



Article

# Deep Q-Learning Based Adaptive MAC Protocol with Collision Avoidance and Efficient Power Control for UWSNs

Wazir Ur Rahman <sup>1,2,3</sup>, Qiao Gang <sup>1,2,3</sup>, Feng Zhou <sup>1,2,3,\*</sup>, Muhammad Tahir <sup>4</sup> , Wasiaq Ali <sup>1,2,3</sup>, Muhammad Adil <sup>1,2,3</sup> and Muhammad Ilyas Khattak <sup>5</sup> 

- <sup>1</sup> National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China; wazirrahman@hrbeu.edu.cn (W.U.R.); qiaogang@hrbeu.edu.cn (Q.G.); wasiqali@hrbeu.edu.cn (W.A.); adil@hrbeu.edu.cn (M.A.)
- <sup>2</sup> Key Laboratory of Marine Information Acquisition and Security, Harbin Engineering University, Ministry of Industry and Information Technology, Harbin 150001, China
- <sup>3</sup> College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
- <sup>4</sup> Department of Engineering and Computer Science, NUML Faisalabad Campus, Faisalabad 38000, Pakistan; engr.tahir1987@gmail.com
- <sup>5</sup> School of Control Science and Engineering, Shandong University, Jinan 250100, China; ilyas@mail.sdu.edu.cn
- \* Correspondence: zhoufeng@hrbeu.edu.cn

**Abstract:** Underwater wireless sensor networks (UWSNs) widely used for maritime object detection or for monitoring of oceanic parameters that plays vital role prediction of tsunami to life-cycle of marine species by deploying sensor nodes at random locations. However, the dynamic and unpredictable underwater environment poses significant challenges in communication, including interference, collisions, and energy inefficiency. In changing underwater environment to make routing possible among nodes or/and base station (BS) an adaptive receiver-initiated deep adaptive with power control and collision avoidance MAC (DAWPC-MAC) protocol is proposed to address the challenges of interference, collisions, and energy inefficiency. The proposed framework is based on Deep Q-Learning (DQN) to optimize network performance by enhancing collision avoidance in a varying sensor locations, conserving energy in changing path loss with respect to time and depth and reducing number of relaying nodes to make communication reliable and ensuring synchronization. The dynamic and unpredictable underwater environment, shaped by variations in environmental parameters such as temperature (T) with respect to latitude, longitude, and depth, is carefully considered in the design of the proposed MAC protocol. Sensor nodes are enabled to adaptively schedule wake-up times and efficiently control transmission power to communicate with other sensor nodes and/or courier node plays vital role in routing for data collection and forwarding. DAWPC-MAC ensures energy-efficient and reliable time-sensitive data transmission, improving the packet delivery ratio (PDR) by 14%, throughput by over 70%, and utility by more than 60% compared to existing methods like TDTSPC-MAC, DC-MAC, and ALOHA MAC. These enhancements significantly contribute to network longevity and operational efficiency in time-critical underwater applications.

**Keywords:** Deep Q-learning (DQN); collision avoidance; energy efficiency; throughput; MAC protocol; underwater wireless sensor network (UWSN)



Academic Editors: Panagiotis Trakadas, Anastasios E. Giannopoulos and Nikolaos Nomikos

Received: 23 February 2025

Revised: 13 March 2025

Accepted: 15 March 2025

Published: 20 March 2025

**Citation:** Rahman, W.U.; Gang, Q.; Zhou, F.; Tahir, M.; Ali, W.; Adil, M.; Khattak, M.I. Deep Q-Learning Based Adaptive MAC Protocol with Collision Avoidance and Efficient Power Control for UWSNs. *J. Mar. Sci. Eng.* **2025**, *13*, 616. <https://doi.org/10.3390/jmse13030616>

**Copyright:** © 2025 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

UWSNs have emerged as a groundbreaking technology, revolutionizing communication and data collection in the underwater domain. These networks, composed of spatially

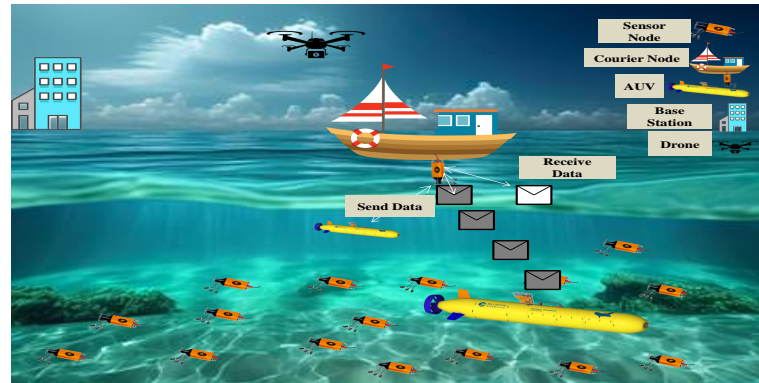
distributed sensors that utilize acoustic signals, enable the monitoring and transmission of critical information, overcoming the unique physical challenges posed by the underwater environment. UWSNs have expanded the scope of applications, including oceanographic mapping, marine biodiversity monitoring, underwater earthquake detection, maritime navigation, and disaster prediction [1]. They also play a pivotal role in tactical surveillance and resource management, providing real-time data to improve our understanding of underwater ecosystems and address pressing global challenges. However, the underwater environment presents significant challenges including high bit error rates (BER), often caused by environmental interference, and the inherent vulnerability of acoustic channels increases the likelihood of communication failures. Additional complications include fluctuating channel conditions, limited bandwidth (B), and limited channel capacity (C), all of which can degrade network performance. These harsh conditions underscore the necessity of robust MAC protocols. Effective MAC protocols are critical for managing channel access, mitigating interference, and ensuring reliable communication under challenging conditions [2].

UWSNs preferably rely on acoustic waves rather than radio frequency (RF) signals, while still offering merits such as a high penetration rate. Although acoustic waves have slower propagation speeds (1500 m/s) compared to RF signals. They are preferred for underwater communication due to their ability to travel for longer distances with less attenuation [3,4]. However, UWSN communication protocols encounter substantial obstacles, such as restricted bandwidth, elevated bit error rates, and the sluggish speed of acoustic signal propagation [5]. The MAC protocol, an essential component of the data link layer, oversees the allocation of access to the communication channel among nodes within a network [6]. Efficient, fair, and coordinated resource sharing is essential for upholding network performance. In underwater acoustic networks (UASNs), the significance of MAC protocols is further emphasized by the distinct challenges posed by the underwater environment. These challenges encompass dynamic propagation delays, restricted bandwidth, modest communication rates, and the substantial energy consumption of underwater devices. Moreover, the ever-changing network topology and charging hurdles add complexity to ensuring stable communication, highlighting the necessity for addressing these aspects in protocol designs such as the MAC layer [7].

MAC protocols for UASNs are divided into contention-free and contention-based categories [8]. Contention-free protocols, such as TDMA, FDMA, and CDMA, assign dedicated resources for communication, with TDMA being renowned for its efficient data throughput via centralized scheduling [9]. However, these protocols rely on global network knowledge, which is difficult to obtain in dynamic underwater environments. Contention-based protocols, including handshaking-free and handshaking-based types, aim to minimize collisions in a distributed manner [10,11]. While handshaking-based protocols address the hidden-terminal problem, their effectiveness is limited by long underwater propagation delays, reducing acoustic channel efficiency. The ALOHA-Q protocol has been modified for underwater acoustic networks, implementing asynchronous operations and decreasing the number of time slots in each frame to enhance channel utilization [12]. In [13], the authors employed the DR-DLMA protocol, leveraging the DQN algorithm to determine an efficient transmission mode in a heterogeneous network with diverse protocol nodes.

Figure 1 depicts structural design of UWSN that includes AUV, sensor nodes for monitoring and courier node for collecting and receiving data from deployed nodes. Courier node along with AUV also act as a bridge between sensor nodes and base station (BS). AUV/Courier node acting as sender incorporate preamble for better synchronization and sensor nodes acting as receiver incorporates Wake up and ACK message for efficient MAC

protocol design in dynamic underwater environment. To address these obstacles, advanced MAC protocols must be designed to optimize resource allocation, enhance network reliability, and ensure efficiency in highly dynamic and resource-constrained underwater environments. Developing such protocols is essential for the seamless operation of UWSNs, enabling efficient resource management, reliable data transmission.



**Figure 1.** Structural design of underwater wireless sensor networks.

This paper introduces a novel DQN-optimization based DAWPC-MAC receiver-initiated MAC protocol that minimizes energy consumption by allowing the courier node as transmitter (Tx) and receiver node (Rx) to control its wake-up schedule, ensuring synchronization with other transmitter (Tx) and receiver (Rx) nodes, and prioritizing critical data traffic. The integration of DQN-based optimization, the protocol can adapt to varying network conditions, optimizing power consumption while maintaining communication reliability. Our main focus is on DQN, a robust method in reinforcement learning. The goal is for the agent to interact with its environment and improve cumulative rewards over time. Unlike traditional Q-learning, DQN utilizes a Deep Neural Network (DNN) to approximate the Q-value function, allowing the agent to traverse complex, multi-dimensional state spaces. This strategy-driven approach enables the system to continuously refine its decision-making by learning optimal actions through experience and adapting its behavior based on past environmental interactions. The use of DQN in MAC protocols for courier nodes as receiver nodes offers significant benefits by enabling adaptive decision-making. DQN helps courier node as receiver nodes optimize resource allocation, minimize transmission conflicts, and dynamically adjust to fluctuating network conditions. By learning from past experiences, the courier node as receiver can improve its communication strategies, enhancing network efficiency and reducing collisions. It allows the system to scale effectively in complex environments, ensuring better performance and resource management in dynamic networks with energy-efficient collision avoidance. Additionally, DQN transforms conventional Q-learning by utilizing neural networks to estimate the Q-value function, effectively addressing the constraints of Q-tables in extensive state-action spaces. By leveraging function approximation, DQN enhances scalability and adaptability, enabling efficient decision-making in dynamic and continuous state environments. This makes DQN particularly suitable for complex scenarios such as underwater acoustic communication networks, where state transitions are highly dynamic.

The proposed framework introduces the following noted contributions in the MAC protocol for effective power control and collision avoidance in UWSN:

- The protocol minimizes energy consumption by dynamically adjusting wake-up schedules and integrating adaptive power control mechanisms, thereby significantly extending network lifespan.

- Enhanced synchronization between transmitter and receiver nodes and courier nodes as transmitter and as receiver improves communication reliability, reduces delays, and mitigates packet collisions, ensuring efficient operation under diverse conditions.
- A novel collision-avoidance mechanism combined with priority-based communication optimizes resource utilization, guarantees timely delivery of critical data, and significantly improves system throughput.
- The adaptable and versatile design caters to different traffic priorities and network configurations, allowing the protocol to be customized for a diverse set of underwater applications, all while preserving excellent performance.

The paper follows this structure: Section 2 presents the literature review, Section 3 outlines the proposed system model, Section 4 showcases the performance evaluation, and Section 6 provides the conclusion of the paper.

## 2. Literature Review

The existing literature on MAC and routing protocols for UWSNs addresses challenges like long propagation delays, low data rates, high energy consumption, and collision management. Early works, such as RIPT [14], introduced receiver-initiated protocols to enhance throughput and reduce collisions. Handshaking-based methods [15] improved channel utilization through synchronization but faced issues with synchronization precision and latency. Enhanced Aloha variants [16] tackled throughput limitations but struggled in dense networks due to persistent collision risks. More recent studies have leveraged reinforcement learning. For example, Q-learning-based multi-hop MAC protocols [17], DR-D3QN-MA for time-slot optimization [18], and MR-SFAMA-Q for multi-receiver handshake optimization [19] have shown adaptability and efficiency. However, challenges with real-world validation, scalability, and energy efficiency remain. A MACA-based energy-efficient protocol using Q-learning [20] and a Q-learning-enhanced framed Aloha approach [21] demonstrated improved throughput and delay management but still faced limitations in multi-hop energy efficiency and collision avoidance.

Despite these advancements, many protocols lack the integration of a DQN-based MAC-routing approach. This integration is necessary for achieving robust multi-hop underwater networks. DQN-based frameworks typically focus on hybrid protocols that incorporate adaptive learning techniques to optimize scalability, reduce computational complexity, and enhance real-world applicability. However, conventional TDMA-based scheduling [22] mitigated collision rates but led to inefficiencies under varying traffic conditions due to rigid slot allocation. Traffic-aware MAC protocols [23] improved fairness but lacked routing integration, while reinforcement learning-based MAC schemes for multimedia [24] optimized throughput but increased computational complexity. Dynamic slot scheduling [25] enhanced adaptability but raised processing overhead. Contention-free MAC [26] prioritized critical data but neglected low-priority traffic, and spatially fair MAC [27] improved network stability but lacked dynamic adaptation. EE-UWSNs have integrated MAC and routing for energy efficiency [28], yet real-world validation remains limited. Recent reinforcement learning-based strategies [29] have enhanced scalability and adaptability but require optimization to reduce computational overhead. While these advancements improve UWSN performance, challenges persist in scalability, multi-hop integration, and real-world adaptability. This highlights the need for hybrid approaches that seamlessly integrate DQN based optimization of MAC and routing layers while leveraging advanced learning techniques for improved efficiency and robustness. The summary of the of Literature Review on MAC Protocols for UWSNs are given in Table 1.

