SCIENCE CHINA
Information Sciences

• Supplementary File •

AUV-assisted water-air cross-medium communication with DQN based path planning

Qiang Chen[1], Wei Li[1,2*], Qinyu Zhang[1,2] & Zhonghan Hao[1]

[1]Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
[2]Pengcheng Laboratory (PCL), Shenzhen 518000, China

Appendix A Variable description

The main variables and their descriptions are listed in table A1.

Appendix B Background and motivation

With rising demands in environmental monitoring, emergency rescue, and underwater archaeology, the deep integration of information transmission technologies across the air, land, and marine domains—specifically between the seawater and air mediums—has become increasingly imperative.

The most common cross-medium transmission methods rely on relays, such as buoys, to bridge the communication gap between air and seawater [1]. These systems typically integrate underwater acoustic (UWA) communication with aerial radio communication and are widely adopted due to their extended transmission range. Direct cross-medium transmission through seawater using electromagnetic waves or lasers has also been investigated [2]; however, this approach is generally limited to short ranges—on the order of several to a few tens of meters—and is therefore not practical for most applications.

Nevertheless, in relay-based acoustic–radio cross-medium transmission, the UWA channel presents far greater challenges than the radio channel. The UWA channel is inherently constrained by limited bandwidth, significant propagation delays, and high levels of environmental noise [3], all of which severely degrade communication quality. Typically, acoustic links experience transmission delays that are five orders of magnitude greater than those of radio links, while their achievable data rates are approximately three orders of magnitude lower [4]. Consequently, acoustic links are prone to congestion, whereas radio links often remain underutilized. A pronounced channel gap thus exists between the UWA and radio links: the radio channel in air supports high throughput and low latency, whereas the UWA channel in water exhibits low capacity, long delays, and severe time variability. This fundamental disparity in channel characteristics poses a significant challenge to achieving efficient and reliable end-to-end cross-medium communication.

Most existing relay-based acoustic-radio cross-medium transmission schemes rely on static surface relays and concentrate primarily on enhancements at the medium access control (MAC) [5, 6] or network protocol layers. However, due to the spatial heterogeneity of underwater environments, better UWA transmission conditions may exist in nearby areas. This makes the problem inherently tied to the physical layer. The simplified physical models commonly adopted in MAC and network layer studies often overlook such spatial variations. Moreover, static surface relays [7] are difficult to adapt to this spatial-varying channel conditions, resulting in increased transmission latency and reduced communication efficiency.

Recently, [8] provides a cross-medium transmission scheme with mobile relay. Since modern AUVs are now capable of carrying both acoustic and radio-frequency communication modules [9], opening up new possibilities for their use in cross-medium applications. An AUV can function as a dual-mode relay—either surfacing to act as an acoustic-to-radio relay or diving to enhance UWA transmission. However, existing research on AUV-enabled cross-medium relaying often treats the UWA and aerial radio channels in isolation. In [8], by optimizing the UWA transmission, the cross-medium transmission can be improved by mobile relay. Very few studies have considered integrated transmission that span both underwater and aerial environments [10].

Although mobile relaying by the AUV has shown promise in cross-medium communication [11], effective path planning for mobile relays remains a significant challenge.

Most existing methods focus on obstacle avoidance with predefined start and end positions in a uniform field, which is not well-suited for cross-medium scenarios where the relay must adapt to the spatial variability and complexity of the UWA channel, and where the start and end points may be loosely constrained or only defined at the surface, and the operational space is fully open. Moreover, due to the lack of a clear analytical expression for the UWA environment map, ergodic-based approaches become computationally prohibitive, as they require exhaustive searches over numerous candidate paths in a fully open space.

Among existing path planning algorithms, intelligent optimization algorithms show the greatest promise for mobile relay systems in cross-medium communications. Existing path planning algorithms fall into four main categories: geometric model search, probabilistic sampling-based, artificial potential field, and intelligent optimization algorithms [12]. Geometric methods [13] suit uniform, static environments but struggle with complex underwater conditions. Probabilistic sampling [14] is efficient for large spaces but can yield

* Corresponding author (email: li.wei@hit.edu.cn)

Table A1 Variable description table

Parameter	Abbreviation	Connotation
$\mathbf{r}$	-	The position coordinates of the receiver
$\mathbf{r}_{\mathrm{BS}}$	-	The position coordinates of the onshore BS as the receiver
$\mathbf{r}_{\mathrm{S}}$	-	The position coordinates of the AUV relay when it is floating on the water surface
$\mathbf{r}_i$	-	The ith position along the AUV path
$h_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}})$	-	Radio channel power gain in the air at location $\mathbf{r}_{\mathrm{S}}$ from $\mathbf{r}_{\mathrm{S}}$
$\mathrm{C}_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}})$	-	The uplink data transmission rate between the AUV at $\mathbf{r}_{\mathrm{S}}$ and the onshore BS at $\mathbf{r}_{\mathrm{BS}}$ in the air
$\mathrm{SNR}(\mathbf{g}, \mathbf{r}_i, f)$	$\mathrm{SNR}(\mathbf{r}_i)$	UWA SNR at position $\mathbf{r}_i$
SL	-	Source level
NL	-	Underwater noise level
$N_t(f)$	-	Turbulence noise
$N_s(f)$	-	Shipping activities noise
$N_\omega(f)$	-	Wave noise
$N_{th}(f)$	-	Seawater noise
$\mathbf{g}$	-	Transmitter position
$\mathrm{TL}(\mathbf{g}, \mathbf{r}, f)$	TL	Acoustic transmission loss
DI	-	Directivity index
$C_{\mathrm{u}}(\mathbf{g}, \mathbf{r}_i)$	$C_{\mathrm{u}}(\mathbf{r}_i)$	UWA channel capacity at location $\mathbf{r}_i$
T_{u}	-	UWA transmission duration
T_{AUV}	-	The time for the AUV to vertically return to the water surface
T_{a}	-	Radio transmission duration
$\mathbf{\Omega}$	-	A series of position points between the initial position $\mathbf{r}_0$ and the transmission end point $\mathbf{r}_N$ of the AUV
N	-	Total number of movement steps for data transmission
h_N	-	Water depth at the transmission endpoint $\mathbf{r}_N$
$\mathcal{F}$	-	The acoustic shadow zone where UWA signals are interrupted or significantly attenuated
S	-	State space
A	-	Action space
F	-	Transition function
ϵ	-	Exploration probability
R_{mid}	-	The intermediate reward
d	-	Distance between the AUV and the underwater transmitter, $d_i = \Vert\mathbf{g}-\mathbf{r}_i\Vert_2$
l	-	The relative distance increment, $l_i = d_i - d_{i-1}$
R_{final}	-	The terminal reward
R_{b}	-	End-to-end data transmission rate
η	-	Radio channel utility

suboptimal paths and is computationally intensive. Artificial potential field algorithms [15] work well for simple, low-dimensional tasks but face challenges in cluttered or high-dimensional environments. Intelligent optimization algorithms [16, 17] adapt effectively to dynamic underwater settings.

The heuristic ant colony optimization (ACO) mechanism which belongs to the intelligent optimization algorithms is a heuristic global optimization algorithm applicable to three-dimensional underwater path planning [18]. Its adaptive characteristics allow it to operate without manual parameter adjustment [12]. However, due to the lack of prior environmental information [12], ACO has a slow convergence rate and is prone to getting trapped in local optima.

Deep reinforcement learning (DRL)-based algorithms [19] among intelligent optimization algorithms learn directly through interaction with the environment, enabling robust adaptability to the uncertain and dynamic conditions characteristic of underwater environments [20]. Within the DRL framework, Q-value-based methods [21] are particularly effective for discrete action spaces, making them suitable for path planning tasks involving discrete movement directions. In [8], the Q learning is used to increase the data length in UWA link in cross-medium transmission.

The Double Deep Q-Network (DDQN) introduces a double Q-network architecture that decouples action selection from value estimation [22]. This mechanism mitigates the overestimation bias inherent in traditional Q-learning approaches, thereby improving learning stability and accelerating convergence. The dueling network [23] architecture decomposes the action-value function into a state-value function and an action-advantage function, allowing the network to separately optimize state evaluation and action selection, thereby reducing redundant computations and enhancing policy efficiency. Multi-step learning [24] accounts for rewards over multiple future time steps starting from the current state, thereby accelerating learning speed and enhancing final performance. The enhanced robustness and adaptive capabilities of Multi-step Dueling DDQN (M3DQN) render it well-suited for complex, spatially varying UWA channels, where environmental parameters change unpredictably. Consequently, M3DQN demonstrates significant potential for optimizing path planning in mobile relay systems, contributing to more reliable and efficient cross-medium communication.

