# Throughput Optimization via RL-Driven AMC with Dynamic Resource Scheduling in LTE


M M Sadman Shafi
*Department of Electrical and Electronic Engineering*
*Islamic University of Technology*
Gazipur, Bangladesh
sadmanshafi@iut-dhaka.edu

Tasnia Siddiqua Ahona
*Department of Electrical and Electronic Engineering*
*Islamic University of Technology*
Gazipur, Bangladesh
tasniaahona@iut-dhaka.edu

Jabir Ahmad Jahin Bhuiyan
*Department of Electrical and Electronic Engineering*
*Islamic University of Technology*
Gazipur, Bangladesh
jabirahmad@iut-dhaka.edu



***Abstract*— Adaptive Modulation and Coding Schemes (AMCS) are essential for maximizing spectral efficiency and maintaining link reliability in modern wireless communication systems such as Long-Term Evolution (LTE). However, conventional threshold-based AMCS approaches rely on static Signal-to-Interference-plus-Noise Ratio (SINR) or Channel Quality Indicator (CQI) mappings, which fail to adapt effectively under rapidly varying channel conditions. To address this limitation, this paper proposes a Reinforcement Learning (RL)-driven AMCS framework integrated with a dynamic proportional fair (PF) scheduling mechanism to jointly optimize Modulation and Coding Scheme (MCS) selection and resource block allocation. A Q-learning-based RL agent is designed to learn optimal scheduling and link adaptation policies directly from environmental feedback by leveraging instantaneous CQI values and their temporal variations, enabling improved decision-making in both stable and highly dynamic mobility conditions. The simulation framework models a single-cell LTE downlink under realistic conditions, incorporating COST-231-Hata path loss, Rician fading, and two distinct mobility scenarios: radial outward and random waypoint movement. At each time step, the RL agent selects the optimal MCS and dynamically allocates resource blocks based on instantaneous SINR, transport block size (TBS), and proportional fairness criteria. Performance evaluation demonstrates that the proposed RL-enhanced AMCS consistently achieves higher throughput compared to the baseline static AMC model across both mobility patterns, showing significant gains in adapting to channel fluctuations while maintaining efficient spectrum utilization. This work establishes a unified framework where RL-driven AMCS and dynamic scheduling operate synergistically, enabling scalable and intelligent resource allocation strategies for LTE and future wireless networks.***



***Keywords*— Adaptive Modulation and Coding, Reinforcement Learning, Q-learning, Proportional Fair Scheduling, LTE, Dynamic Resource Allocation, Link Adaptation**


## I. Introduction

Adaptive Modulation and Coding (AMC) is a fundamental technique in modern wireless communication systems, particularly in Long-Term Evolution (LTE) networks, where it dynamically adjusts the Modulation and Coding Scheme (MCS) based on instantaneous channel quality conditions. By exploiting favorable channel states with higher-order modulation and stronger coding rates while switching to more robust schemes under poor conditions, AMC ensures an optimal trade-off between spectral efficiency and link reliability. This adaptability is crucial for maximizing system throughput and maintaining quality of service (QoS) in rapidly varying wireless environments where user mobility, interference, and fading significantly influence link performance.

Recent research has increasingly explored the integration of Reinforcement Learning (RL) into AMC frameworks to overcome the limitations of conventional threshold-based MCS selection mechanisms. Traditional AMC relies on predefined Signal-to-Interference-plus-Noise Ratio (SINR) or Channel Quality Indicator (CQI) thresholds, which often fail to capture temporal channel dynamics and cannot adapt efficiently to diverse propagation scenarios. RL, on the other hand, enables agents to learn optimal modulation and coding strategies directly from interaction with the environment, balancing throughput maximization with reliability. Techniques such as Q-learning and Deep Q-Networks (DQN) have demonstrated significant improvements by predicting channel trends and selecting MCS levels that enhance performance under both gradual and abrupt SINR variations.

However, AMC alone is insufficient for achieving end-to-end optimization in LTE systems, where dynamic resource scheduling plays an equally critical role. Since MCS selection directly impacts the Transport Block Size (TBS) and Resource Block (RB) allocation, a tightly integrated AMC-scheduling framework can better exploit channel-aware adaptations. RL-based AMC provides the scheduler with intelligent feedback on link conditions, enabling more efficient spectrum utilization, minimizing retransmissions, and improving user fairness under highly dynamic traffic loads. Such synergy between RL-driven AMC and dynamic scheduling forms the cornerstone for achieving high spectral efficiency, low latency, and robust quality-of-service guarantees in next-generation LTE and beyond networks.

In this paper, we present an RL-enhanced AMC framework that establishes a unified perspective where intelligent link adaptation and resource scheduling converge, paving the way for scalable and efficient wireless systems capable of meeting the growing demands of future networks.

