

Article

# An Adaptive Hybrid Automatic Repeat Request (A-HARQ) Scheme Based on Reinforcement Learning

Shih-Yang Lin <sup>1,\*</sup>, Miao-Hui Yang <sup>2,†</sup> and Shuo Jia <sup>2</sup><sup>1</sup> School of Architecture Engineering, Weifang University of Science and Technology, Weifang 262700, China<sup>2</sup> School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China; miaohuiyang@hotmail.com (M.-H.Y.); peterj1a@hotmail.com (S.J.)

\* Correspondence: shihyang.lin@hotmail.com

† These authors contributed equally to this work.

**Abstract:** V2X communication is susceptible to attenuation and fading caused by external interference. This interference often leads to bit error and poor quality and stability of the wireless link, and it can easily disrupt packet transmission. In order to enhance communication reliability, the 3rd Generation Partnership Project (3GPP) introduced the Hybrid Automatic Repeat Request (HARQ) technology for both 4G and 5G systems. Nevertheless, it can be improved for poor communication conditions (e.g., heavy traffic flow, long-distance transmission), especially in advanced or cooperative driving scenarios. In this paper, we propose an Adaptive Hybrid Automatic Repeat Request (A-HARQ) scheme that can reduce the average block error rate, the average number of retransmissions, and the round-trip time (RTT). It adapts the Q-learning model to select the timing and frequency of retransmission to enhance the transmission reliability. We also design some transmission schemes—*K*-repetition, *T*-delay and [*T*, *K*]-overlap—which are used to shorten latency and avoid packet collision. Compared with the conventional 5G HARQ, our simulation results show that the proposed A-HARQ scheme decreases the system's average BLER, the number of retransmissions, and the RTT to 5.55%, 1.55 ms, and 0.97 ms, respectively.

**Keywords:** hybrid automatic repeat request; reinforcement learning; block error rate; 5G new radio; vehicle-to-everything



**Citation:** Lin, S.-Y.; Yang, M.-H.; Jia, S. An Adaptive Hybrid Automatic Repeat Request (A-HARQ) Scheme Based on Reinforcement Learning. *Electronics* **2023**, *12*, 4127. <https://doi.org/10.3390/electronics12194127>

Academic Editor: Christos J. Bouras

Received: 1 September 2023

Revised: 27 September 2023

Accepted: 29 September 2023

Published: 3 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Advanced vehicle-to-everything (V2X) applications for autonomous vehicles reflect the functional aspects of vehicular communication technology and influence the performance requirements of the communication system. The Society of Automotive Engineers (SAE) classifies autonomous vehicles into six levels, ranging from lower to higher, based on whether a human driver or an automation system is primarily responsible for monitoring the driving conditions. These levels are: 0—No Automation; 1—Driver Assistance; 2—Partial Automation; 3—Conditional Automation; 4—High Automation; 5—Full Automation. Based on the 3rd Generation Partnership Project (3GPP) release 17 TS 22.186 [1], it identifies the business requirements for enhanced V2X scenarios and provides the expected performance for all levels of autonomous vehicles. With the latest 3GPP release 17 TS 22.186 technique specification, we can gain a clear understanding of the impact of communication performance at the level of V2X automation.

The 5G New Radio (NR) technique has characteristics of ultra-low latency, super-strong link capability, and ultra-high broadband, and the performance of the on-board networking system has been greatly improved. Nevertheless, the wireless channel is prone to attenuation and fading caused by external interference. This often results in bit errors of the wireless transmission, poor quality and stability of the wireless link, and easy damage to data packet transmission. In order to address the issue of bit error during transmission, the 3rd Generation Partnership Project (3GPP) presented the Hybrid

Automatic Repeat Request (HARQ) technology for both 4G and 5G systems. Although the 5G HARQ technique guarantees its reliability and meets the delay requirements for V2X communication, it does not easily maintain high communication quality in poor conditions (e.g., heavy traffic flow, long-distance communication). This is particularly true in advanced driving or cooperative driving for vehicle platooning information exchange, cooperative collision avoidance, emergency trajectory alignment, cooperative lane change, sensor information sharing, information exchange between vehicles, etc.

Therefore, it is necessary to explore a more efficient retransmission mechanism for vehicular transmission in 5G NR-enhanced V2X scenarios. This will improve the reliability of information transmission and reduce the retransmission delay. Considering transmission conditions, this study focuses on the retransmission optimization based on the 5G NR HARQ. In this paper, an Adaptive Hybrid Automatic Repeat Request (A-HARQ) scheme is proposed to improve the communication quality, which can reduce the average block error rate, average number of retransmissions, and round-trip time (RTT). The A-HARQ scheme includes not only the 5G HARQ retransmission mechanism, but also the  $K$ -repetition mechanism, which can shorten the RTT and improve the transmission success rate.

To illustrate our research, we have organized the rest of this paper as follows: In this section, we provide a brief introduction to the background and motivation. In Section 2, we discuss the functioning of HARQ and survey some related works. In Section 3, we state some problems and provide an overview of the system. The proposed Adaptive Hybrid Automatic Repeat Request (A-HARQ) scheme is studied in Section 4. We evaluate its performance through simulation and provide detailed descriptions in Section 5. Finally, we summarize the work presented in this article and discuss future research directions related to HARQ in the Section 6.

## 2. Related Works

In practice, packet errors and losses are inevitable due to ubiquitous noise, signal interference, and channel fading. As an incorrectly decoded message does not bring about fresh awareness, these packet errors and losses will result in stale information, leading to uncontrollable residual errors, system instability, and incorrect decisions.

HARQ retransmission, as a standard technique for improving transmission reliability, has been adopted in various wireless standards [2]. HARQ is a physical-layer mechanism that employs feedback to transmit at higher target block error rates (BLERs) while achieving robustness of the transmission by providing retransmissions based on the feedback of Acknowledgement or Negative Acknowledgement (ACK/NACK). However, the HARQ procedure poses a bottleneck for achieving the previously mentioned latencies because conventional HARQ allows retransmissions only upon receiving a NACK. The Base Station (BS) requires a few time units for detection when it receives the packet for the first time, and then it issues the feedback. In order to improve the success rate of HARQ retransmission, more HARQ iterations are needed to enhance the overall HARQ RTT. Hence, optimizing HARQ to improve the accuracy of information transmission and reduce transmission delay becomes a critical issue.

A decoder is responsible for a minimum of 60% of the time required for the user equipment (UE) to receive and process data. Therefore, to fulfill the communication requirements of time-sensitive applications in V2X, it is feasible to predict the decoding result before the successful decoding process by the decoder. In their study, the authors in [3] proposed an Early HARQ (E-HARQ) technique based on decoder result prediction. The accuracy of ACK/NACK feedback for predicting the uncoded bit error rate and indicating block errors was estimated, based on the likelihood ratio of information bits, as approximately 90%. Additionally, the delay is reduced by approximately 50%. However, early feedback errors can significantly affect the interrupt rates and latency. Authors in [4] also conducted both early transmission feedback and regular feedback in an attempt to correct the early incorrect prediction, reducing the influence caused by the inaccurate estimation. In [5], the authors proposed a new spatially coupled code, which is formed by

sending a low-density parity block code (LDPC-BC) through block Markov superposition transmission (BMST) to form the BMST-LDPC code. The BMST-LDPC scheme is integrated with the Hybrid Automatic Retransmission Request (HARQ) over the block-fading channel, improving the throughput performance of up to 10% compared with the conventional HARQ scheme.

Because the probability of retransmission increases with the increase in the number of receiving ends, it is challenging to apply the traditional HARQ technique. Authors in [6] proposed an external-code-based HARQ, in which the receiver only provides feedback on the number of destroyed code blocks (CBs), while the transmitter utilizes an external code to generate parity CBs and then employs an internal code to transmit these parity CBs for retransmission. This scheme can improve the reliability and resource efficiency of the fading channel and effectively reduce the retransmission rate. Authors in [7] proposed a dynamic Hybrid Automatic Repeat Request scheme that is different from the traditional HARQ. This scheme allows for the dynamic adjustment of the maximum retransmission times based on the last situation. In delay-sensitive applications where channel state information is unavailable to the sender, the scheme demonstrates improved better performance in terms of packet error rate and throughput.

In order to achieve the requirement of  $10^{-5}$  BLER as defined by the Ultra-Reliable Low-Latency Communication (URLLC) specification, the authors in [8] proposed a sub-code-based Early HARQ (SC EHARQ) scheme based on the Low-Density Parity Check (LDPC) subcode. It takes advantage of the LDPC subcode structure to provide faster feedback, achieving earlier retransmission and exhibiting better false positive energy. This results in fewer transmission failures compared with when E-HARQ predicts NACK as ACK. However, choosing the appropriate threshold is a critical issue for its performance. The authors in [9] improved the SC EHARQ scheme, which was the first exploratory study on the method for predicting quantitative improvement. It uses machine-learning methods to predict the decodability of received messages, using more advanced classification methods to predict decoding results before the final decoder iteration. More complex input features are utilized to further improve the classification performance, and appropriate methods for distinguishing between different classifiers are discussed. In [10], the authors conducted a mathematical analysis of packet errors, retransmissions, and delays related to bandwidth. They applied the M/G/1 queuing model to the delay analysis. The minimum bandwidth required was determined, as well as the relationship between the allocated bandwidth and the total delay, which established the adaptive control of the maximum number of retransmissions, thereby improving the performance of the URLLC.