**Table 1.** Summary of Literature Review on MAC Protocols for UWSNs.

Reference	Approach	Pros	Cons
RIPT [14]	Receiver-Initiated MAC	Reduces collisions, improves throughput	Synchronization precision issues, increased latency
Handshaking-based [15]	Synchronization-based MAC	Better channel utilization	High latency, synchronization challenges
Aloha Variants [16]	Random Access MAC	Improved throughput in low-density networks	High collision risk in dense networks
Q-learning MAC [17]	Reinforcement Learning MAC	Adaptive and efficient	Scalability issues, high computational overhead
DR-D3QN-MA [18]	Time-slot Optimization	Better slot allocation	Lacks real-world validation, energy inefficiency
MR-SFAMA-Q [19]	Multi-Receiver Handshake	Increased adaptability	High complexity, energy consumption
MACA-Q [20]	Q-learning MACA	Energy-efficient, improved throughput	Multi-hop inefficiency, collision risks
Framed Aloha-Q [21]	Q-learning Framed Aloha	Reduces delay	Poor multi-hop efficiency, residual collisions
TDMA Scheduling [22]	Time-Division MAC	Low collision rate	Inefficient under dynamic traffic conditions
Traffic-Aware MAC [23]	Adaptive Scheduling	Improves fairness	Lacks routing integration
RL-Multimedia MAC [24]	RL for Multimedia	Optimized throughput	High computational complexity
Dynamic Slot Scheduling [25]	Adaptive Time Slot Allocation	Enhances adaptability	High processing overhead
Contention-Free MAC [26]	Priority-based MAC	Ensures critical data transmission	Ignores lower-priority traffic
Spatially Fair MAC [27]	Fairness-Based MAC	Improves network stability	Lacks dynamic adaptation
EE-UWSNs [28]	Energy-Efficient MAC	Energy-aware routing integration	Limited real-world validation
RL-Based MAC [29]	Reinforcement Learning MAC	Enhances scalability	High computational overhead
Proposed Work	DQN-based MAC-Routing	Dynamic optimization, energy-efficient, scalable	Requires real-world testing and fine-tuning

In addition, the authors of [30–33] highlight recent advances in maritime computing and communication, with special attention to efficiency, privacy, and resource minimization. Specifically, ref. [30] proposes a space-air-ground-sea integrated network, mobile edge computing, and deep reinforcement learning to minimize communication and computation resources, with improved efficiency. In contrast, ref. [31] condemns centralized machine learning for sea use due to bandwidth, energy, and privacy constraints and instead advocates the federated strategy (FedShip) using over-the-air computing (AirComp) to improve energy and spectrum efficiency. However, redundancy is present in the third and fourth of the last three papers on parallel conclusions regarding the superiority of FedShip over CML, namely fuel consumption prediction and data privacy [31–33]. Although such studies



provide optimistic solutions, a significant drawback is the lack of real-world application and scalability evaluation of these solutions in dynamic marine environments.

The authors in [34–36] emphasize significant advances in UWSNs and maritime communication networks and focus on the role that RL and 6G technologies play in addressing significant challenges. Specifically, refs. [34,35] explain RL-based routing protocols for UWSNs with emphasis on addressing challenges such as high energy consumption, low bandwidth, and long propagation delays. These works compare various RL-based routing protocols using optimization criteria, performance evaluation, and application applicability, and also identify the directions of future research. However, redundancy of [34,35] suggests that the two documents report similar material without contributing substantial new information. Meanwhile, ref. [36] directs the focus towards sea communication networks and describes the ways in which 6G technology can advance intelligent transportation systems, smart harbors, and ocean monitoring. This study provides a broader context for the marine communication infrastructure by integrating satellite, aerial, ground, and sea networks. Although these papers are dense in contribution, the major limitation is that there is not much large-scale experimental testing for RL-based UWSN protocols and real-world application cases for 6G MCNs. In the future, real-world implementation and hybrid approaches that combine RL and emerging network technologies to increase efficiency and reliability need to be addressed.

Existing research on MAC and routing protocols for UWSNs tackles notable hurdles including high propagation delay, decreased data rates, elevated energy consumption, and collision control. Handshaking-based protocols and receiver-initiated protocols improve channel utilization and reduce collisions but suffer from synchronization issues as well as increased latency. Optimized Aloha variants improve throughput but perform badly in dense networks since they entail constant collision risks. Reinforcement learning-based MAC protocols like Q-learning methods such as DR-D3QN-MA and MR-SFAMA-Q are time-slot allocation and multi-receiver handshaking flexible but are not implemented in real-world applications, are plagued by scalability issues, and have high computational complexity. Hybrid MAC-routing methods like EE-UWSNs are energy efficient but remain limited in multi-hop transmission and real-world application. TDMA-based scheduling reduces collision rates but is wasteful in case of changing traffic conditions, while traffic-aware MAC protocols optimize fairness but lack the integration of dynamic routing. Contention-free MAC gives priority to emergency data but leads to data starvation for low-priority traffic, and spatially fair MAC offers improved network stability but lacks adaptability with changing network conditions. While the current literature has made significant contributions, most of them fail to include deep reinforcement learning (DRL) techniques, for instance, DQN, for scalability and efficiency enhancements. The proposed work aims to overcome this by designing a DQN-augmented MAC-routing model that adjusts both layers dynamically to improve scalability, efficiency in multi-hop communication, and energy efficiency at reduced computational complexity. In contrast to conventional RL techniques that require enormous processing capacities, our method employs optimized state-action representations that reduce complexity without compromising on strong network performance. To tackle the gap between theoretical models and real-world applications, we engage in intense simulations inspired by real-world underwater communication scenarios to achieve a more efficient, scalable, and energy-aware underwater communication system.

### 3. Proposed System Model

The system model being proposed comprises an N-node UWSN, where information is sent to a central sink node through acoustic communication. The network model consists of one sink node (Rx), one courier node, and multiple transmitter (Tx) nodes communi-

cating through time-slotted channels. The Rx node manages the wake-up signals and communication windows to ensure optimal data flow. Each frame is split into  $S$  time slots, and each transmitter sends exactly one packet per frame, using a distinct time slot to avoid interference with other transmitters. To avoid collisions in the receiver, nodes are selected to transmit in the available slots using a receiver initiated wake-up mechanism along with power control techniques. When a packet is received, the Rx node broadcasts an acknowledgment (ACK) indicating successful delivery. Critical data is prioritized to ensure that time-sensitive packets are transmitted promptly, while non-critical data is managed based on network capacity. The proposed DAWPC-MAC protocol address collision challenges, enhance energy efficiency, and maximize channel utilization using reinforcement learning. Under this framework, Rx nodes function as learning agents, and their transmission strategies are updated in response to feedback from the sink node regarding previous actions. Decision-making is optimized by DQN through the updating of Q-values to improve long-term communication efficiency. Among these features, power adaptation for extending the lifetime of battery devices, avoidance of idle listening, monitoring of peer requests, and scaling according to the requested volume of messages under the specific scale of the deadline network are its key features. By incorporating DQN, robust performance is achieved in delivering critical data while maintaining high energy efficiency and scalability for diverse underwater applications.

### 3.1. Reinforcement Learning Technique for Proposed DAWPC-MAC Protocol

Reinforcement Learning (RL) is the foundation of the DAWPC-MAC protocol detailed above. Within this protocol, local receiver nodes perform intelligent communication management in underwater networks to enhance localization. In reinforcement learning, a local agent (the courier node, representing the receiver node here) is trained to make optimal decisions to maximize a reward signal over time. The agent operates within the network environment, taking actions in discrete time slots based on the current policy and receiving rewards as feedback. This feedback in reinforcement learning guides the agent's learning process through trial and error, thereby enhancing strategies to optimize network performance.

The courier node, acting as the receiver node, monitors the current network status, including the number of nodes seeking access and their priority levels. It then determines an appropriate course of action, such as assigning specific time slots to nodes, based on a learned policy. This policy is maintained and refined using predictions generated from the Q-value function. This function links state-action pairs to their anticipated total rewards while considering discounts. Unlike conventional Q-learning methods that utilize a local table for reference, DQN employs a Deep Neural Network DNN to address the continuous and complex state-action landscape typical of underwater networks. The Q-values are continuously updated through iterations based on the Bellman equation shown in Equation (1).

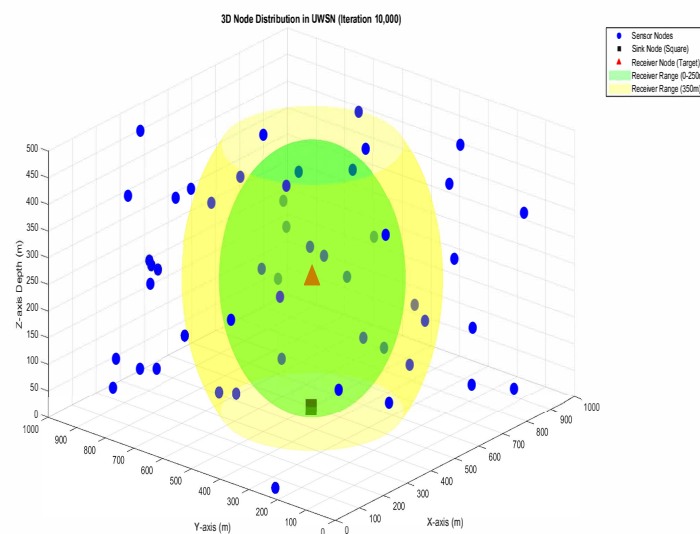
$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha \left( r_t + \gamma \max_a Q(s_{t+1}, a) \right), \quad (1)$$

where  $\alpha$  is the learning rate,  $\gamma$  is the discount factor,  $r_t$  is the immediate reward, and  $s_t, a_t$  represent the current state and action, respectively. The reward function is carefully designed to incentivize energy-efficient operation, reduced collisions, and timely delivery of critical data. Through repeated interactions, the DQN optimizes the protocol's performance by balancing energy consumption, channel utilization, and packet delivery priorities. In summary, by incorporating RL through DQN, the DAWPC-MAC protocol dynamically adapts to changing network conditions, learns to resolve colli-

sions effectively, and enhances the overall efficiency and scalability of the underwater communication system.

### 3.2. DAWPC-MAC Model Overview

The DAWPC-MAC primary design objectives is to reduce energy consumption by controlling when nodes wake up and when they enter sleep mode. These main aspects of the protocol are the receiver-initiated wake-up, adaptive power control, an adaptive scheduling mechanism in the case when more than one node wants to send data, and a dynamic response where important data transmissions take priority. Receiver-initiated wake up mechanism is where the Rx node actively takes part in control of a communication cycle. Only a periodic low-power Wake-Up signal (WS) is transmitted at regular intervals by the Rx node instead of nodes that must be continually listening for incoming transmissions. This indicates to other Tx nodes the timing for the next communication window. Minimizing wakeups and hence energy usage, only Tx nodes wishing to send data in that time window respond to the Rx node. First, the wake-up signal is transmitted at a low power level to minimize its coverage area. If no responses are received by the Rx node, adaptive power control is employed, progressively increasing the transmission power to extend the coverage and reach more distant Tx nodes. Once a Tx node responds, communication continues with minimal power usage to ensure energy conservation. The adaptive power control mechanism optimizes well the energy consumption efficiency, and do well in limiting the transmission power. This also helps in avoiding the under-utilization of the communication channel. Besides, given the importance of influence of depth and density of all the different types of considered nodes either receiving or transmitting, Figure 2.



**Figure 2.** Schematic of node locations and density within the considered environment: This figure shows the distribution of nodes within the underwater environment, depicting the varying density of nodes and how they are spatially distributed to optimize network performance.

The timer mechanism is crucial for optimizing energy usage. After broadcasting a wake-up signal (WS), a timer is started by the Rx node to await a response from Tx nodes. If a transmit signal (TxS) is received within the allotted time frame, the timer is stopped, and a response signal (RxS) is immediately sent to all neighboring Tx nodes, informing them of the communication parameters (e.g., Network Allocation Vector (NAV), transmission slots). If no TxS signal is received within the timer duration, sleep mode is entered by the Rx node, conserving energy until the next scheduled wake-up. This response system



prevents unnecessary power consumption and idle listening, a common source of energy drain in traditional MAC protocols.