## II. Related Works

AMC is a critical technology for optimizing performance in modern wireless networks like LTE and OFDM. A review of recent literature reveals a strong trend towards applying RL to automate MCS selection, yet many of these efforts are characterized by their incremental contributions. For instance, several RL-based schemes for LTE aim to adapt MCS using past throughput to overcome CQI limitations, but their novelty is considered modest given existing Q-learning approaches [1]. This pattern is mirrored in OFDM systems, where RL is used for online MCS adaptation, but the contribution is limited as it largely reuses established concepts to replace conventional lookup tables [2], [3]. General Q-learning frameworks that map CQI to MCS also primarily serve to replace fixed tables with standard RL adaptation, marking an incremental step [4]. Even when more advanced Deep Reinforcement Learning (DRL) is applied to MCS selection in HetNets [5] or for video rate allocation[6], the novelty is often constrained by the adaptation of standard DRL frameworks to well-studied optimization problems.

This trend of limited novelty is not exclusive to RL-based AMC. A significant body of work focuses on performance analysis and minor refinements of conventional systems. Studies analyzing AMC performance for OFDM under various channel models frequently reproduce standard BER–SNR trade-off analyses without introducing novel methods [7, 8]. Similarly, other analyses largely confirm well-known results comparing fixed and adaptive schemes across different channels [9] or offer incremental improvements through parameter tuning rather than new techniques [10]. Research into coding and resource allocation often follows a similar path, presenting basic comparative evaluations of coding schemes [11], reiterating known throughput reliability compromises [12], or repackaging standard error-control techniques for marginal gains [13]. Furthermore, approaches to resource allocation may rely on heuristics that trade optimality for simplicity [14], while other adaptations refine existing threshold-based methods [15]or apply standard machine learning techniques with modest impact [16]. An RL-based scheduler for IoT traffic also highlights this theme, showing only incremental gains over existing heuristics [17].

In synthesis, the literature points to two clear trends: the application of RL to AMC often results in modest improvements, and non-RL approaches frequently yield only incremental refinements. While RL is applied to scheduling [17] and resource allocation [14], the novelty is often limited. This reveals a distinct gap for research that moves beyond simply replacing components and instead investigates the deeper integration of RL-based AMC within a dynamic scheduling framework to unlock more substantial performance gains.

## III. METHODOLOGY

This section describes the simulation setup, channel parameter configuration, mobility models, path loss model, dynamic resource scheduling and link adaptation, Block Error Rate (BLER) calculation and the RL-based AMC framework used in this study. Figure 1 illustrates the adopted system architecture for the study.

### *A. Simulation Setup*

The simulation framework evaluates the performance of AMC in a single-cell network scenario with one active user. At each time step, the signal-to-interference-plus-noise ratio (SINR) is calculated considering realistic propagation conditions, followed by the dynamic selection of modulation and coding scheme (MCS), transport block size (TBS), and resource block (RB) allocation. The overall setup can be summarized as follows:

**(i) System Initialization:**
The simulation environment is designed to reflect a typical urban cellular scenario. A base station (BS) height of 30 meters and a cell radius of 500 meters are chosen to model realistic urban coverage, while a carrier frequency of 2.1 GHz represents a common LTE band. The system is simulated over 1000 time steps, providing sufficient temporal resolution to capture the evolution of channel conditions and link adaptation behavior.

A transmit power of 43 dBm ensures sufficient coverage across the cell, while an antenna gain of 10 dBi models a directional base station antenna enhancing effective radiated power. The thermal noise power spectral density is set to 174 dBm/Hz, which is standard for communication systems. These parameters provide a realistic baseline for evaluating the impact of propagation effects and channel variability on link adaptation performance. Table I summarizes the key simulation parameters.

Table I: General Simulation Parameters

| Parameter | Value |
|---|---|
| Simulation time | 1000 steps |
| Carrier frequency | 2.1 GHz |
| Bandwidth | 10 MHz |
| BS height | 30 m |
| Cell radius | 500 m |
| Noise PSD | –174 dBm/Hz |
| Transmit power | 43 dBm |
| Antenna gain | 10 dBi |

**(ii) Channel Parameter Configuration:**
The simulation incorporates both large-scale and small-scale channel effects to create a realistic propagation environment for evaluating adaptive modulation and coding schemes. Large-scale fading is modeled using log-normal shadowing, which accounts for signal attenuation caused by obstacles such as buildings and foliage. The shadowing values are generated with a standard deviation of 4 dB, reflecting moderate variability typical in urban areas, and are spatially correlated over a distance of 100 meters to mimic the fact that closely spaced locations experience similar attenuation. To prevent unrealistic extremes, the shadowing values are constrained within ±8 dB.

Small-scale fading is captured using Rician fading, which is suitable for scenarios where a line-of-sight (LOS) path coexists with multiple scattered components. A concise description of Rician fading is provided below.

*Rician Fading:*
Rician fading models the small-scale variations of the received signal in environments where a strong LOS component exists alongside multiple scattered paths. The relative strength of the LOS path compared to the scattered multipath components is quantified by the K-factor, defined as the ratio of LOS power to the power of the scattered components. This fading model is particularly relevant for urban macro- and microcell deployments, where the UE may maintain partial visibility of the base station, and it allows realistic simulation of rapid signal amplitude variations superimposed on large-scale path loss and shadowing effects.