As delays exist by nature and play a critical part in affecting the freshness of information, the authors in [11] comprehensively considered various types of nontrivial system delays and derived unified closed-form average Age of Information (AoI) and average Peak AoI expressions for the HARQ.

In order to support the new URLLC service, which aims to facilitate the transmission of small packets with strict requirements for latency and reliability, the authors in [12] proposed a spatiotemporal analytical framework for analyzing contention-based grant-free (GF) (i.e., configured grant) access schemes. It analyzed three GF access schemes with HARQ retransmissions, i.e., Reactive, K-repetition, and Proactive. It defined the latent access failure probability to characterize the URLLC reliability and latency performances. The results showed that under shorter latency constraints, the Proactive scheme provided the lowest latent access failure probability, whereas under longer latency constraints, the K-repetition GF transmission scheme achieved the lowest latent access failure probability.

Although current research has enhanced the HARQ in various aspects, they have not considered the reasonable spatiotemporal allocation of transmission on limited channel resources based on real-time wireless conditions. This paper focuses on optimizing the HARQ scheme by addressing the delay and reliability of information transfer between 5G NR-V2X UEs. In order to reduce the average number of retransmissions, transmission delay, and average BLER, and to improve the reliability of information transmission, an Adaptive-

HARQ transmission scheme is designed. This scheme includes  $T$ -delay,  $K$ -repetition, and  $[T, K]$ -overlap methods according to the factors that affect wireless conditions. Finally, we consider the external wireless channel conditions for adopting the Q-learning algorithm, with optimization based on reinforcement learning for improving the transmission of HARQ packets.

### 3. Problem Statement and System Framework

#### 3.1. Problem Statement

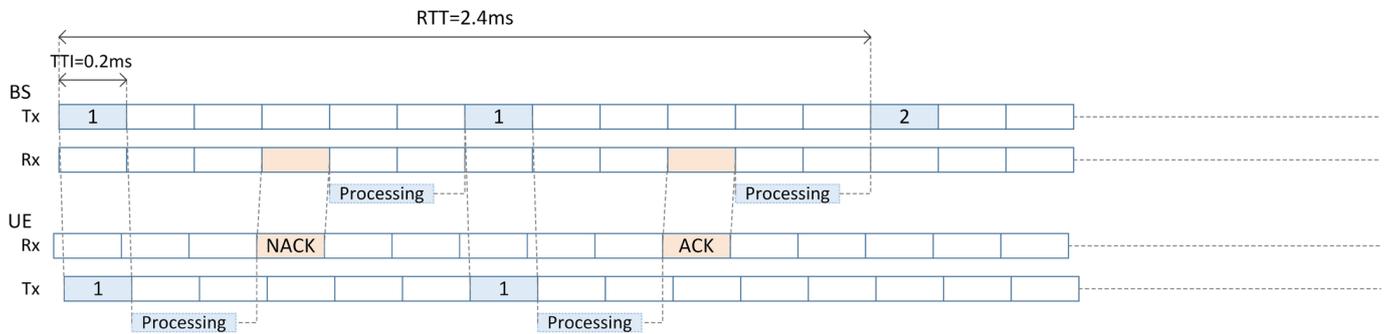
The Stop-and-Wait Protocol of HARQ (SW\_ARQ) was used to send data as an error correction technique in the 3GPP LTE protocol standard. In the SW\_ARQ, the transmitter (TX) sends a data frame and stops to wait for the acknowledgment, and then the receiver uses the 1-bit information for ACK or NACK acknowledgment of the data frame. However, the stopping method for confirming each transmission results in low throughput. Hence, the HARQ has been improved in the 5G system to achieve efficient and high-performance error correction. The MAC layer makes up for the waste of air interface resources of a single SW\_ARQ process by using multiple SW\_ARQ processes, without affecting its timelines [13]. The base station assigns the HARQ process number to the indoor terminal to determine the buffer at the physical layer. The physical layer combines the last decoded failed bit stream with the current retransmitted bit stream through the soft-combining technique to enhance the demodulation gain, offering benefits such as reduced latency and improved the decoding and demodulation gain. However, the disadvantage is that the physical layer of the receiving terminal requires a relatively large cache to store the bit information that is not properly decoded.

The complete HARQ procedure requires the MAC layer and the physical layer to work together. The MAC layer implements the SW\_ARQ protocol. If MAC frames need to be retransmitted as called by the feedback mechanism, it resends the MAC frames to the receiver (RX). The retransmitted MAC frames will be encoded according to the redundant version number when encoding at the physical layer. The physical layer stores the original bits of the empty interface that are not correctly decoded in the buffer. After the retransmitted data frame is received, it is integrated with the original bit stored in the buffer and decoded it. The introduction of soft-merge technology effectively reduces the proportion of error bits caused by channel interference, so that the correct rate of decoding is improved. The overall probability of decoding and the anti-interference ability are enhanced.

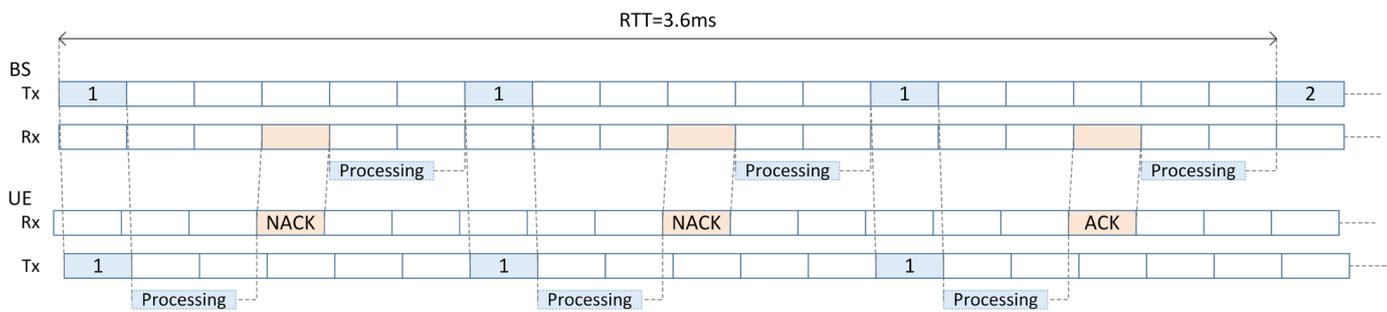
Figure 1a and Figure 1b respectively depict the downlink transmission process of a retransmission and double retransmission in a normal HARQ. Assuming the transmission time interval (TTI) is 0.2 ms, the RTT through the 5G HARQ to complete a retransmission and double retransmission is 2.4 ms and 3.6 ms, respectively. Note that here, the UE is sent with a negative time offset relative to the base station (BS) based on the final timing-advance (TA) settings. After the TX sends the data, physical forward error correction (FEC) reduces the number of retransmissions by adding redundant information so that the RX can correct some of the errors. For errors that cannot be corrected by forward error correction, the RX requests the sender to resend the data through the SW\_ARQ mechanism. The RX employs an error detection code, typically a cyclic redundancy check (CRC), to detect whether the received packet is wrong. If there is no error, the RX will send a positive ACK to the sender, and the TX will process this and send the next packet upon receiving the ACK. If an error occurs, the RX drops the packet and sends a NACK to the TX. Subsequently, the TX retransmits the same data in turn after receiving the NACK.

In addition, the 5G standard incorporates an asynchronous HARQ for both uplink and downlink transmissions. This feature enhances the flexibility of scheduling the timing and resource allocation, particularly in the context of the time-division duplexing (TDD) mode. Multiple parallel HARQ processes are allowed in the 5G standard, and while one HARQ process is awaiting an acknowledgment, the sender simultaneously carries out another HARQ process to transmit data. These aforementioned processes of HARQ form

a HARQ entity, which integrate the stop protocol together. Additionally, the 3GPP NR Release-15 [14] supports the  $K$ -repetition (Krep) scheme, which allows for a predefined number of consecutive replicas of the same packet without waiting for feedback.



(a)



(b)

**Figure 1.** 5G HARQ scheme of retransmission once (a) and retransmission twice (b).

On the other hand, there exist numerous factors that exert an influence on the BLER. The Channel Quality Indication (CQI) is transmitted from the terminal to the BS with the information measurement. It primarily represents the quality of the downstream channel. The LTE protocol defines the quality of the channel as CQI and quantizes it into a sequence of 0–15. The larger the CQI value, the better the channel quality, and the higher the utilization rate of the modulation coding method. It also means a greater efficiency and a larger corresponding transmission block, thus providing a higher downstream peak throughput. The opposite is true for small CQI values.

