

REINFORCEMENT LEARNING-BASED 5G NETWORK OPTIMIZATION USING BER SIMULATION

Tupakula Sunayana¹, Dr. S. Swarnalatha²

¹M.Tech Student, Department of Electronics and Communication Engineering,

Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh, India

²Professor, Department of Electronics and Communication Engineering, SVU College, Tirupathi, A.P., India

ABSTRACT

5G Wireless Communication Based on Reinforcement Learning Essay 5G wireless communication technologies are predicted to be able to maintain dependable communication while providing the highest possible data rates and minimum delays over time-varying channels. The paper explores the feasibility of employing adaptive modulation as a Markov decision process. The paper trains a Deep Q Network (DQN) controller to make decisions on the most appropriate modulation, out of QPSK, 16-QAM, and 64-QAM options, based on the estimated Signal-to-Noise Ratio (SNR). The communication system has a 20 MHz bandwidth, subject to Rayleigh fading and additive white Gaussian noise. The transmitter modulates binary information data with selected encoding methods. The receiver demodulates and decodes the input data. The BER, latency, throughput, and spectral efficiency are evaluated for the feedback process to the DQN agent. The reward function in the paper's experiments prioritizes lower BER, lower latency, and higher throughput for the selected action. The results provided by the authors demonstrate that the BER for all three modulation schemes decreases with increasing SNR as expected. The results also confirm that higher-order modulations have higher latency than lower-order ones on average. Most importantly, the experiments have shown that the DQN agent learned to switch between different modulations depending on the conditions and follow the optimal policy of switching from QPSK to 16-QAM and eventually to 64-QAM with rising SNR. The study therefore proves the feasibility of adopting a learning-based approach to modulation control and serves as a valuable basis for future research into the 5G physical layer.

Keywords: 5G wireless communication, reinforcement learning, Deep Q-Network, adaptive modulation, QPSK, 16-QAM, 64-QAM, SNR, BER, Rayleigh fading, AWGN, latency, spectral efficiency.

1. INTRODUCTION

Wireless communication has evolved through generations to meet growing demands for connectivity, capacity, reliability, and data rate. In a 5G-oriented system, the radio link is still highly sensitive to propagation effects, multipath fading, receiver noise, interference, and user movement. Since the received signal quality is not constant, a transmission mode that is appropriate at one moment may be unnecessarily conservative or unreliable at another.

Adaptive modulation is one of the practical approaches to address such a variation. The central idea is simple: it is possible to adopt a robust modulation scheme when the channel is poor and a higher-order scheme when the channel is good. The project evaluates QPSK, 16-QAM, and 64-QAM. As for the number of bits carried per symbol, QPSK is 2, for 16-QAM it is 4, and for 64-QAM, 6. In terms of the largest noise margin, QPSK is likely to perform best of these three options. It is worth noting that 64-QAM can exploit a good channel more efficiently.

The problem emerges when the decision is made on the modulation based only on fixed rules. An adaptive modulation is conventionally designed upon predefined SNR thresholds or lookup tables. Such rules are easy to implement, but the thresholds are fixed at the design time and do not learn from the communication experience at run-time. In a learning approach, the agent observes the channel and picks up a

modulation, evaluates the outcome, and updates the policy.

The present project implements this idea in the form of a deep Q-network. This DQN uses SNR as the main channel state and selects one of the 3 possible modulation schemes. The signal is then passed through a Rayleigh fading channel with AWGN, received, demodulated, and assessed for BER and latency. The BER, latency, and throughput or spectral efficiency are the inputs to the learning feedback, thus forming a closed loop between the environment (channel) and the agent (DQN). The contribution of the study is providing an interpretable simulation framework, which ties the channel quality, the selected modulation scheme, the resulting BER, latency, and the DQN learning into a single process. The work limits itself to a small action space for the purpose of easy visualization of the learned policy. As such, the work can be used as a baseline for future works, which will incorporate more information about the channel, coding schemes, MIMO, power control, and standardized 5G NR MCS selection algorithms.

