International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

Deep Reinforcement Learning-Based AUV Path Planning for Energy-Efficient Data Collection in Underwater Wireless Sensor Networks

1Mr. V. Balasubramaniyam, 2Dr. P. Srimanchari
1Research Scholar, 2Assistant Professor,
1,2Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu
1subramaniyamjothi@gmail.com

Abstract -Underwater Wireless Sensor Networks face unique challenges including severe acoustic channel impairments, high propagation delays and node energy constraints. Traditional Autonomous Underwater Vehicle data collection schemes rely on static waypoint traversal, which fails to adapt to dynamic ocean environments and varying channel conditions. In this paper, we propose a deep reinforcement learning based AUV path planning framework that leverages Proximal Policy Optimization and Soft Actor-Critic algorithms to learn energy-optimal, channel-aware collection trajectories. The AUV agent observes real-time acoustic Signal-to-Noise Ratio, residual energy of sensor nodes, and ocean current vectors to make intelligent navigation decisions. Simulation results demonstrate that the proposed HPSA-AUV achieves a 94.2% data collection rate while reducing energy consumption by 25.5% and improving average acoustic SNR by 29.9% compared to greedy waypoint baselines. These results validate the efficacy of DRL for adaptive AUV mission planning in complex underwater environments.

Keywords: *Underwater Wireless Sensor Network, Autonomous Underwater Vehicle, Deep Reinforcement Learning, Proximal Policy Optimization, Path Planning, Acoustic Channel, Hybrid Proximal-Soft Actor*

I.INTRODUCTION

Underwater Wireless Sensor Networks have emerged as a critical infrastructure for oceanographic monitoring, offshore oil pipeline inspection, marine biodiversity assessment, and underwater tactical surveillance [1]. Unlike terrestrial sensor networks that leverage radio frequency communication, UWSNs predominantly rely on acoustic signals for data transmission owing to the rapid attenuation of RF and optical waves in seawater [2]. Acoustic channels, however, introduce severe impairments including frequency-dependent path loss, multipath fading, Doppler spread from node mobility and ocean currents, and propagation delays on the order of milliseconds per meter [3].

To overcome the limitations of static network topologies and extend the operational lifetime of sensor nodes, Autonomous Underwater Vehicles have been proposed as mobile data mules [4]. An AUV navigates through the deployment area, collecting data from sensor nodes via short-range acoustic links before returning to a surface buoy or dock for data offloading. The efficiency of this paradigm critically depends on the AUV's trajectory: a suboptimal path incurs unnecessary energy expenditure, misses data from energy-depleted nodes, and fails to exploit favourable channel windows.

Existing AUV path planning approaches broadly fall into three categories: (i) greedy heuristic methods that visit the nearest unvisited node [5], (ii) optimization-based methods that solve variants of the Travelling Salesman Problem offline [6], and (iii) reinforcement learning based methods that learn adaptive policies online [7]. The first two categories are inherently static and cannot react to real-time changes in channel quality or node energy. RL-based methods offer the promise of adaptivity, yet most prior work considers simplified channel models or ignores energy heterogeneity among sensor nodes.

The main contributions of this paper are:

- A Markov Decision Process formulation of the AUV data collection problem that jointly models acoustic channel quality, node residual energy, and ocean current disturbances in the state space.
- A deep RL agent trained with Proximal Policy Optimization and Soft Actor-Critic that learns

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

energy- optimal, channel-aware trajectories entirely from simulation experience, with no pre-defined waypoints.

- A comprehensive comparative evaluation against greedy waypoint and random waypoint baselines using NS3-Aqua-Sim with the Thorp acoustic channel model, demonstrating significant gains in data collection rate, energy efficiency, and link quality.
- An ablation study isolating the contribution of individual reward components, providing design guidance for practitioners implementing DRL-based AUV systems.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the system model. Section 4 describes the MDP formulation and DRL framework. Section 5 details the simulation setup. Section 6 presents and analyses results. Section 7 concludes the paper.

II.RELATED WORK

2.1 AUV-Assisted Data Collection

Early work on AUV-assisted UWSN data collection focused on static planning. Luo et al. [5] proposed a greedy nearest-neighbor heuristic for AUV routing, demonstrating improved collection efficiency over fixed-topology networks but suffering from poor worst-case path lengths. Shi et al. [6] formulated AUV trajectory optimization as a variant of the capacitated TSP and solved it using genetic algorithms, achieving near-optimal static paths. However, both approaches pre-compute trajectories offline without accounting for dynamic channel conditions or in-field node energy depletion.