Distance is another important factor affecting the BLER and transmission latency [15–18]. As the distance between the TX and the RX increases, the Signal-and-Interference-to-Noise Ratio (SINR) continues to decrease, and the probability of packet loss due to packet collision also increases [18]. Moreover, the rate of packet loss experiences a rapid increase once the received signal power reaches its perceived power threshold. In addition, the BLER is also influenced by various network load factors.

Therefore, based on making full use of these characteristics of the 5G standard and the above analysis of the factors influencing the communication performance of the 5G NR-V2X, this study proposes an alternative optimization transmission scheme to address the issues that arise when the BLER exceeds the threshold.

### 3.2. System Overview

As a technical document defined by the 3GPP, the BLER is used to estimate the errors of the physical layer. In this study, the powerful learning ability of Q-learning in reinforcement

learning is utilized to obtain better communication schemes. The interaction between the agent and the environment in Q-learning will be determined by the update process of the Q-Table, which basically employs the TD Bellman Equation. In this study, the sequential difference method is used to set the Reward Calculator mechanism, and the update of the Bellman Equation is adopted to obtain the optimal strategy. The Bellman Update Equation is as in Equation (1):

$$Q(S, A) \leftarrow Q(S, A) + \alpha * [R + \gamma * \max Q(S', A) - Q(S, A)], \tag{1}$$

where  $\alpha$  is the learning rate,  $R$  is the immediate benefits, and  $\gamma$  is the discount (or attenuation) rate. The larger  $\alpha$  is, the less the effect of retaining the previous training. And  $\max Q(S', A)$  is the benefit in memory, and it refers to the maximum utility value in the action of the next state  $S'$ . The agent aims to obtain the maximum reward after choosing the action in the next state  $S'$ , so that the next time in state  $S$ , it can continue to get the reward by choosing the correct action. The larger the value of  $\gamma$ , the greater the role played by  $\max Q(S', A)$ , i.e., the more attention is paid to past experiences. In contrast, the smaller the  $\gamma$ , the more attention is paid to the immediate benefits  $R$ .

Q-learning is a reinforcement learning approach that learns a policy that maximizes the expected reward through training and feedback. In the Adaptive HARQ Q-learning model, we design a retransmission program with a Q-learning model to determine the optimal retransmission time and frequency, thereby enhancing network performance.

The functional component blocks of the adaptive HARQ (A-HARQ) scheme are depicted in Figure 2, where BLER is the average downlink block error rate, CQI is the average CQI, Distance is the distance between UEs, and NumRBs is the number of resource blocks. The A-HARQ scheme takes BLER, CQI, Distance and NumRBs as input linguistic variables. A state classifier is responsible for classifying the current transmission state according to the input linguistic variables. The state identity denoted by  $S_i$  can then be obtained from the state classifier. The optimal action  $A_k$  for  $S_i$  is inferred from the  $\epsilon$ -greedy strategy. By the action decision, a suitable optimized transmission scheme can be determined by the A-HARQ scheme. After the reward  $R(S_i, A_k)$  is generated by the Reward Calculator based on the system feedback of the transmission result, the Q values  $q(S_i, A_k)$  in the Q-Table are updated by the Q-Function Update. Moreover, the learning rate  $\alpha$  of the Q-Table will be adjusted by the Optimizer according to the feedback results of the system when making online decisions. The detailed design is given as Section 4.1.

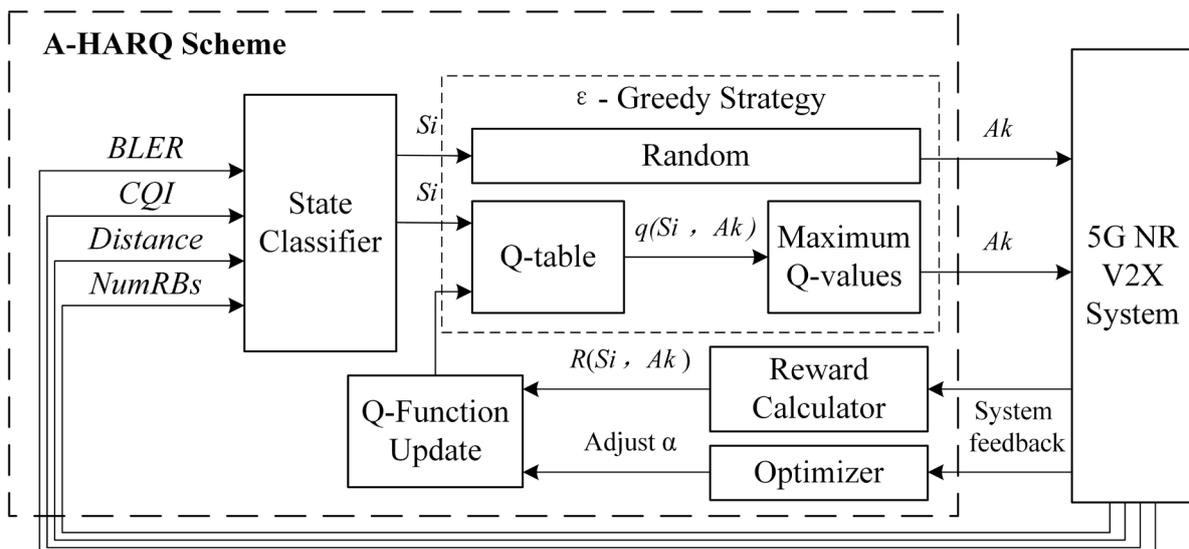


Figure 2. The architecture of the A-HARQ scheme.

## 4. The Proposed Adaptive HARQ Q-Learning Model

### 4.1. Model Description

The major components of the proposed Adaptive HARQ Q-learning model include the agent, state, action, and reward calculator. The agent serves as a representation of the 5G NR system, which performs an action according to the Q-Table obtained by reinforcement learning. The action here refers to when and how many times to transmit or retransmit the packet. The transmission results will affect the system state. The feedback from the system will determine the rewards obtained using the reward calculator and the optimization of the Q-Table. The objective of the proposed Q-learning method is to identify a set that will optimize the overall cumulative reward. The following is a detailed description of the model sections.

#### 4.1.1. 5G NR System

In this study, the “NR TDD Symbol-Based Scheduling Performance Evaluation” module of the 5G Toolbox of MATLAB R2021a is used as the main system simulation environment. Moreover, the visualization function of the BLER in the “NR Cell Performance Evaluation with Physical Layer Integration” module was invoked to form the simulation environment of this study [14,19–22]. The former example models a symbol-based scheduling scheme in the TDD mode and evaluates the network performance. Symbol-based resources schedule shorter transmission durations that span only a few symbols within a timeslot. In the TDD mode, physical uplink shared channel (PUSCH) and physical downlink shared channel PDSCH (PDSCH) transmissions are scheduled in the same frequency band with separation in the time domain. The latter example demonstrates the integration of a high-fidelity 5G Toolbox™ physical layer in a 5G NR node and models a 5G NR cell consisting of a set of user equipment (UE) connected to a 5G Base Station (gNB). The NR stack on the nodes includes radio-link control (RLC), medium-access control (MAC), and physical (PHY) layers.

To avoid any additional delays caused by Q-learning models in real networks, we adopt an asynchronous framework that can separate the training process from the decision-making process of Q-learning. The framework includes two stages: offline training and online decision, and the detailed design is given in Section 4.2.

#### 4.1.2. State Space

The setting of *State* space will determine whether the algorithm can converge and affect the analysis and design of reward. We employ innovative methods to categorize the data transmitted by 5G NR-V2X by *BLER*, *CQI*, *Distance* and *NumRBs*.

As is clear from the previous introduction to the HARQ process, the delay in processing by the receiver decoder has a great impact on the transmission RTT. More NACKs being returned means that more incorrect data are transmitted. It also means that the more times that retransmission is needed, the correspondingly greater the delay accumulation of the receiver decoder. Therefore, reducing the BLER can enhance the reliability of information transmission and reduce the delay. The BLER can be calculated as in Equation (2):

$$BLER = \frac{NACK}{ACK + NACK} \quad (2)$$

In addition to the important index of the downlink BLER (*BLER*), the design state space also takes the average downlink Channel Quality Indication (*CQI*), UE distance (*Distance*), and the number of resource blocks (*NumRBs*) into account. They all belong to discrete and relatively independent data, which together constitute the *State* space of the Q-learning model.

When the *Distance* ranges from 0 to 2000 m, the *NumRB* value ranges from 0 to 120. The UE distance is divided by a section every 100 m for a total of 20 sections. The number of resource blocks is divided into five intervals for a total of 20 intervals. The intervals of the UE distance (*Distance*) and the resource block numbers (*NumRBs*) are arranged and

combined, and the two state elements of the downlink BLER (*BLER*) and the downlink average CQI (*CQI*) are included to jointly establish the state space. Each divided state represents a class of transmission scenarios, including all possible transmission scenarios. Finally, the state space  $S_i$ , ( $i = 1, 2, \dots, 400$ ). The state-space table is shown in Table 1.