This paper proposes a cross-medium communication scheme that employs an AUV as a mobile relay. By effectively utilizing the idle resources of the radio link to improve the performance of the acoustic link, this scheme thereby enhances the end-to-end cross-medium transmission performance. The main contributions are as follows:

- We propose a cross-medium mobile relaying system. In the cross-medium mobile relaying system, we aim to minimize the end-to-end cross-medium transmission duration by the mobile relay, leveraging an acoustic-radio cross-medium channel model. This model effectively captures the channel differences in the cross-medium links. Unlike the radio channel, the UWA channel is spatially varying. Therefore, the movement of the mobile relay during the idle period of the radio link can shorten the UWA transmission duration, thereby improving cross-medium transmission performance.
- AUV path planning method based on the M3DQN (APP-M3DQN) learning, is proposed to minimize the end-to-end cross-medium transmission duration. Unlike traditional path planning algorithms that prioritize obstacle avoidance in the homogeneous medium, our route can maximize transmission performance by allowing flexible destination positions and adapting to spatially varying acoustic environments through interaction with the underwater environment.
- Simulation results demonstrate that the AUV mobile relay scheme achieves superior performance compared to the traditional static relay scheme [7], particularly in terms of the end-to-end cross-medium transmission duration, end-to-end transmission rate, and radio channel utilization. Furthermore, the proposed APP-M3DQN, outperforms other approaches [25]. The simulations are based on a selected region of the South China Sea.

Appendix C Cross-layer modeling for acoustic–radio cross-medium relaying

In this work, the cross-medium uplink relay network comprises an underwater source node, a mobile AUV relay node that can operate underwater and receives acoustic signals and send radio signal at the water surface, and an above-water destination node (e.g., an onshore BS), to which the relay forwards information via radio transmission upon surfacing. Transmission is divided into two hops: UWA transmission from the source to the relay, followed by above-water radio forwarding from the relay to the destination.

Relay networks are typically studied in a layer-by-layer manner; consequently, most existing studies on cross water–air relaying are conducted at the network layer and tend to oversimplify physical-layer effects, particularly by modeling UWA transmission loss using idealized distance-based spreading laws. However, the physical disparity between media is a key issue in cross-medium transmission. Therefore, this work focuses on a cross-layer model for acoustic–radio cross-medium relaying that jointly considers UWA propagation and above-water radio transmission, integrates detailed physical-layer characteristics into network-layer analysis, and leverages AUV relay mobility to enhance cross-medium transmission efficiency.

Appendix C.1 Air radio channel model

When the AUV is floating on the sea surface as a acoustic-to-radio relay, it forwards the underwater data collected to the onshore BS. Generally, the air medium is homogeneous, and the data transmission between the AUV and the onshore BS follows the line-of-sight propagation. The radio channel power gain can be expressed as [26]:

$$h_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}}) = \frac{\chi}{[||\mathbf{r}_{\mathrm{BS}} - \mathbf{r}_{\mathrm{S}}||_2^2]^{\xi}}, \tag{C1}$$

where the parameter χ is the channel power gain related to the reference distance (e.g., 1 m), and ξ is the radio path loss exponent greater than 1, $\mathbf{r}_{\mathrm{BS}}$ represents the position coordinates of the onshore BS, and $\mathbf{r}_{\mathrm{S}}$ represents the position coordinates of the AUV relay when it is floating on the water surface.

The achievable data rate for the uplink transmission between the AUV and the onshore BS, based on Shannon's theorem [27], can be expressed as:

$$\mathrm{C}_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}}) = B_{\mathrm{a}} \log_2\left(1 + \frac{q_{\mathrm{a}} h_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}})}{N_{\mathrm{a}}}\right), \tag{C2}$$

where B_{a} is the bandwidth of the radio channel, q_{a} represents the transmission power between the AUV and the onshore BS, and N_{a} represents the background noise power at the onshore BS, which acts as the receiver.

Appendix C.2 Underwater acoustic channel model

According to the passive sonar equation [28], The SNR at the receiver of the acoustic channel can be expressed as: $\mathrm{SNR} = \mathrm{SL} - \mathrm{NL} - (\mathrm{TL} - \mathrm{DI})$, where SL is the source level, NL is the underwater noise level, TL is the acoustic transmission loss, and DI is the directivity index.

The source level SL is obtained by:

$$\mathrm{SL} = 10 \lg P + 170.77, \tag{C3}$$

where P is the transmission power of the UWA transmitter.

NL is primarily composed of self-noise of the UWA receiver, turbulence, ships, waves, and seawater. The turbulence noise $N_t(f)$, shipping activities noise $N_s(f)$, wave noise $N_\omega(f)$, and seawater noise $N_{th}(f)$ can be expressed in sequence as [29]:

$$10 \lg N_t(f) = 17 - 30 \lg f, \tag{C4}$$

$$10 \lg N_s(f) = 40 + 20(s - 0.5) + 26 \lg f - 60 \lg(f + 0.03), \tag{C5}$$

$$10 \lg N_\omega(f) = 50 + 7.5\omega^{0.5} + 20 \lg f - 40 \lg(f + 0.4), \tag{C6}$$

$$10\lg N_{th}(f) = -15 + 20\lg f, \tag{C7}$$

where s is the factor of the shipping activity, with a range of [0, 1], representing low and high activity. ω is the wind speed, in units of m/s, and f is the sound wave frequency, in units of kHz.

$$\begin{aligned} \mathrm{NL} &= 10\lg(N(f)) \\ &= 10\lg\left(N_{MR} + N_t(f) + N_s(f) + N_\omega(f) + N_{th}(f)\right) \\ &\approx 10\lg\left(N_{MR} + N_\omega(f) + N_{th}(f)\right), \end{aligned} \tag{C8}$$

where N_{MR} denotes the self-noise of the UWA receiver. In this work, the UWA receiver is served by the cross-medium relay, which can be implemented by an AUV. Turbulence noise influences only the very low frequency region ($f < 10$ Hz), while shipping noise dominates in the 10–100 Hz frequency range. In UWA communications, f is at kHz level, thereby $N_t(f) \approx 0$ and $N_s(f) \approx 0$ [30] in this work.

In contrast to the air medium, the underwater medium exhibits complex and heterogeneous properties, resulting in non-line-of-sight propagation. Consequently, the TL demonstrates significant spatial variability in UWA channels. Assuming that the sea depth, seabed topography, and sound velocity profile are known, and based on ray-tracing theory, the UWA TL can be expressed as a function of the transmitter position $\mathbf{g}$, receiver position $\mathbf{r}$, and frequency f, denoted by $\mathrm{TL}(\mathbf{g}, \mathbf{r}, f)$. The bellhop model [31], founded on ray-tracing principles, is employed to numerically evaluate the TL. An example of the spatially varying TL obtained from the bellhop model is shown in Fig. C1. It incorporates the inhomogeneous nature of the underwater medium and outputs the transmission loss distribution after describing the input environment, which can reflect the spatial variation of the UWA channels.

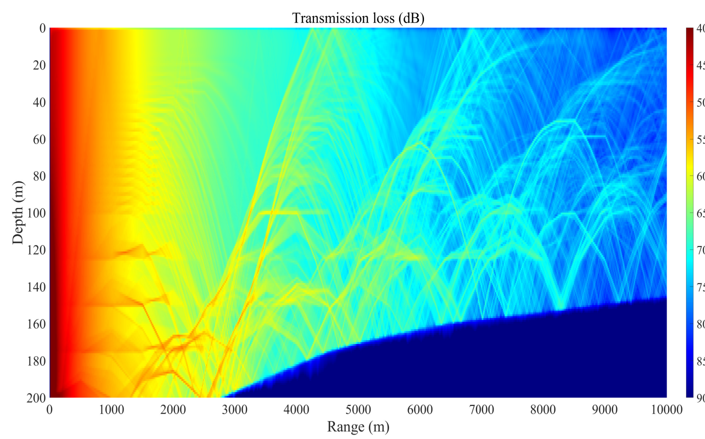


Figure C1 An example of spatially varying TL distribution diagram based on the bellhop model [31] for UWA transmitter at (0, 200 m) with an acoustic wave frequency of 10 kHz. The deep blue regions at the bottom of the figure indicate sound shadow zones, where sound waves cannot propagate directly. The area corresponds to an oceanographic profile at 117.6°E, 21.9°N in the South China Sea, with a water depth of 200 m and a horizontal extent of 10 km. The UWA transmitter is located at (0, 200 m). The blue regions represent acoustic shadow zones, where direct sound propagation is inhibited.