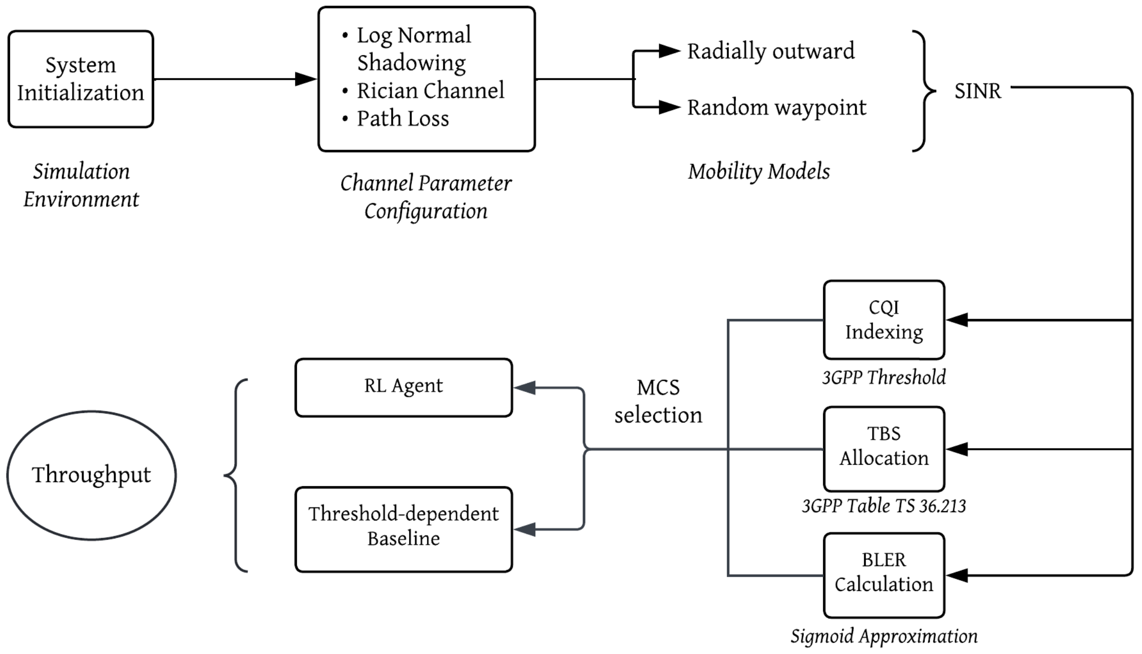


Figure 1: System Architecture.

In this study, the K-factor is set to 6 dB, indicating that the LOS path is dominant but multipath components still contribute to signal variations. The Doppler frequency is calculated based on the UE's instantaneous speed, introducing realistic temporal fluctuations due to mobility. A Rician channel object is instantiated with controlled parameters, including path delays, average path gains, K-factor, and maximum Doppler shift, ensuring that the small-scale fading is constrained and reproducible. This methodology allows the simulation to capture realistic channel dynamics while maintaining control over extreme variations, providing a reliable environment to assess the performance of adaptive modulation, coding, and resource allocation strategies. Table II summarizes the main channel parameters used in this study, including shadowing, fading, Doppler, and path loss characteristics.

Table II: Channel Model Parameters

| Parameter | Value |
|---|---|
| Channel type | Rician fading |
| K-factor | 6 dB |
| Doppler frequency | Based on radial speed |
| Shadowing | σ = 4 dB (bounded ±8 dB) |
| Correlation distance | 100 m |

**(iii) Path Loss, Received Power and Fading Application:** At each time step, the UE-to-BS distance is computed from the UE's current position, and the COST-231-Hata urban path loss model is applied to determine the large-scale attenuation. A brief discussion of COST-231-Hata path loss model is given below.

*COST-231-Hata Path Loss Model:*

The COST-231-Hata model is widely used to estimate large-scale path loss in urban cellular environments, particularly at frequencies between 1.5 and 2 GHz. It considers key parameters such as base station height, mobile station height, and the distance between the transmitter and receiver to provide realistic signal attenuation estimates over a wide coverage area. When combined with shadowing and small-scale fading, the COST-231-Hata model allows the simulation to capture both the predictable large-scale variations and the stochastic fluctuations in received signal strength, providing a robust framework for evaluating adaptive modulation, coding, and resource allocation strategies.

The path loss in dB is calculated as

$$L = 46.3 + 33.9log_{10}(f_c) - 13.82log_{10}(H_{BS}) + (44.9 - 6.55log_{10}(H_{BS}))log_{10}(d_{km}) + 3 \quad (1)$$

where $f_c$ is the carrier frequency in MHz, $H_{BS}$ is the base station height, and $d_{km}$ is the UE-BS distance in kilometers.