**Table 1.** State space.

State	BLER	CQI	Distance	NumRBs
S1	...	...	[0, 100]	[0, 20]
S2	...	...	[0, 100]	[20, 25]
S3	...	...	[0, 100]	[25, 30]
...	...	...	...	...
S400	...	...	[1900, 2000]	[115, 120]

#### 4.1.3. Action Space

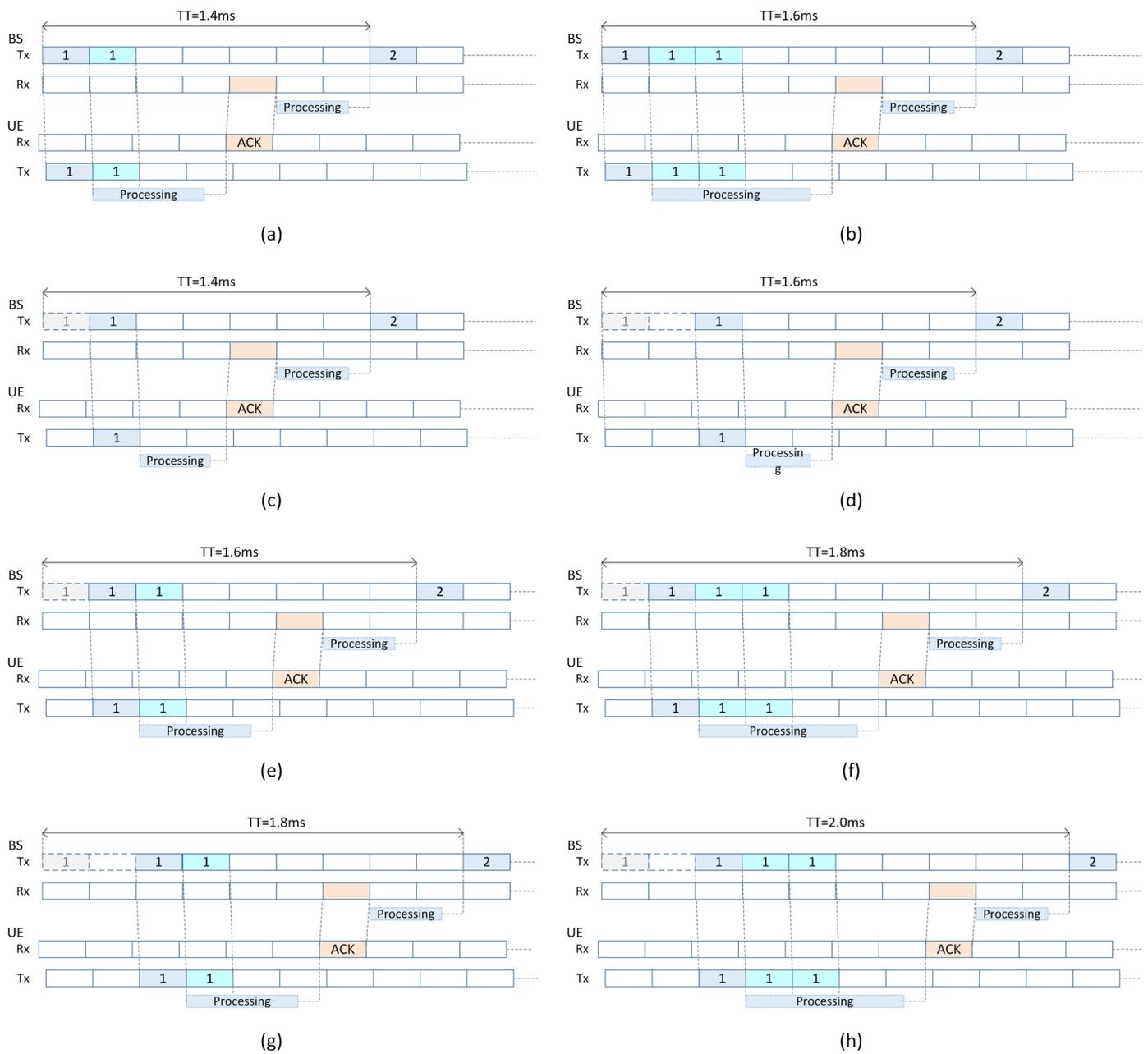
A simple and efficient action-space design can reduce the difficulty of convergence and improve the training speed. The designed action space in this study is  $A_k$ , ( $k = 0, 1, \dots, 8$ ), giving a total of nine discrete program actions, where A0 is the 5G HARQ transmission scheme shown in Figure 1, A1 and A2 are *T*-delay scheme schemes, A3 and A4 are *K*-repetition schemes, and A5–A8 are [T, *K*]-overlap schemes. Details are shown in Table 2, where A1–A8 correspond to the eight HARQ alternative transmission schemes shown in Figure 3.

**Table 2.** Action space.

Action	A-HARQ Scheme
A0	5G HARQ scheme
A1	<i>T</i> -delay scheme where $T = 1$ slot offset
A2	<i>T</i> -delay scheme where $T = 2$ slot offset
A3	<i>K</i> -repetition scheme where $K = 1$
A4	<i>K</i> -repetition scheme where $K = 2$
A5	[T, <i>K</i> ]-overlap scheme where [T, <i>K</i> ] = [1, 1]
A6	[T, <i>K</i> ]-overlap scheme where [T, <i>K</i> ] = [1, 2]
A7	[T, <i>K</i> ]-overlap scheme where [T, <i>K</i> ] = [2, 1]
A8	[T, <i>K</i> ]-overlap scheme where [T, <i>K</i> ] = [2, 2]

Although a larger number of retransmissions increases the success rate of packet transmission, the waiting time for the receiver to confirm the NACK signal and retransmitting will greatly affect the delay during retransmission. So, in order to save the time and resources required for retransmission, we limit the *K*-repetition scheme up to 3 times, that is, 1 normal transmission plus 2 consecutive retransmissions. The correct selection of the designed optimized transmission scheme can avoid the waste caused by the redundancy of channel resources when the channel quality is good and the long delay and high BLER caused by repeated retransmission when the channel quality is poor.

For example, when CQI is poor, the probability of information transmission failure is large. In this case, the information can be sent with the *T*-delay scheme (delay 1 slot or delay 2 slots) to improve the success rate of the information transmission. Because an appropriate delay transmission scheme can avoid an environment with poor CQI, there is a greater probability of reducing the average BLER while releasing more limited channel resources at the same time. It is worth noting that although it seems that the *T*-delay scheme will temporarily waste time, the fact is that, the average number of retransmissions is reduced because it may avoid interference. Consequently, both the overall transmission time and the average delay are subsequently reduced.



**Figure 3.**  $T$ -delay scheme where  $T = 1$  slot (a) and  $T = 2$  slots; (b)  $K$ -repetition scheme where  $K = 1$  (c) and  $K = 2$  (d);  $[T, K]$ -overlap scheme of  $T$ -delay and  $K$ -repetition where  $[T, K] = [1, 1]$  (e),  $[T, K] = [1, 2]$  (f),  $[T, K] = [2, 1]$  (g), and  $[T, K] = [2, 1]$  (h).

When the user is far away, there exists a high probability of information transmission failure. In this case, the information can be sent with the  $K$ -repetition scheme (send 2 or 3 repetitions) to enhance the success rate of the information transmission and reduce the average transmission delay. Because the appropriate repetition transmission mechanism can save the waiting time after the receiver confirms the NACK signal when the original retransmission is required, this reduces the average BLER. It is worth noting that the use of a continuous repetition transmission mechanism will temporarily borrow channel resources; however, due to the improvement in the transmission success rate, the original required number of retransmissions is greatly reduced, and the channel resources will also make up for it.

Moreover, when the distance is far and the CQI is poor at the same time, the  $[T, K]$ -overlap scheme of the repetition sent after delay can be selected. The selection of a trans-

mission scheme cannot focus on a single parameter but needs to comprehensively consider factors such as the CQI, UE distance, and the number of resource blocks. The 5G HARQ and eight designed alternatives together constitute the action space  $A_k, (k = 0, 1, \dots, 8)$  of the A-HARQ scheme in this study.

#### 4.1.4. Reward Calculator

The objective of the A-HARQ mechanism is to select the appropriate transmission scheme for each packet to reduce the entire retransmission in the 5G NR-V2X physical layer. The objective of Q-learning in the A-HARQ employs a learning model to choose the optimal scheme that can reduce the average BLER across various transmission environments.