Therefore, the SNR at the UWA receiver in dB is as follows:

$$\begin{aligned} \mathrm{SNR}(\mathbf{g}, \mathbf{r}, f) &= \mathrm{SL} - \mathrm{NL} - (\mathrm{TL} - \mathrm{DI}) \\ &= 10\lg P + 170.77 - 10\lg(N(f)) - \mathrm{TL}(\mathbf{g}, \mathbf{r}, f), \end{aligned} \tag{C9}$$

where $\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)$ can be approximately expressed as a function of the position of the underwater transmitter $\mathbf{g}$, the position of the AUV relay diving underwater as a receiver $\mathbf{r}$, and the acoustic frequency f, due to the spatial varying TL. We assume that the sonar receiver mounted on the AUV is omnidirectional, i.e., $\mathrm{DI} \approx 0$ dB.

According to Shannon's channel formula [27], if the frequency f is known, the capacity of the UWA channel with spatially varying characteristics is:

$$\begin{aligned} C_{\mathrm{u}}(\mathbf{g}, \mathbf{r}) &= B_{\mathrm{u}} \log_2\left(1 + 10^{\frac{\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)}{10}}\right) \\ &= B_{\mathrm{u}} \log_2\left(1 + 10^{\frac{10\lg P + 170.77 - 10\lg(N(f)) - \mathrm{TL}(\mathbf{g}, \mathbf{r}, f)}{10}}\right), \end{aligned} \tag{C10}$$

where B_{u} is the bandwidth of the UWA channel.

If both media are jointly considered through comparative analysis, the spatial variability of the capacity $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r})$ is inevitable compared with $\mathrm{C_a}(\mathbf{r}_{\mathrm{S}})$.

Appendix D Proposed cross-medium transmission via mobile relay

When comparing the above two channel models in cross-medium transmission, the spatial variation of the capacity in the UWA environment is non-negligible compared to that in the air radio environment within the context of cross-medium communications. As shown in Fig. C1, even slight change in the underwater position can significantly reduce TL. Consequently, using a mobile relay with a reasonable movement can enhance the efficiency of UWA transmission.

The mobile relay scheme proposed in this paper employs an AUV to establish a hybrid acoustic-radio link communication, as illustrated in Fig. D1. The AUV navigates along an underwater path that adapts to the spatial variations of the UWA channel, enabling efficient data acquisition via UWA communication before returning to the surface. Once all data is received, the AUV surfaces to forward the data to an onshore BS as an acoustic-to-radio relay, thereby facilitating cross-medium data transmission.

As illustrated in Fig. D1, which depicts an example of the cross-medium data transmission model with a mobile relay. The proposed mobile relay scheme utilizes an AUV as a dynamic relay. The AUV, acting as a mobile relay, departs from its initial position and navigates along an optimized path to adapt to the spatial variability of the UWA channel, thereby minimizing the cross-medium transmission duration to the greatest extent. It collects data from the underwater transmitter during its trajectory until all data is received. The AUV then ascends vertically to the surface and transmits the data to the onshore BS via radio link.

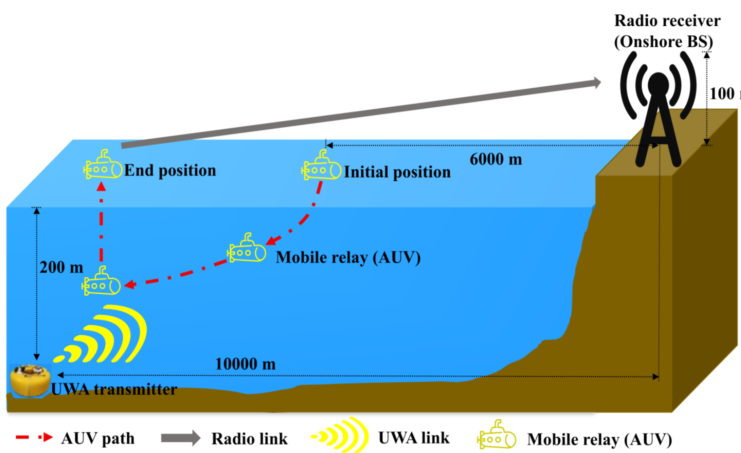


Figure D1 An illustrative example of the proposed cross-medium transmission employing an AUV as the mobile relay. In this example, the UWA transmitter is positioned 200 m underwater, with a horizontal distance of 10,000 m from the onshore BS located 100 m above the water surface. The gray arrow in Fig. D1 indicates the radio link, while the yellow ripple-like arcs represent the UWA link. After diving underwater to receive signals from the UWA transmitter via the acoustic link, the AUV resurfaces to forward the data to the onshore BS through the radio link.

Appendix D.1 Mobile relay path planning for minimizing cross-medium transmission duration

In this section, the cross-medium transmission is formulated as a path-planning problem for the mobile relay. The AUV mobile relay must satisfy the following prerequisites:

- We assume calm sea conditions [32], where the acoustic propagation pattern shown in Fig. C1 is stable and known in advance, under which $\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)$ remains invariant in time [33], accordingly, $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r})$ is invariant and known.
- We focus solely on duration of the data transmission, neglecting link-establishing frames, handshake frames, preamble headers, and other control messages. For simplicity in path planning, we also disregard the AUV's kinematic posture (pitch, roll, and yaw).
- We assume that the AUV relay is capable of omnidirectional movement (up, down, left, and right) at a constant speed, while the transmitter remains fixed at a given location.
- We disregard the potential rate reduction introduced by mobility, since the channel capacity gains arising from spatial variations in the UWA channel are numerically much greater than the mobility-induced losses.
- Since the mobile relay can be easily recalled to be recharged, energy consumption factors are not considered in this work.

In cross-medium transmission, we aim to shorten the end-to-end cross-medium transmission duration to improve the performance of acoustic-radio cross-medium communication systems.

In the scenario of using the AUV as a mobile acoustic-to-radio relay for cross-medium information transmission, the end-to-end cross-medium transmission duration T can be expressed as:

$$T = T_{\mathrm{u}} + T_{\mathrm{AUV}} + T_{\mathrm{a}}, \tag{D1}$$

where T_{u} is the UWA transmission duration for the AUV to move underwater according to the path planning algorithm and efficiently collect data, T_{AUV} denotes the time to immediately return vertically to the water surface after the completion of data collection, and T_{a} represents the radio transmission duration during which the AUV spends on the water surface forwarding data to the onshore BS.

The underwater area is divided into grids along both horizontal and vertical directions, with a spacing of p, as illustrated in the example shown in Fig. D2. Within this grid, the AUV path is discretized into Ω, where Ω consists of a series of positions between the initial position $\mathbf{r}_0$ and the transmission end position $\mathbf{r}_N$:

$$\mathbf{\Omega} = \{\mathbf{r}_0, \mathbf{r}_1, \cdots, \mathbf{r}_N\}, \tag{D2}$$

where $\mathbf{r}_i = (x_i, h_i)$, and represents the specific coordinates of the i-th position on the path with horizontal x_i and vertical h_i, and N represents the total number of step in the path planning. When the AUV move to $\mathbf{r}_i$, we have $\mathrm{SNR}(\mathbf{g}, \mathbf{r}_i, f)$, accordingly, $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r}_i)$.

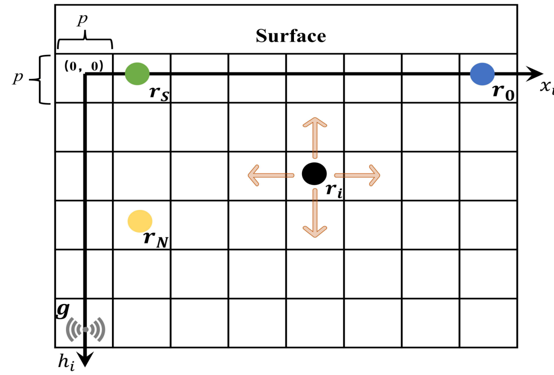


Figure D2 Discretized grid map. The underwater area is divided into grids with spacing p along the horizontal and vertical directions. The mobile relay starts at $\mathbf{r}_0$ (blue circle), completes data reception at $\mathbf{r}_N$ (yellow circle), and moves step by step with four possible directions (orange arrows) at each grid point $\mathbf{r}_i$. Finally, it surfaces vertically at $\mathbf{r}_S$ (green circle) to complete the transmission.

In this scenario, it is assumed that the AUV moves at a constant speed in all directions, with the position of the UWA transmitter and the transmission data volume being fixed. The goal is to optimize the AUV's movement path to minimize the end-to-end cross-medium transmission duration, thereby improving the performance of the cross-medium system. And after AUV mobile finish its receiving information, it comes back to surface at the position where $\mathbf{r}_{\mathrm{S}} = (x_N, 0)$, to relay the information to the onshore BS.