In this study, shadowing is added as a correlated log-normal random variable bounded within ±8 dB to avoid extreme fluctuations. To avoid unrealistically low path loss values when the UE is very close to the base station, a minimum path loss of 80 dB is enforced (i.e., L = max (L, 80)).

The received power in dBm is then computed by combining the transmit power, antenna gain, path loss, and shadowing:

$$P_{rx} = P_{tx} + G_{ant} - L - X_{\sigma} \quad (2)$$

the terms mean the following:
$P_{rx}$ = Received power at the user equipment (UE) in dBm;
$P_{tx}$ = Transmit power of the base station in dBm;
$G_{ant}$ = Antenna gain of the transmitter (base station) in dBi;
and $X_{\sigma}$ represents the shadowing term.

The received power is converted to linear scale for further processing. Small-scale fading is introduced using the Rician fading channel with a K-factor of 6 dB. To ensure realistic but stable variations, the fading amplitude is constrained between 0.5 and 1.5. This combination of path loss, shadowing, and controlled fading allows the simulation to capture both large-scale and small-scale propagation effects in a reproducible manner.

**(iv) SINR Calculation:**

The instantaneous SINR at each time step is computed by taking the ratio of the received signal power after fading to the thermal noise power. In linear scale, the received power after applying the fading amplitude is

$$P_{rx,lin}^{faded} = (fading)^2 \times P_{rx,lin} \tag{3}$$

and the noise power is calculated from the thermal noise spectral density $N_0 = -174$ dBm/Hz. The SINR in dB is then obtained as

$$SINR = 10log_{10}(\frac{P_{rx,lin}^{faded}}{P_{noise}}) \tag{4}$$

where $P_{noise} = 10^{(N0/10)}$. By bounding the fading and carefully combining large-scale and small-scale effects, this methodology ensures that the SINR variations are realistic, stable, and suitable for evaluating link adaptation performance in adaptive modulation and coding schemes.

### *B. Mobility Models*

To capture different types of mobility behavior, two models were implemented in the simulation framework. These models provide complementary perspectives: one representing a controlled, predictable motion, and the other emulating more realistic, random UE movement patterns in a cell.

**(i) Radial Outward Mobility:**

In this model, the UE moves directly away from the base station (BS) along a straight radial path. This mobility pattern produces a smooth and predictable relationship between distance, path loss, and SINR, resulting in a continuous decline in signal quality over time. Such a controlled setup is useful for analyzing baseline performance, as it clearly shows the gradual adaptation of link in response to steadily worsening channel conditions. It also serves as a reference case for comparing performance under more dynamic and random mobility scenarios.

**(ii) Random Waypoint Mobility:**

The Random Waypoint (RWP) model is one of the most widely used mobility models in wireless network simulations because it captures a balance between randomness and realism. In this model, a set of waypoints is randomly distributed within the coverage area, and the UE moves sequentially from one waypoint to another. After reaching a waypoint, the UE may pause for a defined duration before selecting the next target location. This process repeats until the simulation ends, producing a trajectory that consists of straight-line movements interspersed with stationary periods. In our implementation, the UE moves at a constant speed between waypoints, with short pauses at each waypoint to emulate real-world user behavior such as stopping or slowing down. This results in a mobility pattern that includes direction changes, variable distances from the BS, and fluctuating channel conditions. Compared to radial mobility, this model introduces a higher degree of variability in SINR, as both path loss and fading can change abruptly due to direction shifts and pauses. This makes it well-suited for analyzing AMC performance under realistic, non-uniform mobility scenarios where link quality can vary unpredictably over time.

### *C. Dynamic Scheduling & Link Adaptation:*

Dynamic proportional fair scheduling is a resource allocation strategy in which a scheduler in the base station continuously assigns modulation and coding schemes (MCS), resource blocks (RBs) and transport block size (TBS) to users based on their instantaneous channel conditions.

After computing the instantaneous SINR at each time step, the simulation employs a dynamic link adaptation and resource allocation framework to optimize downlink transmission parameters in real time. The SINR is first translated into a Channel Quality Indicator (CQI) using a set of predefined SINR thresholds, which reflects the link's current quality and informs the appropriate modulation and coding scheme (MCS). Each CQI is then mapped to a specific MCS according to standardized 3GPP tables, ensuring compatibility with practical LTE systems.

Resource block (RB) allocation is performed using an ultra-granular SINR-to-RB mapping, where low SINR results in fewer RBs to maintain reliability, and high SINR allows more RBs to maximize throughput. The transport block size (TBS) is determined by indexing a pre-defined TBS table, which follows the specifications of 3GPP TS 36.213, version 14.2.0 Release 14 (Table 7.1.7.2.1-1), covering MCS indices 0–14 across 110 RBs. This table provides the exact number of bits that can be transmitted for any combination of MCS and RB allocation. It includes the mapping of MCS index and number of allocated RBs to the corresponding transport block size, reflecting the achievable data payload for each configuration.