2.2 Reinforcement Learning for AUV Navigation

The application of RL to AUV navigation was pioneered by Carlucho et al. [7], who trained a Q-learning agent for target tracking in simulated water currents. Subsequent work by Zeng et al. [8] applied Deep Q-Networks to multi-AUV cooperative search, demonstrating scalability to larger deployment areas. More recently, actor-critic methods have gained traction: Xu et al. [9] employed Advantage Actor-Critic for AUV obstacle avoidance, and Li et al. [10] used PPO for energy-constrained patrolling. However, none of these works explicitly model acoustic channel quality as a state variable influencing the reward structure, which is the key novelty of our approach.

2.3 Acoustic Channel Modelling in UWSNs

Realistic acoustic channel modelling is essential for credible UWSN simulation. The Thorp model [11] characterizes frequency-dependent absorption loss and is widely adopted for its accuracy across the 1–100 kHz range used in practical deployments. Brekhovskikh and Lysanov [12] provide a comprehensive treatment of multipath propagation due to surface and bottom reflections. Our simulation integrates both effects via the NS3-Aqua-Sim channel stack, providing a more realistic evaluation environment than prior DRL works that use simplified distance-based models.

III.SYSTEM MODEL

3.1 Network Architecture

The three-dimensional underwater sensor network comprising N stationary sensor nodes deployed at varying depths within a bounded ocean volume V = L × W × D meters. Each node i is equipped with an acoustic modem, energy harvesting circuitry, and a rechargeable battery of capacity E_max. Nodes sense environmental parameters (temperature, salinity, pressure) and buffer readings until an AUV comes within communication range r_c. A single AUV departs from a surface dock, collects data from as many nodes as possible within a time budget T_max, and returns to dock. The AUV operates at a nominal cruising speed v_AUV and is subject to time-varying 3D ocean current vectors u(x, y, z, t) obtained from a HYCOM ocean circulation model replay.

3.2 Acoustic Channel Model

The Thorp absorption model for acoustic path loss. The transmission loss TL(d, f) over distance d at carrier frequency f (kHz) is given in Eq.1

$$TL(d,\ f) = k.\ 10\ log10(d) + d.\ \alpha(f)\ 1000\ [db] \quad (1)$$

where k is the spreading factor (k = 1.5 for practical spreading), and α(f) is the Thorp absorption coefficient in dB/km given in Eq 2

$$\alpha(f) = \frac{0.11f2}{1+f2+\frac{44f2}{4100+f2}} + 2.75 \times 10{-}4f2 + 0.003[\text{dB/k}] \quad (2)$$

The received SNR at the AUV from node i at distance d_i is $SNRi = Ptx - TL(di, f) - N(f)$ where P_tx is the transmit power and N(f) is the ambient noise power spectral density derived from the WOSS noise

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

model incorporating shipping, wind, and thermal noise components.

3.3 Energy Model

Each sensor node i has residual energy e_i(t) at time t. A data transmission event of packet size L_pkt bits at data rate R_b consumes transmission energy $Etx = Ptx \times LpktRb$ per packet. Nodes that deplete their energy (e_i

≤ 0) become inactive for the remainder of the mission. The AUV's propulsion energy is modeled as $Eprop = cprop \cdot vAUV2 \cdot dseg$ per path segment of length d_seg, where c_prop is the hydrodynamic drag coefficient.

IV. MDP FORMULATION AND DRL FRAMEWORK

4.1 State Space

At each decision step t, the AUV observes a state vector s_t ∈ S comprising: (i) normalized AUV position [x_AUV, y_AUV, z_AUV] / [L, W, D]; (ii) relative position vectors to all N sensor nodes; (iii) residual energy

ratio e_i(t) / E_max for each node; (iv) estimated SNR_i for each node based on current distance and Thorp model; (v) elapsed mission time t / T_max; and (vi) local ocean current velocity vector u(t). The full state dimensionality is |S| = 3 + 3N + N + N + 1 + 3 = 5N + 7.

4.2 Action Space

The AUV's action a_t ∈ A is a continuous 3D velocity vector [v_x, v_y, v_z] bounded within the AUV's maneuvering envelope: |v_x|, |v_y| ≤ v_max and |v_z| ≤ v_z_max. This continuous action space is a key advantage over discrete waypoint-based formulations, enabling fine-grained channel-aware maneuvering. For the DQN baseline, we discretize into 26 directional primitives for comparability.

4.3 Reward Function

The reward function r_t is designed to simultaneously incentivize data collection, penalize energy waste, and reward high-quality acoustic links:

$$rt = w1 \sum \Delta Di(t) + w2 \cdot SNRi(t) \cdot \mathbf{1}[connected] - w3 \cdot Eprop(t) - w4 \cdot \mathbf{1}[timeout] \quad (3)$$

where ΔD_i(t) is the volume of data successfully received from node i at step t, 1[connected] is an indicator that the AUV is within communication range of at least one node, E_prop(t) is propulsion energy consumed at step t, 1[timeout] is a terminal penalty for mission timeout, and w_1 through w_4 are tuning weights set empirically as {1.0, 0.3, 0.5, 2.0} after grid search.

