Learning-Aided Joint Beam Divergence Angle and Power Optimization for Seamless and Energy-Efficient Underwater Optical Communication

Huicheol Shin, *Student Member, IEEE*, Soo Mee Kim, *Member, IEEE*, and Yujae Song, *Member, IEEE*

***Abstract*—Integrating underwater optical wireless communication (UOWC) with marine applications, such as underwater sensors, buoys, and marine surface vehicles (MSVs), requires the aligning and maintaining of the optical beam between the transmitter and receiver for point-to-point (P2P) UOWC during the data transmission. An additional issue is the difficulty in exchanging batteries for marine applications because of the relatively high costs and risks compared with battery exchanging in terrestrial applications. This study seeks to resolve these issues via joint optimization of the beam divergence angle and transmission power level in an underwater sensor (i.e., transmit node) to maintain a seamless connection with an MSV (i.e., receive node) while minimizing the battery consumption of the sensor. In this regard, we adopt a hybrid underwater acoustic-optical communication system, where acoustic and optical communications are used for low-rate control data transmission and high-rate sensing data transmission, respectively. Under this framework, we propose a two-phase deep reinforcement learning (TPDRL) algorithm considering two agents (inner and outer) that determine different actions using an underwater sensor. Specifically, the primary role of the outer agent is to choose a transmission power level based on the long-term signal-to-noise ratio (SNR) between the underwater sensor and MSV. Next, the inner agent finds the beam divergence angle for the given transmission power (selected from the outer agent) based on the short-term instantaneous SNR. Simulation results demonstrate that the proposed TPDRL algorithm enables seamless and energy-efficient P2P UOWC, performing better than the algorithm with only the inner agent and other existing algorithms.**

***Index Terms*—Beam divergence angle, deep reinforcement learning (RL), energy efficiency, transmission power, underwater optical communication.**

I. Introduction

IN THE last few decades, wireless communication technologies have received consistent attention and have been further developed for extreme conditions, such as polar, space, and underwater environments [1], [2]. There are numerous opportunities for research in the underwater environments because they are still unexplored owing to inaccessibility and other constraints. In particular, underwater wireless sensor networks (UWSNs) for underwater environments have been considered as a solution for realizing the maritime Internet of Things (IoT), which enables the collection of various marine data with high-speed data transmission. Through the development of UWSNs, various marine applications, such as ocean geology data collection, ocean pollution monitoring, ocean exploration, disaster prediction, military surveillance, and reconnaissance, can be achieved.

To realize UWSNs, a variety of researches have been conducted. For example, in [3], the anti-jamming relay scheme for UWSN based on reinforcement learning (RL) was investigated, which enables an underwater relay to not only determine whether to leave the heavily jammed position but also choose the relay power level based on various state information, such as the previous bit error rate (BER), the previous reply power, and the jamming power measured by the relay.

When implementing UWSNs, communication systems based on acoustics and optics in underwater environments have been considered [4], [5], [6]. In the case of an underwater acoustic wireless communication (UAWC) system, many disadvantages, such as low-data rate, high delay owing to low-propagation speeds, and high attenuation from the aquatic medium, remain unaddressed. Moreover, UAWC systems comprising sound navigation and ranging (SONAR) devices are typically bulky, costly, and consume a large amount of energy for transmission [7]. Additionally, UAWC systems can adversely affect aquatic mammals and fish [8]. Compared with acoustic communication, underwater optical wireless communication (UOWC) has the advantages of higher data rate, lower latency, enhanced security, and lower implementation cost [2]. Owing to these advantages, UOWC can be regarded as a low-cost and energy-efficient solution for the realization of UWSNs for real-time marine applications and oceanic big data services.

Manuscript received 14 December 2022; revised 23 May 2023; accepted 8 August 2023. Date of publication 14 August 2023; date of current version 7 December 2023. This work was supported in part by the "Development of Polar Region Communication Technology and Equipment for Internet of Extreme Things (IoET)," funded by Ministry of Science and ICT (MSIT), South Korea, and in part by the Korea Institute of Ocean Science and Technology under Grant PEA0132. *(Corresponding author: Yujae Song.)*

Huicheol Shin and Soo Mee Kim are with the Maritime ICT Research and Development Center, Korea Institute of Ocean Science and Technology, Busan 49111, South Korea, and also with the Marine Technology and Convergence Engineering, University of Science and Technology, Busan 49111, South Korea (e-mail: shc0305@kiost.ac.kr; smeekim@kiost.ac.kr).

Yujae Song is with the Department of Robotics Engineering, Yeungnam University, Gyeongsan 38541, South Korea (e-mail: yjsong@yu.ac.kr).

Digital Object Identifier 10.1109/JIOT.2023.3304655

A. Related Works

However, for the implementation of UWSNs, the following problems should be addressed in UOWC systems: limited energy support, impaired underwater channels compared with terrestrial environments, and misalignment of optical transceivers [7]. To address these problems, numerous studies have been conducted in [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], and [17]. Zeng et al. [7] compared several digital modulation techniques in terms of power consumption and bandwidth required to achieve the same BER. In [9], an energy-efficient and reliable routing protocol was proposed, wherein a source node transmits a packet with local flooding, and an adaptive routing mechanism is applied to find the optimal route with minimum energy consumption. In addition to [9], various energy-efficient routing protocols for UWSNs have been developed to satisfy their respective requirements in [10], [11], [12], [13], [14], [15], and [16]. Furthermore, a comprehensive review of the literature on energy-efficient UWSNs was provided, including energy-saving techniques for the entire protocol stack and hardware in terms of signal processing, circuitry, and batteries, in [17] and references therein.

Optical signals undergo severe absorption, scattering, and turbulence in an underwater environment. Furthermore, multipath fading also occurs because of the interactions of water molecules and particulates in the ocean with photons. To minimize the transmission attenuation, the spectra of light corresponding to blue and green were used for the UOWCs because the light in seawater shows relatively low attenuation in the wavelength range of 450–550 nm, which denotes the blue and green spectra, as demonstrated in [18]. However, the aquatic medium changes rapidly, making it difficult to estimate underwater channel information and provide seamless connections. To address this problem for UOWCs, that is, signal attenuation in underwater channels, there has recently been a growing interest in developing deep learning techniques for UOWCs to compensate for channel attenuation and ensure performance in terms of BER, outage probability, etc [19], [20], [21], [22]. In [19], a novel Gaussian kernel-aided deep neural network (DNN) equalizer was proposed to compensate for the attenuation of underwater optical communication channels. The applied Gaussian kernel can reduce the training time compared with a conventional DNN and improve the BER performance. Chi et al. [19] verified their work by implementing an experimental test bed of optical transceivers. In [20], a progressive growth meta-learning (PGML)-based automatic modulation recognition (AMR) framework was developed, which significantly reduced the learning time and improved the probability of accurate classification. Jiang et al. [21] investigated a signal-detection scheme based on deep learning by exploiting the physical mechanism of a single-photon avalanche diode (SPAD) and prior knowledge of signal processing on the receiver side. A multilayer perception (MLP)-based deep learning network, which consisted of channel compensation layer and a layer that works as a demodulator, was adopted, and the proposed scheme showed better BER performance compared with two conventional logarithmic likelihood ratio (LLR)-based soft-decision methods. In [22], a new DNN for joint channel classification, channel estimation, and signal detection was designed for different marine environments, such as harbor, coastal, clear, and mixed waters. Simulation results verified that the proposed scheme considerably outperformed the existing least-square and linear minimum mean square error schemes in terms of BER without any prior information of the UOWC channel.

In practical UOWCs, the misalignment between optical transceivers can significantly degrade the system performance. Prior works [23], [24], [25], [26] have studied the effect of misalignment to address this problem. In [23], the effect of misalignment between the laser transmitter and receiver was investigated according to the change in link distance and transmit power. For a given BER threshold, the required transmit power of the laser transmitter is presented as a closed-form expression of the link distance and misalignment distance. In [24], the misalignment caused by the underwater absorption and scattering processes with a change in the divergence and elevation angle of the optical transmitter was considered, and a spatial distribution of light intensity was derived. Both [23], [24] analyzed the impact of the misalignment problem in UOWC systems. Vali et al. [25] demonstrated a tradeoff among the maximum acceptable lateral offset distance, power loss, and channel bandwidth using Monte Carlo simulations. In addition, the optimal divergence angle of the transmitted laser beam was obtained while maximizing the acceptable lateral offset distance for clear water and harbor water. In [26], a convolutional neural network (CNN)-aided optical multiple-input and multiple-output (MIMO) receiver architecture was developed to compensate for misalignment. In the proposed receiver, the conventional signal combiner and demodulator of the receiver were replaced with two concatenated CNNs to learn the underwater channel characteristics and detect the transmitted data without channel state information (CSI). Through extensive simulations and experiments, it was shown that the proposed receiver is resistant to misalignment and can achieve a BER performance close to that of maximum ratio combining (MRC) with perfect CSI.

B. Motivation and Contribution

Although a change in the beam divergence angle of the transmit node (e.g., underwater sensor) affects the received power strength of the receive node [e.g., marine surface vehicle (MSV)] in the point-to-point (P2P) UOWC scenario, to the best of our knowledge, there is no existing literature that *jointly* optimizes the beam divergence angle and transmission power to maintain a seamless wireless optical link while minimizing the battery consumption of the transmit node located in the water. In addition, most of the existing literature lack a detailed study on the online adaptive optimization of the beam divergence angle according to the *random shaking and movement of MSV* floating above sea level owing to various external factors (e.g., wind, waves, and the presence of large ships).