By integrating SINR-driven MCS selection with fine-grained RB allocation and using a standardized TBS table, the methodology dynamically adapts both coding rate and spectral resource usage. This approach closely emulates real-world adaptive modulation and coding mechanisms.

### *D. Block Error Rate (BLER) Calculation:*

The simulation estimates the Block Error Rate (BLER) as a function of the instantaneous Signal-to-Interference-plus-Noise Ratio (SINR) and the selected Modulation and Coding Scheme (MCS). BLER represents the probability that a transport block is received with errors and requires retransmission, making it a key metric for evaluating link reliability and adaptive modulation performance.

A sigmoid-based model is employed to approximate BLER behavior for each MCS index. This model captures the sharp transition in error probability that occurs as SINR decreases, which is characteristic of practical wireless communication systems. The BLER for a given MCS and SINR is defined as:

$$BLER(\gamma, MCS) = \frac{1}{1+exp\,[-steepness(-\gamma_{th}-\gamma)]} \tag{5}$$

where $\gamma$ is the instantaneous SINR in dB, $\gamma_{th}$ is the MCS-specific SINR threshold at which BLER is approximately 0.5, and *steepness* is the parameter controlling the sharpness of the transition. The steepness determines how quickly the BLER changes from low to high values around the threshold; a larger steepness value corresponds to a more abrupt transition. Lower-order modulation schemes, such as QPSK, have lower thresholds and gentler slopes, while higher-order schemes like 256-QAM require higher SINR and exhibit steeper transitions. This increase in steepness with higher MCS occurs because higher-order modulations are more sensitive to noise and interference, so a small change in SINR near the threshold causes a rapid change in BLER.

The thresholds and steepness values are pre-determined for each MCS index, reflecting realistic link-level performance under varying SINR conditions. This approach provides a computationally efficient means to estimate BLER.

By combining the instantaneous SINR with the MCS-specific parameters, the model produces BLER values bounded between 0 and 1, ensuring physically meaningful probabilities.

### *E. RL-Based AMC Framework:*

**Background on Reinforcement Learning:**

Reinforcement Learning (RL) is a machine learning paradigm that enables an agent to autonomously learn an optimal policy for interacting with an environment through trial-and-error, with the objective of maximizing cumulative reward. The environment is typically modeled as a Markov Decision Process (MDP), formally defined by a finite set of states $S=\{s_1, s_2,\dots, s_n\}$, a finite set of admissible actions $A(t)=\{a_1(t), a_2(t),\dots, a_m(t)\}$, available at time t, a state transition probability function $P(s, a, s')$ describing the probability of reaching state $s' \epsilon S$ when taking action $a \epsilon A$ in state $s' \epsilon S$ and a reward function $R(s,a)$ mapping each state–action pair to a scalar reward.

The goal of an RL agent is to learn a policy π, i.e., a mapping from states to actions, that maximizes the expected long-term cumulative reward. For a given policy π, the value of being in state $s$ is quantified by the value function:

$$V^{\pi}(s) = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{6}$$

where $r_{t+k}$ is the reward obtained k-steps into the future, and $\gamma \epsilon [0,1]$ is the discount factor controlling the importance of future rewards relative to immediate rewards. The optimal policy $\pi^*$ is the one that maximizes $V^{\pi}(s)$ for all $s' \epsilon S$.

When the transition dynamics $P(s, a, s')$ are unknown, model-free RL techniques are employed to approximate the optimal policy through interaction with the environment. One such widely adopted technique is Q-learning, which learns the state-action value function (Q-function) $Q(s,a)$ representing the expected return from taking action $a$ in state $s$ and thereafter following the optimal policy. Q-learning updates the Q-values iteratively according to:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r(s,a) + \gamma \, max_{a'} Q(s',a') - Q(s,a) \tag{7}$$

where $\alpha \epsilon [0,1]$ is the learning rate, controlling the weight assigned to new information relative to past estimates. Over repeated interactions, Q-learning is guaranteed to converge to the optimal Q-function $Q^*(s, a)$, provided sufficient exploration of the state–action space. The trade-off between exploration (trying new actions to discover potentially better rewards) and exploitation (selecting the currently best-known action) is central to RL design, and exploration strategies such as ϵ-greedy are often employed to balance this dilemma.

**RL-Based AMC Architecture Used in This Study:**

The AMC framework we have used leverages Q-learning to dynamically select the optimal Modulation and Coding Scheme (MCS) under varying channel conditions. The state representation $s=\{CQI, \Delta CQI\}$ captures both the instantaneous Channel Quality Indicator (CQI) value, which is obtained from a pre-defined SINR-to-CQI mapping, and the temporal evolution of the channel quality through ΔCQI, defined as the difference between consecutive CQI estimates. This compact state representation enables efficient learning while reducing the dimensionality of the state space compared to SINR discretization.