4.4 PPO Agent Architecture

We train the primary agent *using* Proximal Policy Optimization (PPO) with a clipped surrogate objective. The policy network π_θ and value network V_φ share a 4-layer fully connected trunk with hidden dimensions [256, 256, 128, 64] and ReLU activations. The actor head outputs a Gaussian policy with state-dependent mean and log-standard deviation. PPO's clipping parameter ε = 0.2, entropy coefficient c_2 = 0.01, and discount factor γ = 0.99. Training runs for 2 × 10^6 environment steps with Adam optimizer at learning rate 3 × 10^{-4}. A SAC agent with identical architecture is trained as a secondary benchmark.

V. SIMULATION SETUP

5.1 Simulation Environment

Simulations are conducted using NS3-Aqua-Sim [13], an underwater networking simulation platform built on NS-3. We deploy N = 30 sensor nodes uniformly at random within a 1000 × 1000 × 200 m volume at depths between 20 and 180 m. The AUV cruises at v_AUV = 2.0 m/s with maximum depth rate v_z_max = 0.5 m/s. Acoustic modems operate at f = 25 kHz with P_tx = 50 W. Mission time budget T_max = 3600 s. Each training episode randomizes node positions and residual energies uniformly. HYCOM current data from the Gulf of Mexico (January 2023) is replayed at 1:10 time scale to introduce realistic disturbances.

5.2 Baseline Methods

- Greedy Waypoint (Greedy WP): AUV always moves toward the nearest unvisited node with sufficient residual energy.
- Random Waypoint (Random WP): AUV selects the next waypoint uniformly at random from unvisited nodes.
- HPSA-AUV: Proposed PPO-based DRL agent (continuous action space).
- SAC-AUV: Soft Actor-Critic agent trained with identical architecture and reward.

5.3 Performance Metrics

- Data Collection Rate (%): Fraction of total buffered data successfully received by AUV. •

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

Energy Consumed (kJ): Total AUV propulsion energy over mission.

- Average Acoustic SNR (dB): Mean SNR across all AUV-node communication events.
- Mission Completion Time (min): Elapsed time until AUV returns to dock.
- Packet Delivery Ratio (%): Ratio of packets received to packets transmitted by nodes.
- AUV Path Length (m): Total 3D distance traversed by the AUV.

VI.RESULTS AND DISCUSSION

6.1 Comparative Performance

Table 1 summarizes the comparative performance of all four methods averaged across 100 test episodes with distinct random seeds. HPSA-AUV achieves the highest data collection rate of 94.2%, outperforming SAC-AUV (91.7%), Greedy WP (78.3%), and Random WP (61.0%). The performance gap is particularly pronounced in energy consumption: HPSA-AUV consumes only 18.4 kJ against 24.7 kJ for Greedy WP, representing a 25.5% reduction. This improvement arises because the PPO agent learns to exploit favourable current directions, effectively surfing ocean currents rather than fighting them.

Table 1- Comparative performance of proposed HPSA-AUV algorithm

Metric	HPSA-AUV	PPO-AUV	SAC-AUV	Greedy WP	Random WP
Data collection rate (%)	96.8	94.2	91.7	78.3	61.0
Energy consumed (kJ)	16.9	18.4	19.1	24.7	31.2
Avg. acoustic SNR (dB)	22.7	21.3	20.8	16.4	13.9
Mission completion time (min)	41.2	43.6	45.2	52.1	67.4
Packet delivery ratio (%)	98.3	96.8	95.4	84.2	73.6
AUV path length (m)	1694	1823	1891	2340	2987

6.2 Acoustic Channel Quality

A notable finding is the significant improvement in average acoustic SNR achieved by DRL agents. HPSA-AUV attains 21.3 dB versus 16.4 dB for Greedy WP — a 4.9 dB gain that directly translates to higher data throughput and lower retransmission rates. Qualitative trajectory analysis reveals that the PPO agent learns to approach nodes from angles that minimize multipath interference from the seafloor, a behavior not explicitly encoded in the reward function but emergent from maximizing the SNR-weighted data term.

6.3 Ablation Study

To isolate the contribution of individual reward components, we train PPO variants with each reward term removed. Removing the SNR bonus ($w_2 = 0$) reduces average SNR by 3.8 dB while collection rate drops to 89.1%, confirming that channel-aware navigation is essential. Removing the energy penalty ($w_3 = 0$) yields an agent that collects 95.1% of data but at 22.3 kJ — a 21.2% energy increase, demonstrating the importance of the energy-efficiency incentive. These results validate each component's contribution to the overall objective.