Based on our design, the selection of transmission scheme (A1–A8) is used in the case that the transmission fails to meet the BLER requirements, while the transmission scheme of the original 5G HARQ (A0) is used for cases that meet the BLER requirement. Therefore, it is possible to initially set the maximum tolerable BLER threshold, denoted as  $BLER_{MAX}$ , which can be adjusted as necessary. In our observation, the BLER threshold is set as  $BLER_{MAX} = 0.098$  in the simulation environment.

The original 5G HARQ BLER parameter is denoted as  $BLER_{HARQ}$ , and the proposed adaptive HARQ BLER parameter is denoted as  $BLER_{A-HARQ}$ . The design of the Reward Calculator is mainly based on whether the BLER is not greater than the  $BLER_{MAX}$ .

When the value of the  $BLER_{HARQ}$  is smaller than the  $BLER_{MAX}$ , two reward rules,  $R_I$  and  $R_{II}$ , can be selected in the Reward Calculator, which are set as follows:

- The reward rule corresponding to the state that meets the requirements of the BLER, i.e.,  $BLER_{HARQ} \leq BLER_{MAX}$ , is  $R_I$ , defined as in Equation (3):

$$R_I = \begin{cases} 10, & Action = A_0 \\ -100, & Action \neq A_0 \end{cases} \quad (3)$$

In this case, choosing the 5G HARQ scheme (A0) will result in a positive reward, while choosing other optimized transmission schemes (A1–A8) will result in a punishment.

- The reward rule corresponding to the state that does not meet the requirements of the BLER, i.e.,  $BLER_{HARQ} > BLER_{MAX}$ , is  $R_{II}$ , defined as in Equation (4):

$$R_{II} = \begin{cases} -100, & Action = A_0 \\ R_0 + R_p, & Action \neq A_0 \end{cases} \quad (4)$$

In the case of the unsatisfied BLER threshold, the real change in BLER ( $BLER_{HARQ} - BLER_{A-HARQ}$ ) after execution of the A-HARQ decision is added as the main basis, and factors such as distance, CQI, and RB number are also considered. In this case, the selection of the 5G HARQ scheme (A0) will be punished, whereas the selection of other optimized transmission schemes (A1–A8) will be rewarded or punished correspondingly according to the key factor of the BLER change. The reward rule of  $R_{II}$  consists of the base reward  $R_0$  and the reinforcement reward  $R_p$ .

The base reward  $R_0$ , reflecting the BLER reduction, can be expressed as Equation (5):

$$R_0 = \begin{cases} -100, & BLER_{A-HARQ} \geq BLER_{HARQ} \\ \tau(BLER_{HARQ} - BLER_{A-HARQ}), & BLER_{A-HARQ} < BLER_{HARQ} \end{cases}, \quad (5)$$

where  $\tau$  is the reward coefficient, set at  $\tau = 100$ . The smaller the  $BLER_{A-HARQ}$  after the A-HARQ scheme, the greater the reward  $R_0$ .

If the transmission scheme only pursues the goal of reducing the BLER while ignoring the actual situation of the current network environment quality and available resources, it will cause a waste of limited time and space resources. Therefore, in the formulation of the reward function in this study, in addition to reducing the BLER, we also need to comprehensively consider the quality of the channel, UE distance, and available resource blocks under various circumstances. Only in this way can we achieve a reasonable allocation

of transmission time slots and limited channel resources, reduce the total transmission time, and improve the reliability of data transmission. Therefore, the reinforcement reward  $R_p$  is set to help the agent to find the optimal strategy and is defined as in Equation (6):

$$R_p = \begin{cases} 0, & BLER_{AHARQ} \geq BLER_{HARQ} \\ R_{CQI} + R_{Distance} + R_{RBs}, & BLER_{AHARQ} < BLER_{HARQ} \end{cases}, \quad (6)$$

where  $R_{CQI}$ ,  $R_{Distance}$ , and  $R_{RBs}$  are reinforcement reward values corresponding to the CQI, UE distance, and available resource blocks, respectively. They comprise reward coefficients and their normalized systems, as shown in Table 3.

**Table 3.** The weight factor of the parameter.

Action	$R_{CQI}$	$R_{Distance}$	$R_{RBs}$
A1	$(1 - \frac{CQI_i}{15}) * 25$	$(1 - \frac{Distance_i}{2000}) * 10$	$(1 - \frac{RBs_i}{120}) * 10$
A2	$(1 - \frac{CQI_i}{15}) * 28$	$(1 - \frac{Distance_i}{2000}) * 8$	$(1 - \frac{RBs_i}{120}) * 8$
A3	$\frac{CQI_i}{15} * 10$	$\frac{Distance_i}{2000} * 25$	$\frac{RBs_i}{120} * 10$
A4	$\frac{CQI_i}{15} * 8$	$\frac{Distance_i}{2000} * 28$	$\frac{RBs_i}{120} * 8$
A5	$(1 - \frac{CQI_i}{15}) * 18$	$\frac{Distance_i}{2000} * 15$	$(1 - \frac{RBs_i}{120}) * 10$
A6	$(1 - \frac{CQI_i}{15}) * 15$	$\frac{Distance_i}{2000} * 18$	$(1 - \frac{RBs_i}{120}) * 8$
A7	$(1 - \frac{CQI_i}{15}) * 15$	$\frac{Distance_i}{2000} * 10$	$(1 - \frac{RBs_i}{120}) * 10$
A8	$(1 - \frac{CQI_i}{15}) * 10$	$\frac{Distance_i}{2000} * 15$	$(1 - \frac{RBs_i}{120}) * 8$

According to the eight different optimal transmission schemes of A-HARQ, the proportions of the CQI, distance, and available resource blocks are different. For example, the channel quality is considered more important when the decision is made to use the action of delay transmission. However, the transmission distance is more important when the repetition transmission action is chosen. Accordingly, the coefficient weights of the more focused factors are the largest. On the contrary, the channel quality is an important factor to consider in the length of the time delay. And the distance is an important consideration when considering the number of repetitive transmissions. At the same time, the amount of available resource blocks is another important factor to be considered to postpone the sending time and/or to reduce the number of repetitive transmissions.

Specifically, for the CQI factor, the worse the CQI is when the  $T$ -delay scheme is selected, the bigger the reward that will be obtained, while the better the CQI is when the  $K$ -repetition scheme is selected, the bigger the reward that will be obtained. And the opposite is true for the factors of distance and the number of resource blocks. The greater the distance and the number of resource blocks, the greater the reward of repeated transmission, while the smaller the distance and number of resource blocks, the greater the reward of the delayed transmission behavior.

#### 4.2. Q-Learning Model with the 5G NR-V2X System

Referring to the study by [23], we adapt it as our asynchronous framework for the proposed A-HARQ architecture. Figure 4 illustrates the two primary stages: offline training and online decision making. From a global perspective, they operate on the same Q-Table. The offline training stage aims at building a basic Q-Table to provide support for the online decision stage. The offline training process is completed in advance, and its outcomes are compiled into the MATLAB logging file. The online decision stage focuses on both the processes of making the decision on what action will be performed and updating the Q-Table to help improve future decisions.