The action space consists of a finite set of integer-valued correction factors $a \epsilon \{-k, \dots, -1, 0, 1, \dots, k\}$ applied to the CQI estimate to determine the final MCS selection. Positive actions correspond to more aggressive MCS selections (higher spectral efficiency), whereas negative actions yield more conservative choices (lower MCS index) to maintain robustness in deteriorating channel conditions. This formulation allows the agent to compensate for CQI reporting delay and imperfect SINR-to-MCS mapping.

At each time step, the RL agent observes the current state $s_t$, selects an action $a_t$ using an ϵ-greedy policy and applies the resulting MCS configuration to the downlink transmission. The reward is defined as the instantaneous link throughput:

$$R(s_t, a_t) = (1 - BLER) \cdot TBS \tag{8}$$

where TBS is the Transport Block Size determined by the selected MCS and number of RBs, and BLER is the estimated Block Error Rate for the given channel conditions. A transmission failure yields a zero reward, thereby incentivizing the agent to avoid overly aggressive MCS levels that lead to packet errors.

The Q-table is updated online using the observed reward and the maximum expected future Q-value from the next state. Over multiple learning episodes, the agent converges toward a policy that maximizes long-term throughput by balancing reliability and spectral efficiency. To ensure sufficient exploration during the early learning phase, the exploration rate ϵ is gradually decayed according to:

$$\epsilon \leftarrow \epsilon \cdot \epsilon_{decay} \tag{9}$$

encouraging more exploitation of the learned policy as training progresses.

Finally, we benchmark the RL-based AMC against a baseline link adaptation scheme. All simulation parameters, channel models, and mathematical formulations remain identical for both the baseline AMC scheme and the proposed RL-based framework; the key distinction lies in the MCS selection strategy, where the baseline employs a conventional static SINR-to-MCS threshold mapping without adaptive correction, while the RL-based model leverages a Q-learning agent that dynamically selects the MCS as an action based on its learned policy derived from continuous interaction with the environment.

## IV. Evaluation

To evaluate the effectiveness of the proposed RL-based AMC scheme, we conduct simulations under two different mobility scenarios and compare its performance with a conventional baseline AMC approach. In particular, we focus on comparing the throughput–distance curves of the RL-based and baseline schemes across both mobility models to assess the robustness and adaptability of the proposed algorithm.

### *A. Performance Evaluation Under Radially Outward Mobility*

Under radial mobility, the SINR decreases progressively as the user moves farther from the base station due to increasing path loss and shadowing effects, which attenuate the received signal power relative to noise. As the channel quality degrades, our dynamic scheduling algorithm responds by selecting lower MCS and allocating fewer resource blocks (RBs). This leads to a corresponding decrease in the TBS with distance. The SINR vs. distance and the TBS vs. distance curves for radial outward movement are shown in Figure 2.

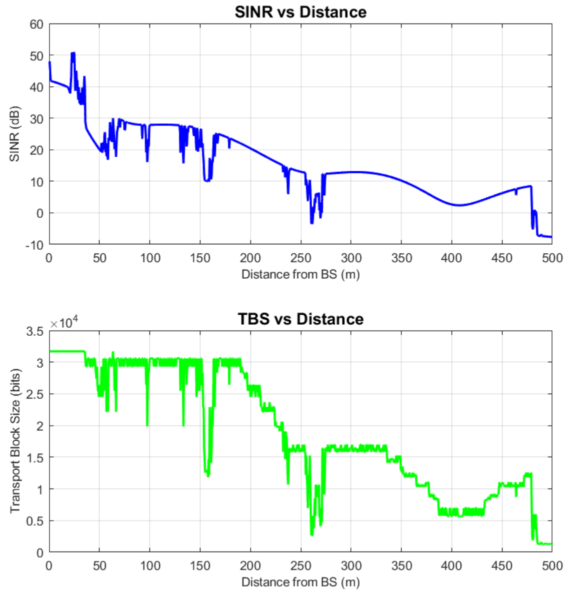


Figure 2: Variation of SINR and TBS with respect to distance from the base station.

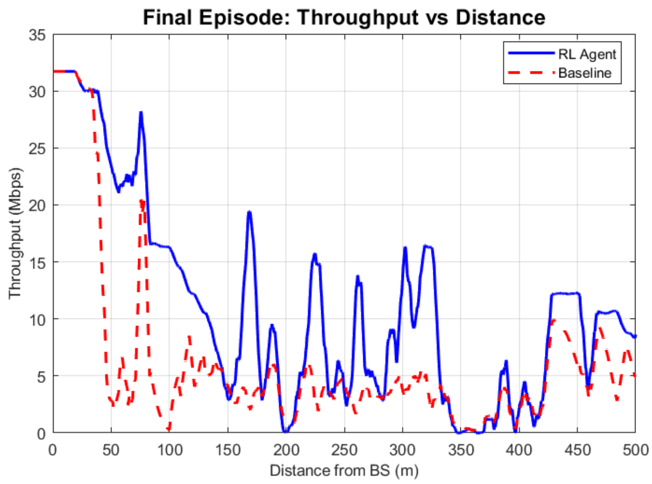


Figure 3: Throughput performance comparison between RL agent and baseline with respect to distance from the base station.