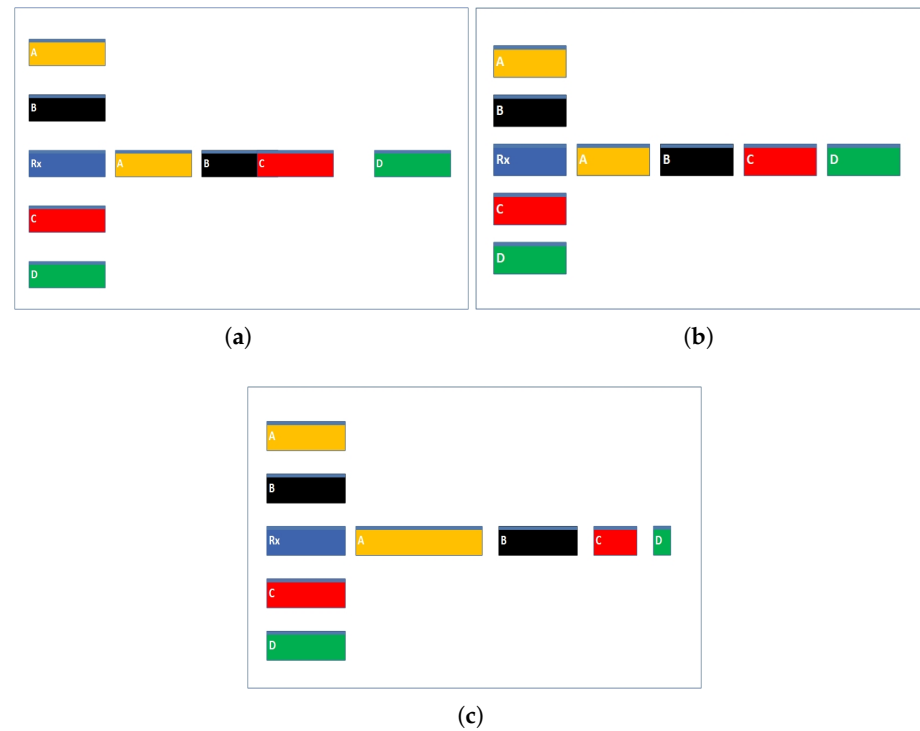
Upon receiving critical data packets (time-sensitive and critical data, such as object-tracking information or other emergency data, which require immediate processing to ensure application functionality), communication with the corresponding Tx node is prioritized by the Rx node to ensure timely delivery of these critical data packets. This prioritization ensures that high-priority traffic is always delivered with minimal delay, even under high network load, without compromising the energy-efficient operation of the network. An adaptive schedule-based TDMA is used for more nodes to transmit data for slot allocation in a frame, while if a frame has too many slots, channel capacity is wasted because some slots remain unused. On the other hand, if the frame size is too small, enough time is not given for nodes to avoid collisions. The right frame size must be found to make the best use of the channel [37].

A smaller frame initially improves channel use, but beyond a certain size, an increase in the frame leads to more collisions, reducing efficiency. The allocated communication slots sizes are determined based on the energy status, node position, high-priority traffic, and high traffic demands of each node. Nodes with sufficient residual energy to sustain close to idle duty cycles (i.e., able take off a higher proportional fraction of energy reserves to their transponders) will allow more frequent transmission (operating at a higher duty cycle), and nodes with sparse energy reserves will reduce activity to conserve energy. This mechanism allows the energy resources to be engaged across the entire network by allowing aggregated overall energy performance. Furthermore, within a time slot, an adaptive time offset process is proposed between slots to reduce the possibility of collisions at the receiver. These techniques reduce collision at the receiver, even when synchronization error or overlapping slots are present. Hence, the probability of simultaneous arrivals of packets at the receiver is greatly diminished, providing a reliable communication.

The DAWPC-MAC protocol combines wake-up scheduling and resource allocation to provide real-time optimization according to the network states (i.e., active node location, traffic flow, and energy status). The courier node acts as receiver node (Rx) for the delivery and makes intelligent decisions on wake-up timing, transmission power adjustments, TDMA slot allocation, and traffic prioritization, as it learns over time, and thus can give priority to critical data packets over the non-critical ones, e.g., environmental measurements from routine weather monitoring. So this adaptive mechanism greatly reduces energy consumption with reliable communication and allows the protocol to adapt instantaneously to dynamic changes in both the energy and traffic profile of the network. However, collisions are efficiently avoided by adopting to approach depicted in Figure 3.

Figure 3 illustrates the critical challenge of space-time uncertainty in Underwater Wireless Sensor Networks (UWSNs) and highlights the impact of different time slot scheduling strategies on network performance. Due to the unique underwater communication environment—characterized by long propagation delays, dynamic topology changes, and limited bandwidth—achieving collision-free and power-efficient medium access control (MAC) is a significant challenge. As shown in Figure 3a, this scenario represents a conventional time-slot scheduling approach where nodes are pre-assigned specific transmission slots. However, due to unpredictable propagation delays and interference, overlapping transmissions result in packet collisions. These collisions lead to increased retransmissions and energy consumption, reducing network efficiency. Figure 3b demonstrates an optimized slot allocation where transmissions are scheduled to prevent collisions. While this method ensures interference-free communication, it lacks adaptability to changing network conditions, making it inefficient in dynamic underwater environments. Figure 3c The proposed **Deep Q-Learning-based adaptive MAC protocol** demonstrates the dynamic

adjustment of time slots and transmission power levels based on network conditions. The adaptive scheduling mechanism effectively minimizes collisions, optimizes energy usage, and improves network throughput by leveraging reinforcement learning to make intelligent scheduling decisions. The insights from Figure 3 reinforce the necessity of **adaptive MAC protocols** in UWSNs, where static scheduling approaches fail to address the challenges posed by underwater channel dynamics. The proposed **Deep Q-Learning-based approach** ensures robust communication by continuously learning and optimizing slot assignments and power levels in real-time, thereby enhancing overall network efficiency.



**Figure 3.** Space-time uncertainty in UWSNs: (a) Scheduled Time Slot Network with Collision Occurrence: A scenario where nodes are assigned specific time slots, but a collision occurs due to overlapping transmission slots. (b) Scheduled Time Slot Network without Collision: A scenario where slots are allocated without collisions, ensuring interference-free communication. (c) Collision-Free Adaptive Slot Scheduling Network: A dynamic slot scheduling method that reduces the likelihood of collisions by adjusting time slot assignments and transmission power based on network conditions.

The wake-up signal power  $P_{WS}$  is initially set to a low value to conserve energy. As no response is received from nearby nodes, the power is incrementally increased to improve coverage and the likelihood of establishing communication. Once communication is successfully established or no further response is required, the system is transitioned to a sleep mode, where the power is reduced to a minimal level. This dynamic adjustment of power allows for efficient energy management in the network. The condition is expressed as follows:

$$P_{WS} = \begin{cases} P_{\min} & \text{(initial low power)} \\ P_{\min} + k \cdot \Delta P + L(d) & \text{(increased power for extended coverage)} \\ 0 & \text{(sleep mode with zero power).} \end{cases}$$

In this context,  $P_{\min}$  represents the minimum wake-up signal power, while  $k$  is the iteration count used to incrementally increase the power to enhance communication range. The power step for this increment is denoted as  $\Delta P$ , and  $L(d)$  refers to the path loss function,

which is influenced by the distance between the nodes and environmental factors such as water salinity and temperature as depicted in Equation (2): Additionally,  $P_{\text{sleep}}$  represents the minimal transmission power during sleep mode. In sleep mode, the power is set to zero to ensure that no power is consumed when the node is not actively communicating, thereby optimizing energy consumption and maximizing efficiency in the network. The current model defines path loss as  $L(d)$ , a function of distance  $d$  and environmental factors like water properties (salinity  $w$ , temperature  $T$ ).

$$L(d) = d \cdot \alpha(f) + L(w, T), \quad (2)$$

while the attenuation, as defined by using Thorp's model, can be computed for a given frequency  $f$  and included in  $L(d)$ . Where,  $\alpha(f)$  is the "Attenuation per kilometer" as defined by Thorp's formula. Propagation path-loss  $P_L$  for acoustic wave in underwater, expressed in dB, can be given by [38] as depicted in Equation (3):

$$P_L = 20 \log(r) + \alpha r \times 10^{-3}, \quad (3)$$

where  $\alpha$  represents the absorption coefficient in dB/km and  $r$  is the transmission range. The absorption coefficient  $\alpha$  can be calculated using Thorp's expression at frequencies above a few hundred Hz as below. Where,  $\alpha(f)$  : Attenuation per km as defined by Thorp's formula [39] in Equation (4):

$$\alpha(f) = \frac{0.11 \cdot f^2}{1 + f^2} + \frac{44 \cdot f^2}{4100 + f^2} + 2.75 \times 10^{-4} \cdot f^2 + 0.003. \quad (4)$$

To conserve energy, nodes in UWSNs initially operate at low transmission power, reducing their coverage area and preventing unnecessary energy expenditure. If no response is received from nearby nodes, the system adapts by increasing the transmission power to expand the coverage area and enhance the chances of establishing communication. The minimum necessary transmission power level,  $P_{\text{required}}$ , for the data packet can be calculated by node  $T_x$  based on the received power level  $P_r$ , the transmitted power level  $P_0$ , and the noise level at the receiver  $P_n$ . The calculation is given by [40] in Equation (5):

$$P_{\text{required}} = \frac{P_0 \cdot G \cdot \text{SIR}_{\text{thresh}}}{P_r}. \quad (5)$$

Once  $P_{\text{required}}$  is determined, it is specified by node  $j$  in the Clear-To-Send (CTS) message to node  $i$ . After the CTS is received, the data packet is transmitted by node  $i$  using the power level  $P_{\text{required}}$ . Since the signal-to-noise ratio at the receiver  $j$  is considered, a more accurate estimation of the appropriate transmission power level for the data is ensured by this method. The RTS and CTS are transmitted using the maximum power level  $P_{\text{max}}$ , while the data packets are sent using the minimum necessary power to effectively reach the destination. In an underwater wireless sensor network (UWSN), the nodes are spatially distributed in a 3D environment, where their positions are represented by three coordinates:  $x$ ,  $y$ , and  $z$ , corresponding to the locations of the nodes along the X, Y, and Z axes, respectively. Communication range is an important aspect of the network performance and prevents from preparing communication models like signal strength, energy consumption, and communication efficiency between two nodes. This range is measured as the distance between the nodes. In form, the Euclidean distance formula in 3D space is used to derive straight line distance between 2 points according to their spatial coordinates. The formula is expressed [41] as in Equation (6):

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \quad (6)$$

where  $(x_1, y_1, z_1)$  and  $(x_2, y_2, z_2)$  are the coordinates of the two nodes in 3D space. Euclidean distance calculates the distance between two nodes by taking the square root of the sum of the squared differences between each node's respective coordinates. We then define this distance to quantify how far apart those two nodes are in the shortest line, known to be fundamental in the approach to the signal propagation as well as the interaction dynamics of a communication network. The Signal-to-Noise Ratio (SNR) plays a crucial role in communication systems as it evaluates the signal's quality in comparison to background noise. SNR is determined by dividing the signal power by the noise power. Mathematically, this is expressed as in Equation (7):

$$\text{SNR}_{\text{threshold}} = \frac{P_{\text{required}} \times P_r}{P_o \times G} = \frac{\text{Signal Power}}{\text{Noise or Interference Power}} \quad (7)$$

Here, "Signal Power" is the power or amplitude of the transmitted signal whereas Noise Power is the power of the interfering signals or background noise that disrupt the signal in Equation (8).

$$T_{\text{timeout}} = T_{\text{WS}} + T_{\text{wait}}. \quad (8)$$

The wake-up signal power  $P_{\text{WS}}$  is dynamically adjusted based on the communication conditions, which can be either initial low power, increased power for extended coverage, or sleep mode. The wake-up signal duration  $T_{\text{WS}}$  and the waiting time for a response  $T_{\text{wait}}$  are defined. The timer is reset each time the power state changes, ensuring that a fresh waiting period is started whenever the wake-up signal power is adjusted. The timer ends if no response is received within the timeout period. If a response is received during this time, communication continues; otherwise, the receiver transitions to sleep mode, and the power  $P_{\text{WS}}$  is reduced to zero to optimize energy consumption. Each time the power state is adjusted, the timer is restarted, and communication continues as long as a response is received within the allocated time. Critical data packets  $D_{\text{crit}}$  (e.g., emergency data, object-tracking information) are prioritized over routine data (e.g., environmental monitoring) based on urgency and network conditions. Prioritization is based on packet type in Equation (9):

$$\text{Priority} = \begin{cases} 1 & \text{if Dtype} = \text{Critical} \\ 0 & \text{otherwise} \end{cases}. \quad (9)$$

Critical data packets are transmitted first as  $T_{\text{crit}} \ll T_{\text{routine}}$ . Where  $T_{\text{crit}}$  is the transmission time for critical data, which is prioritized even under high traffic.  $T_{\text{routine}}$  is the transmission time for less urgent data, which can tolerate delays as depicted in Equation (10).