To minimize the end-to-end cross-medium transmission duration, the optimization problem can be expressed as $\mathcal{P}1$ as follows:

$$
\begin{aligned}
\mathcal{P}1: \quad & \min_{\mathbf{\Omega}} T, && \text{(D3)}\\
s.t.: \quad & \text{(D1)}, &&\\
& T_{\mathrm{u}} = N \times \Delta t, && \text{(D4)}\\
& T_{\mathrm{AUV}} = h_N / v_{\mathrm{AUV}}, && \text{(D5)}\\
& T_{\mathrm{a}} = D_{\mathrm{total}} / C_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}}), && \text{(D6)}\\
& \sum_{i=0}^{N} \Delta t \times C_{\mathrm{u}}\left(\mathbf{r}_i\right) \geqslant D_{\mathrm{total}}, && \text{(D7)}\\
& x_i \in \left[0, X_{\max}\right], && \text{(D8)}\\
& h_i \in \left[0, H_{\max}\right], && \text{(D9)}\\
& \mathbf{r}_i = (x_i, h_i) \notin \mathcal{F}, && \text{(D10)}
\end{aligned}
$$

where D_{total} is the total amount of data (in bits) transmitted and relayed, and $C_{\mathrm{a}}(\mathbf{r}_{\mathrm{S}})$ represents the radio channel capacity at $\mathbf{r}_{\mathrm{S}}$ where the relay reaches the water surface, $C_{\mathrm{u}}(\mathbf{r}_i)$ is short for $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r}_i)$ by assuming a fixed and known $\mathbf{g}$, and represents the UWA channel capacity at $\mathbf{r}_i$. Additionally, the AUV's movement range must remain within the valid range, where the horizontal direction cannot exceed $X_{\max}$ and the vertical direction cannot exceed the water depth $H_{\max}$.

In $\mathcal{P}1$, constraints (D1) and (D4)-(D6) specify the composition of T. Constraint (D7) enforces the data transmission requirement during the mobile relay's underwater duration T_{u}. Constraints (D8)-(D10) capture the spatial limitations of the mobile relay while underwater. $\mathcal{F}$ denotes the acoustic shadow zone where UWA signals are interrupted or significantly attenuated.

In general, transmission duration underwater T_{u} is substantially longer than transmission duration in the air T_{a} owing to the different capacity of UWA and radio channels, while T_{AUV} further prolongs the underwater duration. Consequently, D_{total} must be sufficiently large to justify the dive. Subject to the spatial constraints in (D8)-(D9), adequate spatial variance is required, which is typically ensured when $H_{\max}$ and $X_{\max}$ are sufficiently large.

Furthermore, our path planning problem $\mathcal{P}1$ are designed for calm sea conditions [32], where the environment remains consistent and UWA channel conditions are relatively stable during both the offline computations performed prior to diving and the subsequent dive-to-surface cycle. Their performance, however, may deteriorate in rough sea conditions. The highly dynamic nature of the UWA environment under such conditions [32] falls beyond the scope of this work, as it can introduce challenges such as fluctuations in channel quality, positional jitters of the transmitter or AUV, and other uncertainties, all of which may adversely affect the performance of cross-medium transmission.

Appendix D.2 AUV path planning based on M3DQN for cross-medium mobile relay

To address the cross-medium optimization problem proposed in $\mathcal{P}1$, we propose an APP-M3DQN algorithm. The M3DQN learning, grounded in DRL, integrates the strengths of the dueling network architecture, double DQN, and multi-step temporal difference learning, enabling it to cope with spatial variation characteristics of UWA channels for path planning.

To minimize the end-to-end cross-medium transmission duration T in $\mathcal{P}1$, the optimal path should traverse regions with high $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r})$. To simplify the computation, we approximate this by guiding the path along areas of high $\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)$. Accordingly, $\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)$ is used to define the underwater SNR map.

With APP-M3DQN, the AUV can autonomously plan an optimal path using the underwater SNR map. By dynamically adjusting its position, the AUV can navigate toward high-SNR regions for a high transmission rate while avoiding the blocking effects of acoustic shadow zones, thereby enhancing data transmission efficiency.

The key components in APP-M3DQN, including the state space, action space, and reward function, are designed as follows:

Appendix D.2.1 *State space*

The current state is represented by the AUV's current position and the remaining data to be transmitted. Therefore, the state space S can be expressed as:

$$\begin{aligned} S = \{(x_i, h_i, D_i) | x_i \in [0, X_{\max}],\ h_i \in [0, H_{\max}], \\ D_i \in [0, D_{\text{total}}]\}, \end{aligned} \tag{D11}$$

where the AUV current position $\mathbf{r}_i = (x_i, h_i)$ must satisfy the constraints in (D8) -(D10), and the remaining data to be transmitted is given by $D_i = D_{i-1} - \Delta t \times C_{\text{u}}(\mathbf{r}_i)$.

Appendix D.2.2 *Action space*

Assume that the AUV can choose to move up, down, left, or right, and each move is by a fixed distance. The action space A can be expressed as:

$$A = \{(0,1), (1,0), (0,-1), (-1,0)\}, \tag{D12}$$

where (0, 1) represents move up, (1, 0) represents right, (0, -1) represents down, and (-1, 0) represents left.

Appendix D.2.3 *Transition function*

Since the AUV's movement is deterministic, i.e., given the current state and action, the next state is deterministic.

$$F(S, A, S') = \begin{cases} 1, & \text{if } S' = \text{next}(S, A), \\ 0, & \text{else}. \end{cases} \tag{D13}$$

In the transition function, $\text{next}(S, A)$ refers to the next state computed based on the current state S and the action A.

Appendix D.2.4 *Exploration probability*

The exploration probability is based on the ϵ-greedy strategy for selecting the next action. With probability ϵ, an action is chosen randomly, and with probability $1 - \epsilon$, the action predicted by the main network as the optimal action is selected. And ϵ is defined as:

$$\epsilon = \begin{cases} \epsilon_0 \beta^n, & \epsilon > \epsilon_{\min}, \\ \epsilon_{\min}, & \epsilon \leqslant \epsilon_{\min}, \end{cases} \tag{D14}$$

where $\epsilon_{\min}$ is the minimum exploration probability, ϵ_0 is the initial exploration probability, β is the exploration decaying rate, and n is the current episode.

Appendix D.2.5 *Reward function*

For path planning problem $\mathcal{P}1$, the reward function is designed in three parts: fixed reward R_{fixed}, intermediate reward R_{mid}, and terminal reward R_{final} .

The fixed reward R_{fixed} is defined as a combination of penalties and rewards based on the satisfaction of (D4), (D5), and (D7)–(D10) in $\mathcal{P}1$. It includes four components: the reward R_{des} for successful data reception, which is granted when the AUV receives all data and incentivizes completion of the core task according to (D7); the fixed movement penalty P_1, applied at each step to encourage the AUV to minimize unnecessary movements, thereby reducing the total number of steps N in (D4) and the vertical ascent steps in (D5), which in turn decreases T_{u} and T_{AUV} in T; the boundary penalty P_2, imposed when the AUV's coordinates exceed the maximum horizontal distance or depth specified in (D8)–(D9); and the shadow zone penalty P_3, applied according to (D10).

The intermediate reward R_{mid} optimizes the path selection through dynamic feedback.

$$R_{\text{mid}}(i) = \sigma_1 \times \text{SNR}(\mathbf{r}_i) - \sigma_2 \times d_i - \sigma_3 \times l_i, \tag{D15}$$

where $\text{SNR}(\mathbf{r}_i)$ is a shorthand for $\text{SNR}(\mathbf{g}, \mathbf{r}_i, f)$, under the assumption of fixed $\mathbf{g}$ and f for a single transmission. The distance is given by $d_i = \|\mathbf{g} - \mathbf{r}_i\|_2$, and the relative distance increment by $l_i = d_i - d_{i-1}$. The parameters σ_1, σ_2, and σ_3 denote the weights associated with SNR, distance, and relative distance, respectively.

The SNR reward term encourages the AUV to move to regions with high SNR in each movement. According to equation (C10), higher SNR results in greater UWA channel capacity, thereby helping to satisfy the task constraint (D7) more efficiently, which reduces the number of movement steps N required for successful data reception in (D4) and consequently decreases T_{u}. We also incorporate a global view for the AUV, where smaller values of d_i and l_i facilitate the AUV's movement toward positions with higher $C_{\text{u}}(\mathbf{r}_j)$, considering not just a single step but the overall trajectory. Furthermore, by accounting for the distance variation between two consecutive steps relative to the transmitter l_i, oscillatory path behaviors are effectively suppressed.