2. RELATED WORK

Adaptive modulation and link adaptation have become important research areas in wireless communication because the quality of a wireless channel varies with time. In a conventional communication system, the transmitter cannot always use the same modulation scheme efficiently since a fixed modulation may waste channel capacity or result in a high error rate if the channel condition deteriorates. Adaptive Modulation and Coding (AMC) addresses this issue by adapting the transmission mode based on channel conditions. Conventional AMC methods commonly employ predefined thresholds, channel quality information, or lookup tables. While such methods are simple and computationally efficient, their decisions are predetermined and do not learn directly from prior transmission outcomes. This has led researchers to consider reinforcement learning and deep reinforcement learning for intelligent link adaptation. [1]

Geiser et al. proposed DRLLA, a deep reinforcement learning approach for link adaptation, and identified the limitations of static decision boundaries used in conventional link-adaptation techniques. Their work demonstrated

that deep reinforcement learning can learn the relationships between channel conditions and transmission parameters rather than relying solely on manually defined boundaries. The study emphasized the suitability of neural-network-based reinforcement learning for link adaptation in dynamic radio environments. This work is related to the present study in that both use learning to determine appropriate transmission decisions. However, the present project uses a much smaller and more interpretable state and action space, with SNR as the main state and QPSK, 16-QAM, and 64-QAM as the available modulation actions. [1]

Parsa et al. investigated joint power allocation and MCS selection using deep reinforcement learning for energy-efficient link adaptation. Their study considered the variable nature of wireless channels and employed a learning-based agent to select transmission parameters based on channel conditions. The work indicated that deep reinforcement learning can be beneficial when multiple communication parameters need to be optimized simultaneously. However, the main focus was joint power and MCS control with energy efficiency as an important consideration. In contrast, the present study concentrates on adaptive modulation and examines the interplay among modulation order, BER, latency, and throughput in a controlled Rayleigh fading and AWGN environment. This narrower formulation makes the learning process amenable to analysis and visualization for an M.Tech-level simulation study. [2]

Lee et al. proposed a deep reinforcement learning-based MCS decision model for wireless cellular communication. Their work formulated MCS selection as a learning problem in which the transmission scheme was adapted based on varying channel conditions. The study reported that deep reinforcement learning can lead to enhanced transmission efficiency compared to a conventional rate-control approach. The work is important to the present study as both investigate the use of deep reinforcement learning for adaptive transmission decisions. A significant difference is that the present project does not seek to model the full cellular MCS decision process. Instead, the present project uses SNR as a directly observable state and limits the action space to three modulation schemes. This simplified environment enables the analysis of

the DQN policy directly from SNR-versus-modulation plots and communication-performance graphs. [3]

Ye, Yu, and Fu developed a deep reinforcement learning-based link-adaptation technique for LTE and NR systems. Their work addressed the deleterious effect of outdated channel-quality feedback and used deep reinforcement learning to choose an efficient modulation and coding scheme. The research demonstrates that learning-based link adaptation can be beneficial when channel information is delayed, quantized, or imperfect. This is relevant to the proposed project, as the basic objective is also to adapt the transmission mode based on channel quality. However, the present work does not attempt to reproduce the complete LTE/NR link-adaptation mechanism. Instead, it provides a simpler physical-layer simulation in which SNR is directly provided to the DQN, the signal is transmitted through Rayleigh fading and AWGN, and the resulting BER and latency are used for evaluation. [4]

Cui et al. investigated reinforcement-learning-based adaptive modulation for underwater acoustic OFDM communication. Their work treated adaptive modulation as a reinforcement-learning problem and considered channel information and BER in the decision process. The study compared different learning algorithms, including DQN and other reinforcement-learning methods, and demonstrated the utility of learning-based modulation in a time-varying communication environment. While the application domain was underwater acoustic communication, the basic concept of an agent observing channel information and selecting a modulation action is similar to the present work. The major difference is the channel environment. The proposed project considers a 5G-oriented terrestrial simulation with Rayleigh fading and AWGN as opposed to an underwater acoustic channel. The present study thus concentrates on the behavior of DQN-based modulation selection under a simpler and controlled wireless model. [5]

The reviewed studies showcase a transition from rule-based link adaptation to learning-based decision-making. Traditional

adaptive modulation is heavily dependent on predefined thresholds, while reinforcement learning enables the communication system to enhance its decisions through repeated interaction with the environment. Deep reinforcement learning enables nonlinear relationships between channel observations and transmission actions to be represented using neural networks. These developments provide the theoretical motivation for using a DQN in the present project. At the same time, the reviewed literature predominantly addresses broader link adaptation, joint resource control, LTE/NR MCS selection, or different communication environments.