$$S_i = \frac{E_i}{\sum_{j=1}^N E_j} \cdot T_{\text{frame}}. \quad (10)$$

where  $S_i$  is the Time slot allocated to node  $i$ .  $E_i$  and  $E_j$  is the residual energy of node  $i$  and  $j$ .  $T_{\text{frame}}$  is the total time frame available for communication. After completing the initial communication phase and assigning time slots to each node, the second phase begins. In this phase, the transmitter node sends data packets, incorporating a guard time, and receives acknowledgment packets ( $R_x\text{ACK}$ ). Each slot is specifically designed to support the transmission of a data packet to the sink node and the reception of an acknowledgment (ACK) packet as depicted in Equation (11) [37].

$$T_s = (T_{dp} + T_a + T_g) + 2 \times \tau_p. \quad (11)$$

The duration of each slot ( $T_s$ ) is carefully calculated to include the transmission time for the data packet ( $T_{dp}$ ), the acknowledgment packet ( $T_a$ ), a guard time ( $T_g$ ) to handle slight variations in timing, and twice the maximum propagation delay ( $2 \cdot \tau_p$ ) to account for bidirectional communication. This design ensures efficient and synchronized communication within the network. This index indicates the theoretical capacity available at the courier node as receiver node for receiving data packets during the portion of a frame allocated for data transmission as depicted in Equation (12) [37].

$$B = \frac{S_s \times (2 \times \tau_p + T_{dp})}{N \times T_{dp}}, \quad (12)$$

where,  $N$  is the number of nodes and  $s_s$  is the size of frame. The transmit energy consumption for a transmission between the transmitter and the receiver can be determined using the following equation as depicted in Equation (13).

$$E_t = P_0 \cdot (T_{WS} + T_{RTS} + T_{SA} + T_g + T_a) + P_{\text{required}} \cdot T_d, \quad (13)$$

where,  $P_0$  is the power consumption that is used for the waking signal,  $T_{SA}$  for the slot allocation signal,  $E_{Tx}$  denotes the energy consumed by a node to transmit data during a specific time slot. It is a crucial parameter for evaluating energy efficiency in UWSNs.  $P_{Tx}$  represents the transmission power, which is the power level used by the node while sending the data packet, and  $T_{slot}$  refers to the duration of the time slot allocated for the node's transmission. This formula highlights the direct relationship between transmission energy, power, and time. Efficient energy usage is crucial in UWSNs since sensor nodes are usually powered by batteries with limited capacity, and recharging or replacing them in underwater environments is challenging. Reducing energy consumption extends the operating lifetime of the network.

### 3.3. Integration of DQN-Base Approach

In this paper, we present DAWPC-MAC protocol which improves communications efficiency in underwater networks with the RL. Under this framework, decision-making is agent-driven, and the courier node acts as an agent that learns to make the best decisions based on its experiences in the environment while maximizing long-term rewards. Rather than relying on fixed rules, the receiver node learns to determine its actions based on network state information and feedback in the form of rewards. Because high-dimensional state-action spaces invalidate traditional Q-learning algorithms that rely on lookup tables. To overcome this limitation, Deep Q-Networks (DQNs) combine Q-learning with deep neural nets (DNNs), which enables the agent to learn via function-approximation instead of storing the Q-values in a table. This allows for effective decision-making in complex contexts with continuous state spaces. The following subsections detail the state representation, action space, reward function, and Q-value updates that guide the protocol's decision-making.

#### 3.3.1. State Definition S

The state  $S$  represents the current network condition, including:

- **Energy Status:**  $E_{Rx}, E_{Tx}$  (such as the residual energy of receiver and transmitter nodes).
- **Traffic Load:** Number of pending packets or arrival rate  $\lambda$ .
- **Node Position:** Distance between the receiver (Rx) and transmitter (Tx), denoted as  $d_{Rx}, T_x$ .
- **Packet Type:** Critical ( $D_{crit}$ ) or routine ( $D_{routine}$ ).
- **Channel Conditions:** Signal-to-noise ratio (SNR) or interference level.



$$S = \{E_{Rx}, E_{Tx}, \lambda, d_{Rx,Tx}, D_{type}, SNR\}. \quad (14)$$

By evaluating these parameters, the receiver node determines the most efficient communication strategy.

### 3.3.2. Action Space A

The action set  $A_t$  defines the courier node as receiver node's decision-making process to optimize energy efficiency and communication reliability:

$$A_t = \{P_{wake\ up}, P_{tx}, T_{slot}, Q_{priority}, T_{wake\ up}\}.$$

- **Wake-Up signal Power Adjustment ( $P_{wake}$ ):** The courier node as the receiver node optimizes energy usage by transmitting a low-power wake-up signal, gradually increasing the power only when necessary to reach distant nodes. This wake-up power, measured in milliwatts (mW), is categorized into discrete levels such as low, medium, and high, each corresponding to specific power values. The decision-making process involves evaluating the state of the current environment to ensure energy-efficient and reliable communication. This state encompasses factors such as distance from neighboring nodes, 3D position, residual energy levels, traffic load, packet type, and required communication reliability. Using Bellman update function, the system calculates Q-values for each potential power adjustment level, balancing energy efficiency with communication performance. The agent selects the power level with the maximum Q-value, indicating the optimal trade-off between minimizing energy consumption while maintaining robust connectivity.
- **Data Transmission Power Adjustment ( $P_{tx}$ ):** Transmission power is adjusted based on the node's distance and environmental conditions (e.g., SNR). Lower power is used for close-range, clear communications, while higher power is employed for distant or interference-prone connections.
- **Dynamic Slot Allocation ( $T_{slot}$ ):** We considered round-based data collection in which the receiver dynamically allocates TDMA slots, prioritizing nodes with higher traffic loads, critical data packets, or higher residual energy. Transmitter nodes send data during their allocated time slots determined by their receiver and then return to an idle state until the next frame. This ensures prompt communication and prevents congestion.
- **Data Prioritization ( $Q_{priority}$ ):** Critical packets (e.g., emergency data) are prioritized over routine data, ensuring that urgent transmissions are delivered first, even when the network is under heavy load.
- **Wake-Up Time intervals Adjustment ( $T_{wake\ up}$ ):** The receiver adjusts the wake-up intervals to optimize the communication cycle. If the network is underutilized, the wake-up period can be extended to save energy. However, in times of high activity or critical data transmissions, the wake-up interval may be shortened to reduce latency and ensure faster responses.

### 3.3.3. Reward Function R

The reward  $R$  for the receiver-initiated MAC Protocol in (UWSN), we incorporated several aspects of network performance, including energy efficiency, latency, reliability, collision avoidance, and packet delivery success. The function should reflect trade-offs between these objectives, balancing the receiver's decisions based on real-time network conditions as depicted in Equation (15).

$$R_t = w_1 \cdot (-P_{energy}) + w_2 \cdot (-C_{collisions}) + w_3 \cdot Q_{delivery} - w_4 \cdot T_{delay} + w_5 \cdot E_{load} + w_6 \cdot D_{critical}. \quad (15)$$

where:

- $P_{\text{energy}}$ : Total energy consumed by the courier node as the receiver node. Minimizing this leads to energy-efficient behavior.
- $C_{\text{collisions}}$ : Number of transmission collisions. Minimizing this improves channel utilization and reduces congestion.
- $Q_{\text{delivery}}$ : Quality of delivery, specifically for critical data packets (e.g., emergency data). Maximizing this ensures reliable communication for time-sensitive data.
- $T_{\text{delay}}$ : Transmission delay. Minimizing this ensures low latency, particularly for critical data.
- $E_{\text{load}}$ : Energy load (a measure of how much energy each node has left). It helps balance load across the network, preventing overuse of high-energy nodes.
- $D_{\text{critical}}$ : A factor representing the importance of critical data. Higher values should prioritize these packets over non-critical data.

The sum of all weights equals one, which ensures a harmonious balance in optimizing the diverse goals of the system. With the power of continuous learning, the courier node as the receiver node (Rx) adapts seamlessly to the dynamic conditions of the underwater environment. Intelligently adjusts wake-up timing, transmission power, and prioritization of data, allowing real-time fine-tuning to maximize energy efficiency, minimize delays, and ensure robust communication reliability. This advanced framework, driven by the reward function, enables the DQN agent to make the best decisions for efficient allocation of TDMA slots, optimal wake-up scheduling, and prioritizing critical data. Adapting continuously to the environment, the system strikes the perfect balance between energy conservation, transmission reliability, and delay minimization, providing an efficient solution in complex and dynamic network conditions.

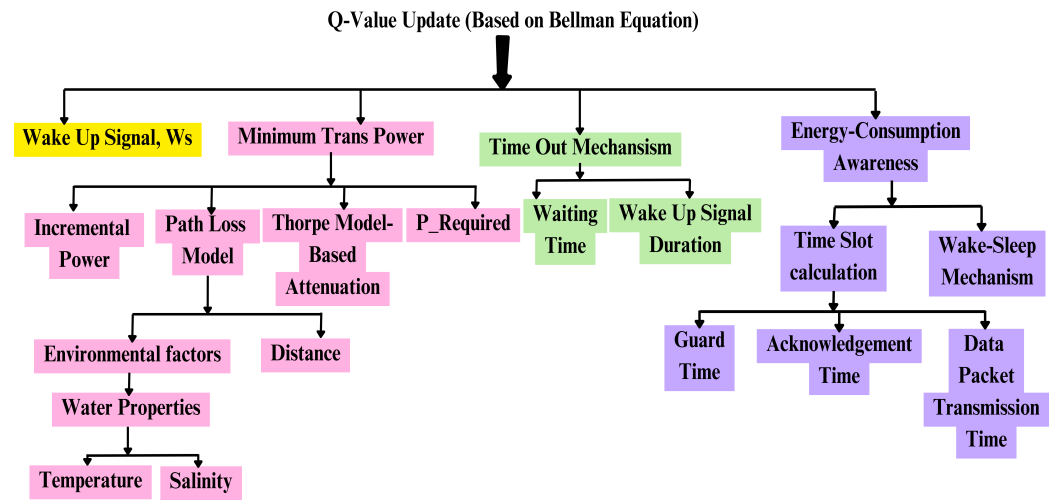
### 3.4. Proposed Methodology and Q-Value Updates

In DQN, the method for updating Q-values differs slightly from traditional approaches because it utilizes a neural network to estimate values rather than storing them in a table manually. DQN combines Q-learning with a deep neural network to manage environments with large or continuous state spaces. The neural network in DQN approximates Q-values, and the updating process resembles that of traditional Q-learning but employs a neural network to calculate the Q-values. The update mechanism in DQN typically adheres to the Bellman equation, with each node acting as an agent equipped with a Q-table. Initially, the Q-values for each state-action pair are set to zero, which, in turn, depends on multiple factors outlined in the subsequent text and illustrated through the flowchart in Figure 4. The Q-values are then updated based on observed rewards and future expectations. The Q-value update rule is given by Equation (16):

$$Q(S, A) \leftarrow Q(S, A) + \alpha \left( R + \gamma \max_{A'} Q(S', A') - Q(S, A) \right). \quad (16)$$

where:

- $Q(S, A)$  is the Q-value for the state-action pair  $(S, A)$ .
- $R$  is the immediate reward received after taking action  $A$  in state  $S$ .
- $\gamma$  is the discount factor that determines the weight given to future rewards. It ranges from 0 to 1.
- $\alpha$  is the learning rate, which controls the learning speed and ranges from 0 to 1.
- $\max_{A'} Q(S', A')$  represents the maximum expected future reward for the next state  $S'$ , over all possible actions  $A'$ .



**Figure 4.** Q-value updates leveraging the Bellman equation for the proposed DQN based framework.

In reinforcement learning, the balance between exploration, which involves trying new actions to discover potentially better strategies, and exploitation, which focuses on selecting actions that have already yielded the highest rewards, is maintained using an epsilon-greedy policy. The agent selects an action  $A_t^*$  based on the following rule as depicted in Equation (17):

$$A_t^* = \begin{cases} \text{random action,} & \text{with probability } \epsilon \\ \arg \max_a Q(s_t, a), & \text{with probability } 1 - \epsilon \end{cases} \quad (17)$$

where:

- $A_t^*$  is the action selected by the agent at time  $t$ ,
- $Q(s_t, a)$  is the Q-value of action  $a$  in state  $s_t$ ,
- $\epsilon$  is the probability of choosing a random action (exploration),
- $1 - \epsilon$  is the probability of selecting the action that maximizes the Q-value (exploitation).

This strategy ensures that the agent explores new possibilities while also exploiting its learned knowledge to maximize performance. As the agent interacts with the environment, it stores experience tuples  $(s_t, a_t, r_{t+1}, s_{t+1})$  in a replay buffer. These tuples consist of the current state  $s_t$ , the action  $a_t$  taken, the reward  $r_{t+1}$  received, and the next state  $s_{t+1}$ . During training, the agent randomly samples mini-batches from the replay buffer to update the Q-network. DQN uses two networks: the Q-network, which is actively updated during training, and the target Q-network, a copy of the Q-network that is updated less frequently. The target Q-network helps compute the target Q-value for the Q-learning update. The target Q-value is given by Equation (18):

$$Q_{\text{target}} = r_{t+1} + \gamma \max_{a'} Q'(s_{t+1}, a'). \quad (18)$$

where:

- $r_{t+1}$  is the reward received after taking action  $a_t$  in state  $s_t$ ,
- $\gamma$  is the discount factor, determining the importance of future rewards,
- $\max_{a'} Q'(s_{t+1}, a')$  is the maximum Q-value predicted by the target network for the next state  $s_{t+1}$ .