Under radial mobility, the RL-based AMC algorithm consistently achieves higher throughput than the baseline scheme across most distances from the base station. Near the cell center, both schemes perform similarly due to uniformly high SINR, leaving little room for adaptive improvement. However, as the user moves toward the cell edge, where SINR gradually decreases, the RL agent demonstrates its learning capability by selecting more aggressive MCS levels than the conservative baseline, leading to improved spectral efficiency while still maintaining acceptable BLER. The resulting curve is smoother and demonstrates better link adaptation under gradual channel quality degradation. Figure 3 shows the throughput-distance curve for radial mobility.

### *B. Performance Evaluation Under Random Waypoint Mobility*

In the random waypoint model, the user's movement causes abrupt variations in distance from the base station, resulting in highly variable SINR values. These fluctuations, caused by rapid changes in path loss and shadowing, affect the TBS. Our dynamic scheduling algorithm continuously adjusts the MCS and allocates resource blocks (RBs) based on the instantaneous SINR. For visualization, the SINR and TBS values are first sorted according to the user's distance from the base station. Figure 4 presents the resulting curves, where the top subplot shows the sorted SINR vs. distance and the bottom subplot shows the corresponding TBS vs. distance. These curves provide a clear depiction of how channel quality and resource allocation evolve with distance.

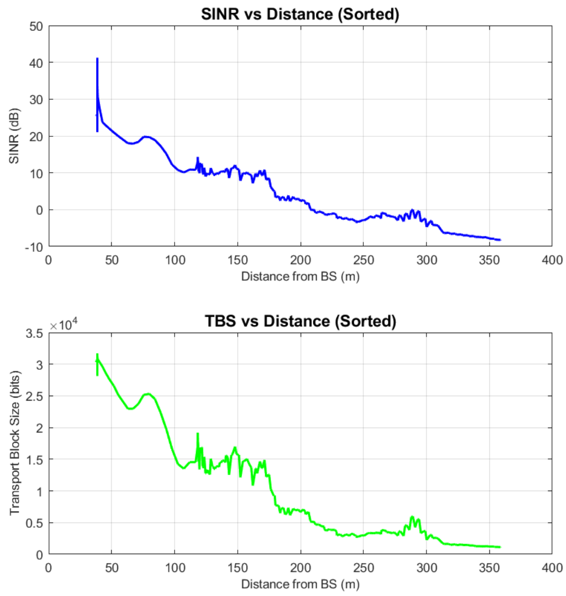


Figure 4: Variation of SINR and TBS with respect to distance from the base station.

Despite the unpredictability of the channel, the RL-based AMC approach consistently outperforms the baseline scheme. The Q-learning agent effectively adapts to rapid SINR fluctuations by leveraging the ΔCQI state variable, which captures the trend of channel quality evolution. This allows the agent to quickly respond to sudden degradations or improvements in SINR by adjusting the CQI correction factor accordingly. The throughput curve remains consistently above the baseline on average, although with more noticeable fluctuations due to the inherent randomness of the mobility model. Figure 5 shows the sorted throughput vs. distance curve for random waypoint mobility model.

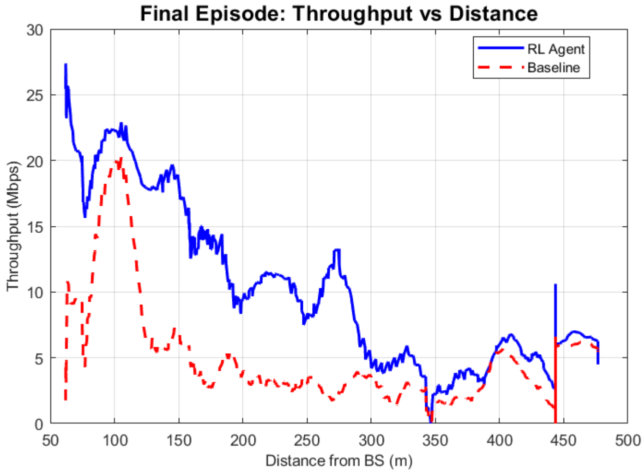

Figure 5: Throughput performance comparison between RL agent and baseline with respect to distance from the base station.

Overall, the results highlight the robustness of the proposed RL-based AMC scheme. In both mobility scenarios, RL-AMC exhibits superior performance compared to the baseline, with the gain being more pronounced under challenging conditions such as at the cell edge or during sudden SINR drops. The use of state variables incorporating both CQI and its temporal evolution allows the learning agent to balance exploitation of favorable channel states and exploration of higher MCS levels.

## V. CONCLUSION

This study aimed to design a dynamic proportional fair (PF) scheduling framework integrated with an RL-based Adaptive Modulation and Coding Scheme (AMC) utilizing Q-learning for optimal Modulation and Coding Scheme (MCS) selection. By jointly adapting MCS decisions and resource allocation, our approach effectively balances throughput maximization and link reliability under varying channel conditions. Simulation results demonstrate that the proposed RL-enhanced scheduling consistently outperforms the baseline static AMC model, achieving superior throughput in both radial mobility and random waypoint mobility scenarios.