The terminal reward R_{final} is to strengthen the overall quality of task completion. R_{final} for $\mathcal{P}1$ can be expressed as:

$$R_{\text{final}} = \rho_1 \times \sum_{i=0}^{N} \left(\text{SNR}\left(\mathbf{r}_i\right) - \text{SNR}\left(\mathbf{r}_0\right)\right) - \rho_2 \times T, \tag{D16}$$

where ρ_1 represents the weight of SNR improvement along the AUV path, and ρ_2 represents the weight of end-to-end cross-medium transmission duration T. R_{final} shows ultimate objective is to shorten the end-to-end cross-medium transmission duration T as in (D1) in $\mathcal{P}1$ by guiding the AUV toward high-SNR regions during its movement.

Algorithm D1 shows the details of the proposed APP-M3DQN.

Then after K_{epi} episodes of iterative learning as in algorithm D1, the algorithm finally outputs the end-to-end cross-medium transmission duration T and the movement trajectory of the AUV.

By applying double Q-learning to the AUV mobile relay path planning task through M3DQN, we can reduce the cumulative errors in DQN. The main network is used to select actions as in step 9, the target network is used to evaluate values in step 17, 19, 21 and 23, and the target network is updated regularly in step 25. The main network dynamically responds to perceived environmental dynamics (e.g., variations in sound field SNR, the AUV's own positional information), while the target network periodically updates its parameters to stabilize value estimation, thereby enhancing resilience to complex environmental disturbances. In long-distance underwater navigation scenarios, by leveraging the dual-network's suppression of cumulative errors in long-sequence decision-making, global deviations in path planning are reduced, thereby further enhancing the AUV's path adaptability and decision-making accuracy in complex underwater environments.

Algorithm D1 AUV path planning based on M3DQN for cross-medium mobile relay

Input: underwater transmitter position $\mathbf{g}$, SNR map $\mathrm{SNR}(\mathbf{g}, \mathbf{r}, f)$, channel capacity of UWA $C_{\mathrm{u}}(\mathbf{g}, \mathbf{r})$, AUV initial state s_0, BS position $\mathbf{r}_{\mathrm{BS}}$;

Output: The end-to-end cross-medium transmission duration T, AUV path $\boldsymbol{\Omega}$;

1: Initialize: the maximum number of episodes K_{epi}, maximum number of steps per episode V, the total amount of data (in bits) transmitted and relayed D_{total}, initial exploration ϵ_0, decaying rate β, set $\epsilon \leftarrow \epsilon_0$

2: Initialize: the fixed reward R_{fixed}; the weight in intermediate reward R_{mid} $(\sigma_1,\ \sigma_2,\ \sigma_3)$; the weight in overall terminal reward R_{final} $(\rho_1,\ \rho_2)$

3: Initialize: the replay memory queue D with capacity M; the dueling DQN network with coefficients θ, the target network with coefficients $\boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$

4: **for** each episode $e = 1, 2, \cdots, K_{\mathrm{epi}}$ **do**

5:   Initialize: a sliding window queue W with capacity K_1

6:   Initialize: the state $\mathbf{s}_0$ using the AUV initial state s_0 from line 1.

7:   Initialize: step counter $n \leftarrow 0$

8:   **repeat**

9:     Choose action $\mathbf{v}_n$ from $\hat{\mathcal{A}}$ based on the ϵ-greedy policy, i.e., $\mathbf{v}_n = \mathbf{v}^{(i^*)}$, where i is randomly selected from $1-4$ with probability ϵ, and takes the value $\arg\max_{i=1,\cdots,4} \hat{Q}(\mathbf{s}_n, \mathbf{v}^{(i)}; \boldsymbol{\theta})$ with probability $1-\epsilon$.

10:     Execute action $\mathbf{v}_n$ and observe the next state $\mathbf{s}_{n+1}$, based on the SNR map and the AUV position $\mathbf{r}_i = [x_i, h_i]$, calculate the intermediate reward R_{mid} under the next state $\mathbf{s}_{n+1}$, and set the reward $R_{\mathrm{n}} = R_{\mathrm{mid}}(s_{n+1}) - P_1$

11:     Store transition $(\mathbf{s}_n, \mathbf{v}_n, R_n, \mathbf{s}_{n+1})$ in the sliding window queue W

12:     **if** $n \geqslant K_1$ **then**

13:       Calculate the K_1-step accumulated return $R_{(n-K_1):n}$ using the stored transitions in W, and store the $(\mathbf{s}_{n-K_1}, \mathbf{v}_{n-K_1}, R_{(n-K_1):n}, \mathbf{s}_n)$ in the replay memory D

14:       Sample random minibatch of K_1-step transition $(\mathbf{s}_j, \mathbf{v}_j, R_{j:j+K_1}, \mathbf{s}_{j+K_1})$ from D

15:       **if** the next state $\mathbf{s}_{j+K_1}$ is terminal (e.g., data reception completed) **then**

16:         calculate T_{u}, T_{AUV},T_{a} and T as defined in equation (D4)-(D6) and (D1).

17:         $y_j = R_{j:j+K_1} + R_{\mathrm{des}} + R_{\mathrm{final}}$

18:       **else if** the position in $\mathbf{s}_{j+K_1}$ is out of boundary **then**

19:         $y_j = R_{j:j+K_1} - P_2$

20:       **else if** the position in $\mathbf{s}_{j+K_1}$ is in shadow zone $\mathcal{F}$ **then**

21:         $y_j = R_{j:j+K_1} - P_3$

22:       **else**

23:         $y_j = R_{j:j+K_1} + \gamma^{K_1} \hat{Q}(\mathbf{s}_{j+K_1}, \mathbf{v}^*; \boldsymbol{\theta}^-)$, where $\mathbf{v}^* = \arg\max_{i=1,\cdots,4} \hat{Q}(\mathbf{s}_{j+K_1}, \mathbf{v}^{(i)}; \boldsymbol{\theta})$

24:       **end if**

25:       Perform a gradient descent step on $\left(y_j - \hat{Q}(\mathbf{s}_j, \mathbf{v}_j; \boldsymbol{\theta})\right)^2$ with respect to network parameters $\boldsymbol{\theta}$

26:       Update $n \leftarrow n+1$, $\epsilon \leftarrow \epsilon\beta$

27:     **end if**

28:   **until** $\mathbf{s}_n$ is terminal (data reception completed) or $\mathbf{r}_i \in \mathcal{F}$ or $x_i \notin [0, X_{\max}]$ or $h_i \notin [0, H_{\max}]$ or $n = V$

29:   After every L episodes, set the target network parameters $\boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$

30: **end for**

Appendix E Simulation

Appendix E.1 Basic settings

In the simulation, the underwater environment size is a 200 m×10000 m profile space with a maximum underwater depth 200 m and a horizontal distance 10000 m, which is the same certain area in the South China Sea as in Fig. C1.

The UWA transmitter is located at (0, 200 m), and we choose the acoustic wave frequency of 10 kHz. The spatially varying TL distribution diagram is as dipicted as in Fig. C1. The model parameters are listed in table E1. The parameter setting of APP-M3DQN is listed in table E2 [34].

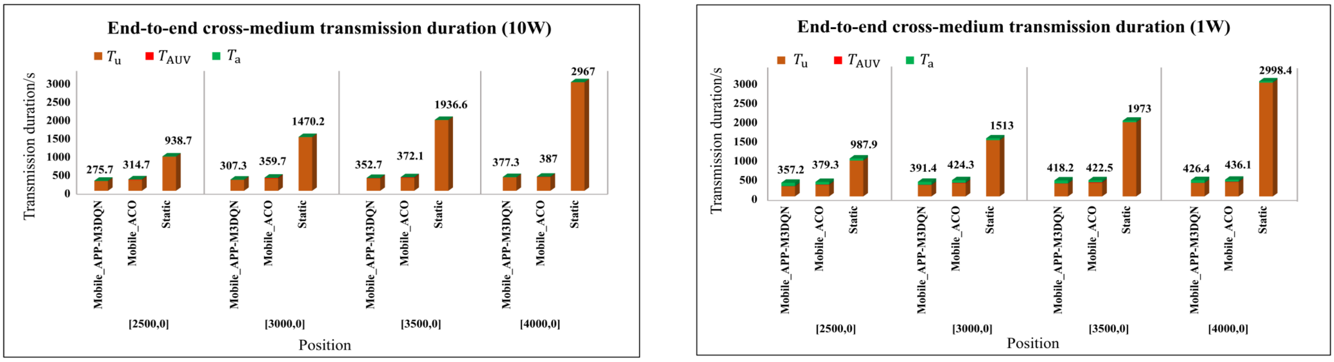
Table E1 Model parameter setting

Parameter	Value
Maximum horizontal distance $X_{\max}$	2500 m, 3000 m, 3500 m, 4000 m
Maximum water depth $H_{\max}$	200 m
BS height Z_{BS}	100 m
BS horizontal distance from the underwater transmitter X_{BS}	10000 m
Transmission data length D_{total}	7 Mbits, 8 Mbits, 9 Mbits, 10 Mbits
AUV speed v_{AUV}	10 m/s
Time interval Δt	1 s
UWA channel bandwidth B_{u}	6 kHz
Underwater transmitter transmission power P	20 W
Wind speed ω	5 m/s
Shipping activity factor s	1
Acoustic frequency f	10 kHz
Air background noise power N_{a}	10^{-10} W
AUV relay power q_{a}	1W, 10 W
Radio channel bandwidth B_{a}	5 MHz
Radio wave TL exponent ξ	1
Radio channel power gain χ	10^{-4}
AUV self-noise [35] N_{MR}	117 dB
Discretized grid spacing of the underwater area p	10 m