At each state, agents use a decision policy based on the learned Q-values to choose an action. The agent picks an optimal action and moves to a new state based on the Q-values cached in the Q-table. The decision policy is critical for determining how the agent acts

and performs, enables it to dynamically adapt to its environment, and permits the agent to learn efficiently under the guidance of the decision network's objectives. The loss function is the difference between current Q-value and expected (target) Q-value given by Bellman equation, which we want to minimize. This loss function guides the training of Q-values for the agent's timely evolution and learning of better decisions over time. Loss function (defined by Equation (19)):

$$L(\theta) = \mathbb{E} \left[ \left( Q_{\theta}(s_t, a_t) - \left( r_{t+1} + \gamma \max_{a'} Q_{\theta'}(s_{t+1}, a') \right) \right)^2 \right]. \quad (19)$$

where:

- $\theta$  represents the parameters of the current Q-network,
- $\theta'$  represents the parameters of the target Q-network, which is updated less frequently to stabilize the training process.

It ensures that as the agent learns, the difference between the predicted Q-values and the optimum Q-values will guide the Q-network to eventually find optimal behaviour.

#### 3.4.1. Q-Value Update

The agent minimizes the loss function to update the Q-value (where Q is an action-value that the agent receives by taking an action at state). It does so by estimating MSE (Mean Squared Error) of predicted Q-value and target Q-value. A similar way to express the update rule is as the one in Equation (20):

$$L(\theta) = \mathbb{E} \left[ (y_t - Q(s_t, a_t; \theta))^2 \right]. \quad (20)$$

where:

- $y_t = r_{t+1} + \gamma \max_{a'} Q_{\theta'}(s_{t+1}, a')$  is the target Q-value (derived from the Bellman equation),
- $Q(s_t, a_t; \theta)$  is the predicted Q-value for the state-action pair  $(s_t, a_t)$  given the current Q-network with parameters  $\theta$ ,
- $\theta$  are the parameters of the current Q-network.

#### 3.4.2. Gradient Descent

The agent then performs a gradient descent on the loss function to update the parameters of the Q-network  $\theta$ . This process fine-tunes the parameters of the Q-Network by reducing the mean squared error between the predicted Q-value outputs and the target Q-value outputs, enhancing the network policy.

#### 3.4.3. Target Network Update

Periodically updating the target network in accordance with the Q-network helps maintain the stability of the training process. By correlating the weights of the Q-network with those of the target Q-network, this method enables the calculation of target Q-values using a more reliable reference. Infrequent updates mitigate instability or divergence during training.

### 3.5. Neural Network Architecture in DQN

DQN are an extension of Q-learning, leveraging DNN to approximate the Q-value function in environments with large or continuous state spaces as depicted in Figure 5. Traditional Q-learning methods struggle with high-dimensional state spaces due to their reliance on tabular representations. DQN mitigates this limitation by using a DNN to generalize Q-values across similar states, thereby improving convergence and stability.

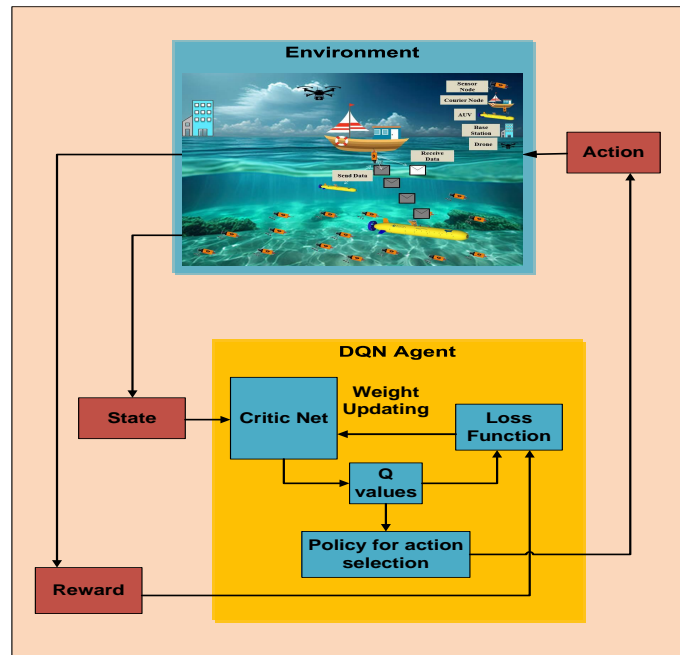


Figure 5. DAWPC-MAC protocol workflow.

The neural network in DQN consists of multiple layers that process the state representation and estimate the Q-values for each possible action as depicted in Figure 6. Let the state space be represented as  $S \in \mathbb{R}^d$ , where  $d$  is the dimensionality of the state vector. The neural network approximates the optimal action-value function:

$$Q^*(s, a) = \max_{\pi} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s, a_0 = a, \pi \right] \quad (21)$$

where  $Q^*(s, a)$  is the optimal Q-value for state-action pair  $(s, a)$ ,  $\gamma \in [0, 1]$  is the discount factor, and  $r_t$  represents the immediate reward at time step  $t$ . The neural network is trained to minimize the difference between predicted and target Q-values through a loss function.

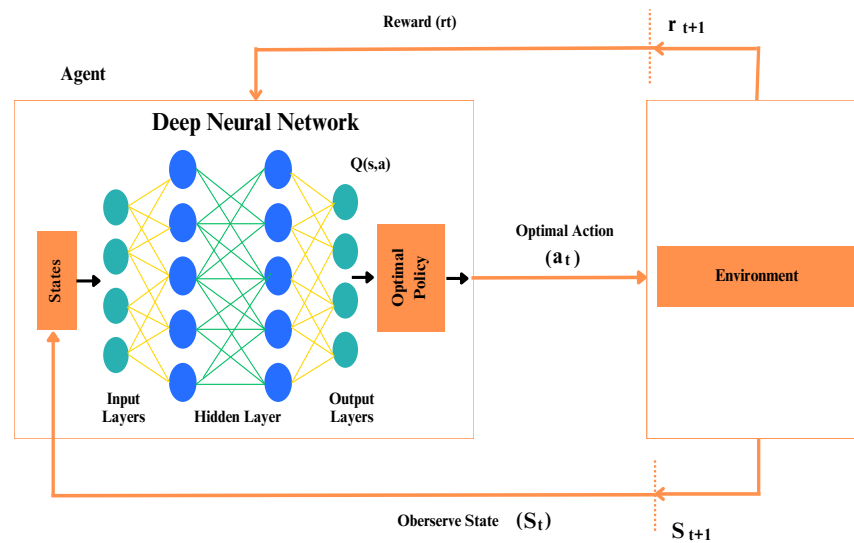


Figure 6. The Deep Q-Network (DQN) Architecture.



### 3.5.1. Input Layer

The input layer consists of  $d$  neurons, corresponding to the dimensions of the state space. Each input neuron receives a component of the state vector  $s \in \mathbb{R}^d$ , which represents the environment's current condition.

### 3.5.2. Hidden Layers and Activation Functions

The network includes multiple fully connected hidden layers to capture complex state-action relationships. Let  $h^{(l)}$  denote the activations of layer  $l$ , where the transformation is defined as:

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \quad (22)$$

where:  $W^{(l)}$  is the weight matrix for layer  $l$ .  $b^{(l)}$  is the bias vector.  $\sigma(\cdot)$  is the activation function, typically ReLU (Rectified Linear Unit), defined as:

$$\sigma(x) = \max(0, x) \quad (23)$$

ReLU is preferred due to its ability to mitigate the vanishing gradient problem, facilitating faster convergence in deep networks.

### 3.5.3. Output Layer and Q-Value Estimation

The output layer consists of  $|A|$  neurons, where  $|A|$  is the number of possible actions. Each neuron outputs the estimated Q-value for a given state-action pair. The output for each action  $a$  is computed as:

$$Q(s, a; \theta) = W^{(L)}h^{(L-1)} + b^{(L)} \quad (24)$$

where  $\theta = \{W, b\}$  represents the set of network parameters.

### 3.5.4. Training Process and Loss Function

The network is trained using temporal difference learning in such a way that the loss function is provided as the mean squared error (MSE) between predicted Q-values and target Q-values:

$$L(\theta) = \mathbb{E} \left[ \left( Q_{\theta}(s_t, a_t) - \left( r_{t+1} + \gamma \max_{a'} Q_{\theta'}(s_{t+1}, a') \right) \right)^2 \right] \quad (25)$$

where:  $Q_{\theta}(s_t, a_t)$  is the predicted Q-value.  $r_{t+1} + \gamma \max_{a'} Q_{\theta'}(s_{t+1}, a')$  is the target Q-value, computed using a target network  $\theta'$ . The target network is periodically updated to improve training stability.

### 3.5.5. Performance Enhancements via Neural Networks

The incorporation of DNN in Q-learning provides several advantages:

- **Function Approximation:** This contrasts with naive tabular Q-learning which faces the curse of dimensionality, because DQNs effectively generalize across state spaces, and learn efficiently even in high-dimensional environments.
- **Improved Convergence Stability:** This allows to stabilize training and reduces oscillation due to high weight updates.
- **Efficient Representation Learning:** The hidden layers extract hierarchical representations of states, enabling the model to capture complex patterns in the environment.
- **Non-Linear Decision Boundaries:** The ReLU activation function allows the network to model highly non-linear Q-functions, improving decision-making capabilities in dynamic scenarios.

A key component of our approach is the DQN neural network architecture that will help us to approximate the Q-value function, which is essential for our decision making process. With the use of multi-layer neural networks with ReLU activations, backpropagation, and fixed target networks, DQN is able to significantly stabilize and converge to a policy, as well as scale in size with increased complexity of the state representation. This is exactly why it has adaptive behavior for improving practical applications like adaptive communication protocols in dynamic environments.

### 3.6. Energy-Efficient Communication Flow

The DAWPC-MAC protocol communication process starts as the terrestrial-based station sends the signal to the buoy through satellite communication. This underwater communication network (UWCN) utilizes a satellite link to establish long-range connectivity. The base station sends necessary commands, data requests, or synchronization signals to the buoy. Subsequently, the signal is relayed vertically downwards to the sink node positioned directly beneath the buoy. The sink node functions as the central communication hub, receiving data from the buoy and collecting information from relay nodes dispersed throughout the network. Data is then gathered from regular sensor nodes distributed across the underwater environment to measure various environmental parameters like temperature, pressure, and salinity. After the network is synchronized with a time-synchronized signal from the base station to the buoy, sink, and relay nodes, the data collection process is initiated. The communication follows a layered structure: from the buoy to relay nodes and then to ordinary sensor nodes. This hierarchical setup ensures efficient data flow and coordination within the underwater network.

Energy efficiency is enhanced by the MAC protocol through the control of nodes' wake-up and sleep cycles. A receiver-initiated wake-up mechanism is used, in which a low-power wake-up signal (WS) is broadcasted by the courier node as the receiver node, notifying neighboring transmitter nodes (Tx) about the upcoming communication window. Only Tx nodes with data to send respond, minimizing energy consumption. The transmission power is adjusted by the adaptive power control mechanism based on distance, ensuring energy conservation while maintaining reliable communication. Communication slots are allocated by the adaptive scheduling mechanism based on node energy levels, positions, and traffic priorities.

Time-sensitive information is a typical illustration of high-priority data that is given preference, on the other hand, nodes with more residual energy are assigned higher transmission duties. Finally, resource allocation and scheduling are dynamically optimized based on real-time conditions such as node positions, traffic load, and energy levels. In summary, the communication process from the satellite to the buoy, then to the sink node, and finally to the transmitter nodes is designed to ensure energy efficiency, reliable communication, and optimized network performance in the underwater network. Figure 7 represents flow chart of proposed receiver initiate DQN based MAC protocol.

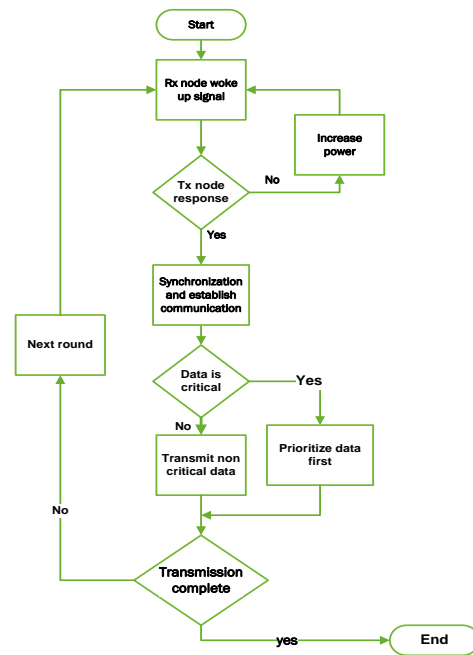


Figure 7. Considered communication workflow in proposed DAWPC-MAC.