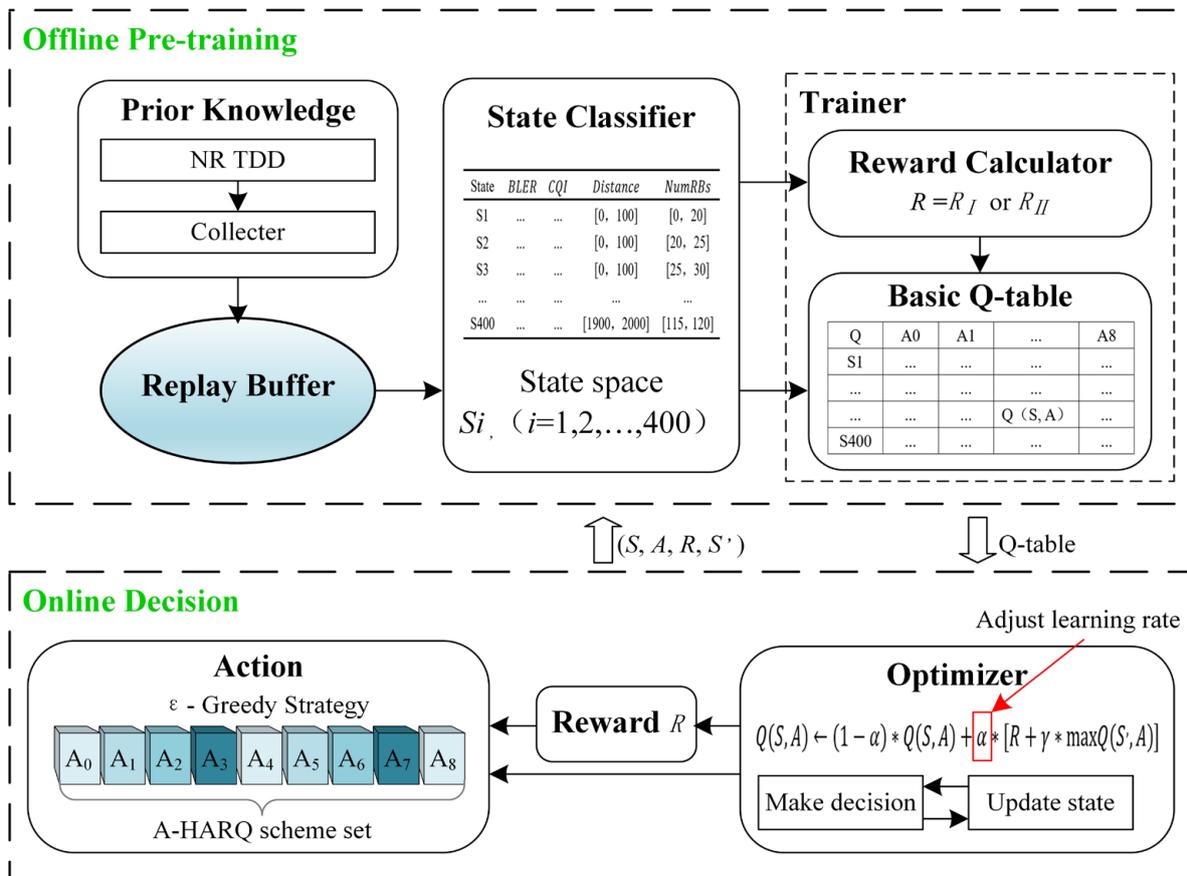


Figure 4. The asynchronous framework of the proposed mechanism.

#### 4.2.1. Offline Training

Offline training in Algorithm 1 employs four modules: *Collector*, *Replay Buffer*, *State Classifier*, and *Trainer*, as illustrated in Figure 4. This stage aims to obtain a basic rule table, which is the basis for the online decision. The rule table, named Q-Table, represents the expected reward of each executable action for all states.

The procedure of Algorithm 1 is presented as follows: Firstly, the state and action information is imported from the replay buffer, as well as the reward calculator function. The *episode*,  $BLER_{MAX}$ , learning rate  $\alpha$ , and discount factor  $\gamma$  are set respectively, and then an all-zero Q-Table is initialized. After the initial state is randomly selected, the training begins, the action is selected according to the  $\epsilon$ -greedy strategy, and the corresponding reward  $R$  is obtained according to the reward rule. Then, enter the next state after executing the current action, obtain the current Q value from the Q-Table and calculate the new Q value, and then update the Q-Table according to the Bellman Update formula. The training is performed until the end of the number of iterations. The following section provides a comprehensive description of each module.

In training, the agent collects data from the environment and stores them in the replay buffer. To keep the agent from making bad decisions, there is no prior knowledge at the beginning; the collector collects data from the environment and subsequently stores them in the replay buffer in the form  $(BLER, CQI, Distance, NumRBs)$ . The parameters represent the average downlink block error rate downlink, average channel quality indication, user distance, and number of resource blocks, respectively.

**Algorithm 1:** Offline Training Algorithm for Agent**Input:** The replay buffer**Output:** A basic Q-TableParameter set:  $episode$ ,  $BLER_{MAX}$ , learning rate  $\alpha$ , discount factor  $\gamma$ ;

Initialize the Q-Table;

Enter state  $S$  randomly from the replay buffer;**for**  $i = 1$  to  $episode$  **do**    **if**  $s < e$  **then**        | Select action  $A$  randomly;    **else**        |  $A = \text{argmax}(Q \text{ value})$ ;    **end if**        Get reward  $R$  according to the Reward Calculator;        Update  $Q$  value for each state and action using Equation (1);        Enter the next state  $S'$ ;    **end if****end for**

Then, the *State Classifier* sorts the data into different categories of the state space. According to the state division and action design in Section 4.1, 400 states (S1–S400) and 9 actions (A0–A8) are obtained to form the matrix of Q-Table of  $400 \times 9$ , as illustrated in Table 4. The initial Q-Table is initialized with all zeros, and it is filled by conducting pretraining with the samples collected from the environment.

**Table 4.** Initial Q-Table.

Q	A0	A1	...	A8
S1	...	...	...	...
...	...	...	...	...
...	...	...	Q(S, A)	...
S400	...	...	...	...

Finally, train the samples and calculate a basic Q-Table, which is also the optimal action decision table. In order to make the algorithm converge and to maintain a certain stability, the learning rate  $\alpha$  is set to the middle value of 0.5 and the discount factor  $\gamma$  is 0.8. In order to make the training effect play a greater role, the  $\epsilon$ -greedy strategy is set to select the action. The model explores the system with probability  $\epsilon$  to take random action and with probability  $\epsilon$  to exploit the Q-Table. Also, set the attenuation factor to prevent the energy from being wasted in randomly selecting the non-optimal action after convergence. The  $\epsilon$  decreases from 0.3 to 0.1 as the number of training episodes increases. The  $\epsilon$ -greedy algorithm can ensure that the state of the state space is fully traversed to ensure that the optimal strategy is finally converged.

**4.2.2. Online Decision**

In the online decision procedure, the system will utilize the Q-Table obtained during the offline training stage to execute the selected action, and it will adjust the learning rate  $\alpha$  according to the system feedback. In addition, the system will continue to collect the data to optimize the Q-Table. The algorithm of the online decision is described in Algorithm 2. The agent first determines the action, and then performs the action within the network. Next, the system computes the reward  $R$  according to Equation (4) or Equation (5) and provides feedback after performing the action. The records of  $S$ ,  $A$ ,  $R$ ,  $S'$  are saved and ready to be delivered to the buffer. The parameters represent the state, action, reward, and next state, respectively. Afterwards, update the Q-Table using Equation (1) based on the

feedback of the environment. The online decision stage includes two modules: the module of decision making and the optimizer module, which will be introduced below.

---

**Algorithm 2:** Online Decision Algorithm

---

**Input:** The basic Q-Table

**Output:** The optimal action decision Table

```

1   Parameter set: episode,  $BLER_{MAX}$ , learning rate  $\alpha$ , discount factor  $\gamma$ ;
2   Initialize the basic Q-Table;
3   for  $i = 1$  to episode do
4       Select actions  $A$  based on  $\epsilon$ -greedy strategy;
5       Compute reward  $R$  and observe next state  $S'$ ;
6       Save record  $S, A, R, S'$  into the replay buffer;
7       Update  $Q$  value for each state and action using Equation (1);
8       Enter the next state  $S'$ ;
9        $S \leftarrow S'$ ;
      Adjust the learning rate  $\alpha$  according to the training results;
10  end for

```

---

At the beginning, the agent will initialize the Q-Table in the same way as for the basic Q-Table that was calculated in the offline training stage. In decision making, we also take actions with the  $\epsilon$ -policy, which is a strategy to weaken the contradiction in reinforcement learning between exploration and exploitation. The model explores the system with a probability  $\epsilon$  of taking random action and with a probability of  $1-\epsilon$  of exploiting the Q-Table. After deciding the action, the 5G NR-V2X system will perform information transmissions according to this action.

The optimizer can adjust the learning rate of the update algorithm according to the BLER optimization. The BLER feedback from each online training system is utilized to adapt the learning rate  $\alpha$  within a certain range of [0.2, 0.8]. The general idea is that when the system feedback regarding the BLER optimization effect performed by this action is satisfactory, the learning rate  $\alpha$  increases, while when the BLER optimization effect is more general, the learning rate  $\alpha$  decreases.

## 5. Performance Evaluation

### 5.1. Simulation Setup

The procedures for constructing the simulation are described as follows. Set the channel bandwidth to 5 MHz and the sub-carrier space (SCS) to 15 kHz, as specified in Section 5.3.2 of the 3GPP TS 38.104 [22]. The simulation employs a lookup table to map the received signal's interference-and-noise ratio (SINR) to the CQI index for a 0.1 BLER [19]. The lookup table corresponds to the CQI table as specified in table of 5.2.2.1-3 of the 3GPP TS 38.214. The complete bandwidth is assumed to be allotted for the PUSCH/PDSCH. The channel quality is periodically improved or deteriorated by 1 every 0.2 s for all RBs of a UE. Whether the channel conditions for a particular UE improve or deteriorate is randomly determined. The initial value of the CQI for each RB and for each UE is given randomly and is limited by the maximum achievable CQI value corresponding to the distance of the UE from the gNB. According to the 3GPP TS 38.323, the maximum radio-link control (RLC) session data unit (SDU) length is 9000 bytes [24]. The complete bandwidth is assumed to be allocated for the PUSCH or PDSCH. The above data will be used as the offline training data for this study. Other parameters in this study are set as follows:

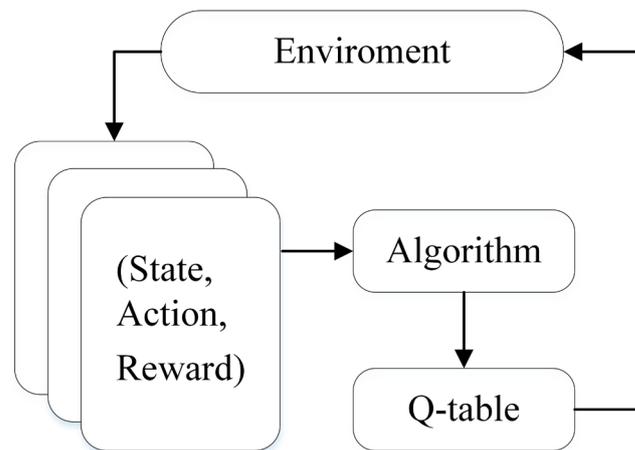
- The subcarrier spacing: 15 KHz
- Periodicity (in ms) at which the UL packets are generated by UEs: 30 ms
- Size of the UL packets (in bytes) generated by UEs: 5000 B
- Periodicity (in ms) at which the DL packets are generated for UEs at gNB: 20 ms

- Size of the DL packets generated (in bytes) for UEs at gNB: 6000 B

We create 4 mobile users in the simulation environment, and an equidistance between each UE and the base station. According to the state division in Section 4.1.2, after modifying and debugging the corresponding user *Distance* and *NumRBs* environment parameters in the built simulation environment, the corresponding downlink average CQI and downlink average BLER are obtained. The above data will be used as the offline training data in this study, and a total of 400 sets of simulation data will be obtained.

The principle of the offline training and the online decision-making algorithm is basically the same. However, the initial Q-Tables of the two are different, and the learning rate in the online decision-making algorithm will be adjusted according to the system feedback. The initial state of each learning sets the random state selection strategy, so that all states have the possibility of traversing it. The Q-Table is updated according to the  $\epsilon$ -greedy strategy, and when the set number of training times of end convergence is reached, the agent will stop training.

Each state space in this study is an independent simulation experiment. The process of training is not a sequential accumulation of continuous actions and feedback, but each training iteration is relatively independent. The state selection process is shown in Figure 5.



**Figure 5.** State selection diagram.

According to the Monte Carlo algorithm, the agent should be able to traverse 9 actions in 400 states to ensure the effectiveness of the algorithm, so the training episodes should meet at least 3600 times. Secondly, in accordance with the principle of Markov chains, it is expected that increasing the number of training times will result in more obvious signs of convergence. In order to make the agent more certain about the choice of the best strategy, the termination condition for the offline training is set at 100,000 training iterations.

## 5.2. Performance Analysis

This section presents an analysis of the performance evaluation of the proposed A-HARQ scheme. The evaluation encompasses simulations conducted within a heterogeneous network environment. The assessment is performed in terms of the BLER, retransmission times, and RTT. The simulation results of the proposed A-HARQ scheme are compared with that of the 5G HARQ scheme. This part only analyzes the results of selecting the optimal transmission scheme for the state that does not meet the requirements of the BLER (i.e., actions A1–A8), because the state that meets the requirements of the BLER maintains the transmission scheme of 5G HARQ (i.e., action A0).

This study guides agents to optimize effective information by calculating rewards according to the feedback of system BLER changes after performing actions and adjusting learning rate strategies. The design of the reward mechanism can overcome the problem of sparse returns and can converge. By utilizing the available prior knowledge data for offline

training, a convergent basic Q-Table is derived. After tens of thousands of times of training, the agent learns the action strategy that can obtain the biggest reward, and the training data in each state will have an obvious tendency to converge to a certain action. Through the offline training simulation, the optimal Q-Table is obtained, and the action corresponding to the maximum Q value in each state is the optimal action. Figure 6 illustrates the ultimate optimal information transmission scheme corresponding to states S1 to S400, respectively.



Figure 6. Action distribution diagram for each state.

According to the final optimal action table, an online decision is made, and performance analysis is conducted on the entire sample space, which consists of 400 states. Figure 7 illustrates the downlink average BLER performance of the A-HARQ scheme and the 5G HARQ scheme. The simulation results demonstrate that the BLER performance of the proposed Adaptive Hybrid Automatic Retransmission Request Technical Solution (A-HARQ) is significantly better than that of the 5G HARQ. The A-HARQ scheme achieved a reduction in the maximum BLER from 11.48% to 10.22% compared with the HARQ scheme. The overall downlink average BLER decreased by 4.8 percentage points from 10.35% to 5.55%, which also means that the average information transmission accuracy performance increased by 46.41%.

Figure 8 depicts the performance comparison of RTT and retransmission times between the A-HARQ scheme and the HARQ scheme. The simulation results demonstrate that that the A-HARQ scheme enables a reduction in the RTT and retransmission instances compared with the HARQ scheme, as shown in Figure 8. The final RTT is shortened by 35.23% and 56.82%, respectively, on average, i.e., from 2.4 ms and 3.6 ms to 1.55 ms, respectively. The average reduction in the number of retransmissions is 2.31% and 51.16% for the first and second cases, respectively. Specifically, the number of retransmissions decreased from 1 to 0.97 for the first case, and from 2 to 0.97 for the second case.

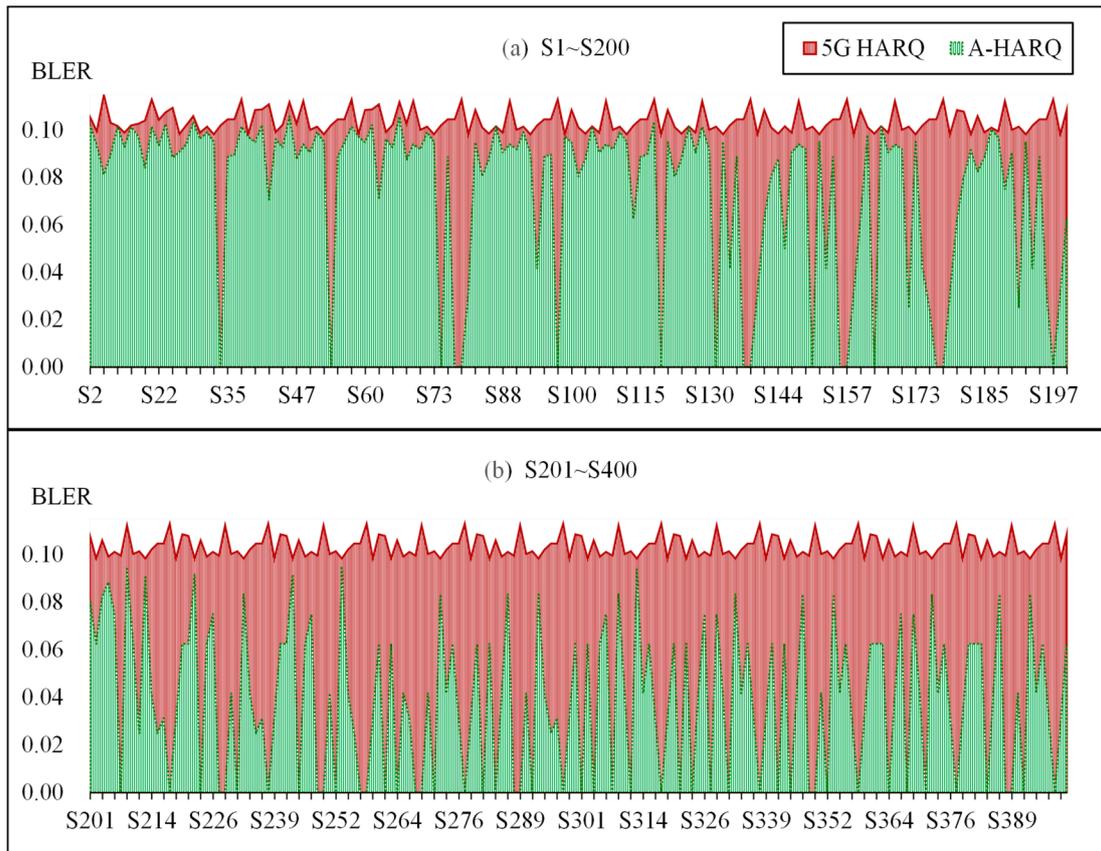


Figure 7. Average downlink BLER performance comparison between 5G HARQ and A-HARQ.