Table E2 Parameters specific to APP-M3DQN

Parameter	Value
Fixed reward for reaching destination R_{des}	600
Movement penalty P_1	1
Boundary crossing penalty P_2	10000
Moving shadow zone penalty P_3	10000
Intermediate reward SNR weight σ_1	10
Distance weight σ_2	0.005
Relative distance weight σ_3	5
Final reward SNR weight ρ_1	0.15
End-to-end cross-medium transmission duration weight ρ_2	0.7
Discount factor γ	0.995
Maximum steps V	500
Number of iterations K_{epi}	3000
Initial exploration probability ϵ_0	0.5
Minimum exploration probability $\epsilon_{\min}$	0.1
Exploration decaying rate β	0.998
Replay memory capacity M	100000
Steps for multi-step learning K_1	30
Update frequency for target network L	5

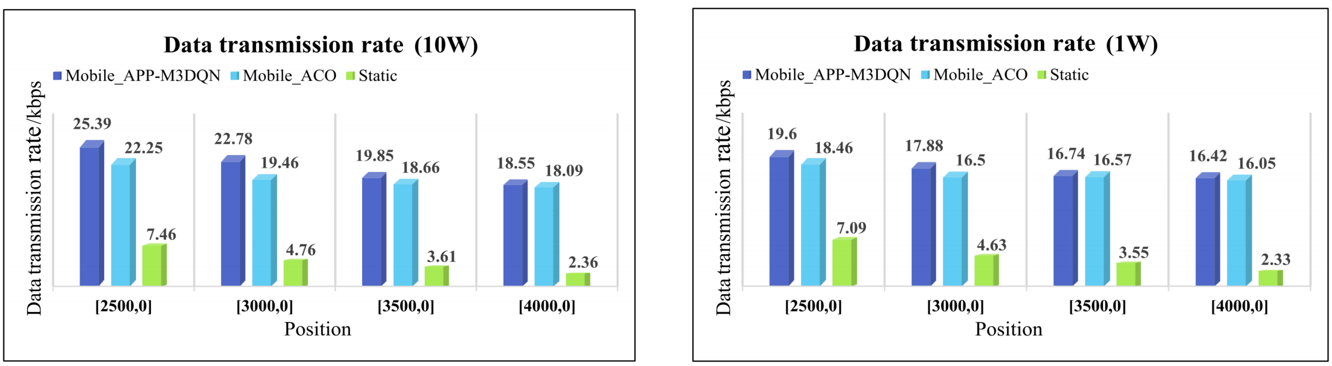
We compare the proposed APP-M3DQN with ACO [25], and static relay scheme in terms of end-to-end cross-medium transmission duration, end-to-end transmission rate, and radio channel utility. We also compare the paths designed by APP-M3DQN and ACO. The ACO algorithm used for comparison is adapted based on [25] to address the problem defined in $\mathcal{P}1$.



(a) AUV relay power $q_a = 10\,\mathrm{W}$

(b) AUV relay power $q_a = 1\,\mathrm{W}$

Figure E1 End-to-end cross-medium transmission duration with different relay initial positions. We compare three relay forwarding schemes: a static relay, a mobile relay using APP-M3DQN learning algorithm, and a mobile relay using the ACO algorithm. $D_{\text{total}} = 7$ Mbits.



(a) AUV relay power for radio transmission $q_a = 10\,\mathrm{W}$

(b) AUV relay power for radio transmission $q_a = 1\,\mathrm{W}$

Figure E2 End-to-end cross-medium transmission rate with different relay initial positions. We compare three relay forwarding schemes: a static relay, a mobile relay using the APP-M3DQN learning algorithm, and a mobile relay using the ACO algorithm. $D_{\text{total}} = 7$ Mbits.

Appendix E.2 Simulation results and analysis

The initial positions of the cross-medium relays are set at $[2500\,\mathrm{m}, 0\,\mathrm{m}]$, $[3000\,\mathrm{m}, 0\,\mathrm{m}]$, $[3500\,\mathrm{m}, 0\,\mathrm{m}]$, and $[4000\,\mathrm{m}, 0\,\mathrm{m}]$, evenly spaced along the axis. We set $D_{\text{total}} = 7$ Mbits. This configuration is likewise adopted in the analyses shown in Figs. E1- E3.

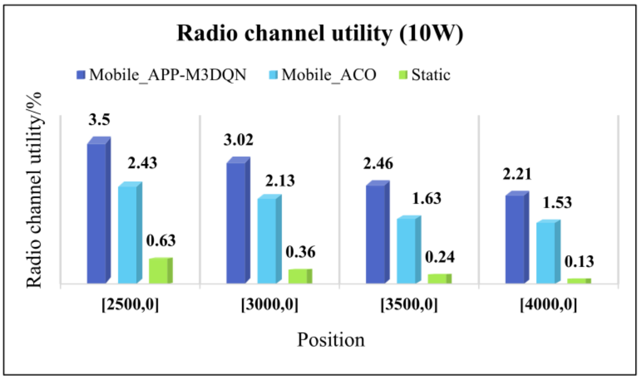
Fig. E1 illustrates the end-to-end cross-medium transmission duration T achieved with different relay initial positions for three relay schemes: a static relay, a mobile relay using the ACO algorithm, and a mobile relay using the APP-M3DQN learning algorithm. In both cases (a) and (b), corresponding to different AUV relay powers for radio transmission, T_{u} accounts for the vast majority of the total transmission duration, particularly in the static-relay scenario. This observation motivates our focus on optimizing the underwater transmission process in this work. The AUV returning time T_{AUV} is negligible in comparison, while the radio transmission time T_{a} is slightly higher in (b) due to the lower radio transmit power. The results in Fig. E1 clearly demonstrate the superiority of the proposed mobile relay scheme, as formulated in $\mathcal{P}1$, in reducing the end-to-end cross-medium transmission duration. The end-to-end cross-medium transmission duration of the static relay scheme is nearly three times longer with increasing distance at $[2500\,\mathrm{m}, 0\,\mathrm{m}]$. This is because static surface nodes may encounter unfavorable transmission conditions and cannot actively avoid weak-signal regions, in contrast to mobile relays. And the APP-M3DQN further shorten the duration by nearly 14%.

As the distance between initial positions increases, the transmission duration increases significantly for the static relay, whereas the mobile relay, as formulated in $\mathcal{P}1$, exhibits a much smaller increase. Moreover, the cross-medium mobile relay employing the APP-M3DQN learning algorithm achieves shorter transmission durations compared to the ACO-based scheme. This is because, through carefully designed rewards and interactions with the UWA environment in APP-M3DQN, the M3DQN-based path planning is less prone to being trapped in local optima.

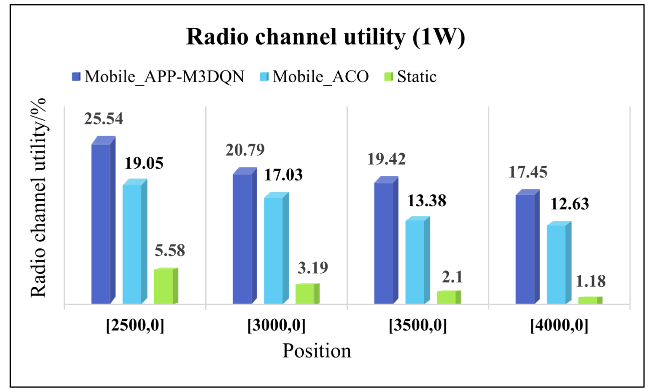
Fig. E2 illustrates the end-to-end data transmission rates R_{b} ($R_{\text{b}} = D_{\text{total}}/T$) of the three relay schemes under different initial positions, with (a) and (b) corresponding to varying AUV relay forwarding power levels. The data rate is a critical metric for transmission performance. Since the total transmitted data D_{total} is the same, the resulting transmission duration shows a trend similar to that in Fig. E1.