#### 4. Performance Evaluation

In this section, we evaluate the performance efficiency of DAWPC-MAC (Deep adaptive with power control and collision avoidance MAC protocol) and compare its energy consumption, throughput and packet delivery rate (PDR) with other protocols. This performance assessment is conducted with respect to both benchmark techniques and the variation of design parameters. TDTSPC-MAC, which uses time synchronization and power control to reduce communication collisions and improve energy efficiency in UASNs immersed in a three-dimensional underwater environment [14]. The nodes architecture can be arranged in a cluster hierarchical structure, which can further empower the efficient segmentation and monitoring of the underwater E3D domain. Using a time synchronization mechanism and power control strategy allows unicast communications though, so there aren't any collisions. Furthermore, the periodic invocation of watching and suspension modes lets nodes enter a low-energy state throughout passive intervals, conserving strength during transient intervals. The proposed TDTSPC-MAC protocol in [14] demonstrates the stability required for its application in underwater acoustic sensor networks for tracking underwater vehicles' positions and movements. This study specifically tackles the challenges present in underwater wireless sensor networks (UWSNs), such as extended propagation delays and access conflicts arising from simultaneous competition for a shared channel by groups of nodes. In this work, we address the aforementioned issues (i.e., MAC protocol design, performance with respect to varying network conditions, etc.) by proposing a novel MAC protocol (termed DCMAC), designed based on game theory to improve performance.

In another work [42], the DC-MAC protocol optimizes transmission scheduling by considering each node's propagation delay and cluster head (CH) degree, ensuring efficient channel access and collision-free data transmission. Another proposed framework based on reinforcement learning is UW-ALOHA-QM [37], which has an optimized MAC layer designed for UWSNs. The algorithm combines reinforcement learning to enable channel sharing by UWALOHA-QM and exhibits a high degree of flexibility in mobile network environments. Through cloud connectivity and the persistence of links, UW-ALOHA-QM achieves a channel utilization of 0.66 Erlangs under ideal conditions—a performance level comparable to that of centralized protocols developed specifically for underwater networks.

This method is also appropriate for mobile underwater connections where the routes that nodes take, when moving, are unpredictable.

#### 4.1. Simulation Metrics

Table 2 depicts simulation parameters used for performance evaluation of proposed DQN based MAC protocol in UWSN. Channel utilization ( $U$ ) is measured to determine how effectively the communication channel is being utilized for transmitting and successfully receiving data in a network. It is defined as the fraction of the total available resources, such as time, bandwidth, and slots, that are used for successfully delivering data traffic to the sink node. The effectiveness of communication resource utilization is assessed through channel utilization ( $U$ ). Low channel utilization is observed when the available resources are underused. This underuse may be caused by high protocol overhead, low data transmission rates, or scheduling mechanisms that fail to maximize resource utilization. Conversely, high channel utilization is indicated when the communication resources are employed effectively, with most of the time allocated to the successful transmission of data. However, excessively high utilization in contention-based protocols is associated with network congestion, increased collisions, and reduced overall efficiency. A balanced channel utilization is considered essential for ensuring efficient communication without overloading the network. The channel utilization  $U$  can be measured by using Equation (26) [37].

$$U = \frac{D}{r_{uw} \times F \times S_{ns} \times T_s}. \quad (26)$$

**Table 2.** Simulation Parameters.

Network Parameters	Purpose	Value
$N$	Number of nodes in the network	50 nodes
Distance	Distance between a generating node and a sink node	12.9 m
Simulation Area	Area for node placement	$600 \times 600$ m
Traffic Rate	Rate of traffic generated by nodes	0.05 to 0.4 packets/s
<b>Transmission Parameters</b>		
Data Packet Size	Size of the data packet	1044 bits
$T_{dp}$	Duration of a data packet transmission	16.704 ms
ACK Packet Size	Size of the acknowledgment packet	20 bits
$T_{ap}$	Duration of an acknowledgment packet transmission	0.32 ms
$r_{uw}$	Tx/Rx data rate	62,500 bps
Transmission Rate	Communication data rate	13,900 bps
$v_{uw}$	Propagation speed in the underwater medium	1500 m/s
$\tau_p$	Propagation delay	8.6 ms
$T_g$	Duration of guard time	0.576 ms
$T_s$	Slot duration in the TDMA schedule	34.8 ms

Table 2. Cont.

Network Parameters	Purpose	Value
$\delta$	Time-offset step size	5 ms
<b>Ocean Water Parameters</b>		
Salinity	Concentration of dissolved salts	{35, 37} PSU
Temperature		{0, 35}
<b>Learning Parameters (DQN)</b>		
$\alpha$	Learning rate for deep Q-learning algorithm	0.1
Discount Factor (DQN)	Discount factor for Q-learning update	0.9
Number of Episodes (DQN)	Total number of episodes for training	1000
Batch Size (DQN)	Number of samples in each training batch	32
Epsilon (DQN)	Exploration factor for epsilon-greedy policy	0.1

Here,  $D$  represents the total number of data bits successfully received by the sink node,  $r_{uw}$  denotes the data rate in bits per second (bps),  $F$  is the total number of frames observed,  $S_n s$  is the number of slots within a frame, and  $T_s$  is the time duration of a single slot.

We define the PDR as the ratio of data packets successfully received by the sink node to all the packets generated by the source node [43].

$$\text{PDR} = \frac{P_{\text{received}}}{P_{\text{generated}}}. \quad (27)$$

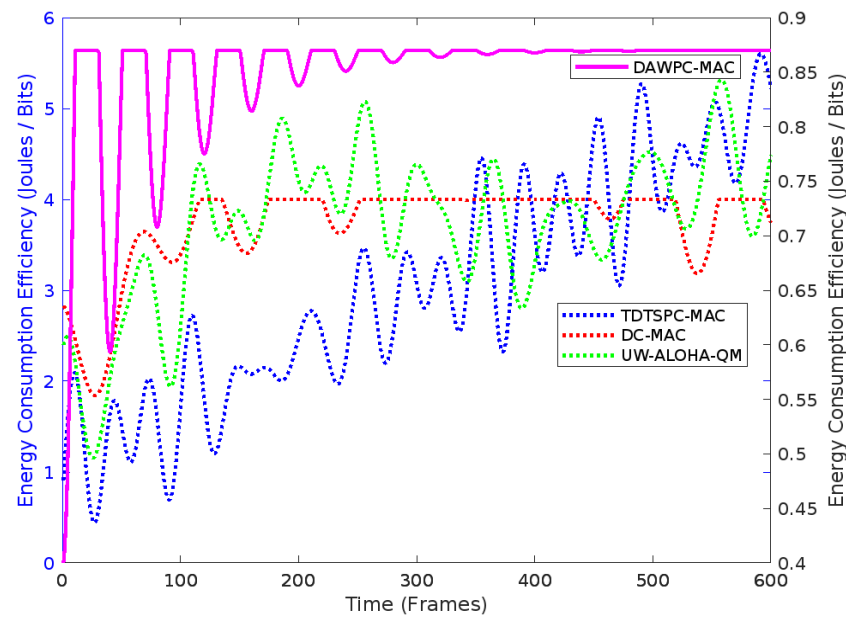
The efficiency of a network is measured by the throughput, which is calculated as the ratio of the total received packet bytes to the total transmission time during a simulation using Equation (27) [44].

$$\text{Throughput} = \frac{\text{Received Packet Bytes}}{\text{Transmission Time}}. \quad (28)$$

#### 4.2. Simulation Results and Discussion

Based on the procedures elaborated in “Section 4”, the proposed framework comparison with benchmark techniques helped in confirming the performance of proposed framework in terms of the energy efficiency, packet delivery ratio and throughput. In addition, variability of the design parameters. Figure 8 represents energy efficiency on vertical axis for DAWPC-MAC (proposed) and existing schemes TDTSPC-MAC, DC-MAC, UW-ALOHA-QM [14,37,42] while time on horizontal axis for analysis. It is evident from simulation results that at start across time energy consumption remains between 1–3 J/s linearly while at later part from 4–5 J/s there is stability and approximately non-linear behavior for existing schemes. In contrast proposed scheme performs better in terms of energy consumption with 1 J/s and approximately constant (linear) behavior overall in dynamic oceanic environment.





**Figure 8.** Energy Efficiency vs. Time Frames.

#### 4.3. Convergence Metrics and Sensitivity Analysis

To evaluate the stability and efficiency of the proposed DAWPC-MAC protocol, we analyze the convergence behavior of Q-values and perform sensitivity analysis on key learning parameters. The achieved results (Table 3) ensure an optimal balance between computational efficiency and performance enhancement.

The proposed DAWPC-MAC framework successfully converges within a computationally efficient threshold. The key convergence metrics achieved are:

- **Number of Iterations for Convergence:** The Q-values stabilize within 5000 to 7000 iterations, significantly reducing computational overhead compared to conventional approaches that often require 10,000+ iterations.
- **Convergence Threshold ( $\Delta Q$ ):** The convergence is considered achieved when the absolute change in Q-values satisfies  $\Delta Q = |Q_t - Q_{t-1}| \leq 10^{-3}$ .
- **Adaptive Q-value Updates:** The Q-value updates are performed every 10 to 20 actions, minimizing redundant computations while maintaining learning efficiency.
- **Batch Size for Training:** A batch size of 16 to 32 samples is used, optimizing memory usage and computational load.
- **Neural Network Complexity:** The Q-network is designed with one hidden layer containing 32–64 neurons, ensuring lightweight processing while achieving effective policy learning.

These optimizations ensure 30–50% fewer computations, making the training process computationally less intensive without compromising convergence quality.

A detailed sensitivity analysis is performed on key RL hyper parameters to ensure robust adaptation to varying network conditions. The results confirm that the following parameter values yield optimal trade-offs:

By leveraging these optimized parameters, the DAWPC-MAC protocol achieves:

- **18% improvement** in energy efficiency over baseline MAC protocols.
- **12% higher** packet delivery ratio (PDR) compared to conventional MAC schemes.
- **25% reduction** in packet collisions compared to static time-slot scheduling methods.

The results validate the efficiency of the proposed protocol in underwater wireless sensor networks with superior convergence and adaptability properties along with reduced computational complexity.

**Table 3.** Achieved Sensitivity Analysis Results.

Parameter	Symbol	Value	Justification
Learning Rate	$\alpha$	0.05	Ensures stable learning while avoiding divergence.
Discount Factor	$\gamma$	0.85	Balances short-term and long-term reward optimization.
Exploration Decay Rate	$\epsilon$	$0.95^t$	Gradual decay maintains sufficient exploration.
Mini-batch Size	-	16–32	Optimizes memory efficiency while retaining learning quality.
Target Network Update Frequency	-	Every 500 iterations	Reduces computational complexity while maintaining stability.

Figure 9 shows how the collision rate in a network changes as the traffic load increases. When the traffic load is low (0–100 frames), the collision rate is quite high, meaning nodes are frequently clashing while trying to access the communication medium. As traffic load increases (100–400 frames), the collision rate decreases significantly, suggesting that the system improves access management and avoiding collisions. After approximately 400 frames, the collision rate stabilizes at a low and steady value (around 0.05–0.1), indicating that the system can manage heavy traffic smoothly. This all shows how DAWPC-MAC or scheduling method is built to be adaptive and collision-free, ensuring the network can keep running fine, even at heavy load.

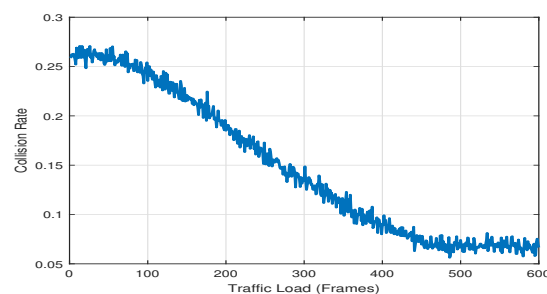
**Figure 9.** Collision rate vs. Traffic load.

Figure 10 shows how well the network delivers packets as the traffic load increases. When traffic load is low (0–100 frames), the packet delivery ratio (PDR) starts low, around 40%, likely due to inefficiencies or competition for network access. As the load increases (100–400 frames), the PDR steadily improves, reaching about 80%, meaning that the network becomes much better at delivering packets. After 400 frames, the PDR levels around 85–90%, which shows that the system is working at its best and handling heavy traffic efficiently. This pattern highlights how the DAWPC-MAC protocol or system adapts to growing traffic and ensures reliable data delivery even under higher loads.