The present project is designed as a focused simulation framework rather than a complete 5G NR implementation. A 20 MHz bandwidth is maintained throughout the experiments so that different modulation schemes can be compared under the same general conditions. SNR is used as the primary state, as it provides a simple representation of channel quality. The DQN has three discrete actions: QPSK, 16-QAM, and 64-QAM. The selected signal is passed through Rayleigh fading and AWGN before reception and demodulation. BER is then calculated from the transmitted and recovered bits, and latency, throughput, and spectral efficiency provide additional metrics of performance.

Another key aspect of the proposed work is the reward design. Rather than evaluating the modulation choice only based on BER, the project uses a reward that considers communication reliability, throughput, and latency together. This enables the DQN to learn a balance between robust transmission and efficient data delivery. A modulation scheme that produces a low BER but unnecessarily increases transmission delay may not always receive the maximum reward, and a high-order modulation that produces excessive errors under poor channel conditions is also penalized. The reward thus provides a common measure through which different transmission behaviors can be compared during the training. The implemented reward in the project is based on BER, throughput, and latency.

The proposed study also emphasizes visual validation of learned policy. After training,

the DQN is tested to observe what modulation is selected at different SNR values. The project results demonstrate that the learned policy transitions from QPSK at lower SNR to 16-QAM at an intermediate range and to 64-QAM at higher SNR. The results are supported by SNR-versus-BER, BER-versus-latency, SNR-versus-latency, and modulation-policy plots. This approach enables the learning behavior to be more interpretable since the output of the neural network can be connected directly to a physical communication decision.

Overall, the reviewed literature provides support for the use of reinforcement learning for intelligent link adaptation and modulation control, but the present work concentrates on a compact and transparent DQN formulation. The study does not aim to replace complete 5G NR link adaptation or standardized MCS selection. Instead, it provides a controlled framework for understanding how a DQN can use channel quality information to select among three modulation schemes while considering BER and latency. This formulation is beneficial for studying the basic behavior of intelligent modulation selection before scaling up the system to more complex scenarios.

The research gap addressed by this work can be summarized as follows. Existing studies have demonstrated reinforcement learning for link adaptation, MCS selection, power allocation, and adaptive modulation in different environments. The present project combines channel modeling, adaptive modulation, DQN learning, BER evaluation, latency analysis, and policy visualization within one compact Python-based simulation. The selected environment is intentionally simple enough to make the learned behavior understandable while capturing the important effects of fading and noise. Future extensions can build on this framework by introducing richer channel states, standardized 5G channel models, additional MCS levels, MIMO, power allocation, interference, and multi-user resource management.

2.1 Research Gap

The project is intended to resolve the following practical gap: The fixed modulation rules do not apply the outcome; optimizing BER introduces too conservative decisions;

maximizing the order of the modulation causes a poor error rate; and the learning policy should be validated with communication-oriented measures instead of rewards. The suggested framework is expected to address the stated gaps by taking the SNR as the state, the modulation as the action, the BER and the latency as the feedback, and the DQN as the decision mechanism.

2.2 Research Objectives

The project objectives are to model the 5G-oriented link, evaluate the QPSK/16-QAM/64-QAM, model the Rayleigh fading and AWGN, develop the adaptation of the modulation as the RL problem, train the DQN using the SNR, evaluate the BER and latency, and validate the learned policy using graphical analysis. The objectives are determined by the scope and implementation section of the project report.