Fig. E3 illustrates the radio channel utility η of three cross-medium relay schemes at different initial positions, and with different AUV relay power at (a) and (b). Radio channel utility η refers to the proportion of the time that radio receiver to effectively receiving data during a transmission round [8], where $\eta = (T_a/T) \times 100\%$. It is a two-link imbalance indicator for cross-medium transmission.

The proposed mobile relay scheme, as formulated in $\mathcal{P}1$, significantly outperforms the static relay scheme at all four positions, demonstrating that the proposed mobile relay can enhance link balance through effective path planning. Additionally, the path planning based on the M3DQN algorithm outperforms the ACO algorithm in terms of radio channel utilization, since the reward design and

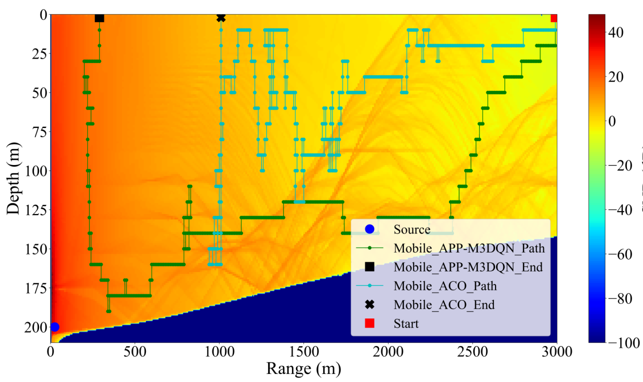


(a) AUV relay power for radio transmission $q_a = 10\,\mathrm{W}$

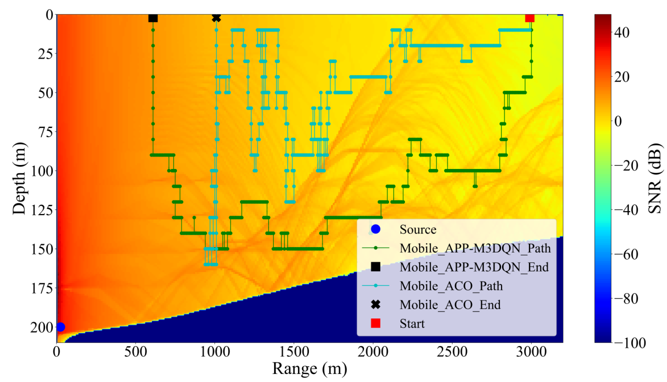


(b) AUV relay power for radio transmission $q_a = 1\,\mathrm{W}$

Figure E3 Radio channel utility (two-link imbalance indicator) with different relay initial positions. We compare three relay forwarding schemes: a static relay, a mobile relay using the APP-M3DQN learning algorithm, and a mobile relay using the ACO algorithm under a total data size of $D_{\text{total}} = 7$ Mbits, with (a) an AUV relay power of 10 W, and (b) an AUV relay power of of 1 W.



(a) Example 1 of AUV paths



(b) Example 2 of AUV paths

Figure E4 Examples of AUV paths planned using APP-M3DQN and ACO, with an initial position of $[3000\,\mathrm{m}, 0]$ and $D_{\text{total}} = 10$ Mbits, are presented.

learning mechanism in APP-M3DQN enable more efficient exploration of high-SNR regions and avoidance of redundant movements. By comparing (a) and (b), it can be observed that when the AUV relay power for radio transmission is reduced to 1 W, the radio channel capacity decreases, while the radio channel utility improves across all relay schemes. However, this improvement becomes marginal if the SNR is sufficiently low, as the radio transmission becomes unreliable. However, the proposed APP-M3DQN-based mobile relay can improve the utilization of the radio channel in both cases.

Fig. E4 illustrates examples of AUV paths planned using M3DQN and ACO with an initial position of [3000 m, 0] and $D_{\text{total}} = 10$ Mbits. The green curves represent the APP-M3DQN algorithm, while the blue curves represent the AUV path based on the ACO algorithm. From Fig. E4, it can be observed that the mobility of the AUV allows for finding better underwater data transmission positions. The APP-M3DQN learning algorithm, grounded in DRL, is more effective than the ACO algorithm in identifying high-SNR regions. This advantage arises from its reward shaping, value decomposition, and exploration strategy, which together guide the AUV toward efficient paths while avoiding redundant movements. In contrast, the ACO algorithm is more susceptible to convergence toward local optima in path planning, as it lacks adaptive learning from environmental feedback.

Figs. E5 and E6 illustrate, respectively, the end-to-end cross-medium transmission duration and data rate, comparing the mobile relay based on APP-M3DQN with the static relay under different data transmission requirements, $D_{\text{total}} \in \{7, 8, 9, 10\}$ Mbits, with initial positions at $[2500\,\mathrm{m}, 0]$, $[3000\,\mathrm{m}, 0]$, $[3500\,\mathrm{m}, 0]$, and $[4000\,\mathrm{m}, 0]$. As the cross-medium data transmission requirement increases, the end-to-end transmission duration increases, whereas the corresponding transmission rate decreases. However, the performance of the static relay scheme declines sharply as the distance and data transmission requirement increase, whereas the mobile relay scheme based on the APP-M3DQN learning algorithm demonstrates stable and efficient performance. This is because the proposed mobile relay scheme based on APP-M3DQN can rapidly adapt to variations in the UWA channel through reward shaping and iterative learning from environmental feedback, enabling the AUV to identify high-SNR regions and avoid unfavorable transmission conditions. As a result, even under the worst simulation condition—transmitting 10 Mbits of data from the initial position $[4000\,\mathrm{m}, 0]$—the end-to-end cross-medium transmission duration remains below 500 seconds.

Figs. E7 presents the learning convergence curve of the APP-M3DQN algorithm and the ACO algorithm, with an initial position of $[3500\,\mathrm{m}, 0]$ and $D_{\text{total}} = 7$ Mbits. Both algorithms reach a certain level of convergence, though both exhibit fluctuations. If we consider one RL episode and one ACO iteration as a single learning cycle, ACO typically requires fewer cycles to converge. However, each ACO iteration generally has higher computational complexity than an RL episode due to its population-based combinatorial search and

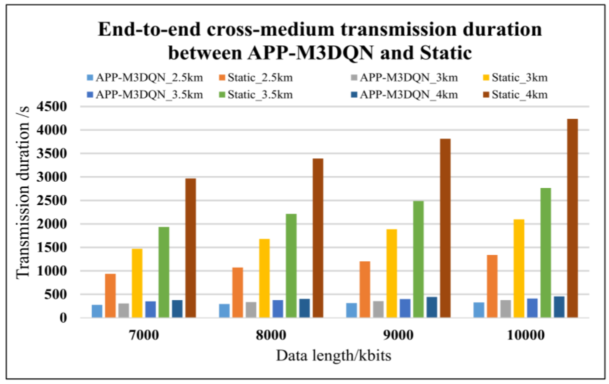


Figure E5 The end-to-end cross-medium transmission duration under different data transmission requirements. It illustrates the impact of varying data length on the performance of the mobile relay scheme based on the APP-M3DQN learning algorithm and the static relay scheme.

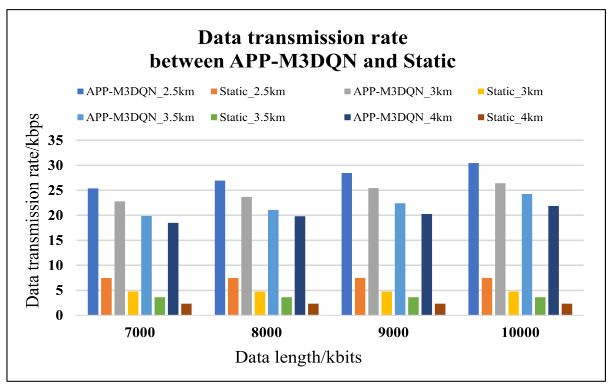
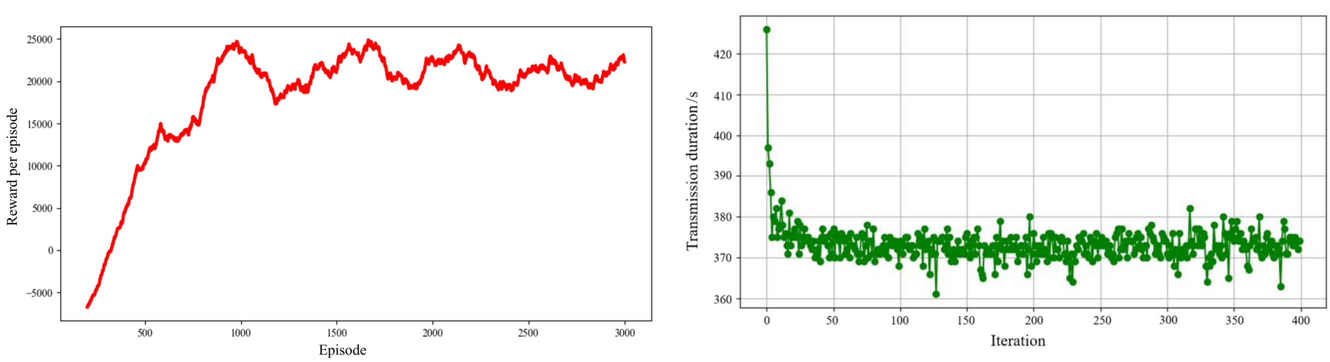


Figure E6 End-to-end cross-medium data rate under different data transmission requirements: the mobile relay scheme based on the APP-M3DQN learning algorithm shows a significant improvement over the static relay scheme.