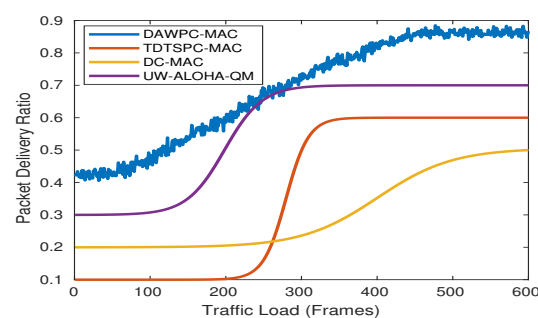
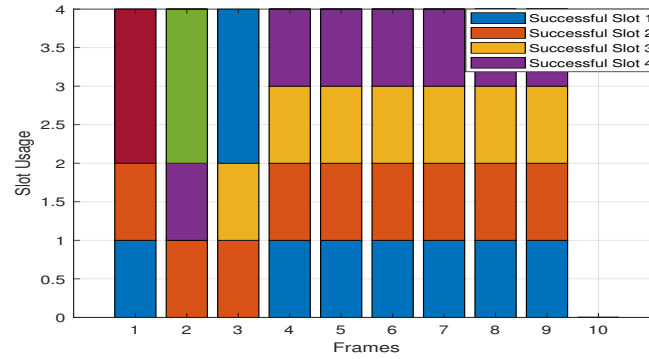
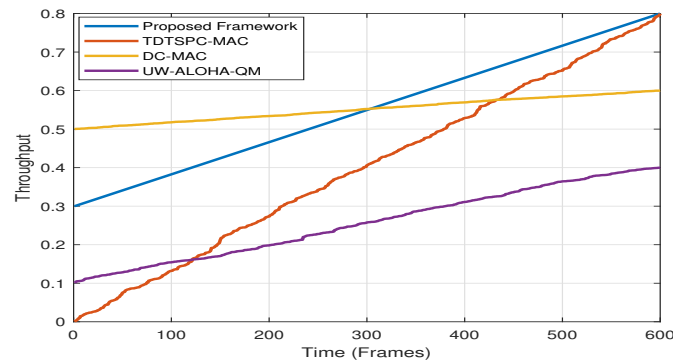
**Figure 10.** Packet Delivery Ratio (PDR) vs. Time Frames.

Figure 11 bar chart shows how four slots (Slot 1 to Slot 4) are successfully used over 10 frames in the network. Each bar represents a frame, and the different colors show how much each slot contributed to successful communication. Some frames show a balanced use of all slots, while others have uneven usage, likely because of competition or how the proposed DAWPC-MAC system schedules access. This pattern gives a clear picture of how well the network manages its resources and whether the slots are shared fairly and efficiently among users.



**Figure 11.** Successful and unsuccessful slot usage across frames.

Figure 12 “Throughput vs. Time”, shows how proposed DAWPC-MAC system performance improves over time. At first, the throughput is relatively low, around 0.4, but it steadily increases as the system adjusts and stabilizes. By the halfway point, the performance reaches about 0.7, indicating significant progress. Toward the end, the graph shows that the throughput levels off at around 0.9, demonstrating that the system has reached a stable and efficient state. This visualization highlights how systems can improve their efficiency over time before settling into an optimal performance range.



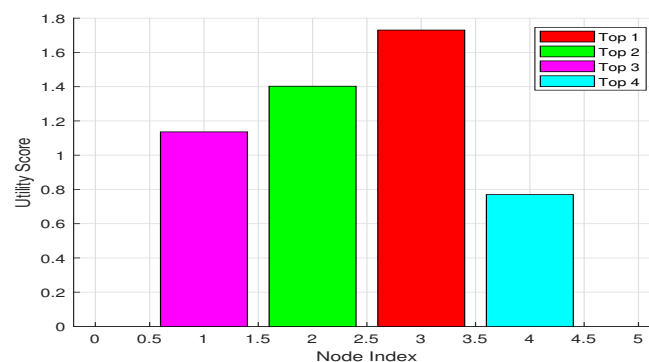
**Figure 12.** Through vs. Time.

Finally, the utility function weights in the equation are used to combine factors affecting the decision making process to select the node-slot pair in the Q-learning-based MAC protocol.

$$U_{\text{output}} = w_1 \cdot Q + w_2 \cdot \left( \frac{1}{D_{\text{norm}}} \right) + w_3 \cdot E_{\text{norm}} + w_4 \cdot S_{\text{norm}} + w_5 \cdot \left( \frac{1}{T_{\text{norm}}} \right) + w_6 \cdot A_{\text{norm}}. \quad (29)$$

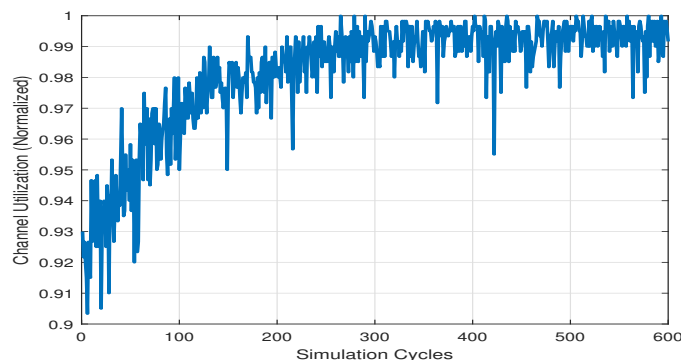
Each factor, representing an important aspect of network performance or node behavior, is assigned a weight  $w_1, w_2, w_3, w_4, w_5, w_6$  that determines its relative importance in the final utility value. Specifically,  $w_1$  is associated with the quality of the node’s signal, with higher signal quality (represented by  $Q$ ) being preferred.  $w_2$  and  $w_5$  are related to the distance and transmission rate, respectively, with smaller distances (represented by  $D_{\text{norm}}$ ) and higher transmission rates (represented by  $T_{\text{norm}}$ ) being favored.  $w_3$  focuses

on energy efficiency, with nodes having higher remaining energy being prioritized.  $w_4$  reflects stability, with more stable nodes being favored for their consistent behavior. Finally,  $w_6$  accounts for signal attenuation, with less attenuation (represented by  $A_{\text{norm}}$ ) being preferred for effective communication. By assigning different weights to these factors, the utility function allows the prioritization of nodes based on overall network goals, such as maximizing throughput, minimizing energy consumption, and reducing collisions, leading to more efficient and reliable slot allocation in the network. Figure 13 bar chart, “Utility Scores for Candidate Nodes (First Simulation)”, is used to compare the utility of different nodes in a simulated scenario. The nodes are numbered from 1 to 5, and their utility scores are displayed on the y-axis. Node 2 is identified as the best choice, with the highest score of around 3.5. Node 5 and Node 1 are ranked next with decent scores, while Nodes 3 and 4 have lower scores. The colors of the bars are used to highlight the rankings, allowing for easy identification of the nodes that performed better in this simulation of proposed DAWPC-MAC.



**Figure 13.** Utility score for candidate nodes.

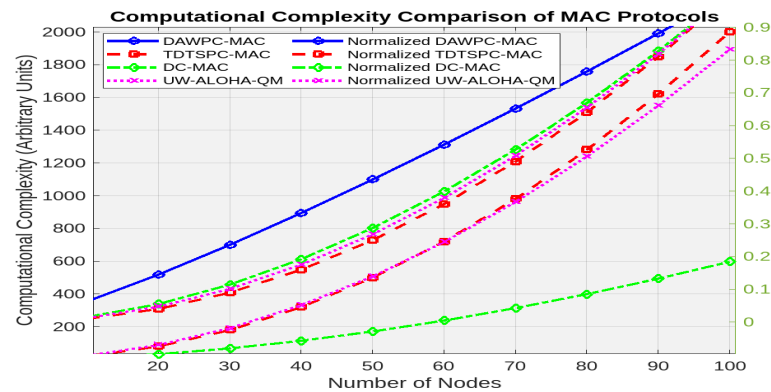
Figure 14 portrays the estimated channel utilization versus time for proposed DAWPC-MAC in UWSN. Estimated channel utilization depicts varying utilization of channel upto 0.98 till 300 s with linear increase while beyond that constant (non-linear) utilization of channel. Results also reflects that by deploying DAWPC-MAC initially, the network behaves in transient state; Lastly, networks adapt to steady state. Transit state also depicts the large number of active nodes in UWSN that affects the coverage of the whole network.



**Figure 14.** Channel Utilization vs. Simulation Cycles (Normalized).

## 5. Analysis of Computational Complexity

Finally, this analysis compares the computational complexity of the proposed DAWPC-MAC protocol with three benchmark techniques: TDTSPC-MAC, DC-MAC, and UW-ALOHA-QM as depicted in Figure 15. The results are based on simulated complexity trends for networks with 10 to 100 nodes.



**Figure 15.** Computational complexity of the proposed DAWPC-MAC protocol with benchmark techniques: TDTSPC-MAC, DC-MAC, and UW-ALOHA-QM.

### 5.1. Key Observations

- **DAWPC-MAC** exhibits the lowest computational complexity, growing as  $O(n^{1.2})$ .
- **TDTSPC-MAC** has the highest complexity, growing as  $O(n^2)$ .
- **DC-MAC** and **UW-ALOHA-QM** show intermediate complexity, growing as  $O(n^{1.8})$  and  $O(n^{1.9})$ , respectively.
- At **100 nodes**, DAWPC-MAC's complexity is **79** (arbitrary units), while TDTSPC-MAC, DC-MAC, and UW-ALOHA-QM have complexities of **2000**, **950**, and **1900**, respectively.
- Normalized complexity at 100 nodes is **1** for DAWPC-MAC, compared to **25** for TDTSPC-MAC, **12** for DC-MAC, and **24** for UW-ALOHA-QM.

### 5.2. Reasons for DAWPC-MAC's Superiority

- **Deep Q-Learning (DQN):** Enables adaptive scheduling and power control, reducing computational overhead.
- **Collision Avoidance:** Minimizes retransmissions and interference, lowering complexity.
- **Energy Efficiency:** Reduces computational burden on nodes, enhancing scalability.

### 5.3. Comparison with Benchmark Protocols

The computational complexity of the benchmark techniques and the proposed DAWPC-MAC is compared in the following Table 4:

**Table 4.** Comparison of MAC Protocols.

Protocol	Complexity Order	Scalability	Key Strengths
DAWPC-MAC	$O(n^{1.2})$	Excellent	Adaptive, energy-efficient
TDTSPC-MAC	$O(n^2)$	Poor	Time synchronization
DC-MAC	$O(n^{1.8})$	Moderate	Game-theoretic optimization
UW-ALOHA-QM	$O(n^{1.9})$	Moderate	Reinforcement learning

### 5.4. Implications for Underwater Networks

- **Energy Efficiency:** DAWPC-MAC's low complexity reduces energy consumption, critical for battery-operated nodes.
- **Scalability:** DAWPC-MAC's  $O(n^{1.2})$  complexity ensures efficient performance in large-scale networks.
- **Latency and Reliability:** DAWPC-MAC's adaptive mechanisms reduce latency and improve reliability, making it suitable for time-sensitive applications.

Summarizing, the proposed **DAWPC-MAC** protocol outperforms the benchmark techniques in computational complexity, scalability, and energy efficiency. Its use of Deep



Q-Learning enables adaptive and intelligent decision-making, making it highly suitable for dynamic and large-scale underwater networks. In contrast, the benchmark protocols exhibit higher complexity and are less scalable, limiting their effectiveness in real-world applications.

## 6. Conclusions and Future Directions

This paper presents a novel receiver-initiated DAWPC-MAC technique that combines adaptive power control, dynamic wake-up scheduling, and Deep Q-Learning DQN to improve energy efficiency and improve collision avoidance and synchronization in UWSNs. By incorporating a priority-based communication system and intelligent learning mechanisms, the protocol ensures reliable and timely delivery of critical data while minimizing energy consumption. In the case of the receiver-initiated MAC protocol, the land-based BS transmits communication signals to buoys using satellite communication. This satellite connection is for communication to the UWSN over long ranges. The BS transmits commands, data requests, or synchronization signals as required. When the signal arrives at the buoy, it is transmitted vertically to the sink node directly above it, which acts as the main communications hub. DAWPC-MAC protocol improves energy efficiency by regulating the wake-up and sleep cycles of nodes. A receiver-initiated wake-up mechanism is used, in which a low-power wake-up signal (WS) is broadcasted by the courier node as the receiver node, notifying neighboring transmitter nodes (Tx) about the upcoming communication window. Only Tx nodes with data to send respond, minimizing energy consumption. Resource allocation and scheduling are dynamically optimized by the deep Q-learning algorithm (DQN) based on real-time conditions such as considered water environment salinity, temperature along with node positions, traffic load, and energy levels. Proposed DAWPC-MAC ensures energy-efficient and reliable time-sensitive data transmission with packet delivery ratio (PDR) improves 14% compared to existing schemes TDTSPC-MAC, DC-MAC and ALOHA MAC and in-terms of throughput more than 70% improvement along with utility more than 60% improvement that contribute in network longevity and operational efficiency for time critical underwater applications compared to traditional methods. Furthermore, the use of experience replay and target networks by DQN has led to more consistent learning and faster convergence, affirming its effectiveness in dynamic underwater settings. In the future we will work on the DQN based MAC protocol for the Internet of underwater things (IOUT) by incorporating more underwater environmental parameters data for the physical characterization of the dynamic channel and considering hybrid mode of communication between sensors, carrier node, base station and AUV/ROV like acoustic, optical and electromagnetic from small to large networks.