3. PROPOSED METHODOLOGY

The proposed methodology employs a Deep Q-network to dynamically choose the best modulation scheme based on the channel state. The channel state is defined by a 20 MHz bandwidth and signal-to-noise ratio (SNR). DQN chooses between three modulation schemes, namely Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (16-QAM), and 64-QAM. The selected signal is sent through the Rayleigh fading channel whose additive white Gaussian noise (AWGN) is applied. The received signal is demodulated at the receiver, and bit error rate (BER) and latency are measured to determine the quality of communication. These metrics are then converted into a reward function that is used to update the DQN weights using an experience replay and target networks. The process is repeated for different channel states until the end of training.

1. Received Signal

The received signal after passing via a wireless channel can be represented as

$$r = hs + n$$

where s denotes the original transmitted signal, h is the Rayleigh fading coefficient of the channel, n is the AWGN noise, and r indicates the received signal.

The received signal undergoes processing at the receiver for demodulation and calculating BER.

2. Signal-to-Noise Ratio

The signal-to-noise ratio is used to define the quality of a communication channel and can be calculated as

$$SNR(dB) = 10 \log_{10} \left(\frac{P_s}{P_n} \right)$$

where P_s indicates the signal power and P_n demonstrates the noise power. In this equation, higher values of SNR suggest better channel quality and worse channel condition for the lesser values of SNR. In the proposed methodology, the value of SNR has been passed to DQN as a state to choose an appropriate modulation scheme.

3. Bit Error Rate

The bit error rate is another important factor to measure the reliability of the received data at the receiver end and can be calculated as

$$BER = \frac{N_{error}}{N_{total}}$$

where N_{error} indicates the total number of erroneous bits and N_{total} is the total number of bits received.

The lower the value of BER, the higher the reliability of the received data. Similarly, the obtained value of BER has been used as an essential parameter for rewarding the DQN agent.

4. Reward Function

To evaluate the efficiency of the chosen modulation scheme, the DQN agent employs a reward function that is explained as

$$R = 100(1 - BER) + 0.1T - 10L$$

where R illustrates the reward, BER indicates the bit error rate, T implies throughput, and L represents the latency. According to this function, the reward will be increased for the higher values of throughput and communication reliability and decreased for higher latency. Therefore, using this function is an appropriate idea to guide the DQN agent for finding a suitable balance between efficiency and reliability.

3.2 DQN-based Decision Process

The DQN agent takes the instantaneous signal-to-noise ratio as input and outputs the Q-values for the three possible modulation actions. The action with the highest Q-value is chosen. During training, the agent explores the environment by taking various actions and learns from the resulting bit error rates, latency, and rewards. The experiences are stored in a replay buffer and used to update the DQN. A target network is used to improve the stability of the training process.

3.3 Proposed Methodology Architecture

The proposed system is a closed learning loop. The channel condition is represented by SNR first. The DQN observes the state and selects one of the available modulation actions. The chosen signal is transmitted through the fading and noise model, received, and evaluated. The performance result is converted into reward feedback, which is stored as experience and used to update the DQN. When training is complete, the network is frozen and the learned policy is evaluated

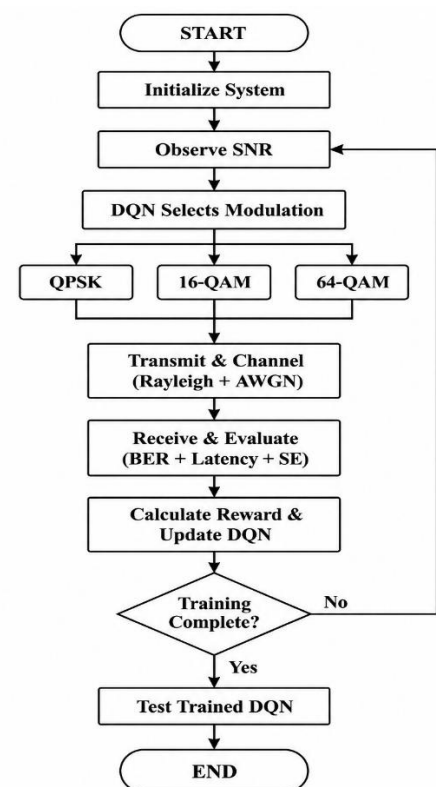


Fig. 1. Proposed methodology flowchart for DQN-based adaptive modulation.