Despite the significant performance gains, this work is currently limited to a single-user LTE downlink environment. Future extensions will focus on multi-user scenarios, where proportional fairness must be jointly optimized across competing users. Additionally, we plan to enhance the state-space representation and reward formulation to better capture temporal and spatial channel dynamics. Further exploration will include benchmarking the RL-based scheduler against Round-Robin and Maximum-CINR allocation strategies, as well as comparing its effectiveness with other dynamic resource scheduling techniques under heterogeneous network conditions.

## REFERENCES

[1] R. Bruno, A. Masaracchia, and A. Passarella, "Robust Adaptive Modulation and Coding (AMC) Selection in LTE Systems Using Reinforcement Learning," *Proc. IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Vancouver, BC, Canada, 2014, pp. 1–6, doi: 10.1109/VTCFall.2014.6966162.

[2] P. H. P. Carvalho, R. D. Vieira, and J. P. Leite, "A Continuous-State Reinforcement Learning Strategy for Link Adaptation in OFDM Wireless Systems," *Journal of Communication and Information Systems*, vol. 30, no. 1, pp. 47–57, Mar. 2015, doi: 10.14209/jcis.2015.6.

[3] J. P. Leite, P. H. P. de Carvalho, and R. D. Vieira, "A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems," *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, 2012, pp. 809–814, doi: 10.1109/WCNC.2012.6214482.

[4] M. Mota *et al.*, "Adaptive Modulation and Coding based on Reinforcement Learning for 5G Networks," 2019, doi: 10.48550/arXiv.1912.04030.

[5] L. Zhang, J. Tan, Y.-C. Liang, G. Feng, and D. Niyato, "Deep Reinforcement Learning based Modulation and Coding Scheme Selection in Cognitive Heterogeneous Networks," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.02868

[6] Z. Zhang, Y. Zheng, M. Hua, Y. Huang, and L. Yang, "Cache-Enabled Dynamic Rate Allocation via Deep Self-Transfer Reinforcement Learning," Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.11334

[7] M. Cate, "Optimization of Adaptive Modulation and Coding Schemes for OFDM Systems: Performance Analysis in Rayleigh, Rician, and AWGN Channels Using MATLAB," 2024. [Online]. Available: https://www.researchgate.net/publication/387896717

[8] L. Tsipi, M. Karavolos, G. Papaioannou, M. Volakaki, and D. Vouyioukas, "Machine learning-based methods for MCS prediction in 5G networks," *Telecommun Syst*, vol. 86, no. 4, pp. 705–728, Aug. 2024, doi: 10.1007/s11235-024-01158-x.

[9] R. Khan, Q. Yang, A. Noor, S. Bin Altaf Khattak, L. Guo, and A. Bin Tufail, "An efficient adaptive modulation technique over realistic wireless communication channels based on distance and SINR," *Frequenz*, vol. 76, no. 1–2, pp. 83–95, Jan. 2022, doi: 10.1515/freq-2021-0066.

[10] R. Khan *et al.*, "Optimizing adaptive modulation technique using standard propagation model for enhanced wireless communication channels," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-16743-6.

[11] A. Al Amin, D. Basak, T. Khadem, M. D. Hossen, and M. S. Islam, "Analysis of modulation and coding scheme for 5th generation wireless communication system," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, Institute of Electrical and Electronics Engineers Inc., Jan. 2017, pp. 1545–1549. doi: 10.1109/CCAA.2016.7813968.

[12] T. Keller and L. Hanzo, "Adaptive Modulation Techniques for Duplex OFDM Transmission," 2000.

[13] A. Sharif, V. M. Potdar, and R. A. Ahmad, "Adaptive channel coding and Modulation scheme selection for achieving high throughput in wireless networks," in *24th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010*, 2010, pp. 200–207. doi: 10.1109/WAINA.2010.197.

[14] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans Wirel Commun*, vol. 4, no. 6, pp. 2726–2736, Nov. 2005, doi: 10.1109/TWC.2005.858010.

[15] R. Fantacci, D. Marabissi, D. Tarchi, and I. Habib, "Adaptive modulation and coding techniques for OFDMA systems," *IEEE Trans Wirel Commun*, vol. 8, no. 9, pp. 4876–4883, Sep. 2009, doi: 10.1109/TWC.2009.090253.

[16] P. P. Devi, Vincent, and J. W. Simatupang, "Machine Learning-based Modulation Format Identification and Optical Performance Monitoring Techniques Implementation," in *2021 Advanced Communication Technologies and Signal Processing, ACTS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ACTS53447.2021.9708283.

[17] S. Chinchali *et al.*, "Cellular Network Traffic Scheduling with Deep Reinforcement Learning."