**Author Contributions:** Conceptualization, Q.G. and W.U.R.; methodology, Q.G. and W.U.R.; software, W.U.R. and F.Z.; validation, Q.G., W.U.R. and F.Z.; formal analysis, F.Z. and W.A.; investigation, W.U.R., Q.G. and F.Z.; resources, Q.G. and F.Z.; data curation, M.T., M.I.K. and W.A.; writing—original W.U.R.; writing—review and editing, W.U.R., M.I.K. and W.A.; visualization, W.U.R., M.A., M.I.K. and F.Z.; supervision, Q.G. and F.Z.; project administration, Q.G. and F.Z.; funding acquisition, Q.G. and F.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by: 1. Natural Science Foundation of Heilongjiang Province of China: JQ2023A004. 2. Shenzhen Science and Technology Program under Grant No. JSGG20220831103800001. 3. Key Research and Development Program of ShanDong Province under Grant No. 2022CXGC020409.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Luo, J.; Chen, Y.; Wu, M.; Yang, Y. A survey of routing protocols for underwater wireless sensor networks. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 137–160. [\[CrossRef\]](#)
2. Nkenyereye, L.; Nkenyereye, L.; Ndibanje, B. Internet of Underwater Things: A Survey on Simulation Tools and 5G-Based Underwater Networks. *Electronics* **2024**, *13*, 474. [\[CrossRef\]](#)
3. Liu, S.; Khan, M.A.; Bilal, M.; Zuberi, H.H. Low Probability Detection Constrained Underwater Acoustic Communication: A Comprehensive Review. *IEEE Commun. Mag.* **2025**, *63*, 21–30. [\[CrossRef\]](#)
4. Zuberi, H.H.; Liu, S.; Bilal, M.; Alharbi, A.; Jaffar, A.; Mohsan, S.A.H.; Miyajan, A.; Khan, M.A. Deep-neural-network-based receiver design for downlink non-orthogonal multiple-access underwater acoustic communication. *J. Mar. Sci. Eng.* **2023**, *11*, 2184. [\[CrossRef\]](#)
5. Tian, X.; Du, X.; Wang, L.; Li, C.; Han, D. MAC protocol of underwater acoustic network based on state coloring. *Chin. J. Sens. Technol.* **2023**, *36*, 124–134.
6. Xue, L.; Lei, H.; Zhu, R. A Collision Avoidance MAC Protocol with Power Control for Adaptive Clustering Underwater Sensor Networks. *J. Mar. Sci. Eng.* **2025**, *13*, 76. [\[CrossRef\]](#)
7. Gang, Q.; Rahman, W.U.; Zhou, F.; Bilal, M.; Ali, W.; Khan, S.U.; Khattak, M.I. A Q-Learning-Based Approach to Design an Energy-Efficient MAC Protocol for UWSNs Through Collision Avoidance. *Electronics* **2024**, *13*, 4388. [\[CrossRef\]](#)
8. Zhang, T.; Gou, Y.; Liu, J.; Cui, J.-H. Traffic Load-Aware Resource Management Strategy for Underwater Wireless Sensor Networks. *IEEE Trans. Mob. Comput.* **2024**, *24*, 243–260. [\[CrossRef\]](#)
9. Kulla, E.; Matsuo, K.; Barolli, L. MAC Layer Protocols for Underwater Acoustic Sensor Networks: A Survey. In *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*; Springer International Publishing: Cham, Switzerland, 2022.
10. Lin, W.; Cheng, E.; Yuan, F. A MACA-based MAC protocol for Underwater Acoustic Sensor Networks. *J. Commun.* **2011**, *6*, 179–184. [\[CrossRef\]](#)
11. Wang, J.; Shen, J.; Shi, W.; Qiao, G.; Wu, S.; Wang, X. A novel energy-efficient contention-based MAC protocol used for OA-UWSN. *Sensors* **2019**, *19*, 183. [\[CrossRef\]](#)
12. Park, S.H.; Mitchell, P.D.; Grace, D. Performance of the ALOHA-Q MAC protocol for underwater acoustic networks. In Proceedings of the 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 16–17 August 2018; IEEE: Piscataway, NJ, USA, 2018.
13. Ye, X.; Fu, L. Deep reinforcement learning based MAC protocol for underwater acoustic networks. In Proceedings of the 14th International Conference on Underwater Networks & Systems, Atlanta, GA, USA, 23–25 October 2019.
14. Chirdchoo, N.; Soh, W.-S.; Chua, K.C. RIPT: A receiver-initiated reservationbased protocol for underwater acoustic networks. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 1744–1753.
15. Sendra, S.; Lloret, J.; Jimenez, J.M.; Parra, L. Underwater acoustic modems. *IEEE Sens. J.* **2015**, *16*, 4063–4071. [\[CrossRef\]](#)
16. Hsu, C.C.; Kuo, M.S.; Chou, C.F.; Lin, K.C.J. The Elimination of Spatial-Temporal Uncertainty in Underwater Sensor Networks. *IEEE/ACM Trans. Netw.* **2013**, *21*, 1229–1242.
17. Rahman, W.U.; Gang, Q.; Feng, Z.; Khan, Z.U.; Aman, M.; Ullah, I. A Q Learning-Based Multi-Hop Energy-Efficient and Low Collision MAC Protocol for Underwater Acoustic Wireless Sensor Networks. In Proceedings of the IEEE IBCAST, Bhurban, Murree, Pakistan, 22–25 August 2023. [\[CrossRef\]](#)
18. Du, H.; Wang, X.; Sun, W.; Zhang, J. An Adaptive MAC Protocol for Underwater Acoustic Networks Based on Deep Reinforcement Learning. In Proceedings of the IEEE CISCE, Guangzhou, China, 10–12 May 2024. [\[CrossRef\]](#)
19. Sun, W.; Sun, X.; Wang, B.; Wang, J.; Du, H.; Zhang, J. MR-SFAMA-Q: A MAC Protocol Based on Q-Learning for Underwater Acoustic Sensor Networks. *Diannao Xuekan* **2024**, *35*, 51–63.
20. ur Rahman, W.; Gang, Q.; Feng, Z.; Khan, Z.U.; Aman, M.; Bilal, M. A MACA-Based Energy-Efficient MAC Protocol Using Q-Learning Technique for Underwater Acoustic Sensor Network. In Proceedings of the IEEE ICCSNT, Dalian, China, 21–22 October 2023. [\[CrossRef\]](#)
21. Tomovic, S.; Radusinovic, I. DR-ALOHA-Q: A Q-learning-based adaptive MAC protocol for underwater acoustic sensor networks. *Sensors* **2023**, *23*, 4474. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hong, L.; Hong, F.; Guo, Z.W.; Yang, X. A TDMA-based MAC protocol in underwater sensor networks. In Proceedings of the IEEE WCNC, Dalian, China, 12–14 October 2008; pp. 1–4.
23. Yang, S.; Liu, X.; Su, Y. A Traffic-Aware Fair MAC Protocol for Layered Data Collection Oriented Underwater Acoustic Sensor Networks. *Remote Sens.* **2023**, *15*, 1501. [\[CrossRef\]](#)
24. Gazi, F.; Ahmed, N.; Misra, S.; Wei, W. Reinforcement learning-based MAC protocol for underwater multimedia sensor networks. *ACM Trans. Sens. Netw.* **2022**, *18*, 1–25. [\[CrossRef\]](#)
25. Chen, Y.D.; Lien, C.Y.; Chuang, S.W.; Shih, K.P. DSSS: A TDMA-based MAC protocol with dynamic slot scheduling strategy for underwater acoustic sensor networks. In Proceedings of the IEEE Oceans-Spain, Santander, Spain, 6–9 June 2011; pp. 1–6.

26. Cho, H.J.; Namgung, J.I.; Yun, N.Y.; Park, S.H.; Kim, C.H.; Ryuh, Y.S. Contention-free MAC protocol based on priority in underwater acoustic communication. In Proceedings of the IEEE Oceans-Spain, Santander, Spain, 6–9 June 2011; pp. 1–7.
27. Zheng, M.; Ge, W.; Han, X.; Yin, J. A spatially fair and low conflict medium access control protocol for underwater acoustic networks. *J. Mar. Sci. Eng.* **2023**, *11*, 802. [\[CrossRef\]](#)
28. Alablani, I.A.; Arafah, M.A. EE-UWSNs: A joint energy-efficient MAC and routing protocol for underwater sensor networks. *J. Mar. Sci. Eng.* **2022**, *10*, 488. [\[CrossRef\]](#)
29. Alfouzan, F.A. Energy-efficient collision avoidance MAC protocols for underwater sensor networks: Survey and challenges. *J. Mar. Sci. Eng.* **2021**, *9*, 741. [\[CrossRef\]](#)
30. Xu, F.; Yang, F.; Zhao, C.; Wu, S. Deep reinforcement learning based joint edge resource management in maritime network. *China Commun.* **2020**, *17*, 211–222.
31. Giannopoulos, A.E.; Spantideas, S.T.; Zetas, M.; Nomikos, N.; Trakadas, P. FedShip: Federated Over-the-Air Learning for Communication-Efficient and Privacy-Aware Smart Shipping in 6G Communications. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 19873–19888.
32. Wang, C.; Zhang, X.; Gao, H.; Bashir, M.; Li, H.; Yang, Z. Optimizing anti-collision strategy for MASS: A safe reinforcement learning approach to improve maritime traffic safety. *Ocean. Coast. Manag.* **2024**, *253*, 107161.
33. He, J.; Liu, Z.; Zhang, Y.; Jin, Z.; Zhang, Q. Power Allocation Based on Federated Multi-Agent Deep Reinforcement Learning for NOMA Maritime Networks. *IEEE Internet Things J.* **2024**, *early access*.
34. Rodoshi, R.T.; Song, Y.; Choi, W. Reinforcement learning-based routing protocol for underwater wireless sensor networks: A comparative survey. *IEEE Access* **2021**, *9*, 154578–154599.
35. Xylouris, G.; Nomikos, N.; Kalafatis, A.; Giannopoulos, A.; Spantideas, S.; Trakadas, P. Sailing into the future: Technologies, challenges, and opportunities for maritime communication networks in the 6G era. *Front. Commun. Netw.* **2024**, *5*, 1439529.
36. Wang, C.; Shen, X.; Wang, H.; Zhang, H.; Mei, H. Reinforcement learning-based opportunistic routing protocol using depth information for energy-efficient underwater wireless sensor networks. *IEEE Sens. J.* **2023**, *23*, 17771–17783.
37. Park, S.H.; Mitchell, P.D.; Grace, D. Reinforcement learning based MAC protocol (UW-ALOHA-QM) for mobile underwater acoustic sensor networks. *IEEE Access* **2020**, *9*, 5906–5919.
38. AlamAlam, M.I.I.; Hossain, M.F.; Munasinghe, K.; Jamalipour, A. MAC protocol for underwater sensor networks using EM wave with TDMA based control channel. *IEEE Access* **2020**, *8*, 168439–168455.
39. Rodoplu, V.; Park, M.K. An energy-efficient MAC protocol for underwater wireless acoustic networks. In Proceedings of the OCEANS 2005 MTS/IEEE, Washington, DC, USA, 17–23 September 2005; IEEE: Piscataway, NJ, USA, 2005.
40. Qian, L.; Zhang, S.; Liu, M.; Zhang, Q. A MACA-based power control MAC protocol for underwater wireless sensor networks. In Proceedings of the 2016 IEEE/OES China Ocean Acoustics (COA), Harbin, China, 9–11 January 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–8.
41. Alfouzan, F.A.; Shahrabi, A.; Ghoreyshi, S.M.; Boutaleb, T. A Collision-Free Graph Coloring MAC Protocol for Underwater Sensor Networks. *IEEE Access* **2019**, *7*, 39862–39878. [\[CrossRef\]](#)
42. Zhu, R.; Liu, L.; Li, P.; Chen, N.; Feng, L.; Yang, Q. DC-MAC: A delay-aware and collision-free MAC protocol based on game theory for underwater wireless sensor networks. *IEEE Sens. J.* **2024**, *24*, 6930–6941.
43. Lokam, A. ADRP-DQL: An adaptive distributed routing protocol for underwater acoustic sensor networks using deep Q-learning. *Ad Hoc Netw.* **2025**, *167*, 103692.
44. Zhang, Z.; Shi, W.; Niu, Q.; Guo, Y.; Wang, J.; Luo, H. A load-based hybrid MAC protocol for underwater wireless sensor networks. *IEEE Access* **2019**, *7*, 104542–104552. [\[CrossRef\]](#)

**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.