4. DQN-BASED ADAPTIVE MODULATION

4.1 State, Action and Decision

The implemented state is $S = [\text{SNR}]$. For each observation, the DQN produces one Q-value for each available modulation action. The exploitation decision is obtained from the action with the largest predicted value:

$$a_t = \operatorname{argmax}_{a \in A} Q(S_t, a; \theta)$$

where the action set is:

$$A = \{\text{QPSK}, \text{16-QAM}, \text{64-QAM}\}$$

Epsilon-Greedy Exploration

During training, the DQN uses an ϵ greedy strategy:

$$a_t = \begin{cases} \text{a random action, with probability } \epsilon \\ \operatorname{argmax}_{a \in A} Q(S_t, a; \theta), \text{ with probability } 1 - \epsilon \end{cases}$$

As training progresses, ϵ is gradually reduced:

$$\epsilon \rightarrow 0$$

This means the agent initially explores different modulation actions and gradually relies more on the modulation policy learned by the DQN.

DQN State Representation

The state observed by the DQN at time t is

$$S_t = [\text{SNR}_t]$$

where S_t represents the current channel state and SNR_t is the observed signal-to-noise ratio.

Q-Value Estimation

For the three possible modulation actions, the DQN thus estimates:

$$Q(S_t, \text{QPSK}; \theta), Q(S_t, \text{16-QAM}; \theta), Q(S_t, \text{64-QAM}; \theta)$$

where θ represents the parameters of the DQN.

During training, epsilon-greedy exploration allows the agent to try out different actions. The exploration probability is then decreased so that the policy becomes increasingly reliant on the Q-network's estimates.

4.2 Neural Network Architecture

The project uses a compact feed-forward DQN because the state is one-dimensional and the action space has only three possibilities. The supplied DQN diagram shows an SNR input followed by dense layers of 64, 64, and 32 neurons with output layers of three neurons representing QPSK, 16-QAM, and 64-QAM,

respectively. This architecture is small enough such that a simulation can be run while also providing nonlinear function approximation.

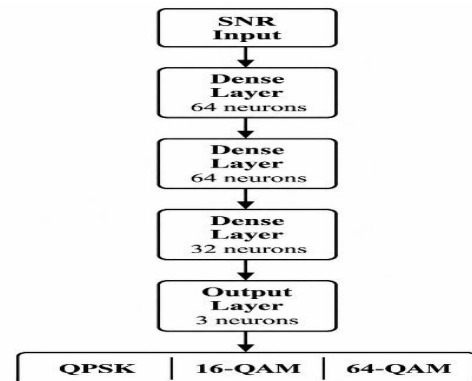


Fig. 2. DQN neural-network architecture used for modulation selection.

4.3 Reward Function

The reward forms the core to the proposed decision process because it determines which modulation behaviors are to be reinforced. The project specifies the implemented reward as

$$\text{Reward} = 100(1 - \text{BER}) + 0.1(\text{Throughput}) - 10(\text{Latency})$$

This structure provides a large positive contribution to reliable transmission, rewards useful throughput, and penalizes for latency. The report explicitly shows that the reward is intended to balance high efficiency, low BER, and low latency.

4.4 Experience Replay and Bellman Update

After an action is taken, the state, action, reward, and next state are stored in replay memory. Mini-batch sampling from the memory reduces the correlation between consecutive experiences and allows previously observed channel conditions to be reused. The DQN then updates the online network using a temporal-difference target:

$$y_t = r_t + \gamma \max_a Q(S_{t+1}, a)$$

where r_t is the reward, γ is the discount factor, and $Q(S_{t+1}, a)$ represents the expected value of the next action.

The loss is the difference between the predicted Q-value and the target. A separate target network is updated less often in order to stabilize the training. The project documentation includes experience replay, Bellman learning,

and periodic target-network updates as explicit stages of the implementation.