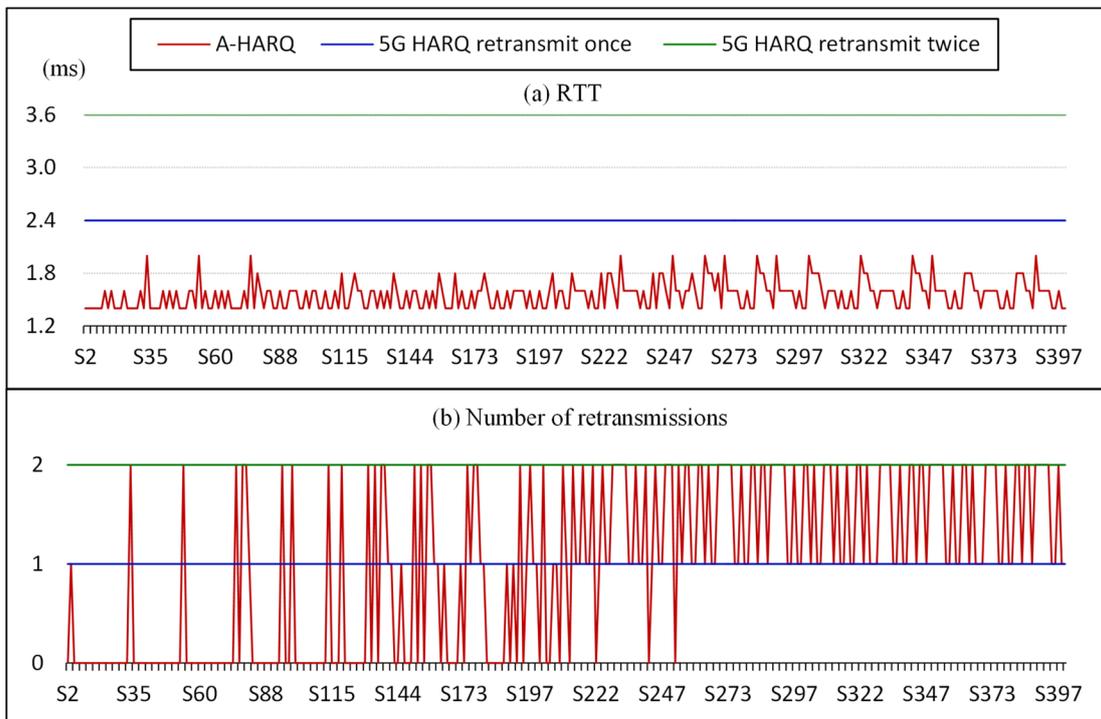


Figure 8. RTT and the number of retransmissions performance comparison between 5G HARQ and A-HARQ.

## 6. Conclusions

In this paper, we examine the communication performance of a 5G NR-V2X within an asynchronous framework that combines offline training and online decision making with a Q-learning model. In order to mitigate the waste caused by redundant channel resources when the channel quality is good and the high delay and large BLERs caused by repeated retransmissions when the channel quality is poor, the  $T$ -delay scheme,  $K$ -repetition scheme and  $[T, K]$ -overlap scheme including 8 optimal transmission schemes are designed. In this study, the A-HARQ mechanism is realized through the decision-making ability of Q-learning's optimal learning model. The proposed A-HARQ for environmental awareness aims to optimize the utilization of limited channel resources and adaptively select the most appropriate actions based on the current system state. Finally, this approach enables the highest effective utilization of existing resources and channel quality and achieves the purpose of reducing the average BLER, average retransmission times, and RTT. In future studies, we will explore the utilization of neural networks to optimize the design of reward mechanisms, with the aim of achieving greater perfection.

**Author Contributions:** Conceptualization, M.-H.Y. and S.J.; methodology, M.-H.Y. and S.-Y.L.; software, M.-H.Y., S.J. and S.-Y.L.; validation, M.-H.Y. and S.-Y.L.; formal analysis, M.-H.Y.; investigation, M.-H.Y. and S.J.; resources, S.-Y.L.; data curation, M.-H.Y. and S.J.; writing—original draft preparation, M.-H.Y.; writing—review and editing, S.-Y.L.; visualization, M.-H.Y.; supervision, S.-Y.L.; project administration, S.-Y.L.; funding acquisition, S.-Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No new data were created during the study.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. 3GPP (3rd Generation Partnership Project). *Enhancement of 3GPP Support for V2X Scenarios*, 3GPP TS 22.186, v17.0.0; 3GPP: Valbonne, France, 2022.
2. Vangelista, L.; Centenaro, M. Performance Evaluation of HARQ Schemes for the Internet of Things. *Computers* **2018**, *7*, 48. [[CrossRef](#)]
3. Berardinelli, G.; Khosravirad, S.R.; Pedersen, K.I.; Frederiksen, F.; Mogensen, P. Enabling Early HARQ Feedback in 5G Networks. In Proceedings of the IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 15–18 May 2016; pp. 1–5.
4. AlMarshed, S.; Triantafyllopoulou, D.; Moessner, K. Supervised Learning for Enhanced Early HARQ Feedback Prediction in URLLC. In Proceedings of the IEEE International Conference on Communication, Networks and Satellite, Batam, Indonesia, 17–18 December 2020; pp. 26–31.
5. Wang, Q.; Cai, S.; Lin, W.; Zhao, S.; Chen, L.; Ma, X. Spatially Coupled LDPC Codes via Partial Superposition and Their Application to HARQ. *IEEE Trans. Veh. Technol.* **2021**, *70*, 3493–3504. [[CrossRef](#)]
6. Yeo, J.; Bang, J.; Ji, H.; Kim, Y.; Lee, J. Outer Code-Based HARQ Retransmission for Multicast/Broadcast Services in 5G. In Proceedings of the IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 22–25 September 2019; pp. 1–5.
7. Shirvanimoghaddam, M.; Khayami, H.; Li, Y.; Vucetic, B. Dynamic HARQ with Guaranteed Delay. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Republic of Korea, 25–28 May 2020; pp. 1–6.
8. Goektepe, B.; Faehse, S.; Thiele, L.; Schierl, T.; Hellge, C. Subcode-Based Early HARQ for 5G. In Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
9. Strodthoff, N.; Göktepe, B.; Schierl, T.; Samek, W.; Hellge, C. Machine Learning for Early HARQ Feedback Prediction in 5G. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6.
10. Jang, H.; Kim, J.; Yoo, W.; Chung, J.-M. URLLC Mode Optimal Resource Allocation to Support HARQ in 5G Wireless Networks. *IEEE Access* **2020**, *8*, 126797–126804. [[CrossRef](#)]
11. Li, A.; Wu, S.; Jiao, J.; Zhang, N.; Zhang, Q. Age of Information with Hybrid-ARQ: A Unified Explicit Result. *IEEE Trans. Commun.* **2022**, *70*, 7899–7914. [[CrossRef](#)]
12. Liu, Y.; Deng, Y.; Elakashan, M.; Nallanathan, A.; Karagiannidis, G.K. Analyzing Grant-Free Access for URLLC Service. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 741–755. [[CrossRef](#)]
13. Fei, Z.; Teng, G. Research on an improved HARQ method based on 5g HARQ. *Chang. Inf. Commun.* **2023**, *36*, 210–212.

14. 3GPP (3rd Generation Partnership Project). 5G; NR; *Physical Layer Procedures for Data*, 3GPP TS 38.214, v15.9.0; 3GPP: Valbonne, France, 2020.
15. Hua, G.D.; Li, J.H.; Wang, Z.; Li, N.; Wang, Z. LTE-V2X Test Based on PC5/Uu Mode. *Automot. Digest* **2022**, *08*, 24–30.
16. Wang, W.L.; Shi, S.H.; Zhang, H.B.; Cai, L. Research on Key Communication Technologies for Vehicle-Road Co-ordination. *J. Highw. Transp. Sci. Technol. (Appl. Technol. Ed.)* **2020**, *16*, 311–315.
17. Lin, J.X.; Zheng, L.; Zong, Y.; Wang, Z.J. Port vehicle-road coordination based on dedicated short-range communication. *J. Shanghai Inst. Shipp. Transp. Sci.* **2021**, *44*, 56–62.
18. Han, W.; Wang, Y.H.; Qin, D. Application of URLLC Technology in Electric Vehicles. *Electr. Meas. Instrum.* **2021**, *58*, 81–86.
19. 3GPP (3rd Generation Partnership Project). NR; *Medium Access Control (MAC) Protocol Specification*, 3GPP TS 38.321, v15.6.0; 3GPP: Valbonne, France, 2019.
20. 3GPP (3rd Generation Partnership Project). NR; *Radio Link Control (RLC) Protocol Specification*, 3GPP TS 38.322, v15.5.0; 3GPP: Valbonne, France, 2019.
21. 3GPP (3rd Generation Partnership Project). NR; *Radio Resource Control (RRC) Protocol Specification*, 3GPP TS 38.331, v15.6.0; 3GPP: Valbonne, France, 2019.
22. 3GPP (3rd Generation Partnership Project). NR; *Base Station (BS) Radio Transmission and Reception*, 3GPP TS 38.104, v16.0.0; 3GPP: Valbonne, France, 2019.
23. Zhong, L.; Ji, X.; Wang, Z.; Qin, J.; Muntean, G.-M. A Q-Learning Driven Energy-Aware Multipath Transmission Solution for 5G Media Services. *IEEE Trans. Broadcast.* **2022**, *68*, 559–571. [[CrossRef](#)]
24. 3GPP (3rd Generation Partnership Project). NR; *Packet Data Convergence Protocol (PDCP) Specification*, 3GPP TS 38.323, v16.2.0; 3GPP: Valbonne, France, 2020.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.