(a) Reward convergence over episodes of the APP-M3DQN algorithm

(b) Transmission duration convergence curve over iterations of the ACO algorithm

Figure E7 Learning convergence curve of the APP-M3DQN algorithm and the ACO algorithm, with an initial position at $[3500\,\mathrm{m}, 0]$ and $D_{\mathrm{total}} = 7$ Mbits.

repeated cost evaluations.

Appendix F Future prospects

This work also raises several open challenges. The study assumes calm sea conditions with a known TL transmission map. However, in many practical scenarios, the TL transmission map is not available in advance. Modern AI techniques may provide a solution by enabling transmission map prediction, similar to radio map prediction. Moreover, in practical transmission scenarios, UWA channels often exhibit highly dynamic and time-varying characteristics due to complex sea conditions. Particularly when encountering strong internal waves or extreme sea states, the channel dynamics become even more pronounced, which may result in relative positional offsets between the AUV and the transmitter. Such offsets can lead the AUV to enter acoustic shadow zones, ultimately degrading the performance of cross-medium transmission. Addressing these challenges requires incorporating ocean physics and developing AUV relay path planning algorithms with stronger adaptability. Finally, exploring cooperative multi-AUV mobile relay mechanisms could further improve data transmission efficiency and optimize the use of network resources.

References

1 Li S, Liu J, Xu H X, et al. Research status of autonomous underwater vehicles in China. Sci China Inf Sci, 2018, 48: 1152–1164
2 Chen Y L, Hu X Y, Wang D Q, et al. Researches on underwater transmission characteristics of blue-green laser. In: Proceedings of Oceans, Taipei: IEEE, 2014. 1–5
3 Liu S Z, Han X, Ma L, et al. Research on channel state information feedback in underwater acoustic adaptive OFDM communication based on sequenced codebook. J Electron Inf Technol, 2024, 46: 2095–2103
4 Jouhari M, Ibrahimi K, Tembine H, et al. Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things. IEEE Access, 2019, 7: 96879–96899
5 Dong W, Yang Q L, Chen Y X, et al. Rhne-MAC: Random handshake MAC protocol based on nash equilibrium for underwater wireless sensor networks. IEEE Sens J, 2021, 21: 21090–21098
6 Guo J N, Song S S, Liu J, et al. Exploring applicable scenarios and boundary of MAC protocols: A MAC performance analysis framework for underwater acoustic networks. IEEE Trans Mob Comput, 2024, 23: 12717–12730
7 Liu Y, Chen F J, Fei J, et al. A new amphibious network infrastructure with hybrid acoustic and radio links. In: Proceedings of the 8th International Conference on Underwater Network Systems (WUWNet), New York: ACM, 2013. 1–2
8 Hao Z H, Li W, Zhang Q Y, et al. Cross-medium transmission with mobile AUV relay path planning: Enhancing underwater acoustic and radio mixed-link performance. IEEE Internet Things J, 2025, 12: 48706-48723
9 Nourizadeh Azar S, Erdemir O, Soomro M, et al. A hybrid acoustic-RF communication framework for networked control of autonomous underwater vehicles: Design and cosimulation. Turk J Electr Eng Comput Sci, 2022, 30: 1475–1491
10 Qu F Z, Qian J Y, Wang J, et al. Cross-medium communication combining acoustic wave and millimeter wave: Theoretical channel model and experiments. IEEE J Oceanic Eng, 2022, 47: 483–492
11 Okereke C E, Mohamad M M, Abdul Wahab N H, et al. An overview of machine learning techniques in local path planning for autonomous underwater vehicles. IEEE Access, 2023, 11: 24894–24907
12 Li D L, Wang P, Du L. Path planning technologies for autonomous underwater vehicles-a review. IEEE Access, 2019, 7: 9745–9768
13 Tovar B, Murrieta C R, LaValle S M. Distance-optimal navigation in an unknown environment without sensing distances. IEEE Trans Robot, 2007, 23: 506–518
14 Nan C, Liu Y G, Xia H, et al. Grid-based RRT for minimum dose walking path-planning in complex radioactive environments. Ann Nucl Energy, 2018, 115: 73–82
15 Khatib O. Real-time obstacle avoidance for manipulators and mobile robots. Int J Robot Res, 1986, 5: 90–98
16 Gautam U, Ramanathan M. Simulation for path planning of SLOCUM glider in near-bottom ocean currents using heuristic algorithms and Q-learning. Defe Sci J, 2015, 65: 220–225
17 Yordanova V, Griffiths H, Hailes S. Rendezvous planning for multiple autonomous underwater vehicles using a Markov decision process. IET Radar Sonar Navig, 2017, 11: 1762–1769
18 Porta Garcia M A, Montiel O, Castillo O, et al. Path planning for autonomous mobile robot navigation with ant colony optimization and fuzzy cost function evaluation. Appl Soft Comput, 2009, 9: 1102–1110
19 Tan J, Yang L G, Li X R, et al. Deep reinforcement learning applications and prospects in industrial scenarios. Chin J Eng, 2025, 47: 768–779
20 Lin Q B, Gou H X, Tian P D, et al. RL-based USV path planning under the marine multimodal features considerations. IEEE Internet Things J, 2025, 12: 15274–15287
21 Chu Z D, Wang F L, Lei T J, et al. Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance. IEEE Trans Intell Veh, 2023, 8: 108–120
22 Zhang K Y, Li Y B, Zhang Z, et al. DDQN-based hybrid routing protocol for UWSNS with void repair mechanism. IEEE Sens J, 2025, 25: 20718–20731
23 Wang Z Y, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning (ICML'16), New York: JMLR.org, 2016. 1995–2003
24 Chandra R, Goyal S, Gupta S. Evaluation of deep learning models for multi-step ahead time series prediction. IEEE Access, 2021, 9: 83105–83123
25 Dahan F, Hindi K F, Ghoneim A, et al. An enhanced ant colony optimization based algorithm to solve QoS-aware web service composition. IEEE Access, 2021, 9: 34098–34111
26 Dai M H, Wu Y H, Qian L P, et al. UAV-assisted multi-access computation offloading via hybrid NOMA and FDMA in marine networks. IEEE Trans Netw Sci Eng, 2023, 10: 113–127
27 Meshcheryakov R, Dushkin A. Applicability of Shannon's theorem to the processing of weakly structured speech information in acoustic noise conditions. In: Proceedings of 8th International Conference on Information, Control, and Communication Technologies (ICCT), Shenzhen: IEEE, 2024. 1–4
28 Wang Y, Fang Z Y, Li J L. Environment-based performance optimization for passive sonar network. In: Proceedings of Oceans, Singapore: IEEE, 2024. 1–6
29 Stojanovic M. On the relationship between capacity and distance in an underwater acoustic communication channel. SIGMOBILE Mob Comput Commun Rev, 2007, 11: 34–43
30 Liu M A, Lian L M. Underwater acoustical engineering. Hangzhou: Zhejiang Science and Technology Press, 2002. 72-73
31 Morozs N, Gorma W, Henson B T, et al. Channel modeling for underwater acoustic network simulation. IEEE Access, 2020, 8: 136151–136175
32 Yang T C. Properties of underwater acoustic communication channels in shallow water. J Acoust Soc Am, 2012, 131: 129-145
33 Li Z, Chitre M and Stojanovic M, Underwater acoustic communications. Nat Rev Electr Eng, 2024, 2: 83-95
34 Zeng Y, Xu X L, Jin S, et al. Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning. IEEE Trans Wireless Commun, 2021, 20: 4205–4220
35 Deng L L, Shi H Q, Shi S G. Numerical and experimental research on the distribution characteristics of self-noise on the flank of an AUV. J Harbin Eng Univ, 2025, 46: 1669–1677