5. RESULTS AND DISCUSSION

5.1 SNR versus BER

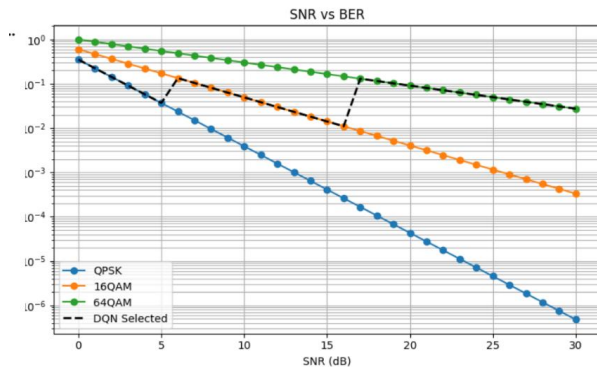


Fig. 3. SNR versus BER for QPSK, 16-QAM, 64-QAM, and DQN-selected operation.

Figure 3 presents the expected improvement in the bit error rate (BER) with an increase in signal-to-noise ratio (SNR). QPSK has the lowest BER among the three fixed options in most of the illustrated range. 16-QAM also has better error characteristics than 64-QAM and gradually improves with SNR. 64-QAM has the worst BER at low SNR due to the close packing of its points in the signal space diagram.

The dashed curve following the selected DQN actions was obtained to assess the agent's flexibility. It follows the modulation curve corresponding to the action selected by the agent at a particular SNR. In particular, in the example shown, the policy tracks QPSK at low SNR, switches to 16-QAM when the first threshold is crossed, and moves to 64-QAM after the second threshold is crossed. Thus, it can be seen that the DQN does not invent a new modulation but rather selects from the available options.

5.2 BER versus Latency

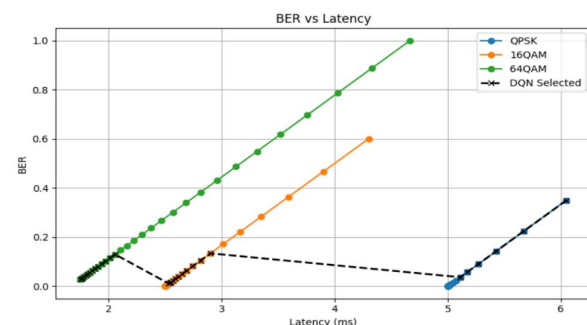


Fig. 4. BER versus latency for the three fixed modulations and the DQN-selected path.

Figure 4 shows the trade-off between reliability and delay and tracks the policy through the modulation curves. QPSK operating points occupy the high-latency region in exchange for improved reliability. On the other hand, the 64-QAM curve reaches the lowest-latency region but has significantly worse BER at poor channel conditions. 16-QAM has mediocre characteristics compared to the two extremes.

The DQN policy switches between the operating points of the three modulations depending on the channel conditions. Thus, the given controller implements a decision-making process that favors robust modulations with good BER characteristics in poor channels and less reliable but latency-efficient options in good channels.

5.3 SNR versus Latency

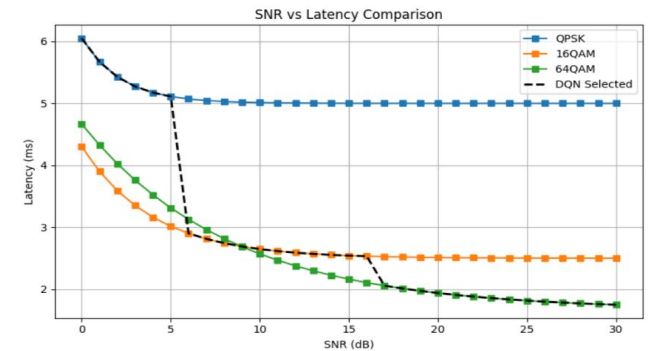


Fig. 5. SNR versus latency comparison of QPSK, 16-QAM, 64-QAM, and DQN selection.

The comparison of the latency helps explain the benefits of using higher-order modulations. In the submitted illustration, the value of QPSK is around 5 ms while the latency of 16-QAM is 2.5 ms and 64-QAM is 1.7-1.8 ms when working with high SNR. It can be noticed that higher-order schemes have a better capacity to carry the information per unit time, which results in fewer symbols and, therefore, reduced latency. For that reason, the DQN policy follows the high-latency QPSK curve at low SNR, shifts to 16-QAM in the middle range, and moves to 64-QAM at high SNR. The agent takes into account the latency as a performance measure and aims to select the best modulation for the current channel conditions.

5.4 Learned DQN Policy

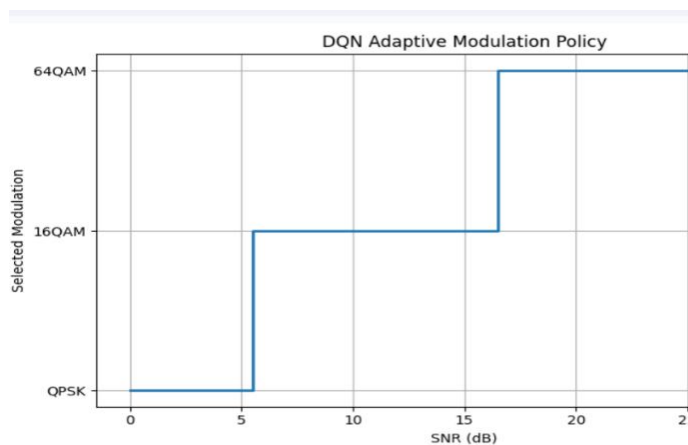


Fig. 6. Learned DQN adaptive modulation policy versus SNR.

Figure 6 shows the most important conclusion from the reinforcement-learning process. As a piecewise function, the policy suggests that the lowest SNR is sufficient to switch to QPSK, whereas 16-QAM is selected for moderate values of SNR, and the highest modulation, 64-QAM, is applied for high SNR. According to the submitted illustration, the first transition happens when the SNR crosses 6 dB and the second transition is at 16-17 dB. The specific values of the thresholds are related to the reward function, state encoding, channel model, and other factors, which are fixed within the problem formulation. From that perspective, the policy found by the agent should not be used as a universal rule for 5G networks since it is tied to the exact set of assumptions that were predefined beforehand.

6. ANALYSIS OF THE PROPOSED APPROACH

6.1 Reliability-Efficiency Trade-off

The main finding from the analysis is that the reliability and efficiency of communication systems push the decision-making process in opposite directions. Thus, in the given illustration, QPSK is the most reliable modulation but also the slowest one, and 64-QAM is the most efficient but also the most error-prone. The middle part is occupied by 16-QAM, and the DQN controller is learning to switch between the available options based on the current SNR.

6.2 Why the Learned Policy Is Useful

The main value of the learned policy comes from the ability to learn the reliability-efficiency trade-off. However, the value function in the

DQN is not a simple linear function that can be hardcoded into a policy. The given policy is useful because it emerges from the RL process, where the agent interacts with the environment. Thus, the approach can be generalized to any other setups with different state, action, and reward spaces, as well as different communication scenarios.

6.3 Practical Meaning of the Results

The most important practical insight from the analysis is that the reliability of communication should be prioritized at poor channel conditions, whereas efficiency becomes more important at higher SNR. Thus, the transition from QPSK to 16-QAM and then to 64-QAM at increasing SNR follows the logic of reducing the transmission overhead for a given level of reliability. The DQN controller can be viewed as an intelligent decision-making element that selects the most appropriate option among the three available ones in real time.

7. LIMITATIONS AND FUTURE WORK

The given research has several significant limitations, which should be noted and can be a subject for further studies. First, the channel is modeled by Rayleigh fading and AWGN noises, which is not suitable for 5 G. Second, SNR is used as the state, while, in reality, a receiver can estimate the channel transfer function to obtain more data. Third, the actions are limited to three modulations, and the coding rate, transmission power, and other attributes of communication systems are not considered. The latency values obtained in the study should be considered relative because they represent the transmission and processing time in a given scenario. Therefore, they cannot be used to estimate the latency of a real-world 5 G network. The conversion of seconds into milliseconds in Figure 1 should be in accordance with the Python implementation used for calculation in the project report. The limitation of the current approach is also noted by the project description, which advises not to compare the results of the simulation with real-world latency. The next version of the future work section should address realistic channel modeling, expanded state and action spaces, and practical implications of the results. The state space can include channel gain, CQI, interference, mobility, latency, BER history, and other factors. The action space can involve a