6.4 Convergence Analysis

HPSA-AUV converges to a stable policy within approximately 1.2×10^6 training steps, while SAC-AUV requires 1.6×10^6 steps for comparable stability. Both DRL agents significantly outperform heuristics from the earliest stages of training (after ~2×10^5 steps), suggesting that even a partially trained DRL policy provides gains over static baselines. Training wall-clock time is approximately 4.2 hours for PPO and 5.8 hours for SAC on an NVIDIA RTX 3090 GPU.

VII.CONCLUSION

This paper presents a deep reinforcement learning framework for AUV path planning in Underwater Wireless Sensor Networks. The key advantage is formulating the problem as a Markov Decision Process with a rich state space capturing acoustic channel quality, node energy heterogeneity, and ocean current disturbances, and training agents with Proximal Policy Optimization and Soft Actor-Critic algorithms in a realistic simulation environment, which demonstrates substantial improvements over conventional heuristic baselines. The proposed HPSA-AUV achieves 94.2% data collection rate with 25.5% lower energy

consumption and 4.9 dB higher acoustic SNR compared to the greedy waypoint baseline. Future work research can focus on multi-AUV cooperative collection, transfer learning from simulation to real AUV hardware, and integration of real-time bathymetric maps to improve depth planning. Extending the framework to handle node failures and adversarial jamming in tactical underwater scenarios is also planned.

REFERENCES

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," Ad Hoc Networks, vol. 3, no. 3, pp. 257–279, 2005.

[2] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 11, no. 4, pp. 23–33, 2007.

[3] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," ACM SIGMOBILE MCR, vol. 11, pp. 34–43, 2007.

[4] G. Han, J. Jiang, N. Bao, L. Wan, and M. Guizani, "Routing protocols for underwater wireless sensor networks," IEEE Communications Magazine, vol. 53, no. 11, pp. 72–78, 2015.

[5] J. Luo, Y. Chen, M. Wu, and Y. Yang, "A survey of routing protocols for underwater wireless sensor networks," IEEE Communications Surveys & Tutorials, vol. 23, no. 1, pp. 137–160, 2021.

[6] Y. Shi, G. Lv, and Y. Ye, "AUV-aided data collection in underwater sensor networks using genetic algorithm," in Proc. IEEE OCEANS, 2019, pp. 1–6.

[7] I. Carlucho, M. De Paula, S. Wang, Y. Petillot, and G. Acosta, "Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning," Robotics and Autonomous Systems, vol. 107, pp. 71–86, 2018.

[8] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for UAV-enabled mobile relaying systems," IEEE Transactions on Communications, vol. 64, no. 12, pp. 4983–4996, 2016.

[9] X. Xu, H. Ge, X. Yan, and C. Li, "Adaptive AUV obstacle avoidance using deep reinforcement learning," in Proc. IEEE ICCA, 2020, pp. 213–218.

[10] R. Li, X. Li, and J. Wang, "Energy-constrained AUV patrolling via proximal policy optimization," Ocean Engineering, vol. 248, p. 110723, 2022.

[11] W. Thorp, "Analytic description of the low-frequency attenuation coefficient," Journal of the Acoustical Society of America, vol. 42, pp. 270–270, 1967.

[12] L. M. Brekhovskikh and Y. P. Lysanov, Fundamentals of Ocean Acoustics, 3rd ed. Springer, 2003.

[13] P. Xie et al., "Aqua-Sim: An NS-2 based simulator for underwater sensor networks," in Proc. IEEE OCEANS, 2009, pp. 1–7. [14] C. Petrioli, R. Petroccia, J. R. Potter, and D. Spaccini, "The SUNSET framework for simulation, emulation and at-sea testing of underwater wireless sensor networks," Ad Hoc Networks, vol. 34, pp. 224–238, 2015.

[15] N. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater sensor networks: Applications, advances and challenges," Philosophical Transactions of the Royal Society A, vol. 370, no. 1958, pp. 158–175, 2012.

[16] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in Proc. International Conference on Machine Learning (ICML), 2018, pp. 1861–1870.

[17] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[18] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," IEEE Communications Surveys & Tutorials, vol. 13, no. 3, pp. 487–502, 2011.

[19] Z. Zhou, Z. Peng, J. H. Cui, Z. Shi, and A. Bagtzoglou, "Scalable localization with mobility prediction for underwater sensor networks," IEEE Transactions on Mobile Computing, vol. 10, no. 3, pp. 335–348, 2011.

[20] F. Campagnaro, R. Francescon, and M. Zorzi, "Multimodal underwater networks: Recent advances and a look at the integration of optical and acoustic communications," in Proc. IEEE WUWNet, 2016, pp. 1–8.