In this work, we consider a joint optimization of the beam divergence angle and transmission power for a P2P UOWC to maintain a seamless connection between an underwater

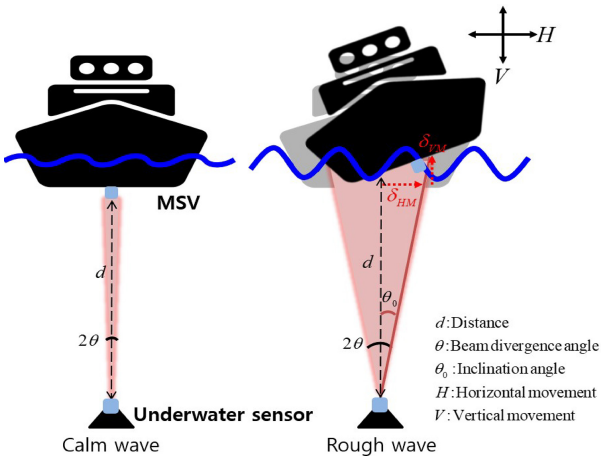


Fig. 1. System model for P2P UWOC between an underwater sensor and an MSV in the ocean.

sensor, that is, a transmit node and a receive node, while minimizing the battery consumption of the sensor. For this, we adopt a hybrid underwater acoustic-optical communication system, where acoustic and optical communications are used for low-rate control data transmission and high-rate sensing data transmission, respectively. Given this scenario, a two-phase deep RL (TPDRL) algorithm is proposed that considers two different types of agents (i.e., inner and outer agents) that determine different actions. Specifically, the primary role of the outer agent is to choose a transmission power level based on the long-term signal-to-noise ratio (SNR) between the underwater sensor and MSV. Next, the inner agent finds the beam divergence angle for the given transmission power (selected from the outer agent) based on the short-term instantaneous SNR. A conceptual diagram of the proposed algorithm is shown in Fig. 4. Simulation results demonstrate that the proposed TPDRL algorithm enables a more seamless and energy-efficient communication between an underwater sensor and MSV compared with the algorithm using only an inner agent and other existing algorithms.

C. Nomenclature

Main symbols in this article are summarized in Table I.

TABLE I
SUMMARY OF MAIN SYMBOLS

δ_{HM}	Horizontal movement of MSV
δ_{VM}	Vertical movement of MSV
d	Communication distance
$d_{initial}$	Initial communication distance
θ_0	Inclination angle
θ	Beam divergence angle
$P_{\rm rx}, P_{\rm tx}$	Received power and transmission power
$\eta_{\rm rx}, \eta_{\rm tx}$	Efficiencies of transmission and reception
L, G	Propagation and geometric loss
H	Fading coefficient
$c(\lambda), a(\lambda), b(\lambda)$	Attenuation, absorption and scattering coefficient
λ	Wavelength
A_r	Aperture size
μ_x, σ_x	Mean and variance of log-amplitude coefficient
σ_H^2	Scintillation index
I_P	Photo current
I_D	Dark current
M	Gain of photodiode
R	Responsivity of photodiode
I_N^2	Total noise power
$\sigma_{thermal}^2$	Thermal noise variance
σ_{shot}^2	Shot noise variance
σ_{dark}^2	Dark current noise variance
B	Bandwidth
q	Electrical charge
F_A	APD noise figure
k	Boltzmann constant
T	Absolute temperature
R_L	Resistance
α	Marginal value
$S^{\rm in}, S^{\rm out}$	State space of inner and outer agent
$A^{\rm in}, A^{\rm out}$	Action space of inner and outer agent
$r^{\rm in}, r^{\rm out}$	reward function of inner and outer agent
$\theta_{\min}, \theta_{\max}$	Minimum and maximum values of angle
$\delta^{\rm in}$	Gap between selectable beam divergence angles
$s_{\rm his}$	Information of previous experience
$\theta_{\rm gap}$	Difference gap between θ and θ_0
$t, \bar{t}$	Update time step of inner and outer agent
$P_{\min}, P_{\max}$	Minimum and maximum values of power
$\delta^{\rm out}$	Gap between selectable transmission powers
$\overline{\rm SNR}$	Average SNR of inner agent during T
$\rm SNR_{req}$	Predetermined SNR constraint
$\overline{\rm GAP}$	Difference gap between $\overline{\rm SNR}$ and $\rm SNR_{req}$

II. SYSTEM MODEL

A. Network Model

In this work, we consider a 3-D UOWC network, wherein multiple underwater sensors communicate with an MSV as illustrated in Fig. 1. Specifically, the underwater sensors were fixed to the seafloor, and a variety of underwater environment data were measured at regular intervals. We note that in practice, most of the underwater environmental sensors are fixed to sea floor or offshore infrastructure as illustrated in [27] and [28], except when the sensors are installed in underwater drones (e.g., remotely operated vehicles). This is because in general, operators utilizing underwater sensors do not want the movement of underwater sensors to gather precise and stable sensing data without any distortion. Further, this work can reflect the movement of underwater sensors without the modification of proposed algorithm which will be explained later, because it already considers the 3-D UOWC network and all values that will derived later reflect this 3-D characteristic. Meanwhile, using global positioning system (GPS) information, the MSV visited specific locations on the sea where the underwater sensors were vertically located according to a predetermined sea route for gathering data from multiple sensors.

For sensing data transfer from an underwater sensor to an MSV, we adopt a P2P UOWC with a hybrid underwater

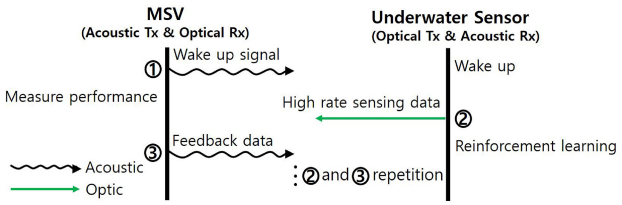


Fig. 2. Communication procedure of hybrid underwater acoustic-optical communication scenario.

acoustic-optical communication modem similar to [4], [5], and [6]. We assume that the hybrid communication modem is installed at the bottom of the MSV and at the top of the underwater sensor to align their beams for optical links. More specifically, for UOWC, laser diode (LD) and light-emitting diode (LED) have been considered as two major and mature light sources at a transmitter side [1]. According to [29] and [30], compared with LED-based UOWC modems, LD-based UOWC modems (e.g., BOLcom-LR and BlueComm 200 provided by BOLSYS and Sornardyne, respectively) can support long direct transmission range up to 150–200 m. Further, according to data recently released by the Korean Government, the average depths of the South Sea and the West Sea are about 71 m (maximum depth of about 198 m) and about 51 m (maximum depth of about 124 m), respectively. This means that a direct communication between a seabed sensor and an MSV via LD-based UOWC can be sufficiently possible in most waters in Korea. For these reason, this work considers LD-based UOWC for high-rate sensing data transmission from an underwater sensor to an MSV. On the other hand, for low-rate control data transmission with high reliability (e.g., wake-up and feedback) from the MSV to the underwater sensor, we consider UAWC because it has relatively high-reception accuracy and reliability, despite the data rate being relatively low compared with underwater optical communication [31]. Thus, an underwater sensor and an MSV can be implemented by optical-transmitter/acoustic-receiver and acoustic-transmitter/optical-receiver, respectively. This means that the underwater sensor does not need to have both optical and acoustic transceivers.

The specific P2P communication procedure between an underwater sensor and MSV is shown in Fig. 2.

1) If the MSV arrives at a specific sea location, it first sends a wake-up signal to the corresponding underwater sensor via UAWC, notifying its arrival and sending data transmission request.
2) Perceiving the wake-up signal reception, the underwater sensor transmits its collected data to the MSV via LD-based UOWC for high-rate communication.
3) After finishing the reception, the MSV sends feedback information to notify the data reception status via UAWC.
4) The procedures 2) and 3) are repeated until all data to be sent has been transmitted. At the last data transmission, the underwater sensor sends not only a sensing data but also a communication completion data.

Regarding energy saving, under the considered communication scenario, reducing power consumption at the side of underwater sensor is very crucial. This is because the underwater sensor not only has very limited batteries but also is very difficult to exchange batteries unlike the MSV which is capable of carrying a lot of batteries on it and easily exchanging them. As such, hereafter, we focus on the power consumption of underwater sensor. Table II presents an example of power draws of commercial underwater acoustic modem (e.g., Popoto S1000 [32]) and optical modem (e.g., Hydromea LUMA X-UV [33]).

TABLE II
POWER DRAWS OF COMMERCIAL UNDERWATER ACOUSTIC MODEM [32] AND OPTICAL MODEM [33]

Modem	Tx (Maximum)	Rx	Active
Acoustic	20 W	1.5 W	1.5 W
Optical	17 W	2 W	2 W

TABLE III
POWER CONSUMPTION OF UNDERWATER SENSOR AT EACH STEP OF THE COMMUNICATION PROCEDURE DESCRIBED IN FIG. 2 UNDER DIFFERENT UNDERWATER COMMUNICATION SYSTEMS

Step	Mode	Standalone UOWC	Hybrid UOWC
1	Rx	2 W	1.5 W
2	Tx	17 W (Maximum)	17 W (Maximum)
3	Rx	2 W	1.5 W

Based on Fig. 2 and Table II, Table III can present brief power consumption of underwater sensor at each step of the communication procedure in Fig. 2 under UOWC with optical transceivers (called standalone UOWC) and UOWC with hybrid underwater acoustic-optical transceivers described above (called hybrid UOWC). From Table III, it is identified that assuming no transmit power control and same control circuit power consumption, both systems consume approximately the same power. Conversely, if transmit power control is possible at the underwater sensor, there can be a possibility to save power while guaranteeing the predetermined communication quality of service. Thus, to extend the battery replacement cycle as much as possible, this work considers an adaptive transmit power control at the underwater sensor, which will be described in Section III.

During the sensing data transmission via the optical link, the MSV floating above sea level might shake randomly because of various external factors (e.g., wind, waves, and the presence of a big ship), even when hovering. This causes changes not only in the MSV location but also in the optical link alignment. In general, the movement of the MSV in the ocean (e.g., USV, ship, and vessel) can be expressed in six values (i.e., surge, sway, heave, yaw, pitch, and roll), and they are expressed by distances (e.g., surge, sway, and heave) or angles (e.g., yaw, pitch, and roll) [34]. Table IV shows an example of the allowable ranges of MSV movement to maintain a stable state depending on the size [35].

Alternatively, the movement distance of the MSV moved from the center of transmit node's signal on the water surface can be presented as horizontal movement δ_{HM} and vertical movement δ_{VM}, as shown in Fig. 1. The horizontal movement is obtained using information on the roll, pitch, sway, and

TABLE IV
CRITERIA FOR STABLE CONDITIONS FOR MSVS IN THE OCEAN

Type	Surge (m)	Sway (m)	Heave (m)	Yaw (°)	Pitch (°)	Roll (°)
Fishing vessel	1.2-1.5	1-2	0.6-1	6	4	8
Freighters	1-2	1.5-2	1-1.5	3-5	2-3	6

surge, as follows:

$$\delta_{\mathrm{HM}} = \sqrt{\left[\sin(\mathrm{roll}) + \mathrm{sway}\right]^2 + \left[\sin(\mathrm{pitch}) + \mathrm{surge}\right]^2} \quad (1)$$

where $\sin(\cdot)$ denotes the sine function. Additionally, the vertical movement is the same as the value of the heave.

With the definition of the horizontal and vertical movements, the communication distance d and inclination angle θ_0, which are illustrated in Fig. 1, can be obtained based on the MSV movement information. The communication distance between the transmit node and receive node is expressed as follows:

$$d = d_{\mathrm{initial}} + \delta_{\mathrm{VM}} \quad (2)$$

where d_{initial} is the initial communication distance. Accordingly, the inclination angle, which refers to the angle change between the center of transmit node's optical signal and the receive node by the shaking of MSV, is calculated as follows:

$$\theta_0 = \arctan\left(\frac{\delta_{\mathrm{HM}}}{d}\right). \quad (3)$$

Furthermore, Fig. 1 shows a change in the beam divergence angle of transmit node in accordance with the shaking and movement of MSV. If the beam divergence angle of the underwater sensor θ is equal to the inclination angle θ_0, the received signal strength at the MSV is the largest. In addition, as the divergence angle increases, the strength of the received signal gradually decreases under the same transmission power level. The process of calculating the intensity of the optical signal is described later in the next section. On the other hand, the rough wave occurs the relatively larger random shaking (i.e., increase in θ_0) and movement of MSV (e.g., increase in δ_{VM} and δ_{HM}), compared with the calm wave. This may result in a high probability of the disconnection of optical link between two nodes.

B. Channel Model

When transmitting an optical signal in an underwater environment, the received power P_{rx} at the receive node can be expressed as follows [36]:

$$P_{\mathrm{rx}} = P_{\mathrm{tx}}\eta_{\mathrm{tx}}\eta_{\mathrm{rx}}LGH \quad (4)$$

where P_{tx} is the transmission power; η_{tx} and η_{rx} are the transmission and reception efficiencies, respectively; L, G, and H are propagation loss, geometrical loss, and fading, respectively.

In UOWC, path-loss consists of attenuation loss L and geometrical loss G. The attenuation loss is determined by absorption and scattering in water, and the geometrical loss occurs due to the spreading of transmitted beam between a transmitter and a receiver. In general, to compute the underwater attenuation loss, the Beer–Lambert (BL) formula is commonly adopted in many existing works [37]. However, the BL formula assumes that all the scattered photons are lost and the transmitter and receiver are perfectly aligned. Since the main purpose of this work is to support communication QoS even if the transmitter (e.g., underwater sensor) and the receiver (e.g., MSV) are not aligned due to random shaking and movement of MSV floating above sea level, we reflect the misalignment effect in the path-loss [38]. First, the propagation loss of an underwater optical signal reflecting the misalignment can be given as

TABLE V
TYPICAL VALUES OF ABSORPTION, SCATTERING, AND ATTENUATION FOR DIFFERENT WATER TYPE WITH $\lambda = 514$ NM

Water type	Absorption (m^{-1})	Scattering (m^{-1})	Attenuation (m^{-1})
Pure seawater	0.0405	0.0025	0.043
Clean ocean	0.114	0.037	0.151
Coastal ocean	0.179	0.219	0.398
Turbid harbor	0.266	1.824	2.09

$$L = \exp\left\{-c(\lambda)\frac{d}{\cos(\theta_0)}\right\} \quad (5)$$

where $c(\lambda)$ is the extinction coefficient which is defined as the summation of the absorption coefficient $a(\lambda)$ and the scattering coefficient $c(\lambda)$, i.e., $c(\lambda) = a(\lambda) + b(\lambda)$. The absorption and scattering coefficients depend on the wavelength of light and the type of water. Table V lists the specific values of absorption, scattering, and attenuation coefficients for different types of water at a specific wavelength.

Furthermore, the geometrical loss of an underwater optical signal reflecting the misalignment can be also given as [38]

$$G = \begin{cases} \frac{A_r\cos(\theta_0)}{2\pi d^2[1-\cos(\theta)]}, & \theta \geq \theta_0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where A_r is the aperture area of the optical modem. As shown in the above equation, G can be obtained only when the beam divergence angle of the optical modem θ at the transmission node is equal to or greater than the inclination angle θ_0 (i.e., $\theta \geq \theta_0$).

Another factor that affects the UWOC channel is underwater turbulence, which is caused by refractive index fluctuations due to salinity and temperature fluctuations [39]. This turbulence causes the received signal intensity to fluctuate over its average also known as fading. In particular, in vertical underwater links, the gradient of salinity and temperature changes with depth [40]. To reflect this characteristic in the channel model, it is assumed that the underwater vertical link is modeled as successive nonmixing total K layers. As such, under the assumption that fading coefficients associated with different underwater layers are independent, the overall fading coefficient spanning K layers can be presented in [40] as

follows:

$$H = \prod_{k=1}^{K} H_k \tag{7}$$

where H_k is the multiplicative fading coefficient of the kth layer in the underwater vertical link. According to [40], in weak and moderate/strong turbulence conditions, the fading coefficient H_k follows log-normal (LN) and Gamma-Gamma (GG) distributions, respectively. This work assumes the weak turbulence condition which is typically observed for short link distances, and this work can be also extended to reflect the moderate/strong turbulence condition by following the GG distribution for modeling H_k. Let $\mu_{x,k}$ and $\sigma_{x,k}$ be the mean and the variance of log-amplitude coefficient $X_k = 0.5\ln(H_k)$ for the kth layer. The probability density function (PDF) of H_k is then presented as

$$f_{H_k}(H_k) = \frac{1}{H_k\sqrt{2\pi\left(4\sigma_{x,k}^2\right)}} \exp\left(-\frac{\left(\ln(H_k) - 2\mu_{x,k}\right)^2}{2\left(4\sigma_{x,k}^2\right)}\right). \tag{8}$$

Let $\sigma_{H,k}^2$ be the scintillation index. Then, the variance of $\sigma_{x,k}$ can be expressed in terms of scintillation index as follows:

$$\sigma_{x,k}^2 = 0.25\ln\left(1 + \sigma_{H,k}^2\right). \tag{9}$$

Since the fading coefficient does not change the average power, its amplitude should be normalized, i.e., $E[H_k] = 1$, such that $\mu_{x,k} = -\sigma_{x,k}^2$ [41]. Under the assumption that H_i and H_j are independent, $i, j \in \{1, 2, \ldots, K\}$, but nonidentically distributed, the PDF of H follows LN distribution as given by

$$f_H(H) = \frac{1}{H\sqrt{2\pi\left(4\sigma_x^2\right)}} \exp\left(-\frac{(\ln(H) - 2\mu_x)^2}{2\left(4\sigma_x^2\right)}\right) \tag{10}$$

where $\sigma_x^2 = \sum_{k=1}^{K} 4\sigma_{x,k}^2$ and $\mu_x = \sum_{k=1}^{K} 2\mu_{x,k}$.

C. Communication Performance

Receiving an optical signal through an underwater channel, the receiver node converts the received signal into an electrical signal using a photodetector. Among the various components of a photodetector, the photodiode plays a key role in converting an optical signal into an electrical signal. According to [42], PIN photodiode and avalanche photodiode (APD) are widely utilized in optical communication applications because of their minute size, high sensitivity, and fast response time. This study considers a photodetector using APD, and accordingly, the photocurrent according to the APD performance can be calculated as follows:

$$I_P = \mathrm{MRP}_{\mathrm{rx}} \tag{11}$$

where M and R are the gain and responsivity of APD, respectively. The values of M and R can vary depending on the manufacturing material (e.g., Si, Ge, or InGaAs) for APD. According to [36], M and R have higher values in the order Si, Ge, and INGaAs. This study considered the use of a Si APD when computing the photocurrent.

Based on (11), the instantaneous SNR can be defined as a communication performance metric

$$\mathrm{SNR}(\theta, P_{\mathrm{tx}}) = \frac{I_P^2}{I_N^2} = \frac{(\mathrm{MRP}_{\mathrm{rx}})^2}{I_{\mathrm{thermal}}^2 + I_{\mathrm{shot}}^2 + I_{\mathrm{dark}}^2} \tag{12}$$

where I_N^2 is the total noise power; it is defined as the summation of the thermal noise power I_{thermal}^2, shot noise power I_{shot}^2, and dark current noise power I_{dark}^2 [43].

More specifically, thermal noise (also known as Johnson noise) is caused by the irregular movement of electrons by thermal energy. The spectral density of the thermal noise is independent of the frequency, which is referred to as a white noise. Thermal noise follows Gaussian statistics, such that it is characterized by its power spectral density $\sigma_{\mathrm{thermal}}^2 = (4kT/R_L)$ [W/Hz], where k is the Boltzmann constant, T is the absolute temperature, and R_L is the load resistance. For a receiver with a bandwidth B, the total thermal noise power I_{thermal}^2 can be expressed by the variance of thermal noise current as follows:

$$I_{\mathrm{thermal}}^2 = \sigma_{\mathrm{thermal}}^2 B = \frac{4kT}{R_L} B. \tag{13}$$

Shot noise (also known as quantum noise) is the noise component of the photocurrent due to irregular electron generation. Shot noise is also a white noise and the noise power spectral density can be given as $\sigma_{\mathrm{shot}}^2 = 2qM^2F_AI_P$ [W/Hz], where q is the electrical charge, and F_A is the APD noise. Given an approximation of Gaussian statistics, the shot noise power can be represented by the variance of the shot noise current as follows:

$$I_{\mathrm{shot}}^2 = \sigma_{\mathrm{shot}}^2 B = 2qM^2F_AI_PB. \tag{14}$$

Dark current noise is the current when the photons are not applied, which is mainly generated in the depletion layer by heat. The dark current noise can also be treated as a white noise with the power spectral density $\sigma_{\mathrm{dark}}^2 = 2qM^2F_AI_D$ [W/Hz], where I_D is the dark current of APD. Same as (13) and (14), the dark current noise power I_{shot}^2 can be presented as follows:

$$I_{\mathrm{dark}}^2 = \sigma_{\mathrm{dark}}^2 B = 2qM^2F_AI_DB. \tag{15}$$

D. Motivation

Using (12), we can compute the instantaneous SNR of the wireless optical link between an underwater sensor and MSV for a given beam divergence angle θ, inclination angle θ_0, and transmission power P_{tx}. For example, the instantaneous SNR is maximized when $\theta = \theta_0$ for a given P_{tx}. However, in practice, the MSV floating above sea level may shake and move constantly and irregularly. This can result in a constant irregular change in the inclination angle θ_0. That is, even though the MSV feeds back the exact value of the inclination angle at a specific time slot, the underwater sensor cannot know the exact inclination angle θ_0 at the time of data transmission due to time causality (i.e., relatively long time delay of the underwater acoustic link). The current study has focused upon this problem. In particular, as shown in Fig. 3, the transmit node is assumed to set beam divergence angle θ to inclination

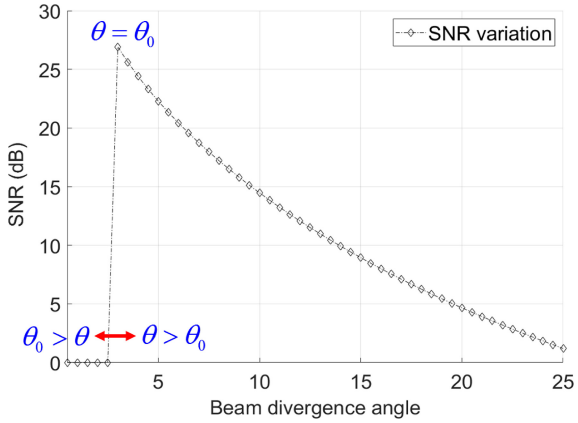


Fig. 3. SNR variation according to beam divergence angle.

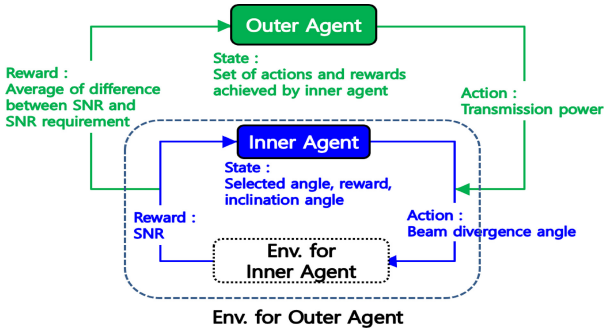


Fig. 4. Illustration of the proposed TPDRL framework.

angle θ_0 (i.e., $\theta = \theta_0$) to maximize the instantaneous SNR. However, at the beginning of or during data transmission, if θ_0 increases owing to the sudden wind fluctuations over the sea, the instantaneous SNR becomes zero, which implies that the wireless optical link is immediately disconnected. In contrast, if θ_0 decreases, the instantaneous SNR also decreases, but this might be acceptable as it does not become zero. From these observations, we can conclude that choosing θ as the summation of θ_0 and marginal value α (i.e., $\theta_0 + \alpha$) can be an effective strategy to support the seamless connection of wireless optical links. Furthermore, the adaptive choice of α is required according to changes in the external environment of the ocean.

Additionally, if there is a marginal gain in the instantaneous SNR, the transmission power of the transmit node P_{tx} can be adjusted to minimize the battery consumption of the transmit node located on the seafloor. This can be crucial in practice because the cost of replacing the battery of underwater sensors is too high compared with that on land.

III. Proposed Algorithm

A. Problem Formulation

The purpose of this work is to jointly optimize the beam divergence angle and transmission power of a transmit node to maintain a seamless connection with a receive node while minimizing the battery consumption of the sensor.

In this regard, we formulated our problem as a Markov game with two agents, which is a multiagent extension of Markov decision process (MDP). Subsequently, we solved the problem using the proposed TPDRL framework, which is briefly illustrated in Fig. 4. Thereafter, with Fig. 4, we presented a detailed and descriptive explanation of our MDP formulation and proposed TPDRL framework.

In our MDP formulation, an underwater sensor is a decision maker with two agents, namely, inner and outer agents, that determine a different action as follows.

1) *Inner Agent:* Beam divergence angle θ to maximize the instantaneous SNR defined in (12).
2) *Outer Agent:* Transmission power P_{tx} to minimize the battery consumption of underwater sensor, while supporting the predetermined average SNR requirement.

B. Inner Agent: Beam Divergence Angle Adaptation

Given transmission power P_{tx}, the role of the inner agent for an underwater sensor is to determine the beam divergence angle θ at every time slot t by learning the inclination angle of the MSV (i.e., θ_0). Note that the inclination angle can change every time slot because of the random shaking and movement of the MSV floating above sea level, as described in the previous section.

To achieve this objective, we formulate the MDP for the inner agent, which consists of tuple $(S^{\text{in}}, A^{\text{in}}, r^{\text{in}})$, where S^{in} is the state space, A^{in} is the action space, and r^{in} is the reward function for the inner agent.

First, an action $a^{\text{in}}(t) = \theta(t) \in A^{\text{in}}(t)$ at time slot t is an element in the action space $A^{\text{in}}(t)$, which represents the set of selectable beam divergence angles supported by the optical modem installed on the underwater sensor

$$A^{\text{in}}(t) = \left\{\theta_{\min}, \theta_{\min} + \delta^{\text{in}}, \ \ldots, \theta_{\max} - \delta^{\text{in}}, \theta_{\max}\right\} \quad (16)$$

where $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum values allowed for the beam divergence angles, respectively, and δ^{in} refers to the predetermined gap between two consecutive beam divergence angles.

Second, the state space at time slot t [i.e., $S^{\text{in}}(t)$] is defined as follows:

$$S^{\text{in}}(t) = \left\{\mathbf{s}_{\text{his}}(t-1), \theta_0(t-1), \theta_{\text{gap}}(t-1)\right\} \quad (17)$$

where $\mathbf{s}_{\text{his}}(t-1)$ includes historical information on the actions and rewards experienced by the inner agent, $\theta_0(t-1)$ is the inclination angle at time slot $t-1$, and $\theta_{\text{gap}}(t-1)$ is the difference between $\theta(t-1)$ and $\theta_0(t-1)$. To limit the size of the state space as learning progresses, we adopted a sliding window of size m in the state space at each time slot. Thus, we define $\mathbf{s}_{\text{his}}(t-1)$ as follows:

$$\mathbf{s}_{\text{his}}(t-1) = \left\{a^{\text{in}}(t-m), r^{\text{in}}(t-m), \ \ldots, a^{\text{in}}(t-1), r^{\text{in}}(t-1)\right\}. \quad (18)$$

The inclination angle at time slot $t-1$ (i.e., $\theta_0(t-1)$) can be fed back along with the reward from the MSV to the underwater sensor via a wireless acoustic link, as illustrated in Fig. 2.

Finally, the reward function of the inner agent r^{in} represents the instantaneous SNR defined in (12) under the selected

action, as follows:

$$r^{\text{in}}(t) = \text{SNR}(\theta(t), P_{\text{tx}}). \tag{19}$$

Note that through receiving the reward value as feedback data via UAWC, the underwater sensor can perceive whether the MSV is within the range of proper beam divergence angle at each time slot.

C. Outer Agent: Transmission Power Adaptation

In the previous section, we described how the inner agent adjusted the beam divergence angle according to the irregular shaking and movement of the MSV for a given transmission power. Here, we describe how the outer agent determines the proper transmission power by observing the long-term SNR performance of the current transmission power. For this purpose, we also calculate the MDP for an outer agent, which consists of a tuple $(S^{\text{out}}, A^{\text{out}}, r^{\text{out}})$, where S^{out} is the state space, A^{out} is the action space, and r^{out} is the reward function for the outer agent.

Because instantaneous SNR performance is affected by a series of actions taken by the inner agent (i.e., a set of beam divergence angles), some time is required to judge the SNR performance of a given transmission power. As such, the outer agent chooses its action, that is, transmission power, at every time interval $T = nt$. For convenience, we index the update time of the outer agent as $\bar{t} = 1, 2, \ldots$ and the time index $\bar{t}$ corresponds to the time slot $(\bar{t}-1)T+1$. Accordingly, $P_{\text{tx}}(\bar{t})$ refers to the transmission power of the underwater sensor at update time $\bar{t}$, which is maintained during time slots $t \in [(\bar{t}-1)T+1 : \bar{t}T]$.

Let $a^{\text{out}}(\bar{t}) = P_{\text{tx}}(\bar{t}) \in A^{\text{out}}(\bar{t})$ be the action of the outer agent at the update time $\bar{t}$. $A^{\text{out}}(\bar{t})$ is the action space, which is a set of transmission powers that the optical modem of the underwater sensor can support

$$A^{\text{out}}(\bar{t}) = \left\{P_{\min}, P_{\min}+\delta^{\text{out}}, \ldots, P_{\max}-\delta^{\text{out}}, P_{\max}\right\} \tag{20}$$

where $P_{\min}$ and $P_{\max}$ are the minimum and maximum values of the allowable transmission powers, respectively, and δ^{out} refers to the predetermined gap between two consecutive transmission powers.

The information in the state space for the outer agent is defined as follows:

$$S^{\text{out}}(\bar{t}) = \left\{(\theta(t))_{t\in[(\bar{t}-2)T+1:(\bar{t}-1)T]},\ \overline{\text{SNR}}(\bar{t}-1),\ \overline{\text{GAP}}(\bar{t}-1)\right\} \tag{21}$$

where $(\theta(t))_{t\in[(\bar{t}-2)T+1:(\bar{t}-1)T]}$ denotes the set of actions of the inner agent achieved through previous updates. Furthermore, $\overline{\text{SNR}}(\bar{t}-1)$ is the average of the rewards of the inner agent (i.e., the instantaneous SNR) through previous updates

$$\overline{\text{SNR}}(\bar{t}-1) = \frac{1}{T}\sum_{t=(\bar{t}-2)T+1}^{(\bar{t}-1)T} \text{SNR}\left(\theta(t), P_{\text{tx}}(\bar{t}-1)\right). \tag{22}$$

Similarly, $\overline{\text{GAP}}(\bar{t}-1)$ is the average difference between the instantaneous SNR and predetermined SNR constraint

$$\overline{\text{GAP}}(\bar{t}-1) = \frac{1}{T}\sum_{t=(\bar{t}-2)T+1}^{(\bar{t}-1)T} \left[\text{SNR}_{\text{req}} - \text{SNR}\left(\theta(t), P_{\text{tx}}(\bar{t}-1)\right)\right] \tag{23}$$

where SNR_{req} denotes the required SNR of the underwater sensor.

Finally, the reward function for the outer agent is expressed as follows:

$$r^{\text{out}}(\bar{t}) = \beta - \overline{\text{GAP}}(\bar{t}) = \beta - \frac{1}{T}\sum_{t=(\bar{t}-1)T+1}^{\bar{t}T} \left[\text{SNR}_{\text{req}} - \text{SNR}\left(\theta(t), P_{\text{tx}}(\bar{t})\right)\right] \tag{24}$$

where β is a hyperparameter that can be used to convert a minimization problem into a maximization one.

D. Proposed TPDRL Algorithm

Algorithm 1 describes the proposed TPDRL algorithm for jointly determining the beam divergence angle θ and transmission power P_{tx} in an underwater sensor (i.e., the DRL decision maker). As mentioned above, when implementing the proposed DRL algorithm, each underwater sensor has two DRL agents (i.e., inner and outer agents) that determine different actions. In the proposed algorithm, both the inner and outer agents adopt a deep Q-network (DQN), where each agent has and learns its own DNN. Owing to space limitation, we omit the detailed description on the structure of DQN [44].

At each time slot t, under a given transmission power, the inner agent first constructs state $S^{\text{in}}(t)$ defined in (17). To construct $S^{\text{in}}(t)$, the inner agent (i.e., underwater sensor) should receive information on the reward and inclination angle at time slot $t-1$ (i.e., $r^{\text{in}}(t-1)$ and $\theta_0(t-1)$) from the MSV as feedback. As mentioned, for feedback signaling, wireless acoustic communication is considered, which can guarantee a data rate (e.g., more than Kb/s) that is sufficient to transmit such information. After constructing $S^{\text{in}}(t)$, the inner agents execute their action (i.e., the beam divergence angle) with probability $1-\epsilon$

$$a^{\text{in}}(t) = \arg\max_{a\in A^{\text{in}}(t)} Q^{\text{in}}\left(s^{\text{in}}(t), a|\Theta^{\text{in}}\right) \tag{25}$$

where $Q^{\text{in}}(s^{\text{in}}(t), a)$ denotes the Q-function achieved by taking action a in state $s^{\text{in}}(t)$, and Θ^{in} is a set of weights for the DNN of the inner agent. Alternatively, with probability ϵ, it randomly chooses an action in action space $A^{\text{in}}(t)$. Moreover, at each time slot, Θ^{in} is updated by minimizing the following mean-squared error (MSE) loss function

$$\text{MSE}\left(\Theta^{\text{in}}\right) = \text{E}\left[y - Q^{\text{in}}\left(s^{\text{in}}(t), a^{\text{in}}(t)|\Theta^{\text{in}}\right)\right]^2 \tag{26}$$

where y is the target value for updating Θ^{in}, which is defined as follows:

$$y = r^{\text{in}}(t+1) + \gamma \max_{a'} Q^{\text{in}}\left(s^{\text{in}}(t+1), a'|\tilde{\Theta}^{\text{in}}\right). \tag{27}$$

Algorithm 1: Proposed TPDRL Algorithm

Initialize replay buffers D^{out}, D^{in}
Initialize DNN$^{\text{out}}$ and DNN$^{\text{in}}$ with weights Θ^{out}, Θ^{in}.
Initialize target DNN$^{\text{out}}$, target DNN$^{\text{in}}$ with weights $\tilde{\Theta}^{\text{out}}$, $\tilde{\Theta}^{\text{in}}$.
for each update time $\bar{t}$ **do**
  The outer agent constructs state $S^{\text{out}}(\bar{t})$
  if *random*() $\leq \epsilon$ **then**
    Randomly select action $a^{\text{out}}(\bar{t})$ from action space $A^{\text{out}}(t)$
  else
    Select $a^{\text{out}}(\bar{t}) = \arg\max_{a \in A^{\text{out}}(\bar{t})} Q^{\text{out}}\left(s^{\text{out}}(\bar{t}), a|\Theta^{\text{out}}\right)$
  end
  Initialize data storage X
  for each time slot $t \in [(\bar{t}-1)T+1 : \bar{t}T]$ **do**
    The inner agent constructs state $S^{\text{in}}(t)$
    if *random*() $\leq \epsilon$ **then**
      Randomly select $a^{\text{in}}(t)$ from action space $A^{\text{in}}(t)$
    else
      Select $a^{\text{in}}(t) = \arg\max_{a \in A^{\text{in}}(t)} Q^{\text{in}}\left(s^{\text{in}}(t), a|\Theta^{\text{in}}\right)$
    end
    Observe $r^{\text{in}}(t)$ and $S^{\text{in}}(t+1)$
    Store the experience sample in D^{in}
    The inner agent randomly samples a mini-batch from D^{in}, and update weight Θ^{in}. In every pre-determined time interval, the inner agent updates the target DNN$^{\text{in}}$ with $\tilde{\Theta}^{\text{in}} = \Theta^{\text{in}}$
    Store $r^{\text{in}}(t)$ in X
  end
  Compute $\overline{\text{SNR}}(\bar{t})$ by averaging the elements in X
  if $\overline{\text{SNR}}(t) \geq \text{SNR}_{\text{req}}$ **then**
    $r^{\text{out}}(\bar{t}) = \beta - \overline{\text{GAP}}(\bar{t})$
  else
    $r^{\text{out}}(\bar{t}) = 0$
  end
  Get $r^{\text{out}}(\bar{t})$ and $S^{\text{out}}(\bar{t}+1)$
  Store the experience sample in D^{out}
  The outer agent randomly samples a mini-batch from D^{out}, and updates weight Θ^{out}. In every pre-determined update time, the outer agent updates the target DQN$^{\text{out}}$ with $\tilde{\Theta}^{\text{out}} = \Theta^{\text{out}}$
end

Meanwhile, the outer agent constructs the state $S^{\text{out}}(t)$ at each update time $\bar{t}$ which corresponds to the time slot $(\bar{t}-1)T+1$. When constructing $S^{\text{out}}(t)$, no additional information should be fed back from the MSV. After constructing $S^{\text{out}}(t)$, the outer agent determines its action (i.e., the transmission power) with probability $1-\epsilon$

$$a^{\text{out}}(\bar{t}) = \underset{a \in A^{\text{out}}(\bar{t})}{\arg\max}\, Q^{\text{out}}\left(s^{\text{out}}(\bar{t}), a|\Theta^{\text{out}}\right). \tag{28}$$

TABLE VI
NETWORK PARAMETERS FOR UOWC

Parameter	Value
Initial communication distance $d_{initial}$	20 m
Transmission wavelength λ	514 nm
Optical efficiency of transmitter η_T	0.9
Optical efficiency of receiver η_R	0.9
Attenuation coefficient c(λ)	0.398 m^{-1}
Aperture size A_r	0.01 m^2
APD gain M	150
APD responsivity R	75
Electrical charge q	1.6×10^{-19}
Noise figure F_A	0.5
Bandwidth B	5 GHz
Boltzmann constant k	1.38×10^{-23}
Dark current I_D	15 nA
Resistance R_L	100 Ω
Temperature T	298 K

In (28), $Q^{\text{out}}(s^{\text{out}}(\bar{t}), a)$ denotes the Q-function achieved by taking action a in state $s^{\text{out}}(\bar{t})$, and Θ^{out} is a set of weights for the DNN of the outer agent. Alternatively, with probability ϵ, it randomly chooses an action in action space $A^{\text{out}}(\bar{t})$. The chosen transmission power $P_{\text{tx}}(\bar{t}) = a^{\text{out}}(\bar{t})$ is used during the time slots $t \in [(\bar{t}-1)T+1 : \bar{t}T]$. To update Θ^{out}, the same steps were adopted by the inner agent.

IV. Performance Evaluations

A. Network Environment

We conducted a numerical evaluation of the proposed algorithm. For simulations, we assumed a scenario wherein one MSV collects data from an underwater sensor through a P2P UOWC. Also, the UOWC channel is assumed to be a 20 m of 1-layer, such that the scintillation index according to the short link distance in the weak turbulence condition is given as $\sigma_H^2 = 0.87$ [45]. The static system parameters adapted for the numerical performance evaluations, summarized in Table VI, were chosen from [36] and [38].

B. Learning Environment

The DQN structure for the inner and outer agents was a fully connected neural network with two hidden layers containing 68 neurons in each hidden layer. The other learning hyperparameters are summarized in Table VII.

We implemented the inner and outer agents of the proposed TPDRL algorithm as follows: the state of the inner agent is set as (17), and the action set of the inner agent (i.e., set of beam divergence angles) is given as $A^{\text{in}} = [1, 2, \ldots, 10]$ with a unit of angle (°). The state of the outer agent consists of the set (21), and the selectable action set is defined as $A^{\text{out}} = [1, 2, \ldots, 10]$ with a unit of power (mW).

C. Generation of Random MSV Movement

To generate the time-varying random movement of MSV, it was assumed that the MSV is a fishing vessel and it is under its stable condition on the sea surface as stated in Table IV.

TABLE VII
LIST OF DQN HYPERPARAMETERS

Hyperparameter	Agent
Episode	100000
ϵ for ϵ-greedy	10^{-4}
Mini-batch size	64
Optimizer	Adam
Activation function	Relu
Learning rate (τ)	10^{-4}
Experience replay size (inner)	10000
Experience replay size (outer)	5000
Target model update cycle (inner)	10000
Target model update cycle (outer)	5000
Discount factor (μ)	0.99

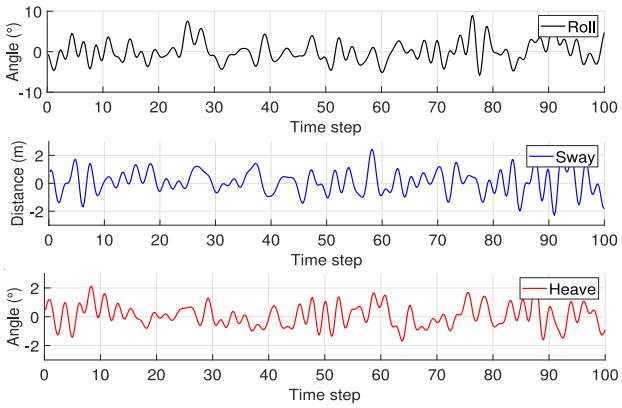


Fig. 5. Examples of randomly generated time-series data of MSV movement: Roll, Sway, and Heave.

Through real-sea experiments conducted by [46] and [47], it was verified that under the stable condition, six values in terms of fishing vessel movement (i.e., surge, sway, heave, yaw, pitch, and roll) change by following random sine waveforms which are bounded within allowable ranges stated in Table IV. To reflect this fact, we have generated the six values in terms of MSV movements by using random sine wave functions which are bounded within the allowable ranges. Fig. 5 illustrates the examples of randomly generated time-series data on roll, sway, and heave which are used for our simulations.

D. Benchmark Algorithms

To validate the proposed algorithm, we have compared its performance with that of the existing optical beam selection algorithms introduced below. Ali [36] proposed an algorithm selecting the widest beam angle that a transmit node can support for guaranteeing the seamless wireless optical connection. The work of [38] investigated the beam selection algorithm considering position uncertainty, wherein the additional position uncertainty is reflected after obtaining the position information of both nodes. In [48], the RL-based beam selection algorithm was presented, where the beam angle increases or decreases by one step; otherwise, the beam angle is maintained. Furthermore, "Random" refers to the algorithm that randomly selects the beam divergence angle, whereas "Optimal" sets the beam divergence angle as the inclination angle. Note that all the existing algorithms have been considered assuming the transmission power as maximum.

E. Computational Complexity Analysis

As illustrated in Fig. 4, the proposed TPDRL algorithm includes independent two DQNs to implement inner and outer agents, respectively. Further, each DQN consists of DNNs, i.e., a Q-network and a target-network with the same structure. Let L^z and m_l^z be the number of layers of DQN for agent $z \in \{\text{in, out}\}$ and the number of neurons in lth layer of DQN for agent z. According to [49], the computational complexity of ϵ-greedy policy based Q-learning algorithm is $O(A)$ where A denotes totally number of steps, and the computational complexity of each training step in DNN using a fully connected network is $O(\sum_{l=1}^{L^z-1} m_l^z m_{l+1}^z)$. Since the target-network of DQN performs only forward propagation at each time slot, the total computational complexity of DQN for agent z (denoted by C^z) can be presented as $C^z = O(3\sum_{l=1}^{L^z-1} m_l^z m_{l+1}^z)$. Note that under the proposed algorithm, the inner agent chooses its action, i.e., beam divergence angle, at every time slot t, whereas, the outer agent determines its action, i.e., transmission power, at every time interval $T = nt$. Thus, the computational complexity of the proposed TPDRL algorithm (denoted by C^{TPDRL}) for a time slot can be presented as $C^{\text{TPDRL}} = O([3n\sum_{l=1}^{L^{\text{in}}-1} m_l^{\text{in}} m_{l+1}^{\text{in}} + 3\sum_{l=1}^{L^{\text{out}}-1} m_l^{\text{out}} m_{l+1}^{\text{out}}]/n)$.

In case of the beam selection algorithm in [48], it is not the DRL-based algorithm but the RL-based algorithm. Thus, for fair comparison, we consider this algorithm implementing DQN where the same state space, the action space, and the reward are reflected as same in [48]. Since the algorithm can implement a single DQN, its computational complexity (denoted by C^{Ω}) for a time slot can be expressed as $C^{\Omega} = O(3\sum_{l=1}^{L^{\Omega}-1} m_l^{\Omega} m_{l+1}^{\Omega})$. Obviously, this algorithm has slightly lower computational complexity compared with the proposed algorithm, but the proposed algorithm can sufficiently execute in real-time. On the other hand, the proposed algorithm achieves better successful probability while it uses the lower transmission power (e.g., about 29% power consumption is reduced in case of SNR requirement = 16). Moreover, other existing algorithms as illustrated in Fig. 10 [i.e., Beam selection ([36], [38], and [48])] are not an algorithm-based method but a deterministic method. Thus, these algorithms also have a relatively lower computational complexity, but their performances (e.g., success probability and transmission power) are significantly lower than the proposed algorithm.

F. Simulation Results

We conducted a performance evaluation of the proposed algorithm using the inner and outer agents. As mentioned in (7), to ensure that the communication link is seamlessly maintained, the beam divergence angle of the transmit node (i.e., underwater sensor) should be set wider than the inclination angle determined by the position change of the receive node (i.e., MSV). In this regard, the inner agent of the proposed algorithm predicts the random movement and shaking of the MSV and selects the beam divergence angle that

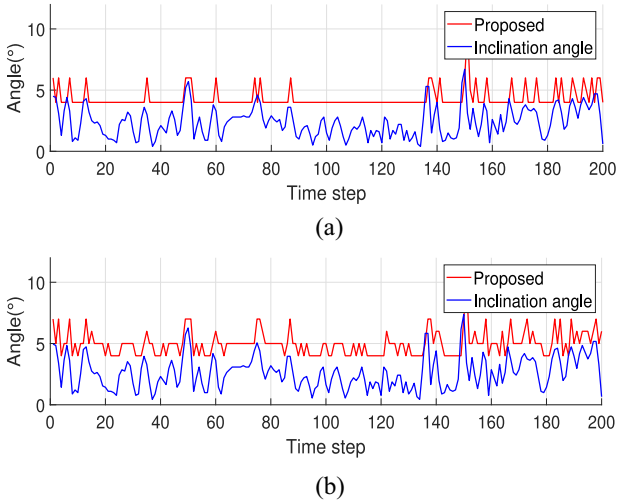


Fig. 6. Beam divergence angle selection according to the learning of inner agent of proposed algorithm. (a) Fluctuation range of inclination angle: 0°–8.4°. (b) Fluctuation range of inclination angle: 0°–9.3°.

maximizes the reward $r^{\text{in}}(t)$ in each time slot. Furthermore, to reduce unnecessary power consumption of the underwater sensor, the outer agent of the proposed algorithm attempts to minimize the transmission power while maintaining the required learning performance of the inner agent. Thus, the transmission power level is selected such that $\overline{\text{GAP}}(\bar{t})$ approaches zero. To analyze the effect of the inner and outer agents in the proposed algorithm, we differentiated the proposed algorithm as the proposed algorithm (inner only) and proposed algorithm (inner–outer). In the proposed algorithm (inner only), without the outer agent, the inner agent only performs learning to choose the beam divergence angle. In contrast, in the proposed algorithm (inner–outer), both the inner and outer agents collaboratively perform joint optimization of the beam divergence angle and transmission power to achieve the objective of this study.

First, we conducted performance evaluation of the proposed algorithm (inner only) to determine how well its performance followed an optimal solution. Fig. 6 presents beam divergence angles achieved from the proposed algorithm (inner only); they are compared with inclination angles, which are optimal values, that the proposed algorithm should follow. From Fig. 6, it can be seen that the beam divergence angles follow the inclination angles while considering additional marginal values, i.e., $\theta = \theta_0 + \alpha$. By considering the additional marginal values, disconnection (i.e., SNR $= 0$ in the case of $\theta < \theta_0$) due to random shaking and movement of the MSV floating above sea level can be avoided. It is observed that as the fluctuation ranges of the shaking and movement of MSV increase (e.g., due to bad weather condition in the ocean), the marginal values also increase.

Fig. 7 shows the moving average of performances with respect to misalignment probability and instantaneous SNR for the proposed algorithm (inner only) and existing algorithms. Here, the misalignment probability indicates the probability that the achieved instantaneous SNR is zero. Fig. 7(a) illustrates that the proposed algorithm (inner only) performs better

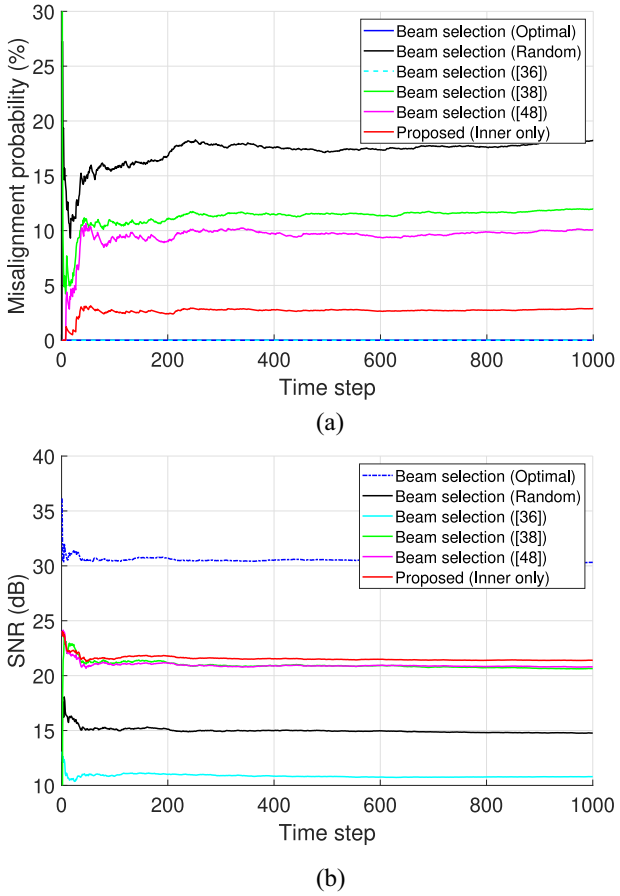


Fig. 7. Moving average of performances of the proposed (inner only) and existing algorithms. (a) Misalignment probability. (b) Instantaneous SNR.

than all other algorithms except for "Beam selection [36]," wherein the misalignment does not occur by setting the beam divergence angle to be always larger than the inclination angle. Meanwhile, because the objective of the inner agent in the proposed algorithm is to maximize the instantaneous SNR while avoiding misalignment, a comparison of the performance on instantaneous SNR is also shown in Fig. 7(b). Fig. 7(b) indicates that the proposed algorithm (inner only) has the highest SNR performance among the existing algorithms. With results from Fig. 7, we can affirm that the proposed algorithm (inner only) maintains the highest instantaneous SNR while avoiding the misalignment between transmit and receive nodes by selecting proper beam divergence angles, according to the changes in the shaking and movement of MSV in the ocean.

Second, we conducted performance evaluation of the proposed algorithm (inner–outer) to identify the necessity of the outer agent in the proposed algorithm. Fig. 8 illustrates the comparison of performances (e.g., average SNR and transmission power) between the proposed algorithm (inner only) and proposed algorithm (inner–outer) under a change in beam divergence angles in an underwater sensor. In Fig. 8, it is shown that the proposed algorithm (inner–outer) can decrease the transmission power while not violating the required SNR constraint, when compared with the proposed algorithm (inner only). This implies that the existence of an outer agent makes

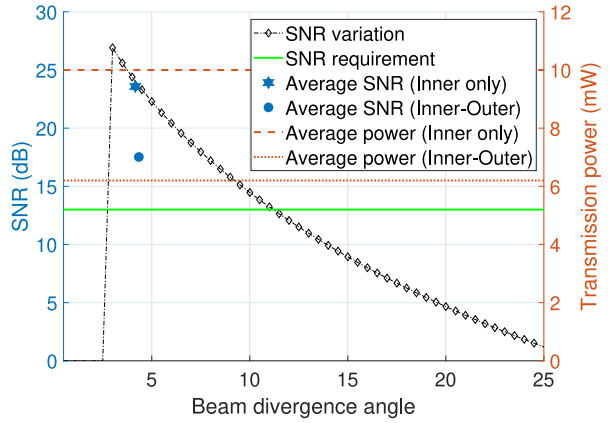


Fig. 8. Performance comparison between the proposed algorithm (inner only) and the proposed algorithm (inner–outer).

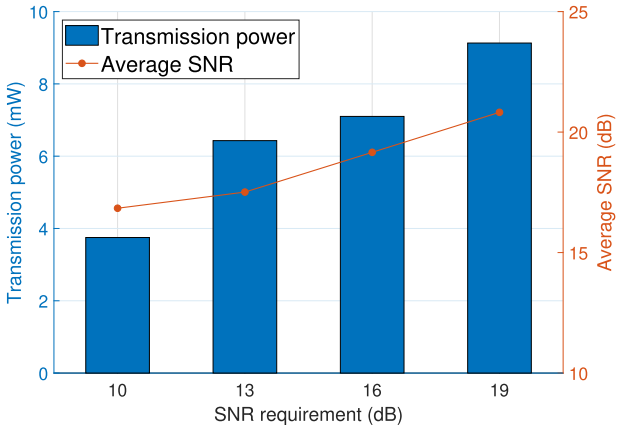
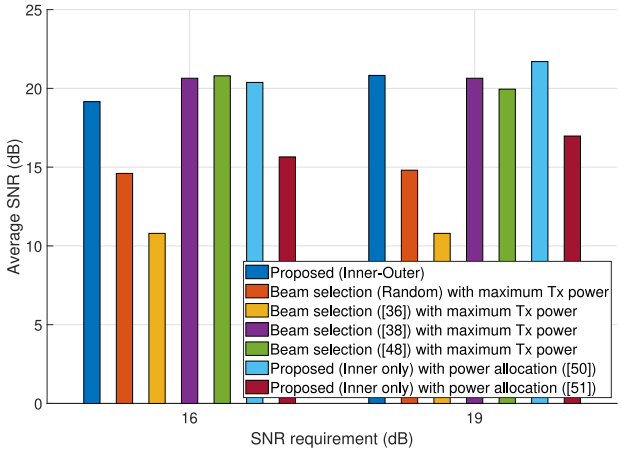
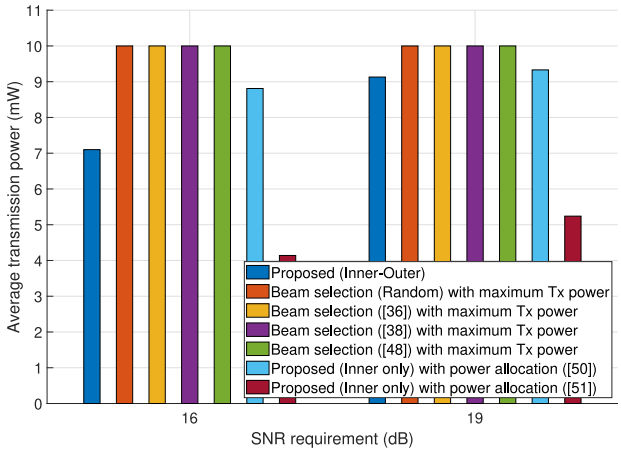


Fig. 9. Average transmission power and SNR under different SNR requirements.



(a)



(b)

Fig. 10. Performance comparison between the proposed (inner–outer) and existing algorithms under different SNR requirements. (a) Average SNR. (b) Transmission power.

an underwater sensor operate for a longer period of time without battery exchange.

Fig. 9 presents average transmission power levels and corresponding average SNRs achieved from the proposed algorithm (inner–outer) under a change in SNR requirements. In Fig. 9, it is seen that the proposed algorithm can satisfy the predetermined SNR requirement SNR_{req}, even though the outer agent adjusts the transmission power level for energy saving depending on SNR_{req}. More specifically, when the SNR requirement $\text{SNR}_{\text{req}} = 10$ dB, a transmission power of approximately 3.9 mW needs to support the SNR requirement, whereas more transmission power (e.g., 9.1 mW) is required to support the relatively high-SNR requirement $\text{SNR}_{\text{req}} = 19$ dB.

Fig. 10 compares performances in terms of the average SNR and the corresponding transmission power between the proposed (inner–outer) and existing UOWC algorithms under different SNR requirements. We note that the existing beam selection algorithms for UOWC (i.e., random [36], [38], and [48]) did not consider the adaptive transmission power control while adjusting the beam divergence angle. As such, they are set to the same transmission power level (i.e., 10 mW) which is the maximum level of transmission power at an underwater sensor considered in this work. In the contrary, the existing power allocation algorithms for UOWC (i.e., [50] and [51]) did not consider the adaptive beam divergence angle control while adjusting the transmission power. As such, for fair comparison, we have combined the inner agent of the proposed algorithm (determining beam divergence angle) with the adaptive power allocation algorithms in [50] and [51]. Fig. 10 illustrates that the proposed algorithm achieves the lowest transmission power while satisfying the predetermined average SNR performance. In particular, it is identified that in Fig. 10(b), the combination of the inner agent of the proposed algorithm and the power control algorithm in [51] consumes the lowest transmission power, but it does not satisfy the average SNR requirements (e.g., 15.65 dB for $\text{SNR}_{\text{req}} = 16$ dB and 16.97 dB for $\text{SNR}_{\text{req}} = 19$ dB, respectively). The reason why the proposed algorithm can achieve the objective of this work is that by the online adaptation of beam divergence angle at the inner agent, the outer agent has the possibility to reduce the transmission power while supporting the SNR requirement. This results show the validity of the collaboration of inner and outer agents in the proposed TPDRL framework to achieve the objective.

V. Conclusion

In this study, we proposed a TPDRL algorithm for jointly optimizing the beam divergence angle and transmission power

level for P2P UOWC environment. The proposed scheme adjusted not only the beam divergence angle by the inner agent to maintain a seamless connection but also the transmission power by the outer agent to minimize the battery consumption of underwater sensor. We conducted the performance comparison between the propose and the existing algorithms to verify the proposed algorithm provides significant gain in terms of power efficiency while achieving a higher instantaneous SNR with a low probability of link misalignment error.

REFERENCES

[1] I. N'Doye, D. Zhang, M.-S. Alouini, and T.-M. Laleg-Kirati, "Establishing and maintaining a reliable optical wireless communication in underwater environment," *IEEE Access*, vol. 9, pp. 62519–62531, 2021.

[2] G. S. Spagnolo, L. Cozzella, and F. Leccese, "Underwater optical wireless communications overview," *Sensors*, vol. 20, no. 8, p. 2261, 2020.

[3] L. Xiao, D. Jiang, Y. Chen, W. Su, and Y. Tang, "Reinforcement-learning-based relay mobility and power allocation for underwater sensor networks against jamming," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 1148–1156, Jul. 2020.

[4] L. J. Johnson, R. J. Green, and M. S. Leeson, "Hybrid underwater optical/acoustic link design," in *Proc. Int. Conf. Transparent Opt. Netw. (ICTON)*, 2014, pp. 1–4.

[5] S. Han, Y. Noh, R. Liang, R. Chen, Y.-J. Cheng, and M. Gerla, "Evaluation of underwater optical-acoustic hybrid network," *China Commun.*, vol. 11, no. 5, pp. 49–59, May 2014.

[6] J. Wang, W. Shi, L. Xu, L. Zhou, Q. Niu, and J. liu, "Design of optical-acoustic hybrid underwater wireless sensor network," *J. Netw. Comput. Appl.*, vol. 92, pp. 59–67, Aug. 2017.

[7] Z. Zeng, S. Fu, H. Zhang, Y. Dong, and J. Cheng, "A survey of underwater optical wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 204–238, 1st Quart., 2017.

[8] N. Saeed, A. Celik, T. Y. Al-Naffouri, and M.-S. Alouini, "Underwater optical wireless communications, networking, and localization: A survey," *Ad Hoc Netw.*, vol. 94, pp. 101935–101969, Nov. 2019.

[9] P. Wang, D.-H. Fu, C.-Q. Zhao, J.-C. Xing, Q.-L. Yang, and X.-F. Du, "A reliable and efficient routing protocol for underwater acoustic sensor networks," in *Proc. IEEE Int. Conf. Cyber Technol. Autom., Control Intell. Syst.*, 2013, pp. 185–190.

[10] V. Rodoplu and M. K. Park, "An energy-efficient MAC protocol for underwater wireless acoustic networks," in *Proc. MTS/IEEE OCEANS*, 2005, pp. 1198–1203.

[11] Y. Wei and D.-S. Kim, "Reliable and energy-efficient routing protocol for underwater acoustic sensor networks," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, 2014, pp. 738–743.

[12] E. Kim, J. Kang, P. K. Chong, S. Yoo, and D. Kim, "Energy efficient local area source routing protocol of underwater sensor networks in the deep ocean," in *Proc. Int. Symp. Commun. Inf. Technol.*, 2007, pp. 948–953.

[13] M. Al-Bzoor, Y. Zhu, J. Liu, A. Reda, J.-H. Cui, and S. Rajasekaran, "Adaptive power controlled routing for underwater sensor networks," in *Wireless Algorithms, Systems, and Applications*, X. Wang, R. Zheng, T. Jing, and K. Xing, Eds. Berlin, Germany: Springer, 2012, pp. 549–560.

[14] R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Modeling the sleep interval effects in duty-cycled underwater sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2016, pp. 1–6.

[15] F. Zorzi, M. Stojanovic, and M. Zorzi, "On the effects of node density and duty cycle on energy efficiency in underwater networks," in *Proc. IEEE OCEANS SYDNEY*, 2010, pp. 1–6.

[16] R. W. Coutinho, A. Boukerche, L. F. Vieira, and A. A. Loureiro, "Modeling and analysis of opportunistic routing in low duty-cycle underwater sensor networks," in *Proc. 18th ACM Int. Conf. Model., Anal. Simulat. Wireless Mobile Syst.*, 2015, pp. 125–132.

[17] K. Y. Islam, I. Ahmad, D. Habibi, and A. Waqar, "A survey on energy efficiency in underwater wireless communications," *J. Netw. Comput. Appl.*, vol. 198, pp. 103295–103324, Feb. 2022.

[18] S. Q. Duntley, "Light in the sea," *J. Opt. Soc. America*, vol. 53, no. 2, pp. 214–233, 1963.

[19] N. Chi, Y. Zhao, M. Shi, P. Zou, and X. Lu, "Gaussian kernel-aided deep neural network equalizer utilized in underwater PAM8 visible light communication system," *Opt. Exp.*, vol. 26, no. 20, pp. 26700–26712, 2018.

[20] J. Zhang, M. M. Wang, T. Xia, and L. Wang, "Maritime IoT: An architectural and radio spectrum perspective," *IEEE Access*, vol. 8, pp. 93109–93122, 2020.

[21] R. Jiang, C. Sun, L. Zhang, X. Tang, H. Wang, and A. Zhang, "Deep learning aided signal detection for SPAD-based underwater optical wireless communications," *IEEE Access*, vol. 8, pp. 20363–20374, 2020.

[22] H. Lu, M. Jiang, and J. Cheng, "Deep learning aided robust joint channel classification, channel estimation, and signal detection for underwater optical communication," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2290–2303, Apr. 2021.

[23] S. Tang, Y. Dong, and X. Zhang, "On link misalignment for underwater wireless optical communications," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1688–1690, Oct. 2012.

[24] H. Zhang and Y. Dong, "Link misalignment for underwater wireless optical communications," in *Proc. Adv. Wireless Opt. Commun. (RTUWO)*, 2015, pp. 215–218.

[25] Z. Vali, A. Gholami, D. G. Michelson, Z. Ghassemlooy, M. Omoomi, and H. Noori, "Use of Gaussian beam divergence to compensate for misalignment of underwater wireless optical communication links," *IET Inst. Eng. Technol.*, vol. 11, no. 5, pp. 171–175, 2017.

[26] H. Lu, W. Chen, and M. Jiang, "Deep learning aided misalignment-robust blind receiver for underwater optical communication," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1984–1988, Sep. 2021.

[27] G. Mangano, A. D'Alessandro, and G. D'Anna, "Long term underwater monitoring of seismic areas: Design of an ocean bottom seismometer with hydrophone and its performance evaluation," in *Proc. IEEE OCEANS*, 2011, pp. 1–9.

[28] N. F. Farr, J. D. Ware, C. T. Pontbriand, and M. A. Tivey, "Demonstration of wireless data harvesting from a subsea node using a 'ship of opportunity'," in *Proc. IEEE OCEANS*, 2013, pp. 1–5.

[29] "BOLcom-LR." BORSYS. 2022. [Online]. Available: http://Inborsys.kr/ko/products

[30] "BlueComm 200." Sonardyne. 2022. [Online]. Available: https://www.sonardyne.com/products/bluecomm-200-wireless-underwater-link/

[31] C. M. G. Gussen, P. S. R. Diniz, M. L. R. Campos, W. A. Martins, F. M. Costa, and J. N. Gois, "A survey of underwater wireless communication technologies," *J. Commun. Inf. Syst.*, vol. 31, no. 1, pp. 242–255, 2016.

[32] "Popoto S1000." PopotoModem. 2023. [Online]. Available: https://www.popotomodem.com/products/

[33] "LUMA X-UV." HYDROMEA. 2023. [Online]. Available: https://www.hydromea.com/

[34] L. Bergdahl, *Wave-Induced Loads and Ship Motions*, Chalmers Univ. Technol., Göteborg, Sweden, 2009.

[35] G. Hibbert and G. Lesser, *Measuring Vessel Motion Using a Rapid Deployment Device on Ships of Opportunity*, OMC Int., Abbotsford, VIC, Australia, 2009.

[36] M. A. A. Ali, "Comparison of modulation techniques for underwater optical wireless communication employing APD receivers," *Res. J. Appl. Sci., Eng. Technol.*, vol. 10, no. 6, pp. 707–715, 2015.

[37] C. D. Mobley et al., "Comparison of numerical models for computing underwater light fields," *Appl. Opt.*, vol. 32, no. 36, pp. 7484–7504, 1993.

[38] A. Celik, N. Saeed, B. Shihada, T. Y. Al-Naffouri, and M.-S. Alouini, "End-to-end performance analysis of underwater optical wireless relaying and routing techniques under location uncertainty," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1167–1181, Feb. 2020.

[39] M. Elamassie, F. Miramirkhani, and M. Uysal, "Performance characterization of underwater visible light communication," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 543–552, Jan. 2019.

[40] M. Elamassie and M. Uysal, "Vertical underwater visible light communication links: Channel modeling and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6948–6959, Oct. 2020.

[41] S. M. Navidpour, M. Uysal, and M. Kavehrad, "BER performance of free-space optical transmission with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2813–2819, Aug. 2007.

[42] M. M. Hayat et al., "Gain-bandwidth characteristics of thin avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. 49, no. 5, pp. 770–781, May 2002.

[43] R. Hui, "Chapter 4—Photodetectors," in *Introduction to Fiber-Optic Communications*, R. Hui, Ed. Cambridge, MA, USA: Academic, 2020, pp. 125–154.

[44] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surveys*, vol. 52, no. 1, pp. 1–38, 2019.
[45] M. Elamassie, M. Uysal, Y. Baykal, M. Abdallah, and K. Qaraqe, "Effect of eddy diffusivity ratio on underwater optical scintillation index," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 34, no. 11, pp. 1969–1973, 2017.
[46] G. Hibbert and G. Lesser, *Measuring Vessel Motions Using a Rapid-Deployment Device on Ships of Opportunity*, OMC Int., Abbotsford, VIC, Australia, pp. 1–4, 2014.
[47] P. Naaijen, R. van Dijk, R. Huijsmanx, and A. El-Mouhandiz, "Real time estimation of ship motions in short crested seas," in *Proc. Int. Conf. Ocean, Offshore Arctic Eng.*, 2009, pp. 243–255.
[48] I. Romdhane and G. Kaddoum, "A reinforcement-learning-based beam adaptation for underwater optical wireless communications," *IEEE Internet Things J.*, vol. 9, no. 20, pp. 20270–20281, Oct. 2022.
[49] C. Li et al., "Dynamic offloading for multiuser Muti-CAP MEC networks: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2922–2927, Mar. 2021.
[50] S. A. H. Mohsan et al., "Investigating transmission power control strategy for underwater wireless sensor networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 281–285, 2020.
[51] F. Xing, H. Yin, Z. Shen, and V. C. M. Leung, "Joint relay assignment and power allocation for multiuser multirelay networks over underwater wireless optical channels," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9688–9701, Oct. 2020.

Huicheol Shin (Student Member, IEEE) received the master's degree in engineering from Korea Maritime and Ocean University, Busan, South Korea, in 2021. He is currently pursuing the Ph.D. degree with the Department of Marine Technology and Convergence Engineering, University of Science and Technology, Busan.

His research interests are the establishment and analysis of communication networks in marine/underwater and reinforcement learning.

Soo Mee Kim (Member, IEEE) received the Ph.D. degree in interdisciplinary programs in radiation applied life science major from Seoul National University, Seoul, South Korea, in 2010.

She was a Postdoctoral Researcher studying medical image processing with the Department of Nuclear Medicine, Seoul National University, from 2010 to 2012; with the Department of Radiology, University of Washington, Seattle, WA, USA, from 2012 to 2015; and with the Division of Radiation Instrument Research, Korea Atomic Energy Research Institute, Daejeon, South Korea, from 2015 to 2017. She has been a Principal Research Scientist with Maritime ICT and Mobility Research Department, Korea Institute of Ocean Science and Technology, Busan, South Korea, since September 2017. Her research interests include underwater sensing and signal/image processing for unmanned automation techniques.

Yujae Song (Member, IEEE) received the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2016.

He was a Visiting Scholar in communication systems from KTH Royal Institute of Technology, Stockholm, Sweden, in 2015. From 2016 to 2022, he was a Senior Researcher with the Maritime ICT Research and Development Center, Korea Institute of Ocean Science and Technology, Busan, South Korea. From 2022 to 2023, he was an Assistant Professor with the Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, South Korea. Since March 2023, he has been an Assistant Professor with the Department of Robotics Engineering, Yeungnam University, Gyeongsan, South Korea. His research interests include design, analysis, and optimization of various wireless communication systems, including 5G, maritime/underwater, and smart grid communications.