mapping of MCS to channel conditions and power control. The communication environment can be extended by 3GPP channel models, multiple-input and multiple-output (MIMO) and beamforming configurations, hybrid automatic repeat request (HARQ), and multi-user scheduling, among others. Such extensions will help analyze the practical implications of the approach better and demonstrate the value of the RL-based decision-making in communications.

8. CONCLUSIONS

This paper presents the adaptive modulation framework for 5G-oriented communication based on the DQN RL method. The environment uses the packet BER as the main performance metric under Rayleigh fading and AWGN conditions. The link occupies a 20 MHz spectrum, and the receiver uses the measured BER to estimate the SNR and update the DQN policy. The given setup allows analyzing the reliability-delay trade-off and the decision-making process behind it.

The experimental results provided with the project report illustrate the basic principles of the DQN-based adaptive modulation. Figure 4 shows that the BER follows the theoretical predictions for the three modulations but has lower values at most SNR due to the dynamic policy. Figure 5 highlights the reliability-delay trade-off and the corresponding changes in the packet BER and latency. Finally, Figure 7 presents the main finding from the analysis, where the DQN policy follows the three modulations depending on the SNR. In the submitted illustration, the switch from QPSK to 16-QAM occurs near 6 dB, and the transition from 16-QAM to 64-QAM occurs near 16-17 dB.

Therefore, the proposed framework provides a baseline for intelligent adaptation in wireless communication. It demonstrates the principles of real-time decision-making and the reliability-delay trade-off while also suggesting the limitations of the current approach.

REFERENCES

- [1] F. Geiser, D. Wessel, M. Hummert, A. Weber, D. Wübben, and A. Dekorsy, "DRLLA: Deep Reinforcement Learning for Link Adaptation," *Telecom*, vol. 3, no. 4, pp. 692-705, 2022.
- [2] A. Parsa, N. Moghim, and P. Salavati, "Joint Power Allocation and MCS Selection for Energy-Efficient Link Adaptation: A Deep Reinforcement Learning Approach," *Computer Networks*, vol. 218, Art. no. 109386, 2022.
- [3] A.-H. Lee, H. Bae, Y.-K. Kim, and C.-K. Kim, "Deep Reinforcement Learning Based MCS Decision Model," *Journal of KIISE*, vol. 49, no. 8, pp. 663-668, 2022.
- [4] X. Ye, Y. Yu, and L. Fu, "Deep Reinforcement Learning Based Link Adaptation Technique for LTE/NR Systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 7364-7378, 2023.
- [5] X. Cui, Z. Zhang, J. Li, B. Jiang, S. Li, and J. Liu, "Reinforcement Learning-Based Adaptive Modulation Scheme Over Underwater Acoustic OFDM Communication Channels," *Physical Communication*, vol. 61, Art. no. 102207, 2023.
- [6] A. Parsa, N. Moghim, and S. Shetty, "QoS-Aware Link Adaptation for Beyond 5G Networks: A Deep Reinforcement Learning Approach," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 6368-6382, 2025.
- [7] B. B. Cebecioglu et al., "Deep Reinforcement Learning Based MCS Selection in Open RAN for Vehicular Communications," in *Proc. IEEE PIMRC*, 2025.
- [8] G. Liang, J. Hu, Y. Zhao, and K. Yang, "Intelligent Link Adaptation for Integrated Data and Energy Transfer: An Enhanced DRL Approach for Long-Term Constraints," *IEEE Transactions on Communications*, vol. 72, no. 11, pp. 6956-6972, 2024.
- [9] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529-533, 2015.
- [10] M. Shafi et al., "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201-1221, 2017.
- [11] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617-1655, 2016.
- [12] N. C. Luong et al., "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE*

Communications Surveys & Tutorials, vol. 21,
no. 4, pp. 3133-3174, 2019.