Towards an Intelligent Reliable Connectivity For Underwater Wireless Communication

by

Imene ROMDHANE

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, "JUNE 28, 2026"

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

M. Georges Kaddoum, Thesis supervisor
Electrical Engineering Department, École de Technologie Supérieure,

M. Stephane Coulombe, Chair, Board of Examiners
Software Engineering and IT Department, École de Technologie Supérieure,

M. Christian Desrosiers, Member of the Jury
Software Engineering and IT Department, École de Technologie Supérieure,

M. Jean-François Bousquet, External Independent Examiner
Department of Electrical and Computer Engineering, Dalhousie University

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON "JUNE 2, 2026"

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

FOREWORD

This dissertation is written based on the author's Ph.D. research outcomes under the supervision of Professor Georges Kaddoum from September 2019 to May 2026. The main theme of this dissertation focuses on the emerging topic of UOWC and presenting RL-based solutions to its related challenges. This dissertation is written as a monograph based on three published IEEE journal papers and one published conference paper as the first author.

In this dissertation, the first two chapters present the introduction and the background and literature review of UOWC. Next, the following four chapters are written based on the research articles. Finally, the conclusion and the recommendation for future work are given in the last chapter.

ACKNOWLEDGEMENTS

First, I sincerely thank my supervisor, Professor Georges Kaddoum, for his unconditional support throughout my Ph.D. It was indeed thanks to his supervision and the multiple collaboration opportunities he ensured me that I came to this stage on my research career. This thesis would not have been possible without his guidance.

I would also like to thank my collaborators Dr. Abdulkadir Celik, Dr. Ziyaur Rahman, Dr. Md. Zoheb Hassan, and my dear friend and colleague Nahed Belhadj Mohamed, for the great collaboration experiences we had together. I am deeply honored to have such fruitful collaborations with them. I learned a lot from them and they made research enjoyable.

Moreover, I would like to thank my dear husband Mohamed Said and my parents Abdelmajid Romdhane and Nejia Chtioui for being the great support they are to me throughout my Ph.D. This achievement would not have been possible without your unconditional love and support. I am lucky to have you in my life.

Furthermore, I would like to thank my close friends Ameni Mezni, Kanika Aggarwal, and Faten Guermazi, for being by my side during my hard times in Ph.D and life in general. Your support gave me strength to go forward and achieve my goal.

Finally, I would like to dedicate this thesis to my son Omar. I hope you grow up to be proud of your mama.

Vers une connectivité intelligente et fiable pour la communication sans fil sous-marine

Imene ROMDHANE

RÉSUMÉ

De nos jours, l'environnement sous-marin reçoit une attention croissante pour le déploiement des communications dans différents domaines. Ainsi, trouver un moyen robuste de communication sous-marine face aux conditions difficiles de l'environnement sous-marin devient crucial. Les communications acoustiques et par radiofréquence (RF) ont été initialement déployées pour les communications sous-marines. Cependant, les communications acoustiques souffrent d'une latence élevée et d'une largeur de faisceau limitée, tandis que les communications RF souffrent d'une consommation énergétique élevée et d'une atténuation sévère sous l'eau. En conséquence, les communications optiques sans fil sous-marines (UOWC) ont récemment été proposées comme un candidat prometteur. Grâce aux avantages remarquables des ondes optiques, tels que la haute vitesse de la lumière, les débits de données élevés et la faible consommation d'énergie, elles sont devenues un centre d'intérêt pour de nombreuses recherches. Pourtant, les UOWC souffrent de limitations qui peuvent dégrader la qualité de la communication, telles que des portées de communication courtes et des problèmes d'erreurs de pointage (PE).

L'objectif principal de cette thèse est de construire un réseau autonome auto-organisé (SON) pour les communications optiques sans fil sous-marines (UOWC) qui soit robuste contre les conditions sévères de l'environnement sous-marin. Le premier objectif est de maintenir la connectivité entre les nœuds en résolvant le problème de l'alignement des faisceaux. Le deuxième objectif est d'étendre la durée de vie du réseau en optimisant la consommation d'énergie des nœuds. Le troisième objectif est de résoudre le problème de l'équité dans les réseaux UOWC.

Dans ce cadre, nous proposons des solutions d'adaptation des faisceaux basées sur l'apprentissage par renforcement (RL) pour les communications optiques sous-marines point-à-point (P2P). Ces solutions optimisent la largeur du faisceau lumineux, l'orientation du faisceau, ou les deux. Ces solutions améliorent le taux de succès du lien optique et garantissent une meilleure qualité de lien en termes de rapport signal-bruit (SNR) pour quatre environnements sous-marins différents, y compris l'eau de mer pure, l'océan propre, l'océan côtier et le port trouble.

En plus, nous analysons la probabilité de défaillance pour les UOWC sur les turbulences océaniques avec des PEs. Un système UOWC combiné incluant un lien direct basé sur une combinaison à ratio maximal (MRC) et un lien indirect assisté par une surface réfléchissante optique intelligente (OIRS) est considéré. La distribution gamma exponentielle généralisée (EGG) est adoptée pour modéliser les turbulences océaniques. Le système proposé améliore la connectivité dans les UOWC à grande vitesse sous différentes conditions de canal sous-marin.

En outre, nous étudions une communication directe air-sous-marin P2P où un véhicule aérien terrestre sans pilote (UAV) communique directement avec un nœud sous-marin sans relais à la surface de l'eau. Plus précisément, nous présentons une technique d'adaptation de lien basée sur le gradient de politique déterministe profond (DDPG), en considérant la technique de multiplexage

par division de fréquence orthogonale (OFDM) dans des conditions de l'informations sur l'état du canal (CSI) bruyantes. Cette technique optimise la largeur du faisceau, l'orientation du faisceau et les niveaux de puissance de transmission de l'UAV, en traitant ces paramètres comme des variables continues pour tenir compte du caractère aléatoire inhérent aux scénarios réels. La méthode proposée prouve qu'elle maintient des taux de succès élevés tout en réduisant la consommation d'énergie par rapport à l'approche basée sur l'apprentissage Q.

Enfin, nous introduisons des solutions d'optimisation du faisceau basées sur DDPG dynamique pour les réseaux UOWC multi-faisceaux. L'accès multiple non orthogonal (NOMA) est adopté au sein de chaque faisceau pour soutenir les communications multi-utilisateurs. Cette technique optimise l'orientation du faisceau et l'allocation de puissance entre différents nœuds récepteurs, en tenant compte de la quasi-stationnarité des nœuds due à la dynamique sous-marine. Des solutions DDPG à agent unique et multi-agents séquentiels sont proposées dans ce chapitre, dans le but de maximiser l'efficacité énergétique (EE) du système tout en garantissant l'équité max-min.

Mots-clés: Communication optique sans fil sous-marine, communications directes air-sous-marine, apprentissage par renforcement, adaptation de liaison, qualité de liaison, allocation de puissance, efficacité énergétique, formation de faisceaux.

Towards an Intelligent Reliable Connectivity For Underwater Wireless Communication

Imene ROMDHANE

ABSTRACT

Nowadays, the underwater environment is receiving increasing attention for deploying communications in different fields. Thus, finding a robust mean of underwater communication against the harsh underwater environment conditions is becoming crucial. Acoustic and radio frequency (RF) communications were initially deployed for underwater communications. However, acoustic communications suffer from high latency and limited beamwidth, and RF communications suffer from high energy consumption and severe attenuation underwater. Consequently, underwater optical wireless communications (UOWC) were lately proposed as a promising candidate. Thanks to the remarkable advantages of optical waves, such as the high speed of light, high data rates, and low power consumption, it became a center of interest for many researches. Yet, UOWC suffer from limitations that can degrade the communication quality, such as short communication ranges and pointing errors (PEs).

The main objective of this thesis is to construct an autonomous self-organized network (SON) for UOWC that is robust against the harsh underwater environment. The first goal is to maintain the connectivity between the nodes by solving the beam's alignment problem. The second goal is to extend the network's lifetime by optimizing the nodes' power consumption. The third goal is to solve the fairness problem in UOWC networks.

In this vein, we first propose reinforcement learning (RL)-based beam adaptation solutions for UOWC point-to-point (P2P) communication. These solutions optimize the light beamwidth, beam orientation, or both. These solutions improve the optical link's success rate and guarantee better link quality in terms of signal-to-noise ratio (SNR) for four different underwater environments, including pure seawater, clean ocean, coastal ocean, and turbid harbor.

Then, we analyse the outage probability for UOWC over oceanic turbulence with PEs. A combined UOWC system including a maximum ratio combining (MRC)-based direct link and an optical intelligent reflecting surface (OIRS)-assisted indirect link is considered. The exponential-generalized gamma (EGG) distribution is adopted to model the oceanic turbulence. The proposed system proves to enhance the connectivity in high-speed UOWC under different underwater channel conditions.

Furthermore, we consider a direct air-to-underwater P2P communication where a terrestrial unmanned aerial vehicle (UAV) is communicating with an underwater node directly with no relay on the water surface. Specifically, we present a deep deterministic policy gradient (DDPG)-based link adaptation technique, considering the orthogonal frequency division multiplexing (OFDM) technique under noisy channel state information (CSI) conditions. This technique optimizes the beamwidth, beam orientation, and transmitting power levels from the UAV, treating these parameters as continuous variables to account for the inherent randomness in real-world

scenarios. The proposed method proves to maintain high success rates, while reducing the energy consumption compared to Q-learning based approach.

Finally, we introduce dynamic DDPG-based beamforming optimization solutions for multi-beam UOWC networks. Non-orthogonal multiple access (NOMA) is adopted within each beam to support multi-user communications. This technique optimizes the beam orientation and power allocation among different receiving nodes, considering node's quasi-stationarity due to underwater dynamicity. Both a single-agent and a sequential multi-agent DDPG-based solutions are proposed in this chapter, in the aim of maximizing the system's energy efficiency (EE) while guaranteeing the max-min fairness.

Keywords: Underwater optical wireless communication, direct air-to-underwater communications, reinforcement learning, link adaptation, link quality, power allocation, energy efficiency, beamforming.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

2D	Two-dimensional
ACPS	Alternating composite phase shift
AF	Amplify-and-forward
AP	Access point
AUV	Autonomous underwater vehicles
AWGN	Additive white Gaussian noise
BA	Beamwidth adaptation
BER	Bit error rate
BOA	Beamwidth and orientation adaptation
BL	Bubble level
CDF	Cumulative distribution function
CM	Cox and Munk
CNN	Convolutional neural network
CSI	Channel state information
DDPG	Deep deterministic policy gradient
DF	Decode-and-forward
DNN	Deep neural network
DQN	Deep Q-learning
DRL	Deep reinforcement learning

EE	Energy efficiency
EW	Exponential Weibull
EGG	Exponential-generalized gamma
FOV	Field of view
GG	Generalized gamma
IM/DD	Intensity modulation/direct detection
i.ni.d	independent and non-identically distributed
ISI	Inter-symbol interference
K	Kelvin
LEDs	Light-emitting diodes
MAC	Medium access control
MGFs	Moment-generating functions
MARL	Multi agent reinforcement learning
MC	Monte-Carlo
MDP	Markov decision process
MIMO	Multiple-input-multiple-output
MISO	Multiple-input-single-output
MRC	Maximum ratio combined
NOMA	Non-orthogonal multiple access
NLoS	Non-line of sight

OA	Orientation adaptation
OBS	Optical base station
OFDM	Orthogonal frequency division multiplexing
OIRS	Optical intelligent reflecting surface
OOK	On-off-keying
OWC	Optical wireless communication
P2P	Point-to-point
PAPR	Peak-to-average power ratio
PAT	Pointing, acquisition, and tracking
PDF	Probability density function
PEs	Pointing errors
QoS	Quality of services
RF	Radio Frequency
RL	Reinforcement learning
RSMA	Rate splitting multiple access
SC	Selection combining
SD	Spatial diversity
SDMA	Space division multiple access
SE	Spectral efficiency
SIC	Successive interference cancellation

SINR	Signal-to-interference-plus-noise ratio
SLIPT	Simultaneous lightwave information and power transfer
SNR	Signal-to-noise ratio
SON	Self-organized network
TD	Temporal difference
UAV	Unmanned aerial vehicle
UACN	Underwater acoustic communication networks
UOWC	Underwater optical wireless communications
UOWSNs	Underwater optical wireless sensor networks
UWSN	Underwater wireless sensor networks
VLC	Visible light communication
VLCP	Visible light communication and positioning
WMMSE	Weighted minimum mean square error
WMMSE-EE-MinRate	WMMSE-based EE maximization with min-rate constraint
QT	Quadratic Transform
QT-EE-MinRate	QT-based EE maximization with min-rate constraint
ZF	Zero-forcing

INTRODUCTION

0.1 Context and Motivation

In the last decade, underwater wireless communications have received a considerable scholarly attention. Considering that 70% of earth surface is covered with water, this type of communication became crucial in many fields, such as the remote control of instruments in oceans, pollution monitoring, oceanography research, oil-controlling, and military and surveillance applications. Many of these applications require real-time transmission, high data rates, and energy-efficiency. Considering the numerous novel applications based on underwater communications, substantial effort has been invested into investigating the optimal way to transmit information under the sea. Of note, because of the differences between marine and terrestrial environments, communication techniques designed for atmospheric communications cannot be adopted in underwater scenarios. Accordingly, new types of communication and new challenges were defined. For instance, acoustic and radio frequency (RF) waves were initially used for underwater communications. However, limitations of these approaches are that acoustic waves suffer from high latency and limited bandwidth, while RF communications suffer from high energy consumption and severe attenuation underwater. Hence, in order to address the aforementioned limitations, underwater optical wireless communications (UOWC) were recently introduced.

0.2 UOWC for Future Underwater Networks

UOWC are based on exchanging information in marine environment using light, where blue-green wavelengths are often preferred due to their relatively lower attenuation in water. UOWC offer various advantages, including high data rates in the order of Gbps, high bandwidth, enhanced security due to narrow optical beams, cost-effectiveness, and low latency due to the high speed of light (approximately $2.55 \times 10^8 m/s$) (Kaushal & Kaddoum (2016); Celik *et al.* (2022)).

Despite the remarkable advantages of UOWC, this technology faces several challenges. While propagating underwater, the light beam undergoes high water absorption and scattering effects, which limits communication range to approximately 50 to 100 meters. Furthermore, light directivity requires a good alignment of the light beam to ensure a successful communication. This is a major challenge in the context of the dynamic underwater environment, including turbulence and sudden movements due to waves or coexisting entities, which may lead to pointing errors (PEs). Furthermore, underwater nodes are usually fed by batteries that are expensive to charge or replace; accordingly, an optimized energy consumption to extend the network lifetime is needed. Moreover, inter-node interference and fairness are key challenges to consider in the case of UOWC networks. Hence, resource allocation among users should be optimized to improve the reliability and fairness of the network.

Considering the dynamicity existing underwater, fixed transmission strategies are often insufficient. Future UOWC networks require self-adaptive mechanisms capable of dynamically adapting beam parameters, transmission power, and resource allocation according to the current channel conditions, node distribution, and service requirements. This thesis aims to contribute to this direction by proposing intelligent reliable connectivity solutions for UOWC systems under various channel and resource constraints.

0.3 Problem Statement

Taking the aforementioned challenges into account, this thesis investigates how to improve the reliability, fairness, and efficiency of UOWC links and networks. The four main problems addressed in this thesis can be summarized as follows:

- **Problem 1: Reliability in UOWC links under channel impairments:** UOWC links face major challenges, namely absorption, scattering, oceanic turbulence, background noise, and pointing errors. These impairments lead to signal fading, limits the received power, and increases outage probability.

- **Problem 2: Beam misalignment and pointing errors:** Optical beams are directional. Such characteristic can be useful for highly confidential information transmission. However, it makes UOWC links highly sensitive to transmitter-receiver misalignment and pointing errors, especially considering the dynamic marine environment. To this end, the beamwidth and beam orientation should be well defined in order to guarantee reliable communications. While considering narrow beams enhances the received power and increases the link range, considering wide beams improves the coverage and robustness of the link. Hence, optimizing the range-beamwidth tradeoff under dynamic environment is an important problem to consider in constructing reliable UOWC links.
- **Resource allocation and interference in multi-user UOWC networks:** UOWC networks can include multiple underwater nodes served simultaneously. This raises the resource allocation and inter-node interference management problems. NOMA is a well-known multi-access technique that ensures a simultaneous access to the network to multiple devices. To ensure better performance of the network, the beam adaptation and power allocation should be optimized in order to maximize network's fairness and energy efficiency, under dynamic network condition.

0.4 Thesis Objectives

The main objective of this thesis is to construct a SON for UOWC robust against the challenging marine environment. This objective can be split into the following four sub-objectives as listed below.

1. UOWC system's connectivity: The first sub-objective is to maintain connectivity between nodes. As explained above, the directivity characteristic of light introduces misalignment problems and PEs. This can prove to be fatal to the system, as it causes information loss

and nodes disconnection from the system. Therefore, by performing beam alignment, we aim to maintain the connectivity within the UOWC system.

2. Turbulence and PEs effects on UOWC: Two of the main challenges faced in UOWC are the oceanic turbulence and PEs. These two challenges can significantly degrade performance of such systems and can even prove to be fatal in some cases. Accordingly, under such conditions, it is important to enhance performance of UOWC systems. To this end, different techniques to overcome these issues were proposed in literature, including the MRC scheme and the introduction of OIRS as a relay. Consequently, we aim to evaluate performance of such techniques under oceanic turbulence and PEs.
3. UOWC system's lifetime: The UOWC nodes are usually fed with batteries that are difficult to recharge or replace. Hence, in order to extend the battery function before it gets exhausted, a good power management is needed. For that reason, we aim to extend the network's lifetime by optimizing nodes' power consumption.
4. Fairness in UOWC: The UOWC fairness is an important point that needs to be factored in to avoid resource starvation or redundant allocation within the network. Accordingly, we aim to address the fairness problem in UOWC networks to improve its performance.

0.5 Thesis Contribution

To reach these objectives in this thesis, we use RL techniques that are assumed to be capable of efficiently solving problems in dynamic environments like the marine environment. Accordingly in this thesis, we propose novel solutions to reach the aforementioned sub-objectives. The structure of thesis is summarized in Fig. 0.1 and discussed in further details in this section. Chapters 3-6 outline the four articles published throughout the period of my Ph.D. program.

First, we present novel RL-based beam adaptation methods for P2P UOWC system. These methods can be used to meet the first sub-objective, or maintaining connectivity between two

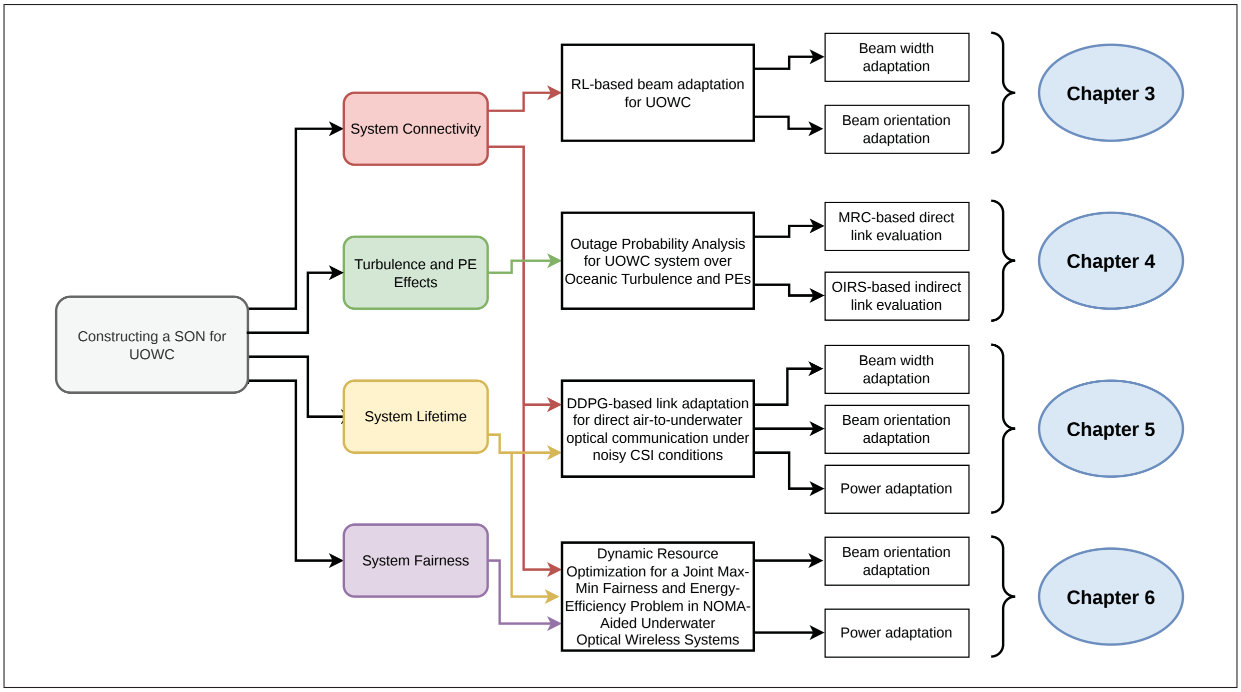


Figure 0.1 Thesis structure

underwater nodes. This is achieved by adapting beam width, beam orientation, or both, using Q-learning and SARSA algorithms. As revealed by the results, the proposed methods prove to improve succes rate of the communication link and SNR as compared to the benchmark for different underwater environments.

Second, we evaluate the performance of UOWC system over oceanic turbulence and PEs. In this work, we focus on the second sub-objective, where the effect of turbulence and PEs is evaluated on two types of UOWC links -namely, the MRC-based direct link and OIRS-based indirect link. These two techniques are widely used underwater to overcome the problems of connectivity and information loss in the presence of obstacles and dynamicity of the marine environment. In this case, the EGG distribution is adopted to model oceanic turbulence; outage probability is taken as the assessment metric.

Furthermore, for direct air-to-underwater optical communication under noisy CSI conditions, we introduce a DDPG-based link adaptation method . This method focuses on the first and third sub-objectives, i.e. system connectivity and lifetime. Unlike in Chapters 3-4, in this case, the link is not fully underwater. However, one UAV situated in terrestrial environment is communicating with an underwater node. Therefore, further challenges are taken into consideration, such as the atmospheric and near-the-surface fog attenuation, and power loss at the air-water interface. The link adaptation in this method includes beam width, beam orientation, and power adaptation. As demonstrated by the results, the proposed method maintains high success rates, while remarkably extending the system lifetime as compared to Q-learning-based benchmark solution.

Finally, we treat the joint max-min fairness and energy efficient resource optimization in NOMA-aided UOWC network. This work integrates the first, third, and fourth sub-objectives, by maintaining system connectivity, optimizing EE that directly effects system lifetime, and guarantees system fairness. Unlike the previous chapters that focused on P2P communications, this work covers the case of an UOWC network that includes multiple quasi-stationary nodes grouped into clusters and receiving information from a multi-beam transmitter. The proposed solutions include beam orientation adaptation to maintain connectivity, and power allocation between nodes to optimize EE. The proposed solutions are found to improve fairness and EE as compared to conventional benchmarks.

0.6 Thesis Outline

The organization of this thesis, which consists of six chapters, is structured as follows. Chapter 1 presents the technical background and literature review necessary for our work. We first introduce the basics of UOWC and its related challenges. This is followed by a discussion of the classic RL, deep reinforcement learning (DRL), and multi-agent reinforcement learning (MARL) algorithms previously proposed in the literature.

In Chapter 2, we present the thesis objective and related sub-objectives, as well as explain the thesis contribution and interconnection between different chapters.

In Chapter 3, we present the first article focused on the beam adaptation in UOWC, that proposes novel beam adaptation methods based on RL for point-to-point (P2P) UOWC. The first method aims to optimize the light beamwidth; the second focuses on adapting beam orientation, whereas the last one optimizes both the light's beamwidth and beam orientation. The proposed RL-based solutions yield optimal positioning and beamwidth of the light source and improve the success rate of considered communication link. They also ensure better link quality in terms of signal-to-noise ratio (SNR) as compared to the uncertainty disk static method for four different underwater environments -namely, pure seawater, clean ocean, coastal ocean, and turbid harbor.

In Chapter 4, we then present the second article analyzing the outage probability for UOWC systems over oceanic turbulence with PEs. In this article, we propose a combined UOWC system consisting of maximum ratio combining (MRC)-based direct link and optical intelligent reflecting surface (OIRS)-assisted indirect link. Oceanic turbulence is modeled by the mixture exponential–generalized gamma (EGG) distribution. We also evaluate system performance using outage probability. An exact outage probability expression is derived for the links under EGG oceanic turbulence and PEs. Moreover, the asymptotic expression for outage probability is derived to present an analytical expression of the system's diversity order. The results demonstrate that the Monte-Carlo (MC) simulation results perfectly match the derived expressions and closely align with the asymptotic expression at high SNR values. Based on this evidence, we argue that the proposed system ameliorates the connectivity of the high-speed UOWC under different underwater channel conditions.

In Chapter 5, we present the third article focusing on link adaptation for direct air-to-underwater optical communications, using the deep deterministic policy gradient (DDPG) algorithm and the orthogonal frequency division multiplexing (OFDM) technique under noisy channel state information (CSI) conditions. The proposed method encompasses dynamic adaptations of beam orientation, beam width, and power levels; to account for the inherent randomness in real-world scenarios, these parameters are treated as continuous variables. To evaluate the effectiveness of the proposed approach, we compare the achieved success rates and consumed energy with our DDPG-based model against a Q-learning-based link adaptation strategy, as well as two static benchmarks. According to the results, the proposed model consistently achieves high success rates, nearly approaching 100%, while concurrently reducing energy consumption by a factor of 10 compared to the Q-learning-based approach, all within the first 100 episodes.

In Chapter 6, we present the fourth article focusing on resource allocation to optimize fairness and energy efficiency (EE) in non-orthogonal multiple access (NOMA)-aided UOWC. In this publication, a dynamic beamforming optimization for multi-beam UOWC is proposed. The UOWC transmitter concurrently transmits multiple beams, and NOMA is applied within each beam to support multi-user communications. The aim is to maximize the system EE of a multi-beam UOWC network while guaranteeing the max-min fairness. To this end, seeking to optimize beam orientations and power allocation and considering nodes' quasi-stationarity in the underwater environment, we propose two DDPG-based beamforming solutions. The first solution is a single-agent DDPG-based approach, while the second solution is a multi-agent DDPG-based one. We also introduce the sequential feature to the multi-agent approach to improve the results, including sequential learning of beam orientation and power allocations tasks.Through extensive simulations, we find that the proposed single and multi-agent DDPG solutions achieve improved fairness and EE as compared to the equal power allocation, weighted minimum mean-square error (WMMSE)-based EE maximization with min-rate constraint (WMMSE-EE-MinRate), and QT-based EE maximization with min-rate constraint (QT-EEMinRate) benchmarks. Moreover, the multi-agent DDPG is also found to outperform the single-agent DDPG solution.

Finally, the last chapter concludes this thesis, lists the different publications published throughout the period of the present Ph.D. period, and discusses further research directions.

CHAPTER 1

BACKGROUND AND LITERATURE REVIEW

In this chapter, we will explain the main tools and concepts used in this project. We will start by defining different underwater communication types and review the challenges faced by each type, followed by a brief presentation of RL fundamentals, its main algorithms, and their use cases in wireless communications-related literature.

1.1 Underwater Wireless Communications Technologies

The waves used for underwater wireless communications can be divided into the following three types: Acoustic, RF, and optical. In what follows, we discuss these three types of waves in more detail.

1.1.1 Acoustic Waves

Acoustic waves are the first type of communication proposed for underwater environments. This is due to their low absorption and high coverage area, amounting up to tens of kilometers Saeed *et al.* (2019a). Yet, acoustic waves suffer from high latency and low bandwidth, which makes them unable to maintain a high quality of services (QoS) and unsuitable for large data transmission, as their achievable data rates are in the order of kbps. Moreover, acoustic nodes are quite expensive and energy-consuming, which makes these networks' cost of deployment a burden. Finally, acoustic waves can disturb marine mammals like dolphins and whales. The acoustic waves are suitable for long-range underwater communications. But considering their limited bandwidth and high latency, other complementary technologies for high-speed communications should be investigated.

1.1.2 RF Waves

RF waves are an alternative that can ensure a larger bandwidth and improve data rates (Kaushal & Kaddoum (2016)). In addition, RF communications can ensure faster transmissions, since they are less sensitive to turbulence and turbidity effects of underwater environments (Saeed *et al.* (2019a)). Yet, RF waves can only be used in shallow water, since they are sensitive to water conductance and absorption. Moreover, this technology cannot be used for high data rate transmissions, as it is limited to extremely low frequency bands. Finally, RF antennas for underwater communications are usually huge, expensive, and induce significant transmission losses, which makes their deployment difficult and power-consuming.

1.1.3 Optical Waves

In recent years, optical waves have attracted significant attention due to their numerous advantages (Saeed *et al.* (2019a); Celik *et al.* (2022)). First, optical waves ensure higher data rates over short distances, i.e in the order of tens of meters. Furthermore, data rates can reach several Gbps in clear water with little to no scattering. Second, owing to the fast propagation of light, latency in such networks is significantly lower than in acoustic and RF-based networks. Moreover, the small size and low cost of optical nodes makes their deployment much easier, while the use of energy conservative lasers and photodiodes significantly reduces energy consumption. Furthermore, optical waves improve physical-layer security due to the use of narrow light beams. Finally, optical waves are environmentally friendly and do not harm underwater species. UOWCs are suitable for real-time AUV communications, underwater image/video transmission, and short to medium-range underwater IoT links. Yet, optical waves face absorption, scattering, oceanic turbulence, and background noise, which may lead to multi-path fading and limits the communication range. They are also sensitive to misalignment problems leading to pointing errors. These challenges will be further studied in the following sections.

Table 1.1 summarizes the aforementioned advantages and limitations of different underwater wireless communication technologies. These technologies are complementary rather than mutually exclusive. While acoustic waves can support long range and low rate communications, optical waves are more suitable for high bandwidth and fast short-range communications. Hence, UOWC is particularly attractive for next-generation's real-time and high data-rate applications.

Table 1.1 Different underwater wireless communications

	Advantages	Limitations
Acoustic waves	-Low absorption -Long coverage	-High latency -Low bandwidth -Low data rates -Low QoS -High cost -High energy consumption
RF waves	-Large bandwidth -Fast transmission	-Restricted to shallow waters -Low data rates -Huge and expensive antennas -High energy consumption
Optical waves	-High data rates (Gbps) -Fast tranmission -Small size -Low cost -Low energy consumption	-Low coverage -Misalignment problems -Multi-path fading

1.2 Fundamentals of UOWC

In this section, we present the basics on UOWC systems, including the link architecture, the optical channel characteristics, and discussing key network structures such as the OIRS-assisted communications and direct air-to-underwater communications.

1.2.1 UOWC Link Architecture

UOWC uses light to transmit information between two underwater devices. The transmitter is usually equipped with laser diode or LED, while the receiver is equipped with a photodiode, APD, or optical detector. The information is modulated and transmitted in the form of optical beam through the underwater channel to the receiver, where the detection is performed. Here we can state two type of links: the line-of-sight, where the light reaches the receiver directly without being blocked, and the non-line of sight where the light is reflected, by an OIRS for example, before reaching the receiver.

UOWC systems can include one transmitter and one receiver, known as P2P communication, and can be extended to a network. These networks include underwater sensors, multiple AUVs, underwater relay nodes, optical intelligent reflection surfaces (OIRS), and surface stations communicating with submerged nodes (see Fig. 1.1). Optical communications can also be used to interconnect underwater devices with terrestrial ones using a direct air-to-underwater optical links where these devices directly interconnect using light without any relay on the water surface.

The performance of UOWC systems is commonly assessed using link-level and network-level performance metrics. First, the optical received power is the amount of power collected at the receiver side after traversing the underwater channel. It is strongly dependent on the transmitted power, beam parameters such as beam width and beam orientation, propagation distance, receiver's aperture and field-of-view, channel attenuation, and pointing losses. Second, the SNR that evaluates the strength of the received signal compared to noise. Outage probability is another important performance metric for UOWC systems that quantifies the probability that the received SNR falls below a threshold. Moreover, the achievable rate measures the maximum reliable data rate supported by the UOWC link under givn channel conditions. Finally, energy efficiency is a key metric for energy-constrained underwater networks, which is defined as the ratio of useful energy output over total energy input. Hence, to ensure a reliable UOWC system, we should not only maximize the received power, SNR, and achievable rate, but also reduce the outage probability while maintaining efficient energy consumption.

1.2.2 Direct air-to-underwater optical communication

Optical communications can be integrated in direct air-to-underwater systems, where a terrestrial node communicates with an underwater node using light, with no relay on the surface. In such communication, the optical signal travels two different media, namely the air and water. Such communication is more challenging than the UOWC, as it undergoes challenges related to atmospheric impairments including fog, rain and pointing errors, challenges at the air-to water-interface including refraction and multi-path interference, and the underwater challenges previously stated. Reaching the air-water surface, part of the light is reflected while the rest is

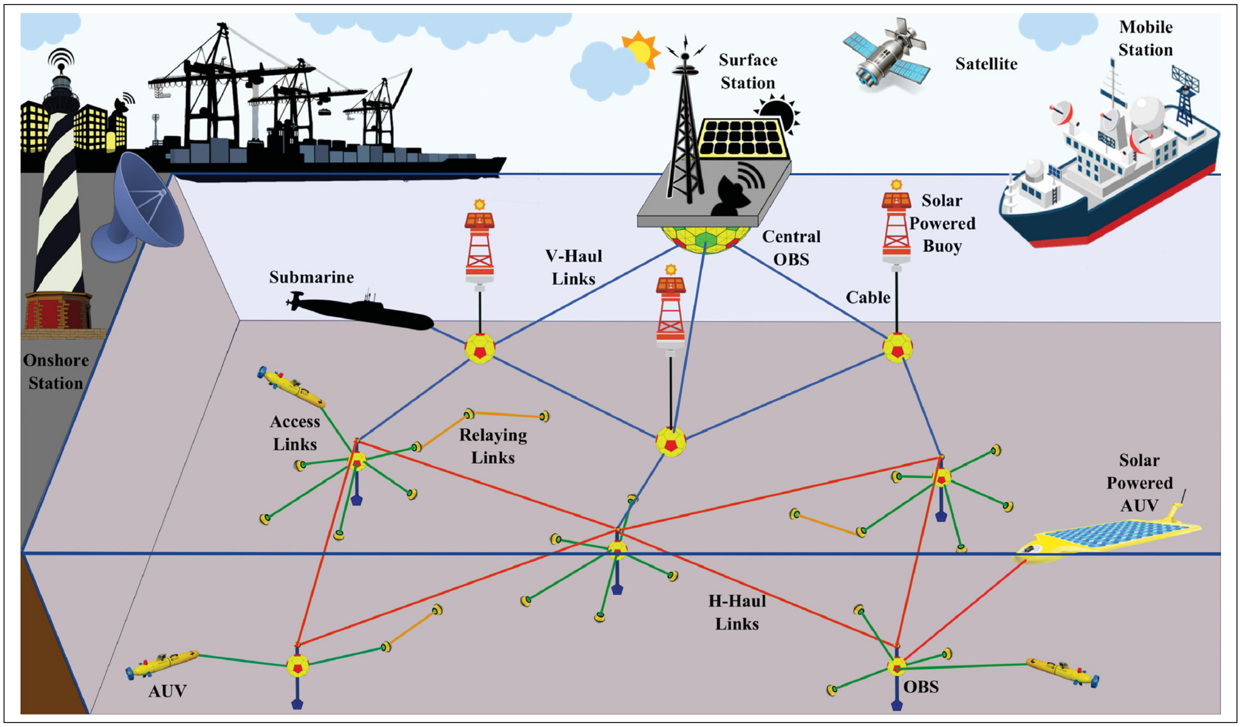


Figure 1.1 Illustration of a generic underwater optical wireless network (UOWN) architecture. Taken from: Saeed *et al.* (2019a)

refracted and transmitted through water according to snell's law. The performance of the direct air-to-underwater link depends strongly on the surface conditions, wavelength, and sea-surface dynamics. Surface waves in practical environments introduce misalignment, beam spreading, and fluctuations in the received optical power.

1.2.3 Optical Intelligent Reflecting Surface (OIRS)

In recent years, the OIRS has emerged as a new solution for channel blockage and obstacles in the aim to enhance underwater connectivity. The OIRS mainly consists of a planar array of reflecting surfaces forwarding the signal in a specific direction. This technology was initially introduced for RF communications. However, it was recently proved to be equally convenient for optical communications, by improving system coverage and reliability. Considering the fact that underwater optical beams are directional and that the light cannot penetrate through objects like RF waves, this increases the risk of signal blockage and misalignment. Hence, OIRS comes as a potential solution that creates a non-line of sight link as a backup to the main link. Eventually, OIRS provides an additional degree of freedom for optical beam adaptation, but beam reflection and alignment needs to be carefully modeled under underwater channel impairments to ensure a practical integration of OIRS.

1.3 Challenges in UOWC

Since water imposes its own optical properties, the underwater world is very different from the terrestrial one. The properties of the underwater environment can be divided into he following two categories: inherent and apparent. While inherent optical properties depend solely on the medium, apparent optical properties depend on both the medium and the geometric structure of light. In what follows, we define and describe the main factors affecting UOWC.

1.3.1 Absorption and Scattering

Absorption is the loss of intensity of light during propagation, while scattering is the change in the light's direction during propagation. These two phenomena are shown in Fig. 1.2, where an element volume of water is modeled by a cube of dimension Δr and volume ΔV. When an incident light of power P_i and wavelength λ goes through it, a small fraction of this light is absorbed by water (P_a), and another fraction is scattered (P_s). The rest, P_t, remains unaffected (Zeng, Fu, Zhang, Dong & Cheng (2017)).

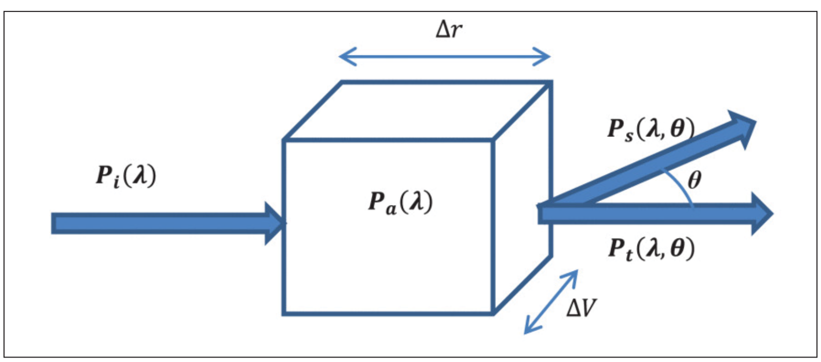


Figure 1.2 Geometry of inherent optical property. Taken from: Kaushal & Kaddoum (2016)

The absorption and scattering effects attenuates the transmitted power exponentially with distance, which significantly limits the communication range. This makes the UOWC suitable for short range communications. While absorption represents the dominating loss factor in pure sea water or in clear oceans, scattering is more important in coastal sea. In addition, the optimum wavelength window for a minimum absorption in open oceans is the blue-green band; however, in coastal waters, it is the green-yellow band. Accordingly, it is advised to use a multi-wavelength adaptive scheme for transmission in wavelength dependent underwater environments.

1.3.2 Oceanic Turbulence

Turbulence in water is caused by quick variations of the medium's refractive index due to fluctuations in density, salinity, temperature, and air bubbles. This leads to large fluctuations in the intensity of the propagating signal, known as scintillation, and temporal channel variations (Kaushal & Kaddoum (2016)). Oceanic turbulence affects the UOWC performance by increasing the outage probability and reducing the reliability of the system.

1.3.3 PEs and Misalignment Problems

Due to the narrow beamwidth of light and in order to achieve connectivity, transmitting and receiving nodes have to face each other. Yet, dynamicity of water due to water currents, wave motion, and mechanical imperfections, makes it challenging to guarantee a continuous connectivity, which introduces PEs and misalignment problems. Here, two main problems can occur -namely, bore-sight and jitter. While bore-sight is a fixed displacement between the center of the beam and the detector's center, jitter is a random dislocation of the beam center at the detector potentially caused by oceanic turbulence, deep currents, or random movements of the sea surface (Saeed *et al.* (2019a)).

Unlike acoustic waves, optical beams are highly directional. Hence, any small angular deviation can cause severe power loss, link interruption, and reduced coverage probability. The PEs and misalignment effects are strongly dependent on the beam divergence angle, link range, and receiver aperture size. To ensure reliable communication, beam steering and adaptive beamwidth control are required to mitigate PEs and misalignment effects.

1.3.4 Background Noise

A propagating light can be modified by background noise in the medium, mainly coming from ambient optical radiation and receiver electronics. Background noise depends on multiple factors. For example, solar noise decreases in high depths, at night and in unclear waters. Also, wider receiver FOVs and wider filter bandwidths collect more noise. Furthermore, background noise is more highlighted in blue-green wavelengths because the transmitted light may overlap with penetrating solar radiation. Finally, the background noise is strongly dependent on the photodetector type, as each type have a different noise sensitivity. The noise is usually assumed to follow a Gaussian distribution, i.e. the sum of contributions of shot noise and thermal noise (Komine & Nakagawa (2004)). Background noise leads to lower SNR, higher outage probability, and shorter communication ranges. Practical mitigation techniques include narrowband optical filtering, smaller field of view at the receiver side, and robust modulation and coding techniques.

1.3.5 Inter-node Interference

Inter-nodes interference is an important challenge in multi-user UOWC systems where interference occurs between different optical links. With an increase of underwater devices, this challenge becomes more salient, as UOWC systems are becoming denser. Inter-nodes interference reduces the network's throughput and introduces fairness issues. To mitigate the interference effect on system performance, in previous research, many multiple access schemes were proposed, such as the space division multiple access (SDMA), the NOMA, and the rate splitting multiple access (RSMA) (Qian, Chen, Du & Liu (2023); Elsayed *et al.* (2022); Rahman, Romdhane & Kaddoum (2024b); Rahman, Hassan & Kaddoum (2024a)). Along with adopting these schemes, optimizing the beamforming and power allocation is important to maintain good performance of the system.

1.3.6 Multipath Interference and Dispersion

Multipath interference and dispersion occurs when the signal reaches the receiver after undergoing reflections on different objects underwater. This results in time dispersion that decreases the data rate due to inter-symbol interference (ISI). The multipath interference effect is dependent on the water depth, surface roughness, receiver field-of-view, and link distance. Several mitigation techniques are proposed in literature, such as OFDM, equalization techniques, and narrow FOV receivers.(Kaushal & Kaddoum (2016)).

1.3.7 Range/Beamwidth Tradeoff

The range/beamwidth tradeoff is one of the main dilemmas in UOWC. This tradeoff is mainly caused by the low range of optical nodes. On the one hand, opting for directive light can be a solution to the issue of increasing the range, and hence network connectivity. However, this may lead to a total loss of signal due to misalignment problems. On the other hand, choosing an omnidirectional light source solves the misalignement problem, but significantly decreases the light range. Therefore, an optimal range/beamwidth tradeoff should be found to ensure reception of the light source (Saeed *et al.* (2019a); Derakhshandeh, Pachnicke & Hoeher (2023)).

The achievable range of UOWC systems are strongly dependent on the real-world scenario. In practice, they depend on the water type, optical wavelength, transmitted optical power, beam divergence, receiver sensitivity, field of view, alignment accuracy, ambient light level, and required link reliability. In clear or moderately clear water, optical communications ensure high data-rate and low-latency communication over short-to-moderate distances. However, in coastal or turbid water, the absorption and scattering affects are highlighted, which limits the communication range. Therefore, UOWC generally complements the acoustic communication: While acoustic links are suitable for long-range and low-rate underwater connectivity, optical links are deployed for short-range, high data-rate, and low-latency communication.

1.3.8 Limited Energy Budget

Due to the corrosive nature of sea water and the expensive cost of battery replacement, generating energy for underwater networks is a challenging task. Accordingly, energy consumption is an important factor to optimize in order to extend network lifetime and maintain a reliable communication. To this end, several energy-aware techniques were proposed in literature to improve the performance of the network. For instance, sleep scheduling technique allows sensor nodes to alternate between active and sleep modes to reduce circuit-level energy consumption. Another technique is to efficiently distribute the available transmit power among nodes in the network according to channel conditions, interference levels, and QoS requirements. Furthermore, simultaneous lightwave information and power transfer (SLIPT) has attracted attention recently as a promising energy harvesting solution by using light signals for both information transmission and energy harvesting (Thai, Nguyen, Le & Pham (2025); Li, Shang, Kong & Tang (2024a)).

1.3.9 Practical Feasibility of Underwater Optical Wireless Communication

Although UOWC is often presented from a theoretical channel-modeling perspective, its practical feasibility is strongly linked to the availability of suitable optical transmitters, sensitive photodetectors, and robust alignment mechanisms. The transmitter is usually equipped with light emitting diodes (LEDs) and laser diodes. LEDs are characterized by low-cost, simple driving circuits, eye-safety advantages, and relatively wide beamwidth that alleviates the pointing error and misalignment problems. However, they suffer from limited modulation bandwidth and transmission range compared to laser-based sources. In contrast, laser diodes (LDs) ensure higher optical power, narrower beamwidth, and larger modulation bandwidth. This makes it more suitable for high data rate applications and longer range communications. Yet, the narrow beam characteristic of LDs highlights the misalignment and pointing error problems. , especially when the transmitter or receiver is mounted on a mobile underwater platform. In both cases, The best spectral region for both light sources is the blue-green spectral region, since seawater exhibits relatively low attenuation in this band (Kaushal & Kaddoum (2016)).

Table 1.2 COMMERCIAL OWC-IOWT PRODUCTS. Taken from Celik *et al.* (2022)

<table>
<tr><th>Company</th><th colspan="2">Product</th><th>Ref.</th><th>Size [cm]</th><th>Weight [g]</th><th>Depth [m]</th><th>Range [m]</th><th>Rate [bps]</th><th>Tx/Rx Type</th><th>Power Tx. [W]</th><th>Wavelength [nm] (Color)</th><th>Angle Tx/Rx</th><th>Comm. & Netw. Interface</th></tr>
<tr><td rowspan="4">Hydromea</td><td rowspan="4">LUMA</td><td>100</td><td>[339]</td><td rowspan="3">10 × 5 × 3</td><td rowspan="3">250</td><td rowspan="4">6000</td><td>2</td><td>100 K</td><td rowspan="4">LED/PD</td><td>1–2</td><td rowspan="4">N/A (B)</td><td rowspan="4">60/60</td><td rowspan="4">RS 232
RS 485
Ethernet</td></tr>
<tr><td>250LP</td><td>[340]</td><td>7</td><td>250–600 K</td><td>2–5</td></tr>
<tr><td>500ER</td><td>[341]</td><td>> 50</td><td>500 K</td><td>2–5</td></tr>
<tr><td>X</td><td>[342]</td><td>10 × ⌀6</td><td>650</td><td>> 50</td><td>10 M</td><td>2–5</td></tr>
<tr><td rowspan="6">Sonardyne</td><td rowspan="6">BlueComm</td><td>100</td><td>[343]</td><td>24 × ⌀12</td><td>5200</td><td rowspan="6">4000</td><td>15</td><td>1–5 M</td><td>LED/PD</td><td>6</td><td>450 (B)</td><td>60/60</td><td rowspan="6">DL/UL,
TDMA,
UDP,
TCP/IP,
Ethernet</td></tr>
<tr><td>200 (Tx)</td><td rowspan="2">[344]</td><td>20 × ⌀14</td><td>3600</td><td rowspan="2">150</td><td rowspan="2">2.5–10 M</td><td>LED</td><td>6</td><td rowspan="4">400–800
(B/W)</td><td rowspan="2">Omni D.</td></tr>
<tr><td>200 (Rx)</td><td>38 × ⌀14</td><td>7100</td><td>PD</td><td>N/A</td></tr>
<tr><td>200UV (Tx)</td><td rowspan="2">[345]</td><td>20 × ⌀14</td><td>3600</td><td rowspan="2">75</td><td rowspan="2">2.5–10 M</td><td>LED</td><td>6</td><td rowspan="2">Omni D.</td></tr>
<tr><td>200UV (Rx)</td><td>38 × ⌀14</td><td>7100</td><td>PD</td><td>N/A</td></tr>
<tr><td>5000</td><td>[346]</td><td>N/A</td><td>N/A</td><td>7</td><td>500 M</td><td>LD/PD</td><td>N/A</td><td>N/A</td><td>Focused</td></tr>
<tr><td rowspan="2">Aquatec</td><td rowspan="2">Aqua Modem</td><td>Op2</td><td>[347]</td><td rowspan="2">28 × ⌀7</td><td>4000</td><td>3500</td><td rowspan="2">1</td><td rowspan="2">80 K</td><td rowspan="2">LED/PD</td><td rowspan="2">9–
36</td><td rowspan="2">N/A (B)</td><td rowspan="2">N/A</td><td>RS 232</td></tr>
<tr><td>Op2L</td><td>[348]</td><td>1000</td><td>500</td><td>RS 485</td></tr>
<tr><td>UON</td><td colspan="2">OMM</td><td>[349]</td><td>21 × ⌀14</td><td>N/A</td><td>4000</td><td>15</td><td>115 K</td><td>LED/PD</td><td>3.3</td><td>N/A (B)</td><td>Omni D.</td><td>RS 232</td></tr>
<tr><td>Ambalux</td><td colspan="2">1013C1</td><td>[350]</td><td>28 × ⌀10</td><td>2500</td><td>61</td><td>40</td><td>10 M</td><td>LED/PD</td><td>N/A</td><td>N/A (B)</td><td>25/25
±5</td><td>Ethernet/TCP
IP/UDP</td></tr>
<tr><td>Shimadzu</td><td colspan="2">MC100</td><td>[351]</td><td>25 × ⌀13</td><td>2850</td><td>3500</td><td>> 10</td><td>95 M</td><td>LD/PD</td><td>N/A</td><td>450–640 (B/G)</td><td>Focused</td><td>N/A</td></tr>
</table>

As for the receiver, it should be equipped with a photodetector to ensure signal collection. The choice of photodetector is critical for enabling practical UOWC systems. While the PIN photodiodes are cheap and simple to implement and ensure good linearity, their sensitivity is limited for weak received optical signals over long distances. Whereas for the avalanche photodiodes (APDs), they ensure internal gain and improved sensitivity, at the cost of higher complexity, biasing requirements, and noise. A for the photomultiplier tubes (PMTs) and single-photon avalanche diodes (SPADs), they are mostly suitable for very low-light conditions, since they provide high sensitivity and photon-counting capabilities, making them attractive for long-range or low-power underwater optical links. The performance of the receiver is further improved in practical deployments through optical bandpass filtering, optical concentrators, narrow or adaptive field-of-view design, and signal-processing techniques that reduce the impact of background light, scattering, and misalignment (Sun *et al.* (2020)).

Followed by the remarkable research interest in UOWC, we notice the upsurge of many successful commercial products for UOWC systems. Optical modems have been developed for short-range high-speed data transfer, AUV/ROV control, data harvesting from underwater sensors, and video transmission. The table 1.2 lists some of these products. These products prove that current UOWC can ensure data rates in the order of Mbps in practical underwater deployments, although the achievable range remains strongly constrained by water clarity, background light, and line-of-sight alignment.

1.4 Self-Organized Networks for UOWC

Self-organized networks (SON) refer to networks that are capable of autonomously perform self-configuration, self-optimization, self-healing, and self-adaptation. The concept of SON was initially defined for terrestrial cellular networks, where it is used for network planning, handover optimization, interference coordination, and load balancing. But recently, this concept was also integrated into UOWCs because such systems require autonomous adaptation as well. In UOWC SON, autonomous configuration includes defining the initial beam alignment, node discovery, and link establishment. Whereas self optimization covers the beam adaptation, power allocation and interference management. As for self-healing, it includes link recovery after misalignment, and relay/OIRS-assisted reconnection.

UOWC SON require autonomous beam alignment, adaptive resource allocation, interference management, and energy-aware operations. Considering the dynamicity existing in UOWC systems and the high complexity of optimization problems due to the multiple factors affecting the link performance, learning-based methods, specifically reinforcement learning (RL) methods, shows to be a good candidate to establish a robust UOWC SON, since it learns from interaction with the environment and can easily adapt to changing environments.

1.5 Reinforcement Learning

Despite potential advantages of UOWC, this technology still faces significant challenges that can limit the performance of such networks. One of the main causes behind such challenges is dynamicity of the underwater environment and its dependency on many uncontrollable factors. The envisaged solution must be dynamic and capable of adapting to any sudden change in the environment. In this context, RL can be a promising candidate for such applications. In this section, we will introduce RL, its different features, and main algorithms.

RL is a machine-learning approach based on the existence of an agent that learns from its environment and takes actions so as to maximize a total reward. The learning process of RL is mainly based on trial-and-error, does not need any prior knowledge of the environment, and is not based on any collected data. This is exactly the case for UOWC, particulary considering dynamicity of the environment and its dependence on many uncontrollable factors, which makes it inadequate to apply traditional methods for such problems. RL systems consist of the following four main elements: policy, reward, value function, and model of the environment (Sutton & Barto (2018)).

Policy π: The policy defines how the agent will act at any given time. The aim of the RL algorithm is to find an optimal policy that maximizes the cumulative payoff.

Reward r: The reward function maps the couple $(state, action)$ to a value, called reward. It reflects the agent's performance. The bigger the reward, the better the agent is performing. In a communication system, the reward definition may include metrics such as the link SNR, throughput, success rate, spectral efficiency (SE), and EE.

Value function V: Unlike the reward that gives an immediate evaluation of the taken action, the value function gives a long-term reward. A current state can give a high immediate reward, but lead to a relatively low value if the next steps give low rewards. Therefore, it can sometimes be wiser to opt for a state with a lower reward if it guarantees a higher value.

Model: In some RL problems, the environment can be approximated by a model. This model can be used to either predict the reward or find possible actions at a certain state. In a given RL problem, if policies are selected according to a given model, it is called a "model-based method". Otherwise, the policy is gradually learned through trial and error, in a process known as "model-free method".

1.5.1 Markov Decision Process (MDP)

MDP is a description of the RL environment, where the current state fully characterises the process. The MDP obeys the Markov property where the current state depends only on the previous state. The MDP is a tuple $< S, A, P, r >$ where S is a finite set of states and A is a finite set of actions. $P(s_t, s_{t+1})$ is the probability that action a_t in state s_t at time t will lead to state s_{t+1} at time $t+1$, also known as transition function $T(s_t, a_t, s_{t+1})$. As for $r(s_t, s_{t+1})$, it is the expected immediate reward received after moving from state s_t to state s_{t+1} when taking action a_t. The aim of the agent is to maximize the reward function and potential reward values in the future. The interaction between the agent and the environment can be modeled as shown in Figure 1.3. The RL problem can be classified into value-based, policy-based, and actor-critic RL agorithms. In what follows, we define the main algorithms for each of these RL classes.

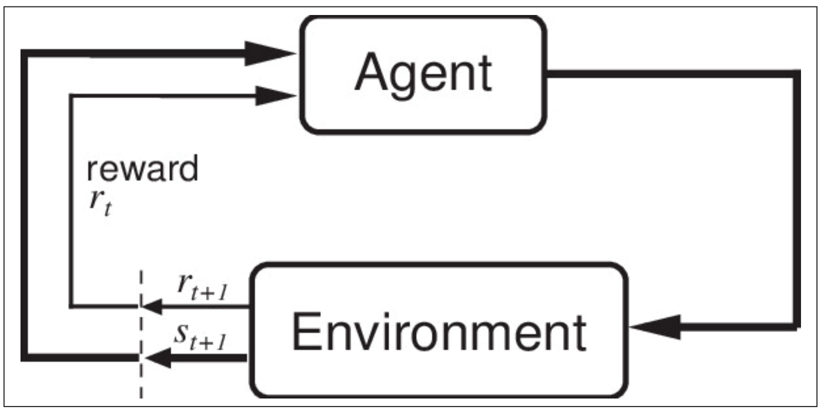


Figure 1.3 Agent-environment interaction in reinforcement learning. Taken from: Sutton & Barto (2018)

1.5.2 Value-Based RL Algorithms

Value based RL algorithms aim to maximize the expected long-term future reward. In this section, we define classical value-based algorithms such as the temporal difference (TD) learning, SARSA, and Q-learning.

1.5.2.1 TD Learning

TD learning is a value-based RL algorithm based on bootstrapping techniques. Said differently, instead of waiting until the end of the episode, a TD learning algorithm updates the estimate in a faster way by considering only n steps into the future. The definition of a TD algorithm is joined with parameter λ ranging between 0 and 1. In what follows, we start by the simplest one, TD(0), where the algorithm looks one step only into the future. The one step return is given by

Eq. (1.1).

$$G_t^{(1)} = r_{t+1} + \gamma v(s_{t+1}) \tag{1.1}$$

The update is shown in Eq. (1.2).

$$v(s_t) \leftarrow v(s_t) + \alpha[r_{t+1} + \gamma v(s_{t+1}) - v(s_t)] \tag{1.2}$$

For non-zero λ, the algorithm combines all n-step returns G_t^ns weighted by λ (see Eq. (1.3)).

$$G_t^\lambda = (1-\lambda)\sum_{n=1}^{\infty} \lambda^{n-1} G_t^{(n)} \tag{1.3}$$

The TD(λ) update can then be described by Eq. (1.4).

$$v(s_t) \leftarrow v(s_t) + \alpha[G_t^\lambda - v(s_t)] \tag{1.4}$$

1.5.2.2 SARSA

SARSA, which is basically an abbreviation of state-action-reward-state-action, is an on-policy TD control algorithm. Its name comes from the fact that the action-value function $Q(s_t, a_t)$ is updated using the next state and action tuple (s_{t+1}, a_{t+1}) obtained using the same policy that determined current action a_t. This can be expressed using Eq. (1.5).

$$Q(s_t, a_t) \leftarrow Q(s_t, s_t) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \tag{1.5}$$

The term $r_{t+1} + \gamma Q(s_{t+1}, a_{t+1})$ represents the temporal-difference error, which combines the immediate reward and the discounted value of the next state-action pair. α is the learning rate that controls this error's effect on the current estimate.

Recently, SARSA was introduced into solving underwater communication challenges. For instance, SARSA was deployed to improve the throughput and harvest power in an underwater acoustic communication network (UACN) (Omeke *et al.* (2024)). Work in Ghazy, Kaddoum, Talhi, Iqbal & Muqaibel (2025) deployed SARSA to overcome the security and reliability problems in UACN. Furthermore, Shi, Yang, Huang, Li & Huang (2023) proposed a SARSA-based adaptive routing protocol for route selection that maintains QoS requirements in an underwater network.

1.5.2.3 Q-learning

Q-learning is an off-policy TD control algorithm. Here, not only the adopted policy is considered, but also the alternative policies along the way. In other words, action-value function $Q(s_t, a_t)$ is updated using the maximum action-value function over all possible actions taken at state s_t. This can be expressed as shown in Eq. (1.6).

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a') - Q(s_t, a_t)) \tag{1.6}$$

The term $r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a')$ represents the best possible expected return from the next state. Hence, the Q-learning algorithm updates the current state-action value to the best achievable future value, independently of the policy-based action.

To date, Q-learning has been widely used in solving wireless communication-related issues. It was first introduced in 2010 in Hu & Fei (2010) who proposed the QELAR protocol, an adaptive routing protocol for underwater wireless communications. Later on, different Q-learning-based routing protocols for underwater wireless communications were proposed (Zhang, Chen & Xie (2025); Rahmani *et al.* (2025); Gao, Wang, Gu & Shi (2024)). For instance, in Shen, Yin, Jing, Liang & Wang (2022); Shen *et al.* (2023), authors proposed a Q-learning-based routing protocols for hybrid optical-acoustic underwater wireless communication networks. In another relevant study, Han, Gong, Wang, Martínez-García & Peng (2021) proposed a Q-learning-based multi-autonomous underwater vehicles (AUV) collaborative data collection technique for UACN to decrease the load of data collection task on a single AUV.

1.5.3 Policy-Based and Actor-Critic RL algorithms

In real-world scenarios, many UOWC control variables are continuous, such as transmit power and beam parameters. However, previously discussed value-based algorithms are unable to support continuous action and state values. This justify the need for policy-based and actor-critic RL algorithms. The policy-based algorithms directly optimizes the policy, and accepts continuous action sets. Meanwhile, actor-critic algorithms combine a policy/actor network and a value/critic network, where actor network chooses actions, while critic network evaluates them. The actor-critic networks supports continuous action and state sets, making it suitable for continuous resource allocation and beam control problems.

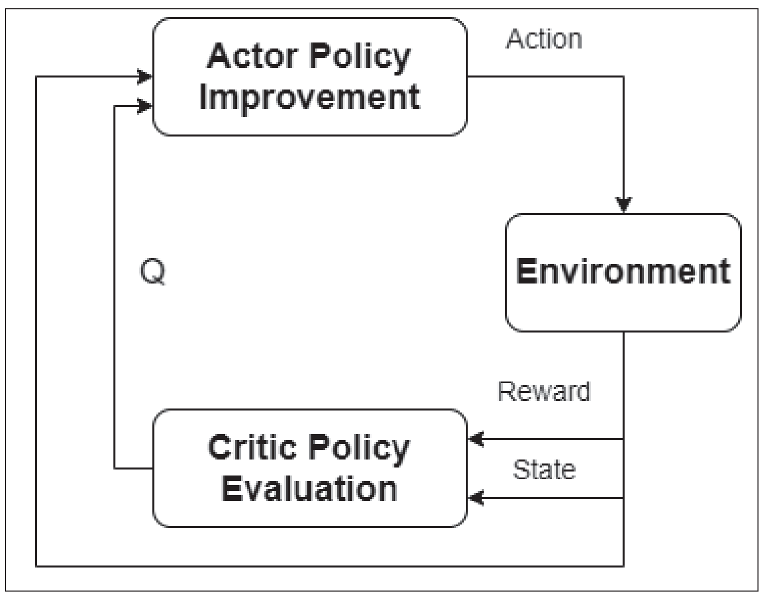


Figure 1.4 Policy iteration and actor–critic learning.
Taken from: Xu *et al.* (2007)

1.5.4 Deep Reinforcement Learning

In case of large dimensions or continuous states and actions, RL algorithms would require a significantly large amount of storage to construct the value function and policy tables. Hence, deep reinforcement learning (DRL) was proposed as an extension of RL to overcome the high-dimensionality challenge, by combining RL with deep neural networks. Deep neural networks are basically used as function approximators to estimate value functions, action-value functions, or policies. This makes DRL particularly suitable for complex UOWC problems, where state set may include the channel conditions, beam parameters, node positions, or battery states. Once trained, the DRL policy can provide fast online decisions compared to traditional iterative optimization methods, which is a promising solution for dynamic underwater scenarios and real-time underwater applications. We state here some DRL algorithms from literature.

1.5.4.1 Deep Q-learning

The deep Q-learning (DQN) approximates the Q-table as a deep neural network (DNN), known as the trained network, of parameters θ (Ge, Liang, Joung & Sun (2020)). This network takes as input the set of states and outputs the resulting actions. In this case, the Q function becomes parameterized, which we denote as $q(s, a, \theta)$. Furthermore, to improve the convergence of the algorithm while maintaining stability in the training phase, the DQN is equipped with the following three other key components:

- Target network: Having the same architecture as the trained network and different parameter θ', the target network is defined as a copy of the trained network updated every T steps. Using the target network along with the trained network improves stability of the learning process;
- Memory: In memory B, all experienced tuples (s_t, a_t, r_t, s_{t+1}) are stored. This memory is useful for experience replay;
- Experience replay: The DQN is trained using previous experiences from the memory. The experiences subset, called mini-batch, is randomly selected at each training step to make better use of the memory and recall rare occurrences. The loss function is consequently calculated as shown in Eq. (1.7),

$$L(\theta) = \frac{1}{2} \sum_{(s_t, a_t, r_t, s_{t+1}) \in B} (r' - q(s_t, a_t, \theta))^2 \tag{1.7}$$

  where $r' = r + \gamma max_{a'} q(s_t, a', \theta')$ is the target value. Then, the trained network weights, θ, are updated using the stochastic gradient descent optimizer.

DQN was also used in numerous previous studies on underwater wireless communications. For instance, Geng & Zhang (2023) proposed DQN-based routing protocols for UACN. These protocols take into consideration energy and latency and aim to extend network lifetime and reduce average time delay. Furthermore, Geng & Zheng (2023) proposed a DQN-based medium access control (MAC) protocol for UACN that improves the system's sum throughput and packet success rate by adjusting the transmission start time and decreasing the time slot length.

1.5.4.2 Deep Deterministic Policy Gradient (DDPG)

All algorithms briefly reviewed above deal with discrete states and actions sets with a defined number of elements in each set. In the case of continuous state and action sets, these algorithms are unable to proceed. Hence, the DDPG algorithm was proposed as an actor-critic algorithm that specifically deals with these cases. Specifically, this algorithm includes the following four main networks (Mnih *et al.* (2015)):

- An online actor network parameterized by θ^{μ}: This network takes as input the states and outputs the actions;
- An online critic network parameterized by θ^{Q}: This network takes as input the states and actions decided by the online actor network, and outputs the Q-function Q(s,a) that evaluates the given state-action tuple;
- A target actor network parameterized by $\theta^{\mu'}$: This network is a previous copy of the online actor network used to track the learned network to ensure stability;
- A target critic network parameterized by $\theta^{Q'}$: It is a previous copy of the online critic network.

First, the agent discovers the environment by taking random actions and storing the resulting experiences in its memory. It then learns the action using the actor network. A certain exploration noise $\mathcal{N}_t$ is added to the learned action to cover more values of the continuous action set. Afterwards, the critic network is updated by minimizing the loss function as shown in Eq. (1.8).

$$L(\theta^{Q}) = \frac{1}{N_B} \sum_{t} \left[Q(s_t, a_t|\theta^{Q}) - y_t \right]^2 \tag{1.8}$$

where N_B is the size of the mini-batch sampled from the memory and y_t is calculated using the Bellman equation defined as shown in Eq. (1.9)

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'}\right) \tag{1.9}$$

The actor network is updated using the sampled policy gradient (see Eq. (1.10)

$$\nabla_{\theta^\mu} \approx \frac{1}{N_M}\sum_t \left[\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s=s_t}\right] \tag{1.10}$$

Finally, the target networks are updated using Eq. (1.11a)-(1.11b).

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \tag{1.11a}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{1.11b}$$

where τ is the update rate of target networks.

DDPG shows to be efficient in solving underwater wireless communications challenges. In Li *et al.* (2024b), a DDPG-based cooperative movement strategy is proposed for collaborative multi-AUV optical communication. The DDPG was also employed in Liu *et al.* (2024) to perform beam alignment for water-air direct optical wireless communication. In another relevant study, Wang *et al.* (2024b) designed the DDPG-based controller for direct air-to-underwater optical wireless communication; this was done to drive the uncrewed aerial vehicle in tracking the AUV movement so as to keep it in its connectivity region. Finally, in Li *et al.* (2023a); Ma *et al.* (2023), DDPG was used to control the AUV direction and speed to achieve a reliable UOWC.

1.5.5 Multi-agent Reinforcement Learning (MARL)

In the previous definition of RL, the existence of a single agent responsible for making decisions is assumed. However, an RL system can contain multiple agents, known as MARL. This solution can alleviate the work and improve convergence by dividing the tasks among different agents that cooperatively work. These agents learn by dynamically interacting together in their shared environment (Busoniu, Babuska & De Schutter (2008)). In MARL, the state transition, the reward, and the action-value function Q depend on joint actions taken by all the agents.

MARL technique was widely used in the underwater communications literature. In Wang *et al.* (2024a), a MARL-based routing protocols for underwater wireless networks was proposed. In another relevant study, Lin, Han, Zhang, Shah & Peng (2023) used MARL to detect underwater pollution using an AUV-based network. In addition, MARL-based routing protocols (Li, Hu, Zhang & Yang (2020b)), and channel access schemes (Zhang, Zhang & Chen (2021)) are proposed in literature for UOWC networks.

In contrast to a single-agent RL, MARL brings valuable advantages, but still faces some challenges. In the one hand, when the agents exploit the decentralized structure of the task, parallel computation can be realized, which speeds up MARL execution. Moreover, agents with similar tasks can learn faster and better by sharing experience. Furthermore, if an agent fails the other agents can overtake the task instead, thereby enhancing the robustness of such systems. Similarly, if a new agent is inserted in the system (e.g. a new node to the network), it can learn faster about the environment by imitating the skilled agents. On the other hand, the complexity of the system depends on the number of agents; therefore, MARL is more complex than a single-agent RL system. Furthermore, since all agents are learning simultaneously, each agent faces a non stationary target learning problem, since the best policy changes as the other agents' policies change. Finally, considering that the effect of each agent's action on the environment depends on other agents' actions, a good coordination mechanism is needed.

1.6 Literature Review

In the previous sections, we have introduced the fundamentals of UOWC, stated the challenges faced by such communication, and went through the main RL algorithms that can be used to solve dynamic optimization problems. In this section, we investigate the literature review related to the contribution of this thesis. We consider four main research directions: beam adaptation and link reliability, outage analysis under turbulence and pointing errors, OIRS-assisted and direct air-to-underwater optical communications, and resource allocation for multi-user energy efficient UOWC networks.

1.6.1 Beam Adaptation and Link Reliability in UOWC

The directivity of light ensures high received power and increases transmission ranges. However, it highlights the misalignment and pointing error problems. A narrow beam increases the received SNR, but it makes the link more sensitive to misalignment and pointing errors. In contrast, wider beams makes the link more robust against node mobility and quasi-stationarity, but they limit the communication range and degrades the received signal. Hence, beam parameters should be optimized to ensure a reliable communication under dynamic conditions.

Recently, many researchers studied the beam alignment problem for UOWC (Saeed *et al.* (2019a); Zeng *et al.* (2017)). In Weng *et al.* (2022a), researchers investigated the pointing error control in UOWC on mobile platforms. It mainly focused on pointing control and did not provide an adaptive beam management framework under dynamic channel conditions, considering both reliability and link range. Work in Derakhshandeh *et al.* (2023) proposed a beam adaptation technique that is based on optical phased array antennas for laser-based UOWC system. However, their proposed solution is mainly hardware-based and does not include a dynamic beam adaptation solution. Researchers in Celik, Saeed, Shihada, Al-Naffouri & Alouini (2019) partly considered the dynamicity existing underwater by optimizing beamwidth under node location uncertainty. Basically, they considered an uncertainty sphere that covers all the possible node locations and adapted the beamwidth such that the light beam covers the whole sphere. Although this solution proved to be robust against misalignment problems, it limited the link range and degraded the link quality since it considered wider beams. To date, the work that investigated learning-based beam adaptation solutions for UOWC systems is very limited. For instance, Weng *et al.* (2022b) studied the beam alignment for AUV-based UOWC systems using RL/POMDP based approach under disturbances, noise, and AUV model uncertainty. This work focused mainly on the beam orientation adaptation and did not consider the range/beamwidth tradeoff for UOWC links.

Finally, several studies addressed the beam alignment and pointing error mitigation problems. However they either focused on static links without taking into consideration the dynamic channel conditions,or the focused on specific alignment scenarios. Hence, the beam adaptation considering underwater dynamicity and channel uncertainty requires futher investigation.

1.6.2 Outage Analysis Under Oceanic Turbulence and Pointing Errors

To better understand the performance limits of UOWC systems, analysing the system's outage probability is essential. While various channel impairments affect the performance of the system, oceanic turbulence and pointing errors are particularly important because they introduce random fluctuations in the received optical signal.

To date, numerous statistical models were proposed in the literature to model the oceanic turbulence, such as the lognormal, log-logistic, gamma, exponential Weibull (EW), and generalized gamma (GG) (Baykal, Ata & Gökçe (2022); Ata & Kiasaleh (2023); Rahman, Tailor, Zafaruddin & Chaubey (2022); Jiang, Li, Zhang, Ju & Huang (2023b); Jiang, Qiu *et al.* (2022)). Furthermore, researchers in Zedini *et al.* (2019) proposed the mixture Exponential-Generalized Gamma distribution (EGG) model for turbulence induced fading in the presence of air-bubble, salinity, and temperature gradients. The combined impact of oceanic turbulence and pointing error on UOWC system performance was also investigated in literature. For instance, Ijeh, Khalighi, Elamassie, Hranilovic & Uysal (2022) evaluated the outage probability of vertical UOWC links under oceanic turbulence and pointing errors effects. To mitigate the effect of oceanic turbulence, various techniques are stated in literature, including the maximal-ratio combining (MRC) techniques (Bansal, Rahman & Zafaruddin (Early Access, May 2023) and spacial diversity (Jamali, Salehi & Akhoundi (2017b)).

1.6.3 Direct Air-to-Underwater and OIRS-Assisted Optical Communications

Underwater nodes can communicate with other underwater nodes as well as terrestrial nodes. Such hybrid communications can be done through a relay station on the water surface, or directly using a direct air-to-underwater link. Apart from previously discussed underwater challenges, the direct air-to-underwater communications undergo important challenges related to atmospheric impairments such as fog, rain, and pointing errors. They also face refraction and multi-path interference at the air-to-water surface. Seeking to address the direct air-to-underwater link challenges, several previous studies focused on investigating the performance of direct air-to-underwater optical communications (Rahman, Zafaruddin & Chaubey (2023); Carver *et al.* (2021)). Also, Liu *et al.* (2024) performed beam orientation optimization of direct-air-to-underwater link using DRL. But they did not consider beamwidth and power optimization.

OIRS-assisted optical link was also introduced recently as a promising technique to overcome link blockage in UOWC systems. In this context, OIRS creates a secondary non-line on sight path, by reflecting and redirecting the optical beam toward the intended receiver. Accordingly, several previous studies introduced the OIRS technology in UOWC systems to improve system performance in occlusion or blockage due to obstacles, oceanic turbulence, attenuation, and PEs (Ata, Abumarshoud, Bariah, Muhaidat & Imran (2023); Naik & Chung (2022); Salam, Srivastava, Bohara & Ashok (2023)).

Finally, direct air-to-underwater optical communications and OIRS-assisted UOWCs are newly introduced promising solutions for improving connectivity underwater. However, they still face critical challenges such as air-water interface effects, mobility, OIRS placement, reflected-link modeling, and adaptive configuration. This requires further investigation of the implementation of these solutions.

1.6.4 Resource Allocation, Interference Management, and Energy Efficiency

Real-world underwater applications include various sensors, AUVs, buoys, ships, and underwater IoT devices. This requires the transition from a P2P communication to UOWC networks. In such networks, optical resources are shared among different users, which introduces inter-node interference, fairness issues, and energy-efficiency challenges. Accordingly, several multiple access techniques are proposed in literature. NOMA is a promising multiple access technique that enables users to share the same time, frequency, and optical resources, by applying successive interference cancellation (SIC) at the receiver.

Several work have investigated NOMA-based UOWC systems. For instance, Elamassie, Bariah, Uysal, Muhaidat & Sofotasios (2021) assessed the capacity of NOMA-enabled UOWC networks under lognormal turbulence. Moreover, Chen *et al.* (2022); Liang, Yin, Jing, Ji & Wang (2022) performed power allocation for NOMA-based UVLC networks. Palitharathna, Suraweera, Godaliyadda, Herath & Thompson (2021) maximized the link throughput using LED/PD selection schemes for NOMA-aided UOWC. Li *et al.* (2023c) performed power allocation for multi-beam NOMA-based UOWC networks. However these works focused only on power allocation and throughput maximization and did not treat the fairness and energy efficiency challenge in dynamic UOWC networks.

Finally, the NOMA-based UOWC-related work in literature focused on network capacity evaluation, throughput maximization, and power allocation. However, fairness and energy-efficiency are two important evaluation metrics for UOWC networks that are not yet explored. Also, optimization problems for NOMA-based UOWC networks are usually too complex considering the dynamicity in network topology and channel conditions. Hence, this motivates the use of RL to solve these problems.

CHAPTER 2

A REINFORCEMENT LEARNING BASED BEAM ADAPTATION FOR UNDERWATER OPTICAL WIRELESS COMMUNICATIONS

Imene Romdhane[1] , Georges Kaddoum[1]

[1] Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame west, Montreal, Canada H3C 1K3.

Paper published in IEEE Internet of Things Journal, vol. 9, Issue. 20, pp. 20270 - 20281, October 2022.

2.1 Introduction

Nowadays, humans' interaction with the underwater world is increasing, which makes underwater communication an attractive field to exploit. This type of communication became crucial in many areas, such as the remote control of instruments in oceans, pollution monitoring, oceanography research, oil-controlling, maintenance, etc. These applications explain the remarkable increase of devices deployed underwater lately. Such devices require high bandwidth and capacity while keeping a low power consumption. Previously, acoustic and Radio-Frequency (RF) waves were considered for underwater communications as they offer cheaper and more flexible communications. Many papers in the literature have focused on underwater acoustic communications Pompili & Akyildiz (2009); Chitre, Shahabudeen & Stojanovic (2008); Otnes, Jenserud, Voldhaug & Soldberg (2009). Moreover, RF underwater communication has been a widely investigated topic as well Shi, Zhang & Yang (2012); Che, Wells, Dickers, Kear & Gong (2010); Hunt, Niemeier & Kruger (2010). However, acoustic waves suffer from high latency and limited bandwidth Partan, Kurose & Levine (2007); Riksfjord, Haug & Hovem (2009), while RF waves suffer from high power consumption and severe attenuation Kulhandjian (2014); Frater, Ryan & Dunbar (2006). Given the limitations of underwater acoustic and RF communications, we recently witnessed the upsurge of a new means of communication, namely UOWC Kaushal & Kaddoum (2016); Saeed *et al.* (2019a). The aforementioned technology is based on exchanging information in submarine environments employing light waves. Such communication offers many advantages, such as higher data rates than acoustic waves, lower power consumption than RF waves, a simpler implementation, and lower computational complexity. UOWC has been widely used in scientific, environmental, civil, and military applications. Such applications include marine exploration, detection of human-made and natural environmental disasters, and submarine communications. In order to guarantee reliable and robust underwater communications, basic knowledge of the environment is needed. Nevertheless, accurately modeling underwater optical wireless channels is challenging due to the harsh underwater conditions. Unlike terrestrial optical wireless communications, underwater light propagation undergoes many losses mainly caused by absorption and scattering. This shortens the coverage range of the optical base station (OBS), defined by the underwater area where the nodes can communicate with the OBS. Moreover, it degrades the link quality of the UOWC and the connectivity of the network can be affected. The transmitted light can also be blocked by the passage of marine creatures and plants on its way. Underwater OBSs are not fixed and can change their position or orientation due to turbulence or convulsive water movement. This can lead to misalignment problems and consequently lead to packet loss. The retransmission of packets deteriorates these networks' performance by introducing problems such as the extension of transmission delays, the increase of power consumption, and congestion problems. These aforementioned problems make reliability, which is the probability of performing a successful communication, a fundamental challenge for such communications.

2.1.1 Related work

Recently, numerous works have been developed on UOWC, aiming to better understand the underwater environment and its different parameters. In Miramirkhani & Uysal (2017), the underwater optical channel was modeled considering the characteristics of the environment and the presence of shadowing and blockage effects for various underwater scenarios. In Saeed, Celik, Alouini & Al-Naffouri (2019b), the connectivity of underwater optical wireless sensor networks (UOWSNs) and their effects on the network localization performance have been studied. The work in Jamali, Chizari & Salehi (2017a); Elamassie, Miramirkhani & Uysal (2018); Huang, Tao, Wang & Zhang (2018) focused on improving the link quality of UOWC systems. The authors in Jamali *et al.* (2017a) studied the end-to-end bit error rate (BER) of multi-hop UOWC systems considering light degrading effects during propagation due to absorption, scattering, and water turbulence. In Elamassie *et al.* (2018), the authors worked on maximizing the coverage of a multi-hop UOWC network. A closed-form path loss expression was developed as a function of the transceiver parameters and the water type. This expression was then used to find the maximum achievable link distance and the maximum possible distance satisfying a targeted end-to-end BER in multi-hop systems with a given number of equidistant hops. The authors in Huang *et al.* (2018) improved the error performance of UOWC links using spatial diversity (SD) and proved that SD is efficient against the impairing effects of fading. The work in Alghamdi, Saeed, Dahrouj, Alouini & Al-Naffouri (2019) goes through the existing centralized routing protocols and proposes distributed routing protocols that guarantee ultra-reliable low-latency underwater communications in multi-link UOWC networks.

Many research works based on static models considered an optimistic scenario where the sudden variations underwater leading to node dislocations have been neglected. However, the work in Celik *et al.* (2019) treated the dynamicity issue by evaluating the performance of UOWC under the realistic assumption of uncertain node locations. They evaluated the UOWC link's connectivity considering the uncertainty disk that covered all the node's possible locations instead of its estimated location. Although this paper considered a more realistic scenario, the proposed method is not energy efficient since the transmitted light should always cover the whole disk, leading to degradation of the network's performance. Thus machine learning, specifically RL, appears to be a promising option to solve these problems more efficiently since it can dynamically adapt to the existing situation. RL approaches have been used to solve link adaptation problems in underwater acoustic communications and terrestrial wireless communications. In Wang, Wang, Sun & Fuhrmann (2017); Fu & Song (2018); Wang, Li & Qian (2019a); Wang *et al.* (2019b), authors treated the link quality of underwater acoustic communications. In Wang *et al.* (2017), the energy consumption/transmission latency trade-off is addressed. The work in Fu & Song (2018) proposed an adaptive modulation strategy based on RL for underwater acoustic communications to improve the communication throughput. Authors in Wang *et al.* (2019a) used

RL for the uplink resource allocation problem in underwater acoustic communication networks (UACN). In Wang *et al.* (2019b), authors focused on the power allocation problem in underwater full-duplex energy harvesting relay networks and sought to maximize the long-term end-to-end sum rate. Moreover, authors in Yang *et al.* (2019); Ahmad, Soltani, Safari & Srivastava (2020); Alenezi & Hamdi (2020); Kong, Wu, Ismail, Serpedin & Qaraqe (2019) used RL to solve link adaptation in visible light communications (VLC) and RF communications. In Yang *et al.* (2019) a model-free RL algorithm was used to decide upon the sub-carrier and power allocation for multi-user integrated visible light communication and positioning (VLCP) systems. Authors in Ahmad *et al.* (2020) and Alenezi & Hamdi (2020) focused on maximizing the throughput of hybrid WiFi/VLC networks by providing an RL-based optimal access point (AP) assignment policy while satisfying QoS requirements. In Kong *et al.* (2019), the authors proposed a resource allocation problem for multi-homing hybrid RF/VLC networks. The aforementioned RL-based work defined their objectives according to the challenges faced in the specific environment they considered. In UOWC systems, ensuring the reliability and a good QoS of the communication system should be done by considering the attenuation and scattering problems and the position, acquisition, and tracking PAT challenge. Thus, adapting the light beamwidth and orientation is essential to enhance the link quality in UOWC systems.

The preceding literature can be classified into two parts. The first part focuses on solving UOWC issues using a static approach, while the second part treats the implementation of RL to solve problems in similar systems. These systems can exist in a marine environment, such as underwater acoustic systems, or a terrestrial environment, like VLC and RF systems. Our work aims to use the dynamicity feature of RL to propose a dynamic solution to the link adaption problem in UOWC.

2.1.2 Novelty and Contribution

In this paper, we consider the light beam adaptation problem in UOWC. This work aims to solve PAT problems by ensuring a good link quality while keeping a high success rate. This can be done by physically adapting the beamwidth and orientation of the light beam to target the receiver specifically. Most of the available work in literature focused on solving PAT problems in UOWC using deterministic optimization techniques based on complete knowledge of the environment. However, as evidenced in our previous discussion, such techniques are inapplicable due to the dynamic underwater environment. Thus, a dynamic approach to solving the given problem without assuming full knowledge of the environment. To tackle this challenge, we developed three RL-based beam adaptation methods that guarantee a reliable and stable link quality against the underwater dynamicity. The choice of the most suitable method depends on the system's hardware characteristics, namely the ability of the node to rotate, its range of rotation, and the beam divergence/convergence. Our proposed methods take into account the sudden changes that may occur to the nodes' positions or orientations. For that, we implement two model-free RL algorithms, one off-policy and one on-policy, namely the Q-learning and SARSA algorithms, respectively. The first method is based on beamwidth adaptation, where the state and action sets are defined using the light's beamwidth. The second method is based on beam orientation adaptation, where the state and action sets are defined using the node's orientation with respect to the horizontal. As for the third method, it is based on both beamwidth and beam orientation adaptation simultaneously. These algorithms optimize the link quality by continuously interacting with the receiver node without knowing the link state. Provided simulation results demonstrate that our proposed methods yield better reliability than the random method. We also compare the received signal-to-noise ratio (SNR) to that of the uncertainty disk method proposed in Celik *et al.* (2019). We prove that our proposed methods result in better link quality in P2P UOWC over four different environments: pure seawater, clean ocean, coastal ocean, and turbid harbor.

The main contributions of this paper can be summarized as follows:

- Unlike the previous work on link adaptation for UOWC systems, our work proposes a dynamic solution that continuously adjusts to the current situation.
- We present three RL-based beam adaptation methods for P2P UOWC based on adjusting the light source's optimal beamwidth, beam orientation, and both beamwidth and beam orientation. Followed by a sudden change of the node position and with no prior information

of its new coordinates, the light beam is adapted to face the receiver to maintain a robust P2P communication.

- We evaluate the reliability of the proposed methods by comparing their success rates with the random method.
- We present a comparison of the performances of the Q-learning and SARSA algorithms.
- We assess the link quality achieved by our proposed algorithms compared to the uncertainty disk static method. The comparison is made based on the SNR metric.

2.1.3 Organization and Nomenclature

The remainder of the paper is organized as follows. In Section II, we present the system model for the three proposed methods, where we define the different proposed RL methods and their system parameters. Simulation results and comparative analysis are presented in Section III. Finally, Section IV concludes the paper. The parameters used in this paper are summarized in Table 2.1.

2.2 System model

In this paper, we consider an underwater P2P UOWC as illustrated in Fig. 2.1. Our work aims to solve PAT problems faced in such communications and improve the link quality. As shown in the figure, the transmitter sends the light signal with a transmission power P0. If the receiver is in the line of sight of the transmitter, it will detect the signal with power Pr and send back an acknowledgment. In that case, we say that the transmission has been successful. If the receiver is not situated in the beam direction, the information is missed, and the transmission fails. Due to the short range of underwater optical communication and the directivity of light, the interference problem from other light sources can be avoided in a P2P UOWC where the transmitter and receiver are directly facing each other. The proposed solution is to adapt the light beam sent from the transmitter node with the aim of optimizing the quality of the signal collected at the receiver side. For simplicity, we will consider a 2D representation of the space. The found results can be easily reproduced in a 3D space by considering a 3D coordinates

Table 2.1 Summary of parameters used in UOWC

P_r	Received Power
θ	Light beamwidth
β	Inclination angle (receiver's normal w.r.t. light beam)
d	Distance between the two nodes
P_0	Transmitted power
A_r	Receiver's area (m^2)
A_t	Transmitter's area (m^2)
c	Beam light attenuation coefficient m^{-1}
a	Absorption coefficient
b	Scattering coefficient
α	Learning rate of the RL algorithm
γ	Discount factor of the RL algorithm
Ω	The receiver's normal w.r.t. horizontal axis
I_L	Photo-current generated by the incident light
q	Electric charge
K	Boltzmann constant
T	Temperature (K)
S	Receiver's sensitivity
I_D	Photodiode's dark current
R	Shunt resistance

system of the nodes (x,y,z), calculate a 3D distance, and define the beam orientation using Euler angles (ϕ, θ, ψ) instead of a single angle representation in 2D space. To this end, we propose three RL-based adaptation methods for P2P communication. Because the beam adaptation is performed on the transmitter node, this latter is defined as the agent. The receiver node sends the feedback to the agent in the form of an acknowledgment packet containing the received power. This acknowledgment is used to define the reward function. The first method performs the beamwidth adaptation using a lens placed in front of the light source and provided with a translation movement along the beam axis. The second method performs the beam orientation adaptation translated by a rotational movement of the transmission node. The third method performs beamwidth and beam orientation adaptations simultaneously.

The beam adaptation should be performed in real-time. Due to underwater's dynamicity, it is impossible to provide a Markov Decision Process (MDP) model for such communication.

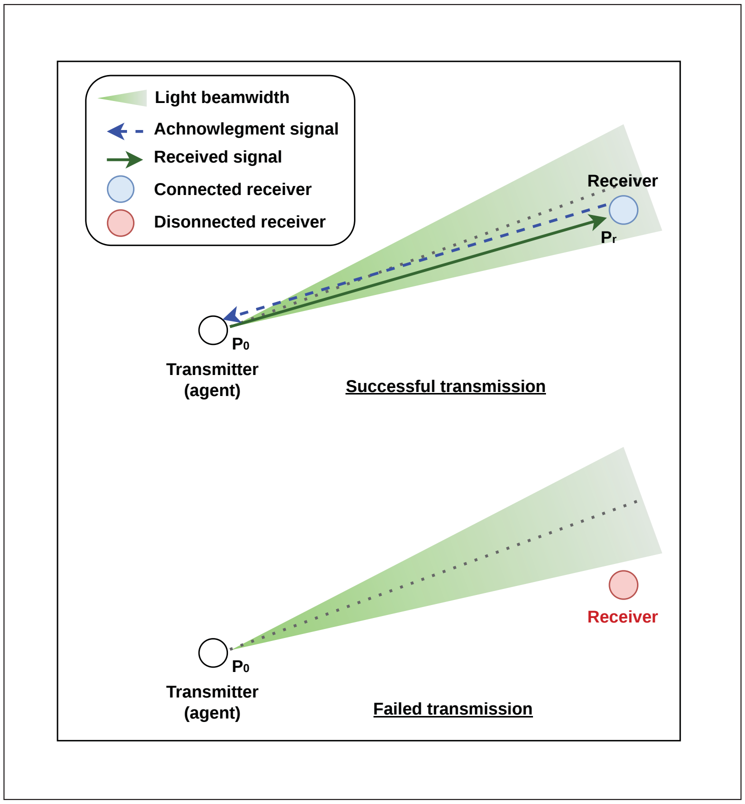


Figure 2.1 System model for a P2P UOWC

Hence, a model-free method is required to solve this problem. For that, two classes of model-free methods exist: on-policy and off-policy. Thus, we implement Q-learning, an off-policy algorithm, and SARSA, an on-policy algorithm. In Q-learning, the agent learns the optimal strategy in real-time by going through a series of discrete episodes, with no prior knowledge of the MDP model at each step Sutton & Barto (2018). In the learning process, our algorithm considers not only the adopted policy but also the alternative policies along the way. In other words, the action-value function $Q(s, a)$ is updated using the maximum action-value function over all possible actions taken at state s. This can be expressed as,

$$\begin{aligned} Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a') \\ -Q(s_t, a_t)), \end{aligned} \tag{2.1}$$

where α is the learning rate, and γ is the discount factor. S and A represent the states and actions sets, respectively. As for the SARSA algorithm, its major difference from Q-Learning is that the maximum reward for the next state is not necessarily used for updating the Q-values. Instead, a new action, and therefore reward, is selected using the same policy that determined the original action. This can be expressed as,

$$\begin{aligned} Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma Q(s'_{t+1}, a_{t+1}) \\ -Q(s_t, a_t)), \end{aligned} \tag{2.2}$$

There is no absolute best choice between these two algorithms. Their performance mainly depends on the situation we are considering. Thus, the motivation behind using both Q-learning and SARSA is to evaluate which algorithm is best to use for each case.

The agent will start with an initial state along which it will send the light beam. As a response, it will get a reward from the receiver node, and accordingly, it will take an action to move to the next state with the aim to improve the total reward and optimize the communication between the two nodes. The direct reward function r is defined as a function of the received power at the destination, collected from the receiver via an acknowledgment. If the receiver is on the transmitter's field of view and the received power is higher than a threshold, an acknowledgment is sent back with the received power value, which represents the reward. This threshold is defined as the minimum power the receiver's photodiode can detect. If no acknowledgment is received, the reward is set to zero. Thus, the reward function can be defined as,

$$r = \begin{cases} P_r & \text{if ACK,} \\ 0 & \text{if no ACK.} \end{cases} \tag{2.3}$$

The attenuation of light underwater can be quantified using the beam light attenuation coefficient c, defined as the sum of the absorption and scattering coefficients a and b, i.e. $c = a + b$, as follows,

$$I = I_0 e^{-cz}, \tag{2.4}$$

where I_0 is the transmitted light power and I is the light power at a distance z from the source. According to the Lambertian law McCluney (2014), the light power received at a particular point in space is

$$P_r = \frac{P_0}{A_t + A_s}, \tag{2.5}$$

where P_0 is the transmission power, A_t is the transmitter's area in m^2, and A_S is the area of the spherical crown formed by the beam at the receiver level. A_S can be calculated using the distance L and the beamwidth θ as

$$A_S = \frac{\pi L^2(1 - cos(\theta))}{2}. \tag{2.6}$$

From (2.4), (2.5), and (2.6), the received power P_r can be expressed as,

$$P_r(\theta, \beta) = \frac{2P_0 A_r cos\beta}{\pi L^2(1 - cos\theta) + 2A_t}.e^{-cd}, \tag{2.7}$$

where β is the angle between the receiver's normal and the light beam, A_r is the receiver's area, and d is the distance between the two nodes. The light distribution model elaborating these aforesaid parameters is shown in Fig. 2.2.

Here the action is taken using the epsilon greedy function defined as,

$$a_t = \begin{cases} \underset{a' \in A}{\operatorname{argmax}}\, Q(S_t, a') & \text{with probability } (1 - \epsilon), \\ \text{any action } a'_t & \text{with probability } \epsilon. \end{cases} \tag{2.8}$$

To improve the decision taking of the epsilon greedy function, we opt for a decaying epsilon as a function of time defined as follows,

$$\epsilon = 0.5/i, \tag{2.9}$$

where

$$i = \begin{cases} i + 10^{-2} & \text{if } t \text{ is multiple of } 10, \\ i & \text{otherwise}, \end{cases} \tag{2.10}$$

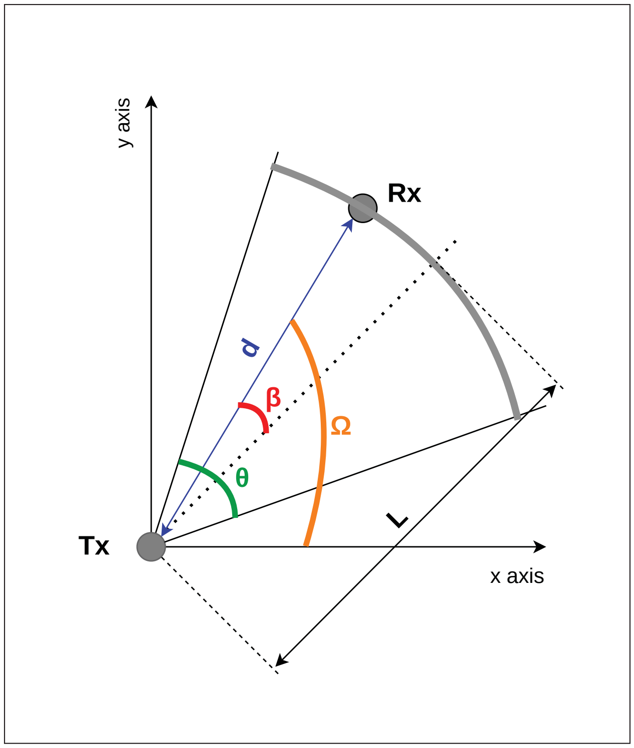


Figure 2.2 Light distribution model of a P2P UOWC system

where t is the iteration of the RL algorithm. The decaying epsilon will favor the optimal policy over time by progressively decreasing the probability of exploration. In what follows, we describe the three proposed adaptation methods and evaluate their performance compared to the uncertainty disk static method.

2.2.1 Beamwidth adaptation for a P2P communication

First, we focus on adapting the light source's beamwidth for a P2P UOWC system using Q-learning and SARSA algorithms. In this context, the agent is located at the transmitter node. Moreover, we define the state set, S_{bw}, to be the light's beamwidth. The action set, A_{bw}, includes three actions: increase the beamwidth, decrease it, or keep it the same. The reward function can be defined as a function of the received power as,

$$r_{bo} = \begin{cases} P_r(S_{bw}, \beta) & \text{if ACK,} \\ 0 & \text{if no ACK.} \end{cases} \tag{2.11}$$

Our Q-learning algorithm is summarized in Algorithms 1. This algorithm uses as input the RL hyperparameters α and γ, the states set S_{bw}, and the actions set A_{bw}. The hyperparameters α and γ represent the learning rate and the decay rate of future reward, respectively. Initially, the Q table is set to zero. The epsilon value of the ϵ-greedy function is initiated at 0.5. At the start of each episode e, the state is initialized to a certain beamwidth value in the set S_{bw}. After every ten episodes, the epsilon value is decreased using the decaying function. The Q-learning algorithm selects an action $a_{bw,t}$ at each step t of the episode using the ϵ-greedy function. A new state, $s_{bw,t+1}$, and reward, $r_{bw,t+1}$, are obtained. The Q table is updated using the Q-learning update equation. Finally, the state is updated to $s_{bw,t+1}$.

The SARSA algorithm is summarized in Algorithm 2. The algorithm initiates the hyperparameters α and γ, the states set S_{bw}, and the actions set A_{bw}. The Q table and the epsilon value are set to 0 and 0.5, respectively. The state is initialized to a certain beam orientation value at the start of each episode e, and the epsilon value is decreased after each 10 episodes. The SARSA algorithm chooses an action $a_{bw,0}$ using the ϵ-greedy function. Then at each step t of the episode, a new state, $s_{bw,t+1}$, and reward, $r_{bw,t+1}$, are obtained. A new action, $a_{bw,t+1}$, is chosen using the ϵ-greedy function, and the Q table is updated using the SARSA update equation.

2.2.2 Beam orientation adaptation for a P2P communication

Here, we assume that the agent is located at the transmitter node, and we implement the Q-learning and SARSA algorithms. Additionally, we chose the set of states, S_{bo}, as beam orientations and the action set, A_{bo}, as moving the beam up, down, or stop it. The reward

Algorithm 2.1 Q-learning algorithm for beamwidth adaptation

Algorithm: Q-learning algorithm for beamwidth adaptation
Input: α, γ, Environment, S_{bw}, and A_{bw}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
  Initialize $s_{bw,0} \in S$
  if *e multiple of 10* **then**
    $i = i + 10^{-2}$
    $\epsilon = 0.5/i$
  end if
  for $t = 0, 1, 2...$ **do**
    if $rand(0, 1) \geq \epsilon$ **then**
      $a_{bw,t} = \underset{a' \in A_{bw}}{\operatorname{argmax}} Q(s_{bw,t}, a')$
    else
      $a_{bw,t}$ randomly chosen
    end if
    Observe $s_{bw,t+1}$
    Calculate $r_{bw,t+1}$
    $Q(s_{bw,t}, a_{bw,t}) = Q(s_{bw,t}, a_{bw,t}) + \alpha(r_{bw,t+1}+$
    $\gamma \max_{a' \in A_{bw}} Q(s_{bw,t+1}, a') - Q(s_{bw,t}, a_{bw,t}))$
    Update $s_{bw,t}$ to $s_{bw,t+1}$
  end for
end for

function is defined as a function of the received power as

$$r_{bo} = \begin{cases} P_r(\theta, \Omega - S_{bo}) & \text{if ACK,} \\ 0 & \text{if no ACK,} \end{cases} \tag{2.12}$$

where Ω is the angle between the receiver's normal and the horizontal axis. The Q-learning and SARSA algorithms for beam orientation adaptation are summarized in Algorithms 3 and 4, respectively. These algorithms proceed similarly to algorithms 1 and 2 defined and explained in the previous section. The decision is made upon the beam orientation instead of the beamwidth.

Algorithm 2.2 SARSA algorithm for beamwidth adaptation

Algorithm: SARSA algorithm for beamwidth adaptation
Input: α, γ, Environment, S_{bw}, and A_{bw}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
    Initialize $s_{bw,0} \in S$
    if *e multiple of 10* **then**
        $i = i + 10^{-2}$
        $\epsilon = 0.5/i$
    end if
    if $rand(0, 1) \geq \epsilon$ **then**
        $a_{bw,0} = \underset{a' \in A_{bw}}{\operatorname{argmax}} Q(s_{bw,0}, a')$
    else
        $a_{bw,0}$ randomly chosen
    end if
    for $t = 0, 1, 2...$ **do**
        Observe $s_{bw,t+1}$
        Calculate $r_{bw,t+1}$
        if $rand(0, 1) \geq \epsilon$ **then**
            $a_{bw,t+1} = \underset{a' \in A_{bw}}{\operatorname{argmax}} Q(s_{bw,t+1}, a')$
        else
            $a_{bw,t+1}$ randomly chosen
        end if
        $Q(s_{bw,t}, a_{bw,t}) = Q(s_{bw,t}, a_{bw,t}) + \alpha(r_{bw,t+1}+$
        $\gamma Q(s_{bw,t+1}, a_{bw,t+1}) - Q(s_{bw,t}, a_{bw,t}))$
        Update $s_{bw,t}$ to $s_{bw,t+1}$
        Update $a_{bw,t}$ to $a_{bw,t+1}$
    end for
end for

Our proposed models provide reliable connectivity between two nodes. However, in some realistic scenarios, the beam extension is limited to a certain degree, due to which the problem of signal loss may occur. Moreover, the node's rotational movement for beam orientation is constrained, which may lead to signal loss as well. Therefore, we propose a more robust scheme in our next subsection.

Algorithm 2.3 Q-learning algorithm for beam orientation adaptation

Algorithm: Q-learning algorithm for beam orientation adaptation
Input: α, γ, Environment, S_{bo}, and A_{bo}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
  Initialize $s_{bo,0} \in S$
  if *e multiple of 10* **then**
    $i = i + 10^{-2}$
    $\epsilon = 0.5/i$
  end if
  for $t = 0, 1, 2...$ **do**
    if $rand(0, 1) \geq \epsilon$ **then**
      $a_{bo,t} = \underset{a' \in A_{bo}}{\operatorname{argmax}} Q(s_{bo,t}, a')$
    else
      $a_{bo,t}$ randomly chosen
    end if
    Observe $s_{bo,t+1}$
    Calculate $r_{bo,t+1}$
    $Q(s_{bo,t}, a_{bo,t}) = Q(s_{bo,t}, a_{bo,t}) + \alpha(r_{bo,t+1}+$
    $\gamma \max_{a' \in A_{bo}} Q(s_{bo,t+1}, a') - Q(s_{bo,t}, a_{bo,t}))$
    Update $s_{bo,t}$ to $s_{bo,t+1}$
  end for
end for

2.2.3 Beamwidth and beam orientation adaptation for a P2P communication

In this subsection, we modify our Q-learning and SARSA algorithms to cater to beamwidth and beam orientation adaptation simultaneously. Hence, this beam adaptation includes two degrees of freedom. The states are defined via the set of beamwidths, S_{bw}, and the set of beam orientations, S_{bo}. The first set of actions, A_{bw}, is to increase the beamwidth, to decrease it, or keep it the same, whereas the second set of actions, A_{bo} is to change the orientation of the beam to up, to down, or to stop and keep it at a particular direction. The reward function is defined as,

$$r = \begin{cases} P_r(S_{bw}, \Omega - S_{bo}) & \text{if ACK,} \\ 0 & \text{if no ACK,} \end{cases} \tag{2.13}$$

Algorithm 2.4 SARSA algorithm for beam orientation adaptation

Algorithm: SARSA algorithm for beam orientation adaptation
Input: α, γ, Environment, S_{bo}, and A_{bo}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
  Initialize $s_{bo,0} \in S$
  if *e multiple of 10* **then**
    $i = i + 10^{-2}$
    $\epsilon = 0.5/i$
  end if
  if $rand(0,1) \geq \epsilon$ **then**
    $a_{bo,0} = \underset{a' \in A_{bo}}{\operatorname{argmax}} Q(s_{bo,0}, a')$
  else
    $a_{bo,0}$ randomly chosen
  end if
  for $t = 0, 1, 2...$ **do**
    Observe $s_{bo,t+1}$
    Calculate $r_{bo,t+1}$
    if $rand(0,1) \geq \epsilon$ **then**
      $a_{bo,t+1} = \underset{a' \in A_{bo}}{\operatorname{argmax}} Q(s_{bo,t+1}, a')$
    else
      $a_{bo,t+1}$ randomly chosen
    end if
    $Q(s_{bo,t}, a_{bo,t}) = Q(s_{bo,t}, a_{bo,t}) + \alpha(r_{bo,t+1}+$
    $\gamma Q(s_{bo,t+1}, a_{bo,t+1}) - Q(s_{bo,t}, a_{bo,t}))$
    Update $s_{bo,t}$ to $s_{bo,t+1}$
    Update $a_{bo,t}$ to $a_{bo,t+1}$
  end for
end for

where Ω is the angle between the receiver's normal and the horizontal axis. Lastly, we summarize all these steps for proposing our robust Q-learning and SARSA methods in Algorithms 5 and 6, respectively. These algorithms proceed in a similar manner to algorithms 1 and 2 defined previously, except for these methods, the decision includes both the beamwidth and beam orientation simultaneously.

Algorithm 2.5 Q-learning algorithm for beamwidth and beam orientation adaptation

Algorithm: Q-learning algorithm for beamwidth and beam orientation adaptation
Input: α, γ, Environment, S_{bw}, S_{bo}, A_{bw}, A_{bo}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
  Initialize $s_{bw,0} \in S_{bw}, s_{bo,0} \in S_{bo}$
  if *e multiple of 10* **then**
    $i = i + 10^{-2}$
    $\epsilon = 0.5/i$
  end if
  for $t = 0, 1, 2...$ **do**
    if $rand(0,1) \geq \epsilon$ **then**
      $a_{bw,t}, a_{bo,t} = \underset{a'_{bw} \in A_{bw}, a'_{bo} \in A_{bo}}{\operatorname{argmax}} Q(s_{bw,t}, s_{bo,t}, a'_{bw}, a'_{bo})$
    else
      $a_{bw,t}, a_{bo,t}$ randomly chosen
    end if
    Observe $s_{bw,t+1}$ and $s_{bo,t+1}$
    Calculate r_{t+1}
    $Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}) = Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}) + \alpha(r_{t+1} + \gamma \max_{a'_{bw} \in A_{bw}, a'_{bo} \in A_{bo}} Q(s_{bw,t+1}, s_{bo,t+1}, a'_{bw}, a'_{bo}) - Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}))$
    Update $s_{bw,t}$ to $s_{bw,t+1}$ and $s_{bo,t}$ to $s_{bo,t+1}$
  end for
end for

2.2.4 Evaluation of the link quality for the proposed methods

This subsection presents a comparison in terms of the link quality between our proposed methods, and the uncertainty disk static method Celik *et al.* (2019). For this latter, we define an uncertainty metric, called ϵ, and construct an uncertainty disk centered at the receiver node's estimated location and with radius $R + \epsilon$, where R is the node radius. The uncertainty metric ϵ is chosen such that the uncertainty disk covers all the possible locations of the node. Moreover, the connectivity is evaluated considering the uncertainty disk instead of the actual location of the

Algorithm 2.6 SARSA algorithm for beamwidth and beam orientation adaptation

Algorithm: SARSA algorithm for beamwidth and beam orientation adaptation
Input: α, γ, Environment, S_{bw}, S_{bo}, A_{bw}, A_{bo}
Initialize $Q, i = 1$
for $e = 0, 1, 2...$ **do**
  Initialize $s_{bw,0} \in S_{bw}, s_{bo,0} \in S_{bo}$
  if *e multiple of 10* **then**
    $i = i + 10^{-2}$
    $\epsilon = 0.5/i$
  end if
  if $rand(0, 1) \geq \epsilon$ **then**
    $a_{bw,0}, a_{bo,0} = \underset{a'_{bw} \in A_{bw}, a'_{bo} \in A_{bo}}{\text{argmax}} Q(s_{bw,0}, s_{bo,0}, a'_{bw}, a'_{bo})$
  else
    $a_{bw,0}, a_{bo,0}$ randomly chosen
  end if
  for $t = 0, 1, 2...$ **do**
    Observe $s_{bw,t+1}$ and $s_{bo,t+1}$
    Calculate r_{t+1}
    if $rand(0, 1) \geq \epsilon$ **then**
      $a_{bw,t+1}, a_{bo,t+1} = \underset{a'_{bw} \in A_{bw}, a'_{bo} \in A_{bo}}{\text{argmax}} Q(s_{bw,t+1}, s_{bo,t+1}, a'_{bw}, a'_{bo})$
    else
      $a_{bw,t+1}, a_{bo,t+1}$ randomly chosen
    end if
    $Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}) = Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}) + \alpha(r_{t+1} + \gamma Q(s_{bw,t+1}, s_{bo,t+1}, a_{bw,t+1}, a_{bo,t+1}) - Q(s_{bw,t}, s_{bo,t}, a_{bw,t}, a_{bo,t}))$
    Update $s_{bw,t}$ to $s_{bw,t+1}$ and $s_{bo,t}$ to $s_{bo,t+1}$
    Update $a_{bw,t}$ to $a_{bw,t+1}$ and $a_{bo,t}$ to $a_{bo,t+1}$
  end for
end for

node. Our evaluation metric is the SNR, defined by the following equation,

$$\text{SNR} = \frac{P_r}{P_n}, \tag{2.14}$$

where P_n is the noise power defined as follows,

$$P_n = 2q(I_L + I_D)BW + 4KTBW, \tag{2.15a}$$

$$I_L = S.P, \tag{2.15b}$$

where I_L is the photocurrent generated by the incident light, q is the electric charge, K is the Boltzmann constant, and T is the temperature in Kelvin (K). The values of the receiver sensitivity S, the photodiode's dark current I_D, and shunt resistance R are usually shared in photodiode's data sheet. The evaluation of the link quality performance is done for four different environments, namely the pure seawater, the clean ocean, the coastal ocean, and the turbid harbor.

2.3 Simulation results and analysis

In this section, we will evaluate our models' performance by simulating four different mediums, namely the pure seawater, the clean ocean, the coastal ocean, and the turbid harbor. These mediums are defined by their attenuation coefficient c, defined as the sum of the absorption and scattering coefficients a and b. The pure sea water has the lowest attenuation coefficient since it is a very clear water with almost no suspended particles or organic matter. Whereas the clean ocean water has slightly more attenuation as it includes some dissolved salts, organic matter, and tiny particles. As for the coastal ocean water, it has stronger attenuation because it contains more sediments, plankton, organic matter, and pollutants. Finally, the turbid harbor water has the highest attenuation because it includes many suspended particles, mud, algae, pollutants, and bubbles. The simulations are realized using python 3.7. First, we present the simulation results for the beam width and beam orientation adaptations separately, then we also show the combined model's simulation results. Finally, we compare the link quality for our proposed methods and the uncertainty disk-based adaptation method Celik *et al.* (2019).

In order to evaluate the performance of our RL-based communication system, we consider a $5m \times 5m$ square area, where the transmitter is located at the origin. The receiver is initially located in the coordinates (2,2) of the space. However, due to the underwater environment's dynamicity, a random movement of the receiver is introduced. We consider a single LED transmitter with an output power $P_0 = 20mW$ and area $A_t = 10m^2$. The receiver has a PDB-S5971 photodiode with transmitter area $A_r = 1.1m^2$. The simulation parameters are summarized in Table 2.2. For each method, we share the results for one random position of the receiver, considering the pure seawater environment. We compare its success rate with the random policy one as a function of the number of simulations. In the random method, the policy is defined randomly at each time step. Eventually, we evaluate the link quality by comparing the SNR found using our RL-based methods to the proposed static method for four different underwater environments.

Table 2.2 Simulation parameters

Simulation area	$5m \times 5m$
P_0	20mW
A_r	1.1 mm^2
A_t	10 mm^2
Received power threshold	0.01mW
γ	0.9
α	0.1
ϵ	Decaying epsilon

2.3.1 Beamwidth adaptation for a P2P communication

Here, we show the results of our proposed beamwidth adaptation scheme to demonstrate its accuracy. For that, we define the set of states as the possible beamwidths. The states and actions sets and reward function for this method are defined in Table 2.3.

Table 2.3 RL model for the beamwidth adaptation method

States	$S_{bw} = [20, 40, 60, 80, 100]$ degrees
Actions	[Increase, Decrease, Keep] beamwidth
Reward	$r = \begin{cases} P_r(S_{bw}) & \text{if ACK} \\ 0 & \text{if no ACK} \end{cases}$

We define the beam orientation to pass from the receiver's estimated location, i.e., 45 degrees. Table 2.4 shows the optimal policy. Also, Fig. 2.3 illustrates the two-dimensional (2D) representation of the resulting beamwidth for the considered situation. According to our adaptation method's results, if the beamwidth is larger than 40 degrees, it decreases. If it is narrower than 40 degrees, then it is increased. Also, a beam narrower than 40 degrees will not cover the node. In contrast, a wider one will cover unnecessary space that will lead to energy dissipation and decrease the link quality. In Fig. 2.4, we evaluate the success rate of our proposed Q-learning and SARSA methods with the random method, where the beamwidth is decided randomly at each simulation run. The Q-learning method gave an increasing success rate reaching about 0.9 after 600 simulations. The SARSA method gave an increasing success rate that converges faster than the Q-learning method and reaches about 0.93 after about 100 iterations. The difference in the speed of convergence between the SARSA and Q-learning algorithms can be explained by the difference in these two algorithms' update equations. While the Q-learning update function includes the maximization function, the SARSA update function does not. The absence of the maximization function in the SARSA update equation ensures a linear Q value update and prevents the value function from divergence, unlike the Q-learning algorithm. On the other hand, the random method's success rate is a constant function around the value of 0.67. This is because this method is not learning with time and hence does not converge to an optimal bandwidth. Thus, for this case, it is better to opt for the SARSA algorithm to realise the beamwidth adaptation. It is important to mention here that the RL algorithm consistently learns the environment by trial and error. Hence in case of a sudden change of the node's position, the RL algorithm learns to adapt to the new situation by converging to the new optimal beamwidth.

Table 2.4 The optimal policy for a random receiver position

States	20	40	60	80	100
Policy	Increase	Keep	Decrease	Decrease	Decrease

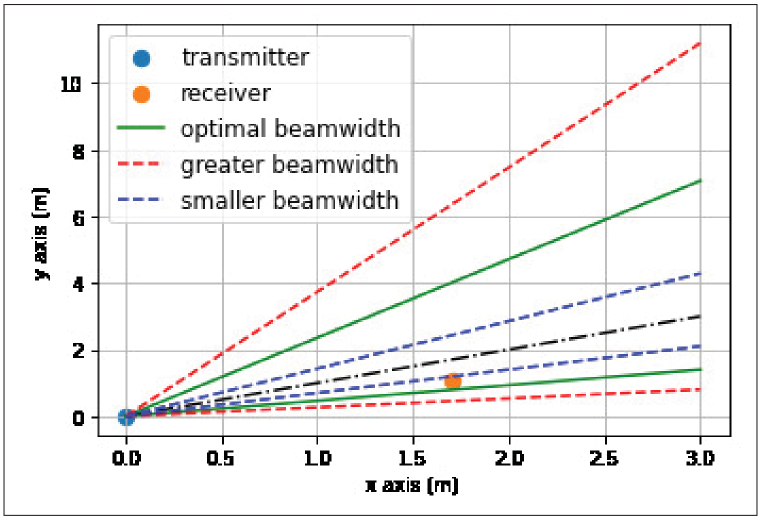


Figure 2.3 A 2D representation of the optimal beamwidth for a random receiver position

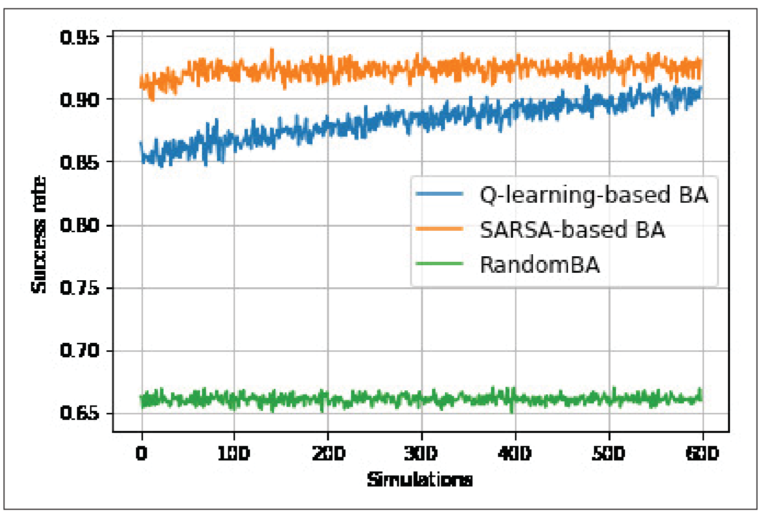


Figure 2.4 A comparison of the success rate between the proposed beamwidth adaptation method and the random method

2.3.2 Beam orientation adaptation for a P2P communication

Second, we consider the results of one run of the beam orientation adaptation system to prove our results' accuracy. We choose the states set, S_{bo}, to be the different possible beam orientations, defined as the angle between the horizontal axis and the beam axis. The states and actions sets and reward function for this method are defined in Table 2.5.

Table 2.5   RL model for the beam orientation adaptation method

States	$S_{bo} = [-80, -60, -40, -20, 0, 20, 40, 60, 80]$ degrees
Actions	[Up, Down, Stop] beam axis
Reward	$r = \begin{cases} P_r(S_{bo}) & \text{if ACK} \\ 0 & \text{if no ACK} \end{cases}$

While adapting the beam orientation, the light beamwidth is fixed to 20 degrees. Table 2.6 shows the optimal policy for a random receiver position, and Fig. 2.5 illustrates the 2D representation of the resulting beam orientation for the considered situation. As we can see from the shared results, if the beam's direction is higher than 40 degrees, it is tilted down. And if it is lower than 40 degrees, then it is taken up. Also, a higher or lower beam than the resulting one by our method will not cover the receiver node, and the packet would be lost. In Fig. 2.6, we evaluate the success rate of our proposed methods with the random method, where the beam orientation is decided randomly at each simulation run. The Q-learning method gave an increasing success rate function that reaches about 0.9 after 600 simulations. The SARSA method's success rate is also an increasing function at converges faster than the Q-learning method reaching 0.85 in around 50 iterations. Although the SARSA converged more quickly, the results found using Q-learning are better. On the other hand, the random method's success rate is a constant function around the value 0.3. This is because this method is not learning with time and hence does not converge to an optimal beam orientation. Thus, for this case, it is a tradeoff between the reliability and the speed of convergence. If the priority is given for the convergence speed, then it is better to opt for the SARSA algorithm to realise the beam orientation adaptation. Otherwise, Q-learning algorithm should be chosen instead. It is worth noting here that the RL algorithm continually learns the environment through trial and error. As a result, in the event of a rapid change in the node's position, the RL algorithm learns to adapt to the new scenario by converging to the new optimal beam orientation.

Table 2.6 The optimal policy for a random receiver position

States	-80	-60	-40	-20	0	20	40	60	80
Policy	Up	Up	Up	Up	Up	Up	Stop	Down	Down

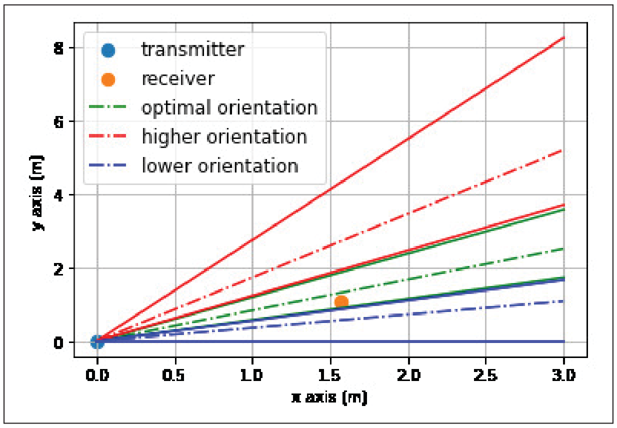


Figure 2.5 A 2D representation of the optimal beam orientation for a random receiver position

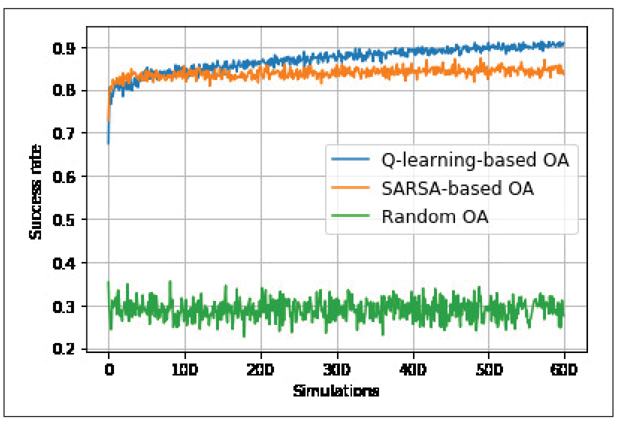


Figure 2.6 A comparison of the success rate between the proposed beam orientation adaptation method and the random method

2.3.3 Beamwidth and beam orientation adaptation for a P2P communication

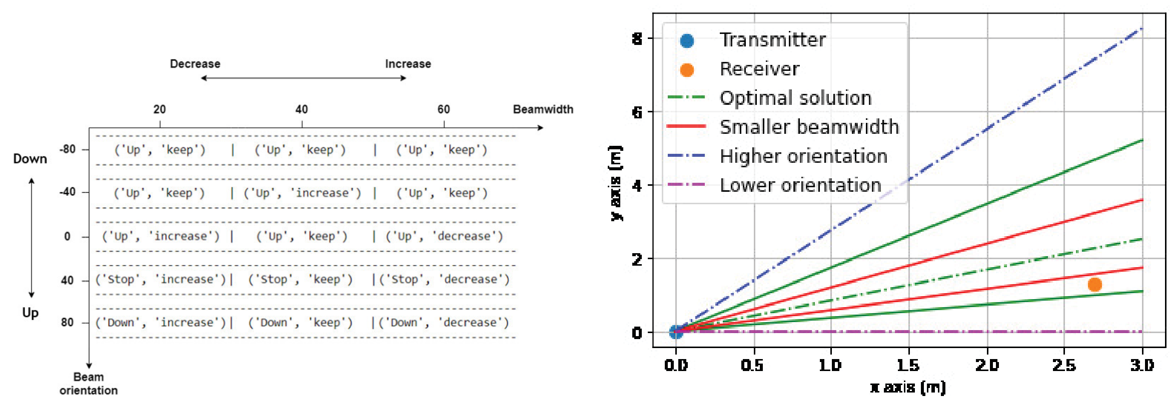
Now, we consider our proposed robust scheme that simultaneously caters for beamwidth and beam orientation. The state is defined using both the beamwidth and beam orientation values. The states, actions, and reward function for this method are defined in Table 2.7.

Table 2.7 RL model for the beam orientation adaptation method

States	S_{bw} = [20, 40, 60] degrees S_{bo} = [−80, −40, 0, 40, 80] degrees
Actions	A_{bw} = [Increase, Decrease, Keep] beamwidth A_{bo} = [Up, Down, Stop] beam axis
Reward	$r = \begin{cases} P_r(S_{bw}, S_{bo}) & \text{if ACK} \\ 0 & \text{if no ACK} \end{cases}$

Fig. 2.7 shows the converged policy and the 2D representation of the resulting beamwidth and beam orientation for the considered situation. The map depicted in Fig. 2.7 (a) can be understood as follows. Changing the beam orientation up or down can be accomplished by moving one step down or up on the map, respectively. While increasing or decreasing the beamwidth can be translated by moving one step right or left in the map, respectively. For instance, if we decide ('Up, 'Increase') at a specific place, then we move one step down and one step right in the map. For the sake of simplicity in Fig. 2.7 (b), we plotted the beam axis of the lower and higher orientation only. As we can see from this figure, the beam axis of the solution given by our method is the closest to the receiver node, compared to the higher and lower directions. Also, a light signal with the beamwidth decided by our method covers the receiver node, while a narrower beam is not in the receiver's light of sight. This proves the optimality of the converged beam. Moreover, from any random initial state of the light beam, it is possible to reach the optimal light beamwidth and beam orientation following our algorithm's optimal policy. In Fig. 2.8, we evaluate the success rate of our proposed methods with the random method, where the beamwidth and beam orientation are decided randomly at each simulation run. The Q-learning success rate is an increasing function that reaches around 0.8 in 600 iterations. The SARSA method also gave an increasing success rate with a slightly better value than the Q-learning

function. This method reached approximately 0.82 in 600 iterations. As for the random method, it gave a constant success rate around the value 0.2. This is because this method is not learning with time and hence does not converge to an optimal bandwidth. Thus, for this case, it is better to opt for the SARSA algorithm to realise the beamwidth and beam orientation adaptation. Similarly, in case of a sudden position change, the RL algorithm learns to adapt to the new situation by converging to the new optimal beamwidth and beam orientation.



a) The optimal policy for a random receiver position.

b) A 2D representation of the optimal beamwidth and beam orientation for a random receiver position.

Figure 2.7 The corresponding results of one run of the beamwidth and beam orientation adaptation system

2.3.4 Evaluation of the link quality for the proposed methods

In this subsection, we compare our methods' link quality with the link quality of the uncertainty disk static method. The adopted performance metric is the SNR. For simulation purposes, we assume an underwater environment of temperature 20 degrees Celsius (293.15K). The parameters I_D, R, S, and BW of the PDB-S5971 photodiode are specified in its paper sheet. The simulation parameters are shared in Table 2.8. As briefly discussed earlier, we consider four simulation environments, namely the pure seawater, the clean ocean, the coastal ocean, and the turbid harbor. The absorption, scattering, and attenuation coefficients of the different underwater environments are shared in Table 2.9, for a chosen wavelength $\lambda = 514nm$. The link quality is

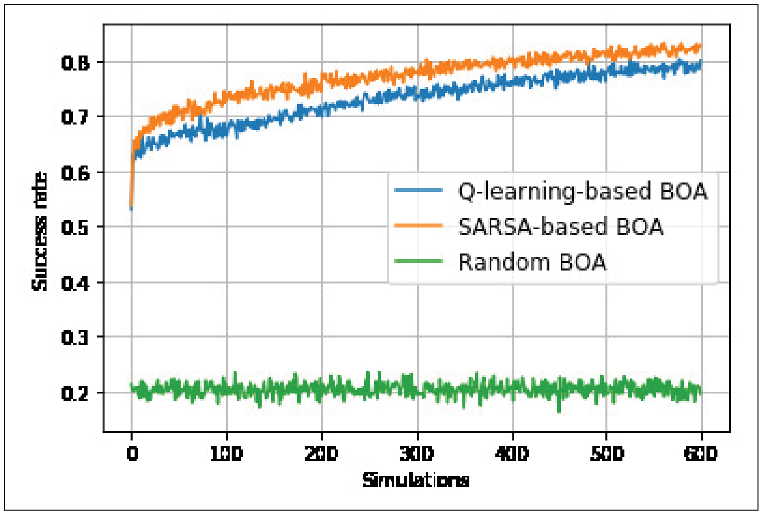


Figure 2.8 A comparison of the success rate between the proposed beamwidth and beam orientation adaptation method and the random method

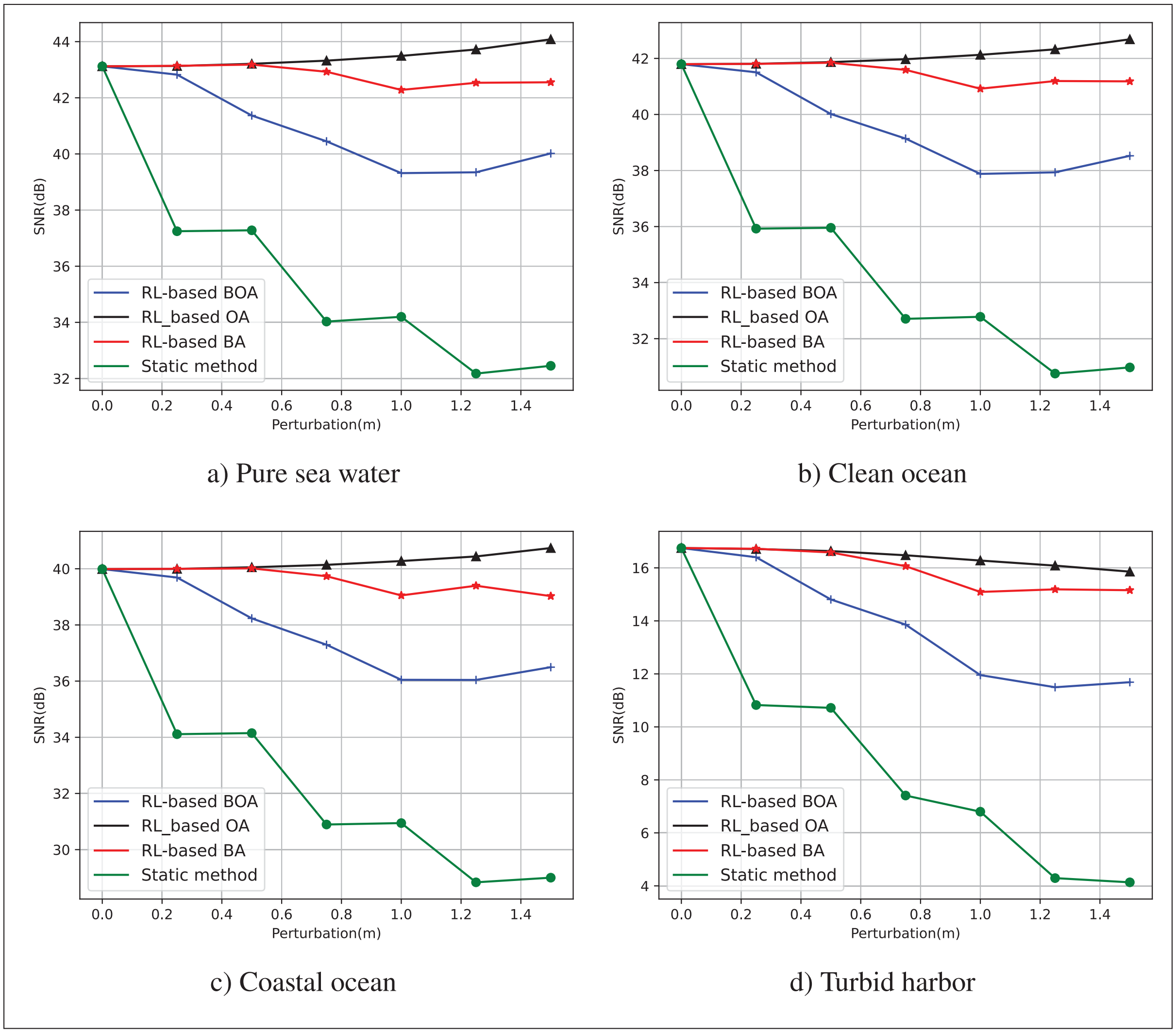
evaluated as a function of the water perturbation, defined as the possible distance the node can travel in meters due to underwater dynamicity. The simulation results are shared in Fig. 2.9.

Table 2.8 Performance evaluation parameters

q	1.6×10^{-19} C
K	1.38×10^{-23} J/K
T	293.15 K
I_D	1 nA
BW	1 GHz
R	$1.43 \times 10^{9}\ \Omega$
S	0.39 A/W

Table 2.9 Absorption, scattering, and attenuation coefficient of various types of water for $\lambda = 514$ nm. Taken from: Mobley (1994)

Water type	$a(m^{-1})$	$b(m^{-1})$	$c(m^{-1})$
Pure sea water	0.0405	0.0025	0.043
Clean ocean	0.114	0.037	0.151
Coastal ocean	0.179	0.219	0.298
Turbid harbor	0.266	1.824	2.19



a) Pure sea water

b) Clean ocean

c) Coastal ocean

d) Turbid harbor

Figure 2.9 Comparison of the Signal quality between the proposed models and the static model for different water types

As shown in Fig. 2.9, the link quality is strongly dependent on the underwater environment. While the pure seawater guarantees a high SNR ranging between 32dB and 44dB, the turbid harbor water gives a significantly lower SNR ranging between 4dB and 17dB. This is mainly because the turbid harbor water's attenuation is remarkably higher than the pure seawater's attenuation. The RL-based beam orientation adaptation (OA) method gave the best results for the four environments, followed by the RL-based beamwidth adaptation (BA) method, then the beamwidth and beam orientation adaptation (BOA) method. The three RL-based adaptation methods outperformed the uncertainty disk static method for all water types. Furthermore, the uncertainty disk static method gave remarkably low SNR for high perturbations of the water for the turbid harbor, while our proposed methods kept an acceptable SNR above 11dB. Thus, our proposed methods outperform the uncertainty disk static method in terms of link quality, demonstrating our proposed schemes' superiority.

2.4 Conclusion

In this paper, we proposed an RL-based beam adaptation method for a P2P UOWC link. This adaptation includes both the beamwidth and the beam orientation adaptations. We aimed to guarantee stable communication under nodes' location uncertainty to ensure high reliability and link quality. To achieve this aim, we proposed three RL-based adaptation methods. The first and second methods find the optimal beamwidth and beam orientation, respectively. The third method combines these two parameters to decide upon the optimal policy. We proposed two model-free algorithms, namely Q-learning and SARSA, for these methods. We evaluated our methods by comparing its success rate evolution with time to the random method. The performance of the two proposed algorithms is mainly dependent on the case. However, the SARSA algorithm proved to converge faster than the Q-learning algorithm. We also compared its resulting link quality to the uncertainty disk method, based on the SNR metric. Our methods gave better results than the uncertainty disk model for four different underwater environments, namely pure seawater, clean ocean, coastal ocean, and turbid harbor. We intend to further assess the effect of oceanic turbulence and PEs on the system performance to ensure more robust UOWC.

CHAPTER 3

OUTAGE PROBABILITY ANALYSIS FOR UOWC SYSTEM OVER OCEANIC TURBULENCE WITH POINTING ERRORS

Imene Romdhane[1] , Ziyaur Rahman[1] , Georges Kaddoum[12] , Ankur Bansal[3] , S. M. Zafaruddin[4]

[1] Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame west, Montreal, Canada H3C 1K3.

[2] Cyber Security Systems and Applied AI Research Center, Lebanese American University,
Beirut 1102 2801, Lebanon

[3] Department of Electrical Engineering, Indian Institute of Technology Jammu,
Jagti Campus, Jammu and Kashmir 181221,India

[4] Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science,
Pilani, Pilani, Rajasthan 333031, India

Paper published in IEEE Communications Letters, vol. 27, Issue. 11, pp. 2988 - 2992, November 2023.

3.1 Introduction

Underwater optical wireless communication (UOWC) systems have recently gained remarkable interest for scientific submarine and oceanic applications Saeed *et al.* (2019a); Celik *et al.* (2022). Compared with radio frequency (RF) and acoustic communications, UOWC ensures a higher throughput owing to the large bandwidth available. Nevertheless, underwater communications experience oceanic turbulence, path loss, and pointing errors (PEs), which degrade the link quality. Oceanic turbulence results from random alternations in the UOWC channel's refractive index, due to random variations in the salinity, water temperature, and air bubbles. Oceanic path loss is caused by molecular absorption and scattering effects, generally modeled by the extinction coefficient. PEs occur in case of misalignment between the transmitter and detector aperture and can lead to information loss. Hence, to enhance the quality of service (QoS), it is important to improve the performance of UOWC systems over different channel impairments.

To better evaluate the performance of UOWC, an exact turbulence-induced fading characterization, considering different underwater conditions, is required. The authors in Salcedo-Serrano, Boluda-Ruiz, Garrido-Balsells & García-Zambrana (2022); Zedini *et al.* (2019); Yang, Zhu, Li, Ansari & Yu (2021); Zedini, Kammoun, Soury, Hamdi & Alouini (2020); Li *et al.* (2021); Rahman *et al.* (2022); Bansal *et al.* (Early Access, May 2023) focused on the assessment of UOWC systems under oceanic turbulence and PEs, considering the temperature gradient, water salinity, and air bubbles. The work in Salcedo-Serrano *et al.* (2022) evaluated the performance of UOWC systems, assuming the Weibull distribution for salinity-induced turbulence. In Zedini *et al.* (2019), Zedini et al. presented the mixture exponential-generalized gamma (EGG) distribution to model air bubbles and temperature-induced oceanic turbulence in UOWC channels. The ability of the EGG model to represent different underwater channel conditions in a unified manner has led researchers to adopt it in evaluating the performance of UOWC systems. Researchers in Yang *et al.* (2021); Zedini *et al.* (2020); Li *et al.* (2021) studied dual-hop UOWC systems based on the EGG turbulence model. In Rahman *et al.* (2022), the case of a multi-layer vertical UOWC link using EGG distribution was investigated. In Bansal *et al.* (Early Access, May 2023), Rahman et. al. assessed the performance of the selection combining (SC) and maximum ratio combining (MRC)-based multi-aperture reception using the EGG model for the UOWC channel irradiance fluctuations. The optical intelligent reflecting surface (OIRS) is a recently emerging technology that comes in handy in confronting channel blockages and obstacles and enhancing underwater connectivity. The OIRS comprises a planar array of reflecting surfaces that forward the signal in a specific direction. Although initially intended for RF communications, new OIRS-assisted UOWC solutions have recently emerged in the literature. In Ata *et al.* (2023); Naik & Chung (2022), the authors proposed an IRS-assisted UOWC to enhance the system's performance in occlusion or blockage due to obstacles, oceanic turbulence, beam attenuation, and PEs.

As mentioned above, the performance of an MRC-based UOWC was investigated in Bansal *et al.* (Early Access, May 2023). However, only the upper bound of the outage probability was derived. In addition, to the best of the authors' knowledge, no work adopting the OIRS-assisted UOWC has used the EGG model for oceanic turbulence. Hence, the main contributions of the presented work can be summarized as follows:

- We derive the exact outage probability expression for the MRC-based UOWC link over EGG oceanic turbulence and PEs effects.
- We also derive the outage probability for the MRC-based UOWC link over high signal-to-noise ratio (SNR) values and obtain an analytical expression of the diversity order.
- In the absence of a direct link between the transmitter and receiver, we obtain the exact outage probability for the OIRS-assisted UOWC link over EGG oceanic turbulence and PEs effects.
- We verify the accuracy of the obtained analytical results with simulations.

3.2 System Model

We consider both the MRC-based direct link and the OIRS-assisted indirect link. We assume an independent and non-identically distributed (i.ni.d) EGG fading channel with transceiver antenna misalignment, leading to PEs. We assume the inter-symbol interference (ISI) in our model to be negligible to provide an in-depth analysis of the oceanic turbulence and PEs effects. The received signal y_i using a noncoherent intensity modulation/direct detection (IM/DD) scheme, is represented as:

$$y_i = h_{l,i} h_i s + n_i, \tag{3.1}$$

where $h_{l,i} = e^{-\alpha l}$ is the underwater path loss coefficient, l is the link distance expressed in meters, and α is the extinction attenuation coefficient. For the MRC-based UOWC link, $h_i = h_{i,MRC} = h_{t,i} h_{p,i}$, and for the OIRS-assisted UOWC link, $h_i = h_{i,OIRS} = \sum_{i=1}^{N_E} h_{i,1} g_{i,2}$, with $h_{i,1} = h_{t,i} h_{p,i}$, $g_{i,2} = g_{t,i} g_{p,i}$. $h_{t,i}$, and $g_{t,i}$, being the fading channel coefficients between the light source and i^{th} aperture ($i \in \{1, \cdots, N_A\}$), the light source and the i^{th} OIRS element, and the i^{th} OIRS element and the destination ($i \in \{1, \cdots, N_E\}$), respectively. $h_{p,i}$, and $g_{p,i}$ define the misalignment errors (i.e., PEs) for each of the previous segments, respectively. The transmitted signal is represented by s, and n_i denotes the additive white Gaussian noise (AWGN) with variance $\sigma_{n_i}^2$. Adopting on-off-keying (OOK) signaling, the instantaneous SNR is defined as

$$\gamma_i = \frac{P_t^2 h_{l,i}^2 h_i^2}{\sigma_{n_i}^2} = \gamma_0 h_i^2, \tag{3.2}$$

where P_t is the optical transmit power such that $s \in \{0, P_t\}$ and

$$\gamma_0 = \frac{P_t^2 h_{l,i}^2}{\sigma_{n_i}^2}. \tag{3.3}$$

$h_{i,MRC}$, $h_{i,1}$, and $g_{i,2}$ follow the proposed EGG oceanic turbulence model with PEsYang *et al.* (2021). The resulting probability density function (PDF) is given as Yang *et al.* (2021):

$$\begin{aligned} f_{h_{tp,i}}(x) = {} & \frac{\omega_i \rho_i^2}{x} G_{1,2}^{2,0}\left(\begin{matrix} \rho_i^2 + 1 \\ 1, \rho_i^2 \end{matrix} \middle| \frac{x}{\lambda_i A_i} \right) \\ & + \frac{(1-\omega_i)\rho_i^2}{x\Gamma(a_i)} G_{1,2}^{2,0}\left(\begin{matrix} \frac{\rho_i^2}{c_i} + 1 \\ a_i, \frac{\rho_i^2}{c_i} \end{matrix} \middle| \left(\frac{x}{b_i A_i}\right)^{c_i} \right), \end{aligned} \tag{3.4}$$

where λ_i, a_i, b_i, and c_i are the exponential distribution parameters, and ω_i is the distributions' mixture coefficient (i.e., $0 < \omega_i < 1$). $A_i = \mathrm{erf}(\upsilon_i)^2$ with $\upsilon = \sqrt{\pi/2}\, r/\omega_{z_i}$, where ω_{z_i} is the beam width and r is the aperture radius. $\rho_i = \frac{\omega_{z\mathrm{eq_i}}}{2\sigma_{s_i}}$ where $\omega_{z_{\mathrm{eq_i}}}$ is the equivalent beam width at the receiver and $\sigma_{s_i}^2$ is the variance of PEs displacement described by the horizontal sway and elevation Farid & Hranilovic (2007). Here, $\Gamma(\cdot)$ denotes the Gamma function, and $G_{p,q}^{m,n}(\cdot)$ is the Meijer G-function.

3.3 Performance analysis

Here, we focus on the statistical analysis of the multi-aperture UOWC system considering the EGG distribution for the oceanic turbulence and PEs. Additionally, we derive analytical formulas for the PDF and CDF to evaluate the performance of MRC-based and OIRS-assisted UOWC links.

3.3.1 MRC-based UOWC Link Analysis

MRC is one of the most used diversity schemes, where the sum of statistical characterization is required J. Zheng *et al.* (2019). In MRC, the same signal is collected at the receiver side from different paths using different antennas. The resulting signal is a weighted sum of all the received signals, where each weight is proportional to its corresponding signal amplitude. We find the exact SNR PDF ($f_{\gamma_{N_A}}(\gamma) = \mathcal{L}^{-1}\prod_{i=1}^{N_A} M_{\gamma_i}(s)$) and CDF ($F_{\gamma_{N_A}}(\gamma) = \mathcal{L}^{-1}\prod_{i=1}^{N_A}\frac{M_{\gamma_i}(s)}{s}$) of SNR of i^{th} aperture UOWC channel $\gamma_{N_A} = \sum_{i=1}^{N_A}\gamma_i$ using the product of moment-generating functions (MGFs) ($M_{\gamma_i}(s) = \int_0^\infty e^{-s\gamma} f_{\gamma_i}(\gamma)d\gamma$) with the inverse Laplace transform.

Theorem 1. *The exact SNR PDF and CDF of N_A-aperture MRC-based UOWC link over an EGG fading channel with PEs are given as:*

$$\begin{aligned} f_{\gamma_{N_A}}(\gamma) &= \prod_{i=1}^{N_A}\frac{\omega_i\rho_i^2}{2\gamma} H^{0,0:2,1;\cdots;2_{N_A},1_{N_A}}_{0,1:2,2;\cdots;2_{N_A},2_{N_A}} \\ &\times\left[\begin{matrix} -:U_1 \\ (1;\frac{1}{2},\cdots,\{\frac{1}{2}\}_{N_A}):U_2 \end{matrix}\middle| \left\{\frac{\sqrt{\gamma}}{\lambda_i A_i\sqrt{\gamma_0}}\right\}_{i=1}^{N_A}\right] \\ &+\prod_{i=1}^{N_A}\frac{(1-\omega_i)\rho_i^2}{2\gamma\Gamma(a_i)} H^{0,0:2,1;\cdots;2_{N_A},1_{N_A}}_{0,1:2,2;\cdots;2_{N_A},2_{N_A}} \\ &\times\left[\begin{matrix} -:U_3 \\ (1;\frac{c_i}{2},\cdots,\frac{c_{N_A}}{2}):U_4 \end{matrix}\middle| \left\{\left(\frac{\sqrt{\gamma}}{b_i A_i\sqrt{\gamma_0}}\right)^{c_i}\right\}_{i=1}^{N_A}\right] \end{aligned} \tag{3.5}$$

$$\begin{aligned} F_{\gamma_{N_A}}(\gamma) &= \prod_{i=1}^{N_A}\frac{\omega_i\rho_i^2}{2} H^{0,0:2,1;\cdots;2_{N_A},1_{N_A}}_{0,1:2,2;\cdots;2_{N_A},2_{N_A}} \\ &\times\left[\begin{matrix} -:U_1 \\ (0;\frac{1}{2},\cdots,\{\frac{1}{2}\}_{N_A}):U_2 \end{matrix}\middle| \left\{\frac{\sqrt{\gamma}}{\lambda_i A_i\sqrt{\gamma_0}}\right\}_{i=1}^{N_A}\right] \\ &+\prod_{i=1}^{N_A}\frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)} H^{0,0:2,1;\cdots;2_{N_A},1_{N_A}}_{0,1:2,2;\cdots;2_{N_A},2_{N_A}} \\ &\times\left[\begin{matrix} -:U_3 \\ (0;\frac{c_i}{2},\cdots,\frac{c_{N_A}}{2}):U_4 \end{matrix}\middle| \left\{\left(\frac{\sqrt{\gamma}}{b_i A_i\sqrt{\gamma_0}}\right)^{c_i}\right\}_{i=1}^{N_A}\right] \end{aligned} \tag{3.6}$$

where $U_1 = \{(1,\frac{1}{2}),(\rho_1^2+1,1);\cdots;\{(1,\frac{1}{2})\}_{N_A},(\rho_{N_A}^2+1,1)\}$,

$U_2 = \{(1,1), (\rho_1^2, 1); \cdots; (1,1), (\rho_{N_A}^2, 1)\}$, $U_3 = \{(1, \frac{c_i}{2}), (\frac{\rho_1^2}{c_1} + 1, 1); \cdots; (1, \frac{c_{N_A}}{2}), (\frac{\rho_{N_A}^2}{c_{N_A}} + 1, 1)\}$, *and* $U_4 = \{(a_i, 1), (\frac{\rho_1^2}{c_1}, 1); \cdots; (a_{N_A}, 1), (\frac{\rho_{N_A}^2}{c_{N_A}}, 1)\}$.

Proof 1. *To find the exact SNR PDF and CDF of i^{th} aperture in the MRC-based UOWC link, we first derive the MGFs ($M_{\gamma_i}(s) = \int_0^\infty e^{-s\gamma} f_\gamma(\gamma) d\gamma$) using* (3.4) *with $x = \sqrt{\frac{\gamma}{\gamma_0}}$. Then applying the definition of Meijer G-function, the resulting $M_{\gamma_i}(s)$ function is given as:*

$$\begin{aligned} M_{\gamma_i}(s) = {} & \frac{\omega_i \rho_i^2}{2} \frac{1}{2\pi\jmath} \int_{\mathcal{L}} \frac{\Gamma(1-r)\Gamma(\rho_i^2 - r)}{\Gamma(\rho_i^2 + 1 - r)} \left(\frac{1}{\lambda_i A_i \sqrt{\gamma_0}}\right)^r \\ & \times \underbrace{\int_0^\infty \gamma^{\frac{1}{2}r-1} e^{-s\gamma} d\gamma}_{I_1} dr + \frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)} \frac{1}{2\pi\jmath} \int_{\mathcal{L}} \\ & \times \frac{\Gamma(a_i - r)\Gamma(\frac{\rho_i^2}{c_i} - r)}{\Gamma(\frac{\rho_i^2}{c_i} + 1 - r)} \left(\frac{1}{(b_i A_i \sqrt{\gamma_0})^{c_i}}\right)^r \underbrace{\int_0^\infty \gamma^{\frac{c_i}{2}r-1} e^{-s\gamma} d\gamma}_{I_2} dr \end{aligned} \tag{3.7}$$

We solve the inner integrals $I_1 = \int_0^\infty \gamma^{\frac{1}{2}r-1} e^{-s\gamma} d\gamma = \frac{1}{s^{\frac{1}{2}r}} \Gamma(\frac{1}{2}r)$ and $I_2 = \int_0^\infty \gamma^{\frac{c_i}{2}r-1} e^{-s\gamma} d\gamma = \frac{1}{s^{\frac{c_i}{2}r}} \Gamma(\frac{c_i}{2}r)$ in (3.8) *using the identity $\int_0^\infty x^{\nu-1} e^{-\mu x} dx = \frac{1}{\mu^\nu} \Gamma(\nu)$ (D. Zwillinger, 2014, 3.381.4) and applying the definition of Fox H-function $H_{p,q}^{m,n} \left[\begin{matrix} (a_1, A_1), (a_2, A_2), \cdots, (a_p, A_p) \\ (b_1, B_1), (b_2, B_2), \cdots, (b_q, B_q) \end{matrix} \middle| z \right] = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j + B_j s) \prod_{j=1}^n \Gamma(1 - a_j - A_j s)}{\prod_{j=m+1}^q \Gamma(1 - b_j - B_j s) \prod_{j=n+1}^p \Gamma(a_j + A_j s)} z^{-s}$ (Mathai, Saxena & Haubold, 2009, Eqn. 1.2) to get:*

$$\begin{aligned} M_{\gamma_i}(s) = {} & \frac{\omega_i \rho_i^2}{2} H_{2,2}^{2,1} \left[\begin{matrix} (1, \frac{1}{2}), (\rho_i^2 + 1, 1) \\ (1,1), (\rho_i^2, 1) \end{matrix} \middle| \frac{1}{\lambda_i A_i \sqrt{s\gamma_0}} \right] \\ & + \frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)} H_{2,2}^{2,1} \left[\begin{matrix} (1, \frac{c_i}{2}), (\frac{\rho_i^2}{c_i} + 1, 1) \\ (a_i, 1), (\frac{\rho_i^2}{c_i}, 1) \end{matrix} \middle| \frac{1}{(b_i A_i \sqrt{s\gamma_0})^{c_i}} \right] \end{aligned} \tag{3.8}$$

To find $f_{\gamma_N}(\gamma) = \mathcal{L}^{-1} \prod_{i=1}^N M_{\gamma_i}(s)$, we use (3.8)*, expand the definition of Fox's H-function, and interchange the order of integration:*

$$
\begin{aligned}
f_{\gamma_{N_A}}(\gamma) = & \prod_{i=1}^{N_A} \frac{\omega_i \rho_i^2}{2} \left(\frac{1}{2\pi\jmath}\right)^{N_A} \int_{\mathcal{L}_i} \frac{\Gamma(1-r_i)\Gamma(\rho_i^2 - r_i)}{\Gamma(\rho_i^2 + 1 - r_i)} \\
& \times \Gamma(\frac{r_i}{2}) \left(\frac{1}{\lambda_i A_i \sqrt{\gamma_0}}\right)^{r_i} (\frac{1}{2\pi\jmath} \underbrace{\int_{\mathcal{L}} s^{-\sum_{i=1}^{N_A} \frac{r_i}{2}} e^{s\gamma} ds}_{I_3}) dr_i \\
& + \prod_{i=1}^{N_A} \frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)} \left(\frac{1}{2\pi\jmath}\right)^{N_A} \int_{\mathcal{L}_i} \frac{\Gamma(a_i - r_i)\Gamma(\frac{\rho_i^2}{c_i} - r_i)}{\Gamma(\frac{\rho_i^2}{c_i} + 1 - r_i)} \\
& \times \Gamma(\frac{c_i r_i}{2}) \left(\frac{1}{(b_i A_i \sqrt{\gamma_0})_i^c}\right)^{r_i} (\frac{1}{2\pi\jmath} \underbrace{\int_{\mathcal{L}} s^{-\sum_{i=1}^{N_A} \frac{c_i r_i}{2}} e^{s\gamma} ds}_{I_4}) dr_i
\end{aligned}
\tag{3.9}
$$

To solve the inner integrals $I_3 = \int_{\mathcal{L}} s^{-\sum_{i=1}^{N_A} \frac{r_i}{2}} e^{s\gamma} ds = \left(\frac{1}{\gamma}\right)^{1-\sum_{i=1}^{N_A} \frac{r_i}{2}} \frac{2\pi\jmath}{\Gamma(\sum_{i=1}^{N_A} \frac{r_i}{2})}$ *and* $I_4 = \int_{\mathcal{L}} s^{-\sum_{i=1}^{N_A} \frac{c_i r_i}{2}} e^{s\gamma} ds = \left(\frac{1}{\gamma}\right)^{1-\sum_{i=1}^{N_A} \frac{c_i r_i}{2}} \frac{2\pi\jmath}{\Gamma(\sum_{i=1}^{N_A} \frac{c_i r_i}{2})}$*, we apply the identity (D. Zwillinger, 2014, 8.315.1) and using the multivariate Fox H-function definition in (Mathai* et al.*, 2009, A.1), we get the exact SNR PDF of i*th *aperture for MRC-based UOWC link in* (3.5)*. Next, to derive the SNR CDF, we use* $F_{\gamma_{N_A}}(\gamma) = \mathcal{L}^{-1} \prod_{i=1}^{N_A} \frac{M_{\gamma_i}(s)}{s}$ *in* (3.8)*, expand the definition of Fox's H-function, and interchange the order of integration:*

$$
\begin{aligned}
F_{\gamma_{N_A}}(\gamma) = & \prod_{i=1}^{N_A} \frac{\omega_i \rho_i^2}{2} \left(\frac{1}{2\pi\jmath}\right)^{N_A} \int_{\mathcal{L}_i} \frac{\Gamma(1-r_i)\Gamma(\rho_i^2 - r_i)}{\Gamma(\rho_i^2 + 1 - r_i)} \\
& \times \Gamma(\frac{r_i}{2}) \left(\frac{1}{\lambda_i A_i \sqrt{\gamma_0}}\right)^{r_i} (\frac{1}{2\pi\jmath} \underbrace{\int_{\mathcal{L}} s^{-1-\sum_{i=1}^{N_A} \frac{r_i}{2}} e^{s\gamma} ds}_{I_6}) dr_i \\
& + \prod_{i=1}^{N_A} \frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)} \left(\frac{1}{2\pi\jmath}\right)^{N_A} \int_{\mathcal{L}_i} \frac{\Gamma(a_i - r_i)\Gamma(\frac{\rho_i^2}{c_i} - r_i)}{\Gamma(\frac{\rho_i^2}{c_i} + 1 - r_i)} \\
& \times \Gamma(\frac{c_i}{2} r_i) \left(\frac{1}{(b_i A_i \sqrt{\gamma_0})_i^c}\right)^{r_i} (\frac{1}{2\pi\jmath} \underbrace{\int_{\mathcal{L}} s^{-1-\sum_{i=1}^{N_A} \frac{c_i}{2} r_i} e^{s\gamma} ds}_{I_7}) dr_i
\end{aligned}
\tag{3.10}
$$

To solve the inner integrals $I_6 = \int_{\mathcal{L}} s^{-1-\sum_{i=1}^{N_A}\frac{r_i}{2}} e^{s\gamma} ds = \left(\frac{1}{\gamma}\right)^{-\sum_{i=1}^{N_A}\frac{r_i}{2}} \frac{2\pi J}{\Gamma(1+\sum_{i=1}^{N_A}\frac{r_i}{2})}$ *and* $I_7 = \int_{\mathcal{L}} s^{-1-\sum_{i=1}^{N_A}\frac{c_i r_i}{2}} e^{s\gamma} ds = \left(\frac{1}{\gamma}\right)^{-\sum_{i=1}^{N_A}\frac{c_i r_i}{2}} \frac{2\pi J}{\Gamma(1+\sum_{i=1}^{N_A}\frac{c_i r_i}{2})}$*, we apply the identity (D. Zwillinger, 2014, 8.315.1) and using the definition of multivariate Fox H-function in (Mathai* et al.*, 2009, A.1), we get the exact SNR CDF of* i^{th} *aperture for MRC-based UOWC link in* (3.6)*.*

Remark 1. *The diversity order is a measure of how effectively the system can combat oceanic turbulence with PEs and improve its performance in challenging oceanic propagation environments. The outage probability* $P_{\text{out}} = P(\gamma \leq \gamma_{\text{th}})$*,* γ_{th} *being the SNR threshold, is an important metric to derive the system's diversity order. We can use the SNR CDF in* (3.6) *with* $\gamma = \gamma_{\text{th}}$ *to come up with an exact formula of the outage probability with EGG turbulence and PEs.*

In the next Lemma, we provide the asymptotic expression of the outage probability of i^{th} aperture of the MRC-based UOWC link.

Lemma 1. *The asymptotic expression of the outage probability of multi-aperture UOWC systems at high SNR* $\gamma_0 \to \infty$ *is given as:*

$$\begin{aligned} P_{\text{out}}^{\infty} = &\prod_{i=1}^{N_A} \frac{\omega_i \rho_i^2}{2\Gamma\left(1+\sum_{i=1}^{N_A} u_i\right)} \prod_{i=1}^{N_A} \frac{\Gamma(1-u_i)\Gamma(\rho_i^2-u_i)\Gamma(\frac{u_i}{2})}{\Gamma(\rho_i^2+1-u_i)} \\ &\times \left(\frac{1}{\lambda_i A_i}\right)^{u_i} \left(\frac{\gamma_{\text{th}}}{\gamma_0}\right)^{\frac{u_i}{2}} + \prod_{i=1}^{N_A} \frac{(1-\omega_i)\rho_i^2}{2\Gamma(a_i)\Gamma\left(1+\sum_{i=1}^{N_A} u_i\right)} \prod_{i=1}^{N_A} \\ &\times \frac{\Gamma(a_i-u_i)\Gamma(\frac{\rho_i^2}{c_i}-u_i)\Gamma(\frac{c_i u_i}{2})}{\Gamma(\frac{\rho_i^2}{c_i}+1-u_i)} \left(\frac{1}{b_i A_i}\right)^{c_i u_i} \left(\frac{\gamma_{\text{th}}}{\gamma_0}\right)^{\frac{c_i u_i}{2}} \end{aligned} \tag{3.11}$$

Proof 2. *Starting from* (3.6) *with* $\gamma = \gamma_{th}$ *and using (Abo Rahama, Ismail & Hassan, 2018, Eqn. (30)), we derive the asymptotic expression of the outage probability at high SNR* $\gamma_0 \to \infty$ *in* (3.11) *and compute the residue of multiple Mellin-Barnes integrals of the corresponding multivariate Fox's H-function at the dominant pole* $u_i = \min\{1, a_i c_i, \rho_i^2\}_{i=1}^{N_A}$*.*

Remark 2. *In order to obtain the diversity order of the i^{th} aperture of the UOWC system, we express* (3.11) *as* $P_{\text{out}}^{\infty} = \gamma_0^{-\mathrm{D_{MRC}}}$. *Considering the dominant pole u_i, $\forall i$ in* (3.11)*, the N_A-product of the term $\gamma_0^{-\frac{u_i}{2}}$ can be given as* $P_{\text{out}}^{\infty} = \gamma_0^{-\sum_{i=1}^{N_A}\frac{u_i}{2}}$*, Therefore, the diversity order of the system is given as* $\mathrm{D_{MRC}} = \sum_{\mathrm{i=1}}^{\mathrm{N_A}} \min\{\frac{1}{2}, \frac{\mathrm{a_i c_i}}{2}, \frac{\rho_{\mathrm{i}}^2}{2}\}$.

3.3.2 OIRS-assisted UOWC Link Analysis

OIRS is an efficient communication alternative in the case of underwater blockage and obstacles. It consists mainly of a planar array of reflecting surfaces that redirect the light beam toward the receiver.

Theorem 2. *The SNR CDF for N_E-element OIRS-assisted UOWC link over EGG fading channels with PEs is given as:*

$$
\begin{aligned}
F_{\gamma_{N_E}}(\gamma) &= \prod_{i=1}^{N_E} \frac{\omega_{i,1}\omega_{i,2}\rho_{i,1}^2\rho_{i,2}^2}{2} H_{0,1:3,4;\cdots;3_{N_E},4_{N_E}}^{0,0:4,1;\cdots;4_{N_E},1_{N_E}} \\
&\times \left[\begin{array}{c} -:V_1 \\ (0;\frac{1}{2},\cdots,\{\frac{1}{2}\}_{N_E}):V_2 \end{array} \middle| \left\{ \sqrt{\frac{\gamma}{\gamma_0}}\zeta_1 \right\}_{i=1}^{N_E} \right] \\
&+ \prod_{i=1}^{N_E} \frac{\omega_{i,1}(1-\omega_{i,2})\rho_{i,1}^2\rho_{i,2}^2}{2\Gamma(a_{i,2})} H_{0,1:3,4;\cdots;3_{N_E},4_{N_E}}^{0,0:4,1;\cdots;4_{N_E},1_{N_E}} \\
&\times \left[\begin{array}{c} -:V_3 \\ (0;\frac{c_{i,2}}{2},\cdots,\frac{c_{N_E,2}}{2}):V_4 \end{array} \middle| \left\{ \left(\sqrt{\frac{\gamma}{\gamma_0}}\zeta_2 \right)^{c_{i,2}} \right\}_{i=1}^{N_E} \right] \\
&+ \prod_{i=1}^{N_E} \frac{\omega_{i,2}(1-\omega_{i,1})\rho_{i,1}^2\rho_{i,2}^2}{2\Gamma(a_{i,1})} H_{0,1:1,6;\cdots;1_{N_E},6_{N_E}}^{0,0:4,1;\cdots;4_{N_E},1_{N_E}} \\
&\times \left[\begin{array}{c} -:V_5 \\ (0;\frac{c_{i,1}}{2},\cdots,\frac{c_{N_E,1}}{2}):V_6 \end{array} \middle| \left\{ \left(\sqrt{\frac{\gamma}{\gamma_0}}\zeta_3 \right)^{c_{i,1}} \right\}_{i=1}^{N_E} \right] \\
&+ \prod_{i=1}^{N_E} \frac{(1-\omega_{i,1})(1-\omega_{i,2})\rho_{i,1}^2\rho_{i,2}^2}{2c_{i,1}\Gamma(a_{i,1})\Gamma(a_{i,2})} H_{0,1:3,4;\cdots;3_{N_E},4_{N_E}}^{0,0:4,1;\cdots;4_{N_E},1_{N_E}} \\
&\times \left[\begin{array}{c} -:V_7 \\ (1;\frac{c_{i,2}}{2},\cdots,\frac{c_{N_E,2}}{2}):V_8 \end{array} \middle| \left\{ \left(\sqrt{\frac{\gamma}{\gamma_0}}\zeta_4 \right)^{c_{i,2}} \right\}_{i=1}^{N_E} \right]
\end{aligned} \tag{3.12}
$$

where $\zeta_1 = \frac{1}{\lambda_{i,1}\lambda_{i,2}A_{i,1}A_{i,2}}$, $\zeta_2 = \frac{1}{\lambda_{i,1}b_{i,2}A_{i,1}A_{i,2}}$, $\zeta_3 = \frac{1}{\lambda_{i,2}b_{i,1}A_{i,1}A_{i,2}}$, $\zeta_4 = \frac{1}{b_{i,1}b_{i,2}A_{i,1}A_{i,2}}$,

$V_1 = \{(1,\frac{1}{2}),(\rho_{i,1}^2+1,1),(\rho_{i,2}^2,1);\cdots;\{(1,\frac{1}{2})\}_{N_E},(\rho_{N_E,1}^2,1),(\rho_{N_E,2}^2,1)\}$,

$V_2 = \{(1,1),(\rho_{i,2}^2,1),(1,1),(\rho_{i,1}^2,1);\cdots;\{(1,1)\}_{N_E},(\rho_{N_E,2}^2,1),\{(1,1)\}_{N_E},(\rho_{N_E,1}^2,1)\}$,

$V_3 = \{(1,\frac{c_{i,2}}{2}),(\rho_{i,1}^2+1,c_{i,2}),(\frac{\rho_{i,2}^2}{c_{i,2}}+1,1);\cdots;(1,\frac{c_{N_E,2}}{2}),(\rho_{N_E,1}^2+1,c_{N_E,2}),(\frac{\rho_{N_E,2}^2}{c_{N_E,2}}+1,1)\}$,

$V_4 = \{(a_{i,2},1),(\frac{\rho_{i,2}^2}{c_{i,2}},1),(1,c_{i,2}),(\rho_{i,1}^2,c_{i,2});\cdots;(a_{N_E,2},1),(\frac{\rho_{N_E,2}^2}{c_{N_E,2}},1),(1,c_{N_E,2}),(\rho_{N_E,1}^2,c_{N_E,2})\}$,
$V_5 = \{(1,\frac{c_{i,1}}{2});\cdots;(1,\frac{c_{N_E,1}}{2})\}$, $V_6 = \{(a_{i,1},1),(\frac{\rho_{i,1}^2}{c_{i,1}},1),(1,c_{i,1}),(\rho_{i,2}^2,1),(\rho_{i,2}^2+1,c_{i,1}),(\frac{\rho_{i,1}^2}{c_{i,1}}+1,1);\cdots;(a_{N_E,1},1),(\frac{\rho_{N_E,1}^2}{c_{N_E,1}},1),(1,c_{N_E,1}),(\rho_{N_E,2}^2,1),(\rho_{N_E,2}^2+1,c_{N_E,1}),(\frac{\rho_{N_E,1}^2}{c_{N_E,1}}+1,1)\}$,

$V_7 = \{(1, \frac{c_{i,2}}{2}), (\frac{\rho_{i,1}^2}{c_{i,1}} + 1, 1), (\frac{\rho_{i,2}^2}{c_{i,2}} + 1, 1); \cdots ; (1, \frac{c_{N_E,2}}{2}), (\frac{\rho_{N_E,1}^2}{c_{N_E,1}} + 1, 1), (\frac{\rho_{N_E,2}^2}{c_{N_E,2}} + 1, 1)\}$,

and $V_8 = \{(a_{i,2}, 1), (\frac{\rho_{i,2}^2}{c_{i,2}}, 1), (a_{i,1}, \frac{c_{i,2}}{c_{i,1}}), (\frac{\rho_{i,1}^2}{c_{i,1}}, \frac{c_{i,2}}{c_{i,1}}); \cdots ; (a_{N_E,2}, 1), (\frac{\rho_{N_E,2}^2}{c_{N_E,2}}, 1), (a_{N_E,1}, \frac{c_{N_E,2}}{c_{N_E,1}}), (\frac{\rho_{N_E,1}^2}{c_{N_E,1}}, \frac{c_{N_E,2}}{c_{N_E,1}})\}$

Proof 3. *The product distribution of* $z_i = h_{i,1}g_{i,2}$ *can be expressed as*

$$f_{z_i}(z_i) = \int_0^\infty \frac{1}{|g_{i,2}|} f_{h_{i,1}}(z_i/g_{i,2}) f_{g_{i,2}}(g_{i,2}) dg_{i,2} \tag{3.13}$$

Substituting the PDF (i.e., (3.4)*) of* $h_{i,1}$ *and* $g_{i,2}$ *over oceanic turbulence with PEs in* (3.13)*, we obtain the product distribution of* z_i*. Using the identity* $G_{p,q}^{m,n}\left(\begin{matrix} a_p \\ b_q \end{matrix} \middle| z\right) = G_{q,p}^{n,m}\left(\begin{matrix} 1-b_q \\ 1-a_p \end{matrix} \middle| z^{-1}\right)$ *and substituting* $g_{i,2}^{c_{i,1}} = t$ *in the fourth integral and then applying the identities (Wolfram function,* 07.34.21.0011.01, 07.34.21.0012.01*) with* $z_i = \sqrt{\frac{\gamma}{\gamma_0}}$*, we get the SNR of a single element OIRS-assisted UOWC link as:*

$$\begin{aligned} f_{\gamma_i}(\gamma) = {} & \tfrac{\omega_{i,1}\omega_{i,2}\rho_{i,1}^2\rho_{i,2}^2}{2\gamma} G_{2,4}^{4,0}\left(\begin{matrix} V_9 \\ V_{10} \end{matrix} \middle| \sqrt{\tfrac{\gamma}{\gamma_0}}\zeta_1\right) \\ & + \tfrac{\omega_{i,1}(1-\omega_{i,2})\rho_{i,1}^2\rho_{i,2}^2}{2\gamma\Gamma(a_{i,2})} H_{2,4}^{4,0}\left[\begin{matrix} V_{11} \\ V_{12} \end{matrix} \middle| \left(\sqrt{\tfrac{\gamma}{\gamma_0}}\zeta_2\right)^{c_{i,2}}\right] \\ & + \tfrac{\omega_{i,2}(1-\omega_{i,1})\rho_{i,1}^2\rho_{i,2}^2}{2\gamma\Gamma(a_{i,1})} H_{4,2}^{0,4}\left[\begin{matrix} V_{13} \\ V_{14} \end{matrix} \middle| \left(\sqrt{\tfrac{\gamma_0}{\gamma}}\tfrac{1}{\zeta_3}\right)^{c_{i,1}}\right] \\ & + \tfrac{(1-\omega_{i,1})(1-\omega_{i,2})\rho_{i,1}^2\rho_{i,2}^2}{2\gamma c_{i,1}\Gamma(a_{i,1})\Gamma(a_{i,2})} H_{2,4}^{4,0}\left[\begin{matrix} V_{15} \\ V_{16} \end{matrix} \middle| \left(\sqrt{\tfrac{\gamma}{\gamma_0}}\zeta_4\right)^{c_{i,2}}\right] \end{aligned} \tag{3.14}$$

where $V_9 = \{\rho_{i,1}^2 + 1, \rho_{i,2}^2 + 1\}$*,* $V_{10} = \{1, \rho_{i,2}^2, 1, \rho_{i,1}^2\}$*,* $V_{11} = \{(1 + \rho_{i,1}^2, c_{i,2}), (\frac{\rho_{i,2}^2}{c_{i,2}} + 1, 1)\}$*,* $V_{12} = \{(a_{i,2}, 1), (\frac{\rho_{i,2}^2}{c_{i,2}}, 1), (1, c_{i,2}), (\rho_{i,1}^2, c_{i,2})\}$*,* $V_{13} = \{(1 - a_{i,1}, 1), (1 - \frac{\rho_{i,1}^2}{c_{i,1}}, 1), (0, c_{i,1}), (1 - \rho_{i,2}^2, c_{i,1})\}$*,* $V_{14} = \{(-\rho_{i,2}^2, c_{i,1}), (-\frac{\rho_{i,1}^2}{c_{i,1}}, 1)\}$*,* $V_{15} = \{(\frac{\rho_{i,1}^2}{c_{i,1}} + 1, \frac{c_{i,2}}{c_{i,1}}), (\frac{\rho_{i,2}^2}{c_{i,2}} + 1, 1)\}$*, and* $V_{16} = \{(a_{i,2}, 1), (\frac{\rho_{i,2}^2}{c_{i,2}}, 1), (a_{i_1}, \frac{c_{i,2}}{c_{i,1}}), (\frac{\rho_{i,1}^2}{c_{i,1}}, \frac{c_{i,2}}{c_{i,1}})\}$.

Table 3.1 EGG Distribution Parameters Li *et al.* (2021)

Salinity	BL	$(\omega, \lambda, a\ , b\ c)$
Fresh water	16.5	(0.5117,0.1602,0.0075,2.9963,216.8356)
Salty water	4.7	(0.2064,0.3953,0.5307,1.2154,35.7368)
PEs parameters	(A_0, ρ)	(0.1641,0.5244) , (0.5076,0.6079)

To find the SNR CDF $F_{\gamma_{N_E}}(\gamma) = \mathcal{L}^{-1}\prod_{i=1}^{N_E}\frac{M_{\gamma_i}(s)}{s}$ *of* i^{th} *element the OIRS-assisted UOWC link in Theorem 2, we follow the same procedure as applied in Theorem 1.*

Remark 3. *To avoid repetition and respect the space limit, the exact outage probability* $P_{\text{out}} = P(\gamma \leq \gamma_{\text{th}})$ *over EGG turbulence with PEs can be derived from the SNR CDF defined in* (3.12) *by replacing* γ *with* γ_{th}.

Remark 4. *To obtain the asymptotic expression of the outage probability for the OIRS-assisted UOWC link at high SNR* $\gamma_0 \rightarrow \infty$*, we can apply the same identity and procedure used in Lemma 1 on each multivariate Fox H-function in* (3.12)*. Consequently, the diversity order of the OIRS-assisted UOWC link can be obtained as* $\mathrm{D_{OIRS}} = \sum_{\mathrm{i=1}}^{\mathrm{N_E}} \min\{\frac{1}{2}, \frac{\mathrm{a_{i,1}c_{i,1}}}{2}, \frac{\mathrm{a_{i,2}c_{i,2}}}{2}, \frac{\rho_{\mathrm{i,1}}^2}{2}, \frac{\rho_{\mathrm{i,2}}^2}{2}\}$.

3.4 Simulation Results and Conclusion

In this section, we assess the outage probability of the UOWC system, for the direct link using the MRC scheme and the indirect link through the OIRS, by considering the EGG turbulence model with PEs. We verify the accuracy of our obtained mathematical expressions against Monte-Carlo (MC) simulations, where we average the results over 10^7 channel realizations. To calculate Meijer's G and Fox's H-function, we utilize the built-in libraries of MATLAB and Mathematica. We consider a distance between the source and receiver of $l = 50m$ and an extinction coefficient $\alpha = 0.056$. We assume gaussian white noise of variance $\sigma_{n_i}^2 = 10^{-4}A^2/GHz$ and threshold SNR $\gamma_{th} = 50dB$. The EGG oceanic turbulence and PEs parameters are summarized in Table 3.1.

First, we evaluate the effect of turbulence, including the water quality and bubble level (BL), and the effect of the number of apertures on the performance of the MRC-based UOWC link.

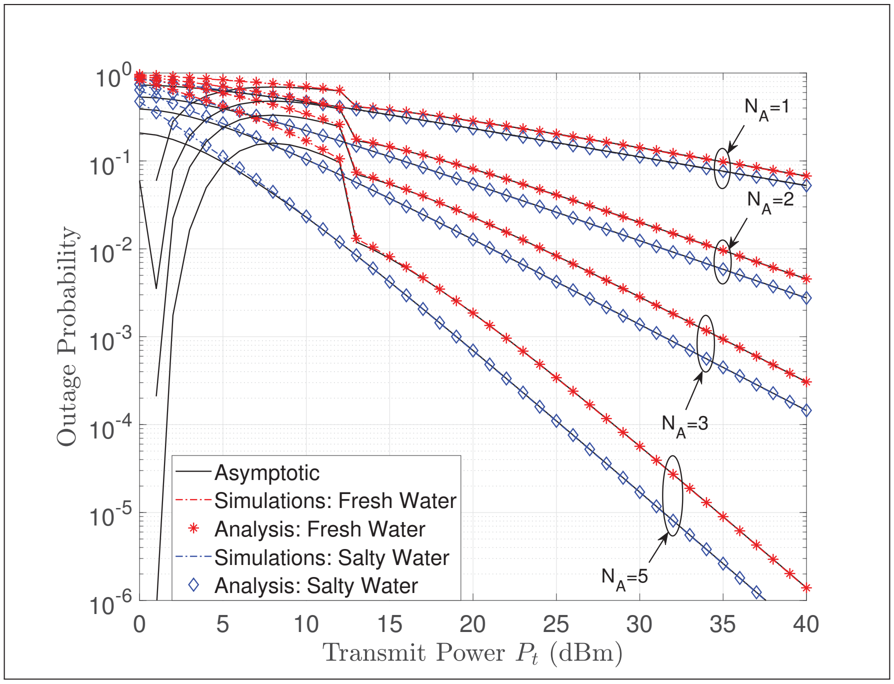


Figure 3.1 Outage probability performance of the MRC-based UWOC link with PEs parameters $A_0 = 0.5076$ and $\rho = 0.6079$

We consider salty water with $BL = 4.7$ and fresh water with $BL = 16.5$ with PEs parameters ($A_i = 0.5076$ and $\rho_i = 0.6079$). We also study the effect of changing the number of apertures from a single aperture ($N_A = 1$) to $N_A = 5$ on the outage probability. As shown in Fig. 3.1, the system performs better in salty water with $BL = 4.7$ than in fresh water with $BL = 16.5$. For example, at 40dBm transmit power, the outage probability of the first environment is 8.5% lower than that of the second environment. In addition, increasing the number of apertures improves the performance of the system. For instance, at 40dBm transmit power, increasing N_A from three to five decreases the outage probability by approximately 40% for fresh water. Furthermore, the fluctuations in the asymptotic plots at low transmit power are due to the asymptotic expression being derived for high power values, which makes its behavior at low transmit power unpredictable. Moreover, the fluctuations witnessed in the outage probability plot for fresh water at low transmit power can be explained by the higher BL values for fresh water (BL=16.5) compared to salty water (BL=2.4). We can also see from Figure. 3.1 that

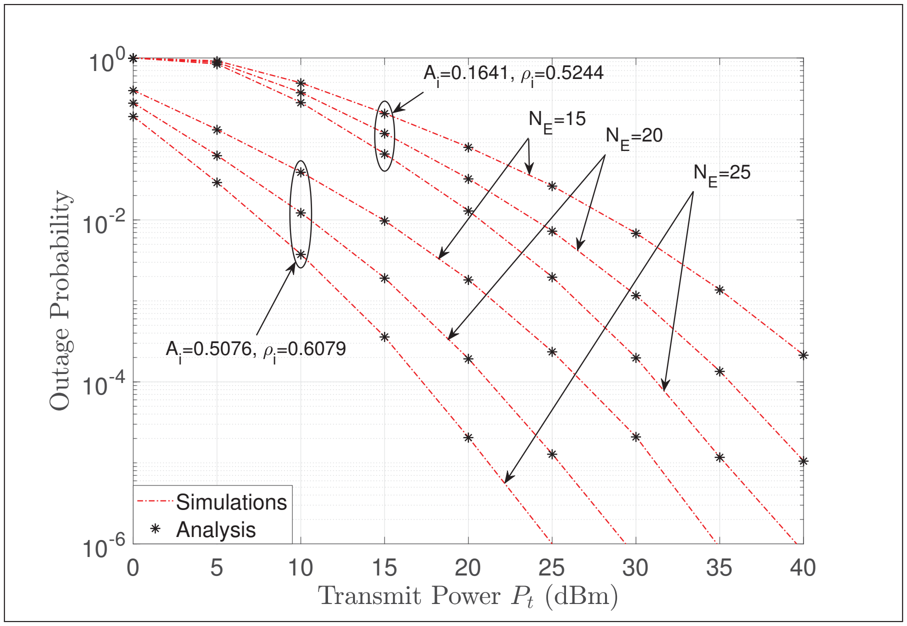


Figure 3.2 Outage probability performance of the OIRS-assisted UWOC link in fresh water with $BL = 16.5$

the slope of the outage probability matches the diversity order of the MRC-based UOWC link $\mathrm{D_{MRC}} = \sum_{\mathrm{i=1}}^{\mathrm{N_A}} \min\{\frac{1}{2}, \frac{\mathrm{a_i c_i}}{2}, \frac{\rho_\mathrm{i}^2}{2}\}$.

Finally, we examine the effect of the PEs and the number of reflecting elements on the performance of the OIRS-assisted UOWC link. We study the case of fresh water with $BL = 16.5$ for two different PEs parameters, as shown in Table 3.1. The number of reflecting elements also varies between $N_E = 15, 20, 25$. As shown in Figure. 3.2, the PEs parameters $(A_i = 0.5076, \rho_i = 0.6079)$ yield better results than the parameters $(A_i = 0.1641, \rho_i = 0.5244)$, since increasing A_i and ρ_i decreases the PEs effects. For instance, at 35dBm transmit power, using the first PEs tuple for $N_E = 15$ decreases the outage probability by approximately 53%, compared to using the second PEs values. In addition, increasing the number of reflecting elements N_E decreases the outage probability. For example, for the same transmit power, increasing N_E from 15 to 25 for $A_i = 0.1641$ and $\rho_i = 0.5244$ decreases the outage probability by approximately 42%. It can be seen from Figure. 3.2 that the slope of the outage probability matches the diversity order of the OIRS-assisted UOWC link $\mathrm{D_{OIRS}} = \sum_{\mathrm{i=1}}^{\mathrm{N_E}} \min\{\frac{1}{2}, \frac{\mathrm{a_{i,1} c_{i,1}}}{2}, \frac{\mathrm{a_{i,2} c_{i,2}}}{2}, \frac{\rho_{\mathrm{i,1}}^2}{2}, \frac{\rho_{\mathrm{i,2}}^2}{2}\}$.

To conclude, the performance of UOWC systems can be enhanced in the presence of oceanic turbulence and PEs by increasing the number of apertures in the MRC-based UOWC link and the number of reflecting elements in the OIRS-assisted link. Hence, the proposed combined system consisting of MRC-based and OIRS-assisted links improves the connectivity of high-speed UWOC systems under diverse channel conditions by increasing the link quality using an MRC scheme and including a backup link using OIRS in case of direct link failure due to obstacles. To further improve the reliability of UOWC links, we aim in future work to jointly optimize beamforming and power allocation in UOWC systems. Another important future work is to assess the performance of hybrid environment-based systems such as air-to-underwater optical communication.

CHAPTER 4

DDPG-BASED LINK ADAPTATION FOR DIRECT AIR-TO-UNDERWATER OPTICAL COMMUNICATION UNDER NOISY CSI CONDITIONS

Imene Romdhane[1] , Ziyaur Rahman[1] , Georges Kaddoum[1]

[1] Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame west, Montreal, Canada H3C 1K3.

Paper published in the 2024 20th International Conference on the Design of Reliable Communication Networks (DRCN),
06-09 May 2024.

4.1 Introduction

In recent times, there has been a growing interest among researchers in underwater communications. These underwater communication systems find applications across various fields, including remote monitoring of oceanic instruments, pollution tracking, and oceanography research. Consequently, the number of underwater devices in use is steadily increasing. In this context, managing power consumption is a significant challenge for system designers because expensive batteries typically power underwater devices. Radio frequency (RF) and acoustic communication methods were initially put forward for underwater systems. Although these communication modes offer cost-effective solutions and deployment flexibility, they have limitations. Specifically, RF communications suffer from high power consumption and signal attenuation issuesKulhandjian (2014). On the other hand, acoustic communications encounter notable transmission delays and bandwidth constraintsErol-Kantarci, Mouftah & Oktug (2011). In light of these limitations, underwater optical wireless communications (UOWC) has emerged as a promising alternative Celik *et al.* (2022). UOWC holds the potential to address the aforementioned challenges associated with RF and acoustic communication methods, making it an area of considerable interest and research activity in recent times.

Recently, many researchers assessed heterogeneous underwater-terrestrial networks as a means to alleviate underwater data traffic through the utilization of optical communication technologies.

These communication technologies offer several advantages, including high data rates in the order of Gbps, cost-effectiveness, and low power consumption. Additionally, they ensure minimal latency due to the high speed of light (approximately $2.55 \times 10^8 \mathrm{m/s}$). However, these technologies face challenges, notably high water absorption and scattering effects, which restrict the range of optical communication underwater to approximately 50 to 100 meters. The combined impact of absorption and scattering effects is quantified through the attenuation coefficient, denoted as c, which is mainly affected by the type of water.

The feasibility of direct air-to-underwater optical communication is mainly governed by the propagation conditions across both media and the air-water interface. In particular, optical signals can penetrate the water surface and propagate underwater when suitable wavelengths are employed, especially within the blue-green spectral window where seawater exhibits relatively low attenuation. However, for shallow underwater receivers, ambient solar radiation may introduce additional background noise at the photodetector, thereby degrading the received signal quality and increasing the probability of detection errors (Sticklus, Hieronymi & Hoeher (2018)). This effect is more pronounced near the water surface and under daylight conditions, while it becomes less significant at larger depths due to the natural attenuation of sunlight in water. In practice, the impact of ambient light can be mitigated by using narrow optical bandpass filters, reducing the receiver field of view, improving transmitter-receiver alignment, and selecting appropriate transmission wavelengths. Therefore, direct air-to-underwater optical links are feasible for short-range and shallow-to-moderate-depth scenarios, particularly when relay deployment is impractical or when low-latency communication between aerial/surface platforms and underwater nodes is required.

Furthermore, direct air-to-underwater systems encounter challenges related to fog near the water surface, aquatic waves, and pointing, acquisition, and tracking (PAT), resulting in misalignment and inter-symbol interference (ISI) issues that affect connectivity and link quality. Consequently, the need to re-transmit lost packets leads to increased transmission delays and congestion issues. Power consumption is another significant challenge, given that these communication systems rely on batteries, which can be costly to recharge or replace. As a result, addressing link quality and power allocation in direct air-to-underwater optical communications represents vital issues that have been the subject of numerous studies in the existing literature. In Li *et al.* (2021), the authors conducted a performance analysis of an unmanned aerial vehicle (UAV)-assisted RF- underwater wireless optical communication (UWOC) system, considering both fixed-gain amplify-and-forward (AF) and decode-and-forward (DF) relaying protocols. The authors in Yang *et al.* (2021) investigated the utilization of optical wireless communication (OWC) transmission in the presence of Gamma-Gamma turbulence, specifically focusing on the application of fixed-gain AF relaying within a mixed UWOC-OWC communication system. In Rahman *et al.* (2023), the authors derived the probability density function (PDF) and cumulative distribution function (CDF) of the signal-to-noise ratio (SNR) for the combined channel effect, considering various fading models such as atmospheric turbulence, random fog, air-to-water interface, oceanic turbulence, and pointing errors.

Additionally, the employment of Reinforcement Learning (RL) techniques has been applied to address the beam adaptation challenge in Underwater Optical Wireless Communication (UOWC). In Zhang *et al.* (2021), Zhang et al. introduced an RL-based routing strategy for multi-hop UOWC, aimed at maximizing network throughput while upholding high reliability standards. Furthermore, Li, Hu, Li & Hu (2019) put forth a Multi-agent Reinforcement Learning (MARL) grounded routing protocol designed specifically for Underwater Optical Wireless Sensor Networks (UOWSN). The protocol's objective is to maximize the network's lifetime. Similarly, researchers in Li *et al.* (2020b) proposed an efficient MARL-based routing protocol with the goal of extending the network's lifetime and enhancing its capacity for UOWSN. In another perspective, Weng *et al.* (2022b) by Weng et al. delved into the alignment challenge within an optical AUV-to-AUV link using RL. Their findings established that the transmitting AUV can successfully perform the alignment task based on parameters such as its relative distance from the receiving AUV, orientation, and velocity of the receiving AUV. Additionally, Romdhane & Kaddoum (2022) introduced an RL-based beam adaptation technique focused on determining the optimal light beamwidth and beam direction. The authors in Qasem, Ali, Deng, Li & Fu (2023) proposed an OFDM-based scheme for UOWC systems that can outperform conventional schemes and offer improvements in SNR and peak-to-average power ratio (PAPR) at the same spectral efficiency (SE).

Motivated to ensure a high success rate through precise beam alignment while concurrently extending the network's lifespan, we develop a cutting-edge deep deterministic policy gradient (DDPG)-based link adaptation model. The main contributions of this paper can be summarized as follows:

- In this paper, we introduce a novel approach for link adaptation in direct air-to-underwater optical communication systems using the DDPG algorithm. This represents a significant advancement in optimizing communication performance in challenging underwater environments.
- The proposed method incorporates dynamic adjustments to key parameters, including beam orientation, beam width, and power levels.
- The study conducts a comprehensive performance comparison of the presented DDPG-based model with two static benchmarks and a Q-learning-based link adaptation technique.
- The results clearly demonstrate the superiority of the proposed model, achieving nearly 100% success rates and 10 times lower energy consumption than the Q-learning-based model, thus indicating its robustness and reliability under challenging conditions.

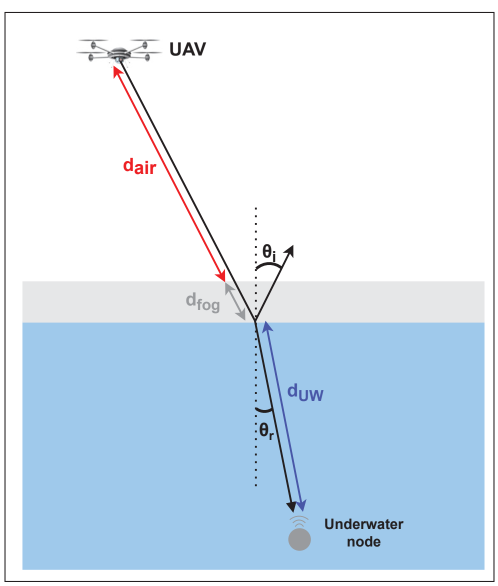


Figure 4.1 Proposed system model

4.2 System Model

We consider a direct air-to-underwater optical wireless communication link, as illustrated in Fig. 4.1. The UAV transmits an optical signal to the underwater node. The link can be divided into three sections: first, the air; second, the foggy layer, situated near the water surface; and last, the underwater section. The signal undergoes absorption, scattering, and multi-path interference while propagating through the different environments. We deploy the OFDM technique and 16-QAM modulation to mitigate the interference and noise effects after propagation. Also, underwater dynamicity and the UAV mobility can lead to misalignment errors that may result in signal loss. Hence, a dynamic link adaptation is required to overcome this issue.

Denoting the transmitted signal from the UAV by X, the received signal Y at the underwater node can be written as:

$$Y = hX + n, \tag{4.1}$$

where $h = [h_0, h_1, \ldots, h_{l-1}]^T$ is the overall end-to-end OFDM link impulse response function with h_{l-1} indicating the channel gain belonging to the largest path delay and l is the length of h. n is the channel noise. The channel effect is removed from the signal by considering a zero-forcing (ZF) equalizer Jiang, Varanasi & Li (2011).

We assume a noisy channel state information (CSI). The received power from the UAV to the underwater node can be defined as Han, Liu, Wang & Li (2022):

$$P_r = (\rho H + \sqrt{1-\rho^2} N_e)\, P_t + N_0, \tag{4.2}$$

where P_t is the transmitted power, H is the air-to-underwater transmittance, and $\rho \in [0, 1]$ is the correlation coefficient that determines the accuracy of the channel estimation. When $\rho = 1$, no channel estimation errors are considered. $N_e \sim \mathcal{N}(0, \sigma_{N_e}^2)$ denotes a white Gaussian random noise. N_0 is an additive white Gaussian noise (AWGN) with variance $\sigma_n^2 = \sigma_{shot}^2 + \sigma_{thermal}^2$, where σ_{shot} and $\sigma_{thermal}$ refer to the standard deviation of shot and thermal noise, respectively Komine & Nakagawa (2004).

The SNR of the considered communication system can be calculated as:

$$\text{SNR} = \frac{P_r}{N_0}. \tag{4.3}$$

4.2.1 Channel characterization

Propagating from air to underwater, the optical signal undergoes various attenuation depending on the environments in which it propagates. Hence, the total air-to-underwater transmittance can be defined as:

$$H = H_{\text{air}} H_{\text{fog}} H_{\text{surface}} H_{\text{uw}}, \tag{4.4}$$

where H_{air}, H_{fog}, H_{surface}, and H_{uw} are the air, fog layer, air/water surface, and underwater transmittance, respectively. According to Kahn & Barry (1997) atmospheric transmittance is defined using the Lambertian emission model as:

$$H_{air} = \frac{m+1}{2\pi d_{air}^2} (\cos\theta_i)^{m+1}, \tag{4.5}$$

$$m = -\frac{\ln 2}{\ln(\cos(\frac{\beta}{2}))}. \tag{4.6}$$

where the distance propagated by light from the UAV to the fog layer is denoted as d_{air}, θ_i is the incident light angle, and β is the beam width.

The atmospheric and fog layer transmittance can be defined as Kim, McArthur & Korevaar (2001):

$$H_{\text{fog}} = e^{-\sigma_{\text{fog}} d_{\text{fog}}}, \tag{4.7}$$

where the distance traversed through the fog layer is defined as d_{fog} and σ_{fog} is the fog attenuation coefficient defined by the following equationsKim *et al.* (2001):

$$\sigma_{\text{fog}} = \frac{3.91}{V_{\text{fog}}}, \tag{4.8}$$

where V is the visibility set as $V_{\text{fog}} = 500$ m.

While penetrating underwater, the light undergoes partial reflection and refraction at the air/water surface. The total surface transmittance can be defined as Aman, Al-Kuwari, Muzzammil, Rahman & Kumar (2023); Guo *et al.* (2018):

$$H_{\text{surface}} = \eta_{\text{deflection}} H_{\text{refraction}}, \tag{4.9}$$

where $\eta_{\text{deflection}}$ is the efficiency resulting from the deflection of light due to surface roughness based on the Cox and Munk (CM) model Cox & Munk (1954), and $H_{\text{refraction}}$ is the total Fresnel transmittance given as:

$$H_{\text{refraction}} = \frac{H_{\perp} + H_{\parallel}}{2}, \tag{4.10}$$

where $H_{\perp} = 1 - \frac{sin^2(\theta_i - \theta_r)}{sin^2(\theta_i + \theta_r)}$ and $H_{\parallel} = 1 - \frac{tan^2(\theta_i - \theta_r)}{tan^2(\theta_i + \theta_r)}$. $\theta_r = sin^{-1}(\frac{n_i}{n_r} sin(\theta_i))$ is the refracted light angle according to Snell's law Shirley (1951), where n_i and n_r are the refractive index of air and water, respectively. Finally, the transmittance in an underwater medium can be defined as Mobley (1994):

$$H_{\text{uw}} = \frac{2A_r cos(\Omega - \theta_i)}{\pi d_{uw}^2 (1 - cos\beta) + 2A_t} e^{-cd_{\text{uw}}}, \tag{4.11}$$

where A_{t} and A_r are the transmitter's and receiver's aperture areas in m^2, respectively. $(\Omega - \theta_i)$ is the angle between the receiver's normal towards the transmitter and the light beam center. The distance from the water surface to the receiver is denoted by d_{uw}, and c is the water attenuation coefficient.

4.2.2 Link Adaptation

The mobility of UAVs and the underwater dynamicity introduce a particular uncertainty in the location and orientation of both the transmitter and receiver, which introduces misalignment problems that lead to link deterioration and even loss of communication. Also, energy consumption is another challenge for wireless communications, where nodes are powered with expensive batteries to implement and maintain. Hence, we aim to extend the battery lifetime by minimizing energy consumption at each transmission. To face these challenges, dynamic link adaptation becomes crucial to guarantee a robust optical communication link.

The objective of the proposed model is to maximize the success rate and minimize the energy consumed at transmission. This can be divided into two optimization problems as follows:

$$\begin{aligned} &\max_{\beta,\theta_i,P_t} \text{SNR} \\ &\text{s.t}\begin{cases} P_t \leq P_{\max} \\ \text{SNR} > \text{SNR}_{\text{th}} \\ \beta_{\min} \leq \beta \leq \beta_{\max} \\ \theta_{i,\min} \leq \theta_i \leq \theta_{i,\max}, \end{cases} \end{aligned} \tag{4.12a}$$

$$\begin{aligned} &\min_{P_t} \frac{E_t}{E_{\max}} \\ &\text{s.t}\begin{cases} E_t = P_t * \tau \\ P_t \leq P_{\max}, \end{cases} \end{aligned} \tag{4.12b}$$

where $\beta_{\min}$, $\beta_{\max}$, $\theta_{i,\min}$, $\theta_{i,\max}$, are the minimum beam width, maximum beam width, minimum beam orientation, and maximum beam orientation, respectively. SNR_{th} is the minimum required SNR to guarantee a successful reception of information. E_t is the consumed energy per transmission, and $E_{\max}$ is the adaptation energy budget. τ is the transmission time defined as $\tau = \frac{packet\ size}{R*N*K}$, where R is the bit rate, N is the number of OFDM subcarriers, and K is the number of bits per symbol.

The optimization problem proposed in (4.12a) is non convex NP hard. Also, the dynamicity related to the location of the transmitter and receiver as well the as dynamicity present underwater justifies solving such a problem using RL techniques. To solve these optimization problems, we propose an RL-based joint beam and power adaptation method where the beam width β, beam orientation θ_i, and transmit power P_t are dynamically allocated to maintain a sustainable communication while extending the UAV lifetime by minimizing energy consumption for each transmission. Since these variables change with a certain randomness in real-world scenarios, we must keep continuous values of such entities. Hence, this justifies the DDPG adequacy for the studied problem.

4.3 Proposed model

4.3.1 DDPG background

DDPG is an extension of the deep Q-learning algorithm (DQN) that deals with continuous states and action sets. It consists of two main networks: The actor and the critic networks. While the actor-network maps states to actions, the critic network evaluates the state-action tuple by simulating the Q-function $Q(s,a)$ using neural networks with no curse of dimensionality. Each of these networks includes an online network and a target network. The online actor and online critic networks are parameterized by θ^{μ} and θ^{Q}, respectively. The target actor and target critic networks are parameterized by $\theta^{\mu'}$ and $\theta^{Q'}$, respectively. The target network is a previous copy of the online network that is used for tracking the learned network to ensure stability and avoid divergence of the network Mnih *et al.* (2015).

4.3.2 DDPG-based link adaptation

The proposed model aims to enhance link connectivity by performing the alignment task to reduce pointing errors while minimizing the UAV's energy consumption to extend the battery lifetime. To do so, we propose a DDPG-based beam and power adaptation method where the beam width, beam orientation, and transmit power are dynamically tuned to cope with the dynamic environment. The agent is defined at the UAV, and we define the states, actions, and rewards as follows:

1. States: The states are defined as the current beam width, beam orientation, and transmit power of the UAV. Hence, the state sets are defined as $\{S_{\theta_i}, S_{\beta}, S_{\mathrm{p}}\}$, where S_{θ_i}, S_{β}, and S_{p} are the sets of possible values of beam width, beam orientation, and transmit power, respectively.

2. Actions: The actions that the agent can take are to increase, decrease, or keep the beam width, beam orientation, and power. Hence, We define the action sets as $\{A_{\theta_i}, A_\beta, A_{\mathrm{p}}\}$, where $A_{\theta_i} = A_\beta = A_{\mathrm{p}} = \{-1, 0, 1\}$ where -1, 0, and 1 refer to decreasing, keeping, and increasing actions, respectively. The action of increasing or decreasing the state variables takes into account a certain randomness introduced in real word scenarios, which requires continuous states sets.
3. Reward: The reward function for the proposed model is defined as a joint of the two optimization problems defined in (4.12a) and (4.12b). It can be written as:

$$r = \begin{cases} \omega_1 \mathrm{SNR} - \omega_2 \frac{E_t}{E_{\max}} & \text{if no misalignment and SNR} > \mathrm{SNR}_{\mathrm{th}} \\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

where ω_1 and ω_2 are summation weights such that $\omega_1 + \omega_2 = 1$. The corresponding algorithm for the proposed model is given in Algorithm 1. First, the online actor and critic networks, Q and μ, the target actor and critic networks, Q' and μ', and the replay buffer B are initialized. For each episode $e \in [1, N_{ep}]$, the states $s_0 = \{s_{\theta_i,0}, s_{\beta,0}, s_{\mathrm{p},0}\}$ are reset to initial values. Then, at each iteration till exhausting the learning energy budget $E_{\max}$, the actions $a_i = \{a_{\theta_i,\mathrm{i}}, a_{\beta,\mathrm{i}}, a_{\mathrm{p,i}}\}$ are selected using the online actor network and executed. The reward is calculated using (4.13) and the next states $s_{\mathrm{i+1}} = \{s_{\theta_i,\mathrm{i+1}}, s_{\beta,\mathrm{i+1}}, s_{\mathrm{p,i+1}}\}$ are observed. The resulting tuple $(s_i, a_i, r_i, s_{\mathrm{i+1}})$ is then stored in B. Next, a random minibatch of N_M transitions is sampled from B, and for each transition (s_i, a_i, r_i, s_{i+1}), y_i is defined as:

$$y_i = r_i + \gamma Q' \left(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'} \right) \text{ for } i \in [1, N_M]. \tag{4.14}$$

The critic network is updated by minimizing the loss, defined as,

$$L(\theta^Q) = \frac{1}{N_M} \sum_i \left[Q(s_i, a_i|\theta^Q) - y_i \right]^2, \tag{4.15}$$

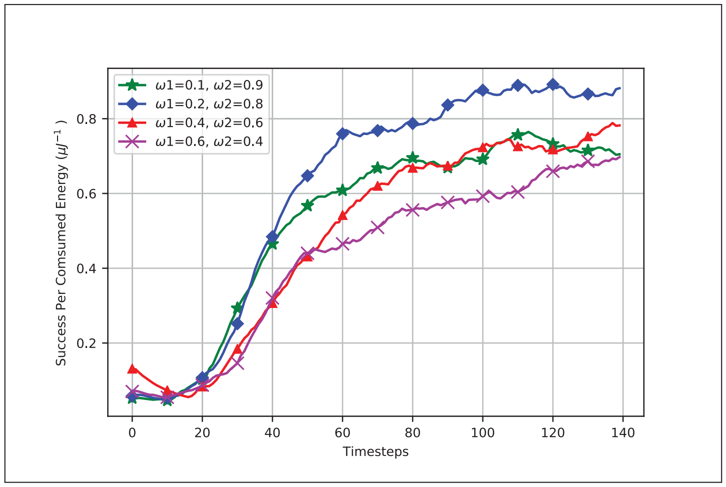


Figure 4.2 Success rate per consumed energy of the proposed DDPG method using different (ω_1, ω_2) tuples for noisy CSI with $N = 256$ and 16-QAM modulation at 200 Mbps

and the actor network is updated using the sampled policy gradient as,

$$\nabla_{\theta^{\mu}} \approx \frac{1}{N_M} \sum_i \left[\nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})|_{s=s_i} \right]. \tag{4.16}$$

Finally, the target networks are updated using the following equations,

$$\theta^{Q'} \leftarrow \alpha\theta^{Q} + (1-\alpha)\theta^{Q'}, \tag{4.17a}$$

$$\theta^{\mu'} \leftarrow \alpha\theta^{\mu} + (1-\alpha)\theta^{\mu'}. \tag{4.17b}$$

4.4 Simulation results and discussion

In this section, we present the performance evaluation of our proposed model. The simulation of the model is performed using Python 3.7 programming language. For comparison, we considered three benchmarks defined as follow:

Algorithm 4.1 DDPG Algorithm for Air-to-Underwater Link Adaptation

Algorithm: DDPG Algorithm for Air-to-Underwater Link Adaptation
Input: S_{θ_i},S_β,S_p, A_{θ_i}, A_β,A_p, Number of episodes N_{ep},γ
Initialize

- Critic network $Q(s_{\theta_i}, s_\beta, s_p, a_{\theta_i}, a_\beta, a_p|\theta^Q)$
- Actor network $\mu(s_{\theta_i}, s_\beta, s_p, a_{\theta_i}, a_\beta, a_p|\theta^\mu)$
- Target networks Q' and μ',$\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
- Replay buffer B
- Done = False

for $e = 0, 1, 2...N_{ep}$ **do**
  Reset $s_{\theta_i,0} \in S_{\theta_i}$,$s_{\beta,0} \in S_\beta$, and , $s_{\mathrm{p},0} \in S_p$
  while *Not done* **do**
    Get $a_i = \{a_{\theta_i,\mathrm{i}}, a_{\beta,\mathrm{i}}, a_{\mathrm{p,i}}\}$ using $\mu(s, a|\theta^\mu)$
    Execute a_i
    Calculate r_i using (4.13)
    Observe $s_{\mathrm{i+1}} = \{s_{\theta_i,\mathrm{i+1}}, s_{\beta,\mathrm{i+1}}, s_{\mathrm{p,i+1}}\}$
    Store $(s_\mathrm{i}, a_\mathrm{i}, r_i, s_\mathrm{i+1})$ in B
    Sample a random minibatch of size N_M from B
    Set y_i using (4.14)
    Update the critic network using (4.15)
    Update the actor network using (4.16)
    Update the target networks using (4.17)
    if *(residual energy $\leq E_{max}$)* **then**
      done = True
    end if
  end while
end for

- The Q-learning-based link adaptation: This is an RL-based benchmark where we replace the DDPG algorithm with the Q-learning algorithm. The states sets in this case become scattered in the same range as the DDPG states sets, with a step of $\frac{\pi}{48}$, $\frac{\pi}{72}$ and 1dBm for the beam orientation, beamwidth, and power, respectively.
- 1^{st} static benchmark: In this benchmark, we opt for a moderate decision by fixing the beam orientation, beamwidth, and power as the mid-value of the states sets defined for the DDPG model, i.e. $\theta_i = 0$, $\beta = \frac{3\pi}{24}$, and $P_t = 32.5$dBm.

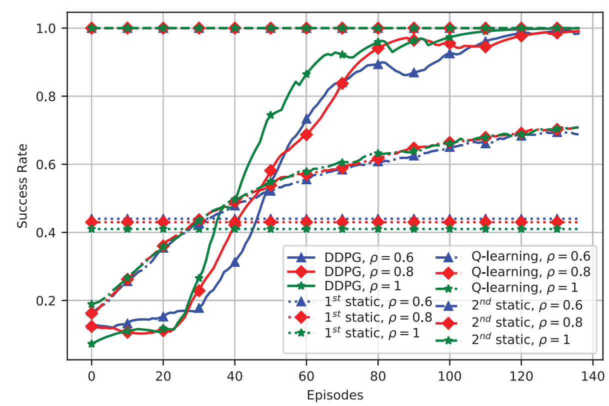
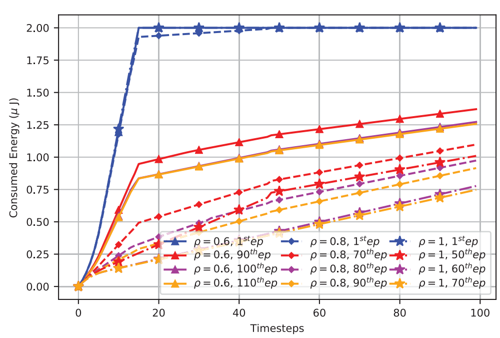
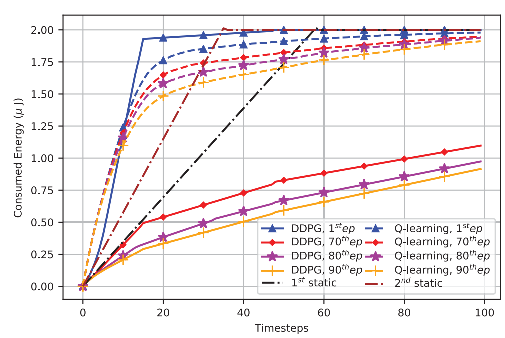


Figure 4.3 Success rate of DDPG vs. Q-learning method for different ρ values with $N = 256$ and 16-QAM modulation at 200 Mbps



a) Consumed energy per episode of DDPG method, $\rho = \{0.6, 0.8, 1\}$



b) Consumed energy per episode of DDPG vs. Q-learning method, $\rho = 0.8$

Figure 4.4 Consumed energy per episode of DDPG and Q-learning methods for different ρ values with $N = 256$ and 16-QAM modulation at 200 Mbps

- 2^{nd} static benchmark: In this benchmark, we focus on maintaining the connectivity by maintaining a maximized success rate equal to 1. For that, we assume that the underwater node moves within an uncertainty sphere of center its estimated location and radius $\epsilon = 1.5$m Celik *et al.* (2019). The UAV is assumed to be located vertically aligned with the estimated location of the underwater node. Hence, we defined the beam orientation $\theta_i = 0$, the beamwidth is defined in a way to cover the whole uncertainty sphere, and the power to ensure the condition $SNR \geq SNR_{th}$.

We set the distance traversed by light from the UAV to the fog layer and through the fog layer, $d_{\text{air}} = 7$m and $d_{\text{fog}} = 1$m, respectively. The distance traversed underwater is set to $d_{\text{uw}} = 5$m. The refraction index of air and water are $n_i = 1$ and $n_r = 1.33$, respectively. We consider pure seawater with an attenuation coefficient $c = 0.043$ for $\lambda = 514$nm. We consider a transmitter equipped with a single LED with an aperture area $A_t = 10mm^2$. The receiver has a PDB-S5971 photodiode with an aperture area $A_r = 1.1mm^2$. The thermal and shot noise variances are set to $\sigma^2_{shot} = 2.2 \times 10^{-15}$ and $\sigma^2_{thermal} = 3.3 \times 10^{-18}$, respectively Komine & Nakagawa (2004), and the CSI noise variance $\sigma^2_{N_e} = 10^{-25}$. The threshold SNR is $\text{SNR}_{\text{th}} = 30$ dB. The link functions under a data rate of $R = 200$ Mbps. In the link adaptation learning phase, we assume a fixed packet of size $K = 4$ Mbits is sent to the underwater node. The signal is modulated using the 16-QAM modulation scheme and then sent using the OFDM technique with $N = 256$ subcarriers. The energy consumption budget set for learning is $E_{\max} = 2\mu J$. The hyper-parameters for the DDPG are set as $\gamma = 0.99$, $\alpha = 0.01$ and the batch size $N_M = 128$. Table 4.1 summarizes the states and action sets. The evaluation metrics adopted to assess the performance of the proposed model are the success rate and the consumed energy per episode.

Table 4.1 States and actions sets

States	Actions
$S_{\theta_i} = [\frac{-\pi}{6}, \frac{\pi}{6}]$	$A_{\theta_i} = \{-1, 1, 1\}$
$S_\beta = [\frac{\pi}{12}, \frac{\pi}{6}]$	$A_\beta = \{-1, 1, 1\}$
$S_{\text{p}} = [25, 40]dBm$	$A_{\text{p}} = \{-1, 1, 1\}$

To guarantee a better performance of the proposed model, it is essential to wisely define the reward weights ω_1 and ω_2. This can be done empirically by evaluating the success rate per consumed energy for different weight tuples. It is important to mention here that various value tuples are tested. But in order to guarantee a clear legible plot, we exhibit only the results of four value tuples. As shown in Fig 4.2, the best results are for $\omega_1 = 0.2$ and $\omega_2 = 0.8$. First, we assess the performance of our model using the success rate metric for different channel conditions, as depicted in Fig. 4.3. Results show that the proposed model converges to almost 100% success rate for different channel conditions. Specifically, it converges in around 60 episodes for perfect CSI ($\rho = 1$), in around 80 episodes for $\rho = 0.8$, and in about 100 episodes for $\rho = 0.6$. On the other hand, the Q-learning algorithm does not exceed 70% success rate after 140 episodes for the different channel conditions. This proves that the proposed method ensures better connectivity compared to Q-learning algorithm with faster convergence. The first benchmark kept a limited success rate in the order of 0.4 for all ρ values, while the second benchmark kept a success rate equal to 1 for all ρ values, aligning with its definition. Second, the proposed method is evaluated in terms of consumed energy per episode. In Fig. 4.4(a), we evaluate the energy consumption of the proposed method for different channel conditions. The convergence for perfect CSI ($\rho = 1$) is the fastest after 60 episodes to around $0.75\mu J$ at 100 timesteps. Then, the convergence slows down for noisy channels to reach $0.9\mu J$ in the 80^{th} episode for $\rho = 0.8$ at 100 timesteps and $1.25\mu J$ in the 100^{th} episode for $\rho = 0.6$ at 100 timesteps. In Fig. 4.4(b), we compare the performance of the proposed DDPG model with the three defined benchmarks for $\rho = 0.8$. The 2^{nd} benchmark exhausts its battery first at 35 timesteps, followed by the 1^{st} benchmark at almost 60 timesteps. Meanwhile, the proposed DDPG method guarantees 52% less energy consumption than the Q-learning model in the 80^{th} episode at 100 timesteps. Hence, we can conclude that the proposed method converges faster and ensures a considerably lower energy consumption of the UAV battery for different channel conditions.

4.5 Conclusion

This paper presented a DDPG-based link adaptation model for a direct air-to-underwater optical link using the OFDM technique under noisy CSI conditions. The proposed method includes dynamic beam orientation, beam width, and power adaptation. Because the change in these variables includes a certain randomness existing in reality, we assumed continuous values for these entities. The success rates of the presented model are compared to the Q-learning-based link adaptation, and two static approaches. The proposed model kept a high success rate, reaching almost 100% while decreasing energy consumption 10 times more than the Q-learning-based solution after 100 timesteps. Hence, our proposed model extends the UAV battery lifetime by minimizing energy consumption while maintaining good reliability of the direct air-to-underwater optical link. In future work, we intend to improve the fairness of multi-user UOWC networks.

CHAPTER 5

DYNAMIC RESOURCE OPTIMIZATION FOR A JOINT MAX-MIN FAIRNESS AND ENERGY-EFFICIENCY PROBLEM IN NOMA-AIDED UNDERWATER OPTICAL WIRELESS SYSTEMS

Imene Romdhane[1] , Ziyaur Rahman[1] , Nahed Belhadj Mohamed[1] , Md. Zoheb Hassan[2] , Georges Kaddoum[13]

[1] Department of Electrical Engineering, École de Technologie Supérieure, 1100 Notre-Dame west, Montreal, Canada H3C 1K3.

[2] Department of Electrical and Computer Engineering, Université Laval, Québec, QC, G1V 0A6, CA.

[3] Cyber Security Systems and Applied AI Research Center, Lebanese American University, Beirut, Lebanon.

Paper published in IEEE Open Journal of the Communications Society, April 2026.

5.1 Introduction

Underwater optical wireless communication (UOWC) is an evolving field that seeks to address communication challenges in underwater environments and opens up new possibilities for underwater exploration, research, and monitoring. As a communication technology, UOWC involves transmitting data through the water using optical signals. Unlike traditional radio-frequency (RF) communication, which is limited in underwater environments due to the high absorption and scattering of electromagnetic waves in water, optical communication employs light to transmit information Kaushal & Kaddoum (2016); Zeng *et al.* (2017); Celik *et al.* (2022). As compared to its traditional counterparts, RF and acoustic underwater communications, UOWC boasts several advantages, including higher bandwidth, enhanced security, lower implementation costs, and reduced time latency Kaushal & Kaddoum (2016). However, despite these benefits over RF and acoustic communications, the link range of UOWC systems is constrained by impairments in the UOWC channel - namely, absorption, scattering, and turbulence Zeng *et al.* (2017). The UOWC channel can be effectively modeled by considering attenuation and fading O. Korotkova & Shchepakina (2012). Said simply, both attenuation and fading contribute to the signal alteration after propagation through the UOWC channel. The attenuation is primarily attributed to absorption and scattering effects Jamali *et al.* (2018). Governed by the Beer-Lambert law, UOWC signal exponentially decays with distance, which makes the link range significantly lower than RF and acoustic link ranges. In addition, optical communications are based on directional beams with a narrow field of view (FOV), which makes the system highly dependent on spatial geometry and nodes positions. Considering the dynamicity underwater, the alignment problem is further highlighted in UOWC, since nodes are considered quasi-stationary underwater. Based on these these characteristics, the UOWC is strongly dependent on nodes locations and beam orientations, which significantly differentiates the considered problem from RF and acoustic-based resource allocation frameworks.

To alleviate signal scintillation induced by turbulence and extend the UOWC range, various multiple-input–multiple-output (MIMO) systems were proposed Yue, Wang, Xu & Xu (2025); Wang, Li, Hao, Lv & Ju (2025); Shetty, Naik, Acharya & Chung (2023). In Yue *et al.* (2025), a non-line of sight (NLOS) UOWC MIMO configuration was proposed to improve the system performance by reducing the path loss and improving the channel impulse response. Furthermore, in Wang *et al.* (2025), the authors investigated the bit error rate (BER) performance of MIMO UOWC systems considering the generalized gamma distribution (GGD) oceanic turbulence, zero-boresight point error, and Elamassie underwater path loss. In Shetty *et al.* (2023), the performance of a MIMO underwater vertical wireless optical communication link was evaluated in terms of BER and outage probability, considering the presence of weak and strong oceanic turbulence, pointing errors, and attenuation losses.

The non-orthogonal multiple access (NOMA) technique was incorporated into the UOWC system to enhance spectrum utilization. A promising multi-access technology, NOMA enables efficient large-scale connectivity and perfectly aligns with the developmental requirements of UOWC Li *et al.* (2023b); Salama, Aly & Amer (2023); Palitharathna, Suraweera, Godaliyadda, Herath & Ding (2022); Liang, Yin, Jing, Ji & Wang (2023); Li *et al.* (2023c). In Li *et al.* (2023b), the authors proposed a straightforward two stage program judgement filter (PJF) for a real-time multi-user successsive interference cancellation (SIC)-free NOMA-based uplink

UOWC system. The results demonstrated that the proposed framework can decrease the BER compared to the standard SIC-based NOMA. Furthermore, in Salama *et al.* (2023), NOMA was applied in conjunction with deep learning convolutional neural network (CNN) to optimize the relay selection and power allocation tasks in UOWC system. The results revealed that NOMA outperformed the orthogonal multiple access system in the studied case. Practical challenges such as attenuation and turbulence, including misalignment at the relay, were considered in Palitharathna *et al.* (2022), with a particular focus on optimizing the instantaneous time-splitting parameter to maximize the uplink achievable rate. To mitigate decoding errors caused by self-interference within NOMA paired groups, in Liang *et al.* (2023), the authors proposed an alternating composite phase shift (ACPS) coding scheme based on constellation optimization. The power distribution optimization model for the UOWC system, constrained by specific light-emitting diode (LED) emission power, was established in Li *et al.* (2023c) for scenarios ensuring user quality of service (QoS) for edge users and maximizing fairness for all users.

However, although communication fairness must be taken into account in nearly all resource allocation challenges in wireless networks, the fairness concern among underwater nodes was not comprehensively addressed in previous research. The absence of such consideration may result in resource starvation or redundant allocation SHI, Prasad, Onur & Niemegeers (2014); Dani, So, Tang & Ding (2022); Jiao *et al.* (2022a); Li, Jiang & Wang (2022); Wang, Wang, Li, Jiang & Chen (2022); Seah, Lee, Lin & Lai (2022); Rahmani *et al.* (2023); Magbool, Kumar & Flanagan (2023); Gou *et al.* (2023, 2024); Li *et al.* (2023c); Agarwal & Krikidis (2025). In SHI *et al.* (2014), the authors highlighted open research challenges, emphasizing interconnectedness of fairness with performance, utility, optimization, and throughput on both network and node levels. The proportional fairness concept was also discussed in Dani *et al.* (2022), where the authors sought to distribute resources among users to ensure a proportional and equitable share based on their individual needs and requirements. This fairness criterion sought to strike a balance between maximizing system throughput and providing a satisfactory level of service to each user. Furthermore, an optimization model was formulated in Jiao *et al.* (2022a); Li *et al.* (2022) to ensure that user nodes meet the fairness criterion, maximizing the minimum rate under the constraint of total transmit power. In Wang *et al.* (2022), the authors addressed the challenge of meeting service requirements for different users within a single beam by categorizing them into the following two types: edge users and fair users, and introduced a power allocation algorithm designed to ensure QoS for edge users and fairness for fair users. In Seah *et al.* (2022), a combined resource scheduler entitled extended weighted fair queuing with latency constraint (EWFQ/LC) was presented for packet scheduling among network slices, considering system fairness and latency constraints. Furthermore, in Rahmani *et al.* (2023), the authors focused on the uplink of a cell-free massive MIMO system employing maximum-ratio combining (MRC) and zero-forcing (ZF) schemes and addressed a power allocation optimization problem to jointly optimize conflicting metrics, namely sum rate, and fairness. Focusing on energy efficiency (EE) and user fairness in systems with numerous antennas, the authors in Magbool *et al.* (2023) proposed a lexicographic-based approach for RIS-assisted mmWave systems that maximizes both EE and user fairness by optimizing power, RIS passive beamforming matrix, and analog precoders. A power management strategy was presented in Gou *et al.* (2023) so as to optimize the reuse, fairness, and capacity of underwater wireless sensor networks (UWSNs). In another pertinent publication Gou *et al.* (2024), the authors investigated the effect of unexpected nodes malfunctions on the performance of imperfect and energy-constrained underwater acoustic sensor networks. To this end, they proposed a semi-cooperative power allocation approach that achieves fairness-effective and robust communication.

However, previous research on fairness in UOWC systems is still very limited, which warrants further investigation. For instance, a power allocation algorithm was proposed in Li *et al.* (2023c) for multiple-beam space division access-based NOMA UOWC systems to maximize max-min fairness. Meanwhile, authors in Agarwal & Krikidis (2025) aimed to maximize fairness while respecting a minimum harvested energy threshold in simultaneous lightwave information and power transfer (SLIPT)-enabled two-user NOMA UOWC systems. However, while the studies briefly reviewed above addressed fairness and EE individually, these two aspects need to be addressed simultaneously, as neglecting the fairness would lead to resource starvation and neglecting the EE would induce energy waste. To the best of our knowledge, the joint optimization of EE and fairness in the context of UOWC networks remains largely unexplored in the literature.

Reinforcement learning (RL) techniques were also employed to overