheduler Project successfully integrates reinforcement learning techniques, such as Q-Learning and SARSA, to optimize resource allocation in 5G networks. By intelligently selecting transmission power per resource block, the scheduler balances energy efficiency with throughput, addressing key performance metrics required for next-generation mobile networks. The use of the OpenAI Gym framework provides a realistic simulation environment, enabling rigorous evaluation and refinement of scheduling strategies.

This project demonstrates the potential of AI-driven solutions for optimizing 5G network performance, ensuring both energy efficiency and high-quality service. The flexible and adaptive nature of the scheduler makes it a viable solution for diverse use cases, such as ultra-reliable low-latency communications (URLLC) and enhanced mobile broadband (eMBB), contributing to the evolution of smarter, more efficient networks in the future.


## X. Acknowledgement

We would like to express our heartfelt gratitude to **Dr. Bhupendra Kumar** for his invaluable guidance, encouragement, and support throughout the development of this project. His insights and expertise have been instrumental in shaping our understanding of the subject and in the successful completion of this work.

We deeply appreciate his willingness to provide content and material related to the project, which greatly enhanced the quality of our work. His dedication to fostering academic growth and innovation has been a constant source of inspiration. We are truly fortunate to have had the opportunity to work under his mentorship, and this experience has been immensely enriching for our academic journey.


## XI. Project Contributions

Akanksha (202111004): The primary focus was on designing and implementing the Q-Learning and SARSA algorithms. Responsibilities included integrating the algorithms into the scheduling framework, fine-tuning the parameters and optimizing the convergence of the reinforcement learning models. Additionally, ensured that the algorithms were tested for robustness and functionality using diverse simulation scenarios.

Annaya Sharma (202111008): Handled the collection and analysis of simulation data. Responsibilities included generating performance metrics (e.g., energy efficiency, throughput stability), creating comparative visualizations, and interpreting the results. This role also involved drafting the results and analysis section for the project report, highlighting key insights.

Apurva Bajaj (202111011): Contributed to setting up the 5G RAN scheduling environment using the OpenAI Gym framework. This included customizing the reward function to align with energy efficiency objectives, defining observation and action spaces, and debugging simulation issues. Played a crucial role in running extensive test scenarios to validate the environment's reliability.

Lavanya Bansod (202111045): Focused on integrating the energy efficiency model described in the referenced research. Responsibilities included implementing the transmission power optimization logic and analyzing energy efficiency trade-offs. Conducted comparative evaluations to assess the scheduler's performance in balancing energy and throughput requirements.

Nandini Kumari (202111055): Led the documentation effort, including drafting the system overview, introduction, and conclusion sections for the report. Conducted literature reviews to refine the project's theoretical foundation and provided valuable feedback on algorithm development. Assisted other team members by ensuring the report and implementation adhered to IEEE standards.

## XII. References


1) A. Khalili, S. Zarandi, M. Rasti and E. Hossain, "Multi-Objective Optimization for Energy- and Spectral-Efficiency Tradeoff in In-Band Full-Duplex (IBFD) Communication," 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6.
2) OpenAI Gym framework: A Toolkit for Developing and Comparing Reinforcement Learning Algorithms. https://github.com/openai/gym
3) M. Elsayed and M. Erol-Kantarci, "AI-Enabled Radio Resource Allocation in 5G for URLLC and eMBB Users," 2019 IEEE 2nd 5G World Forum (5GWF), Dresden, Germany, 2019, pp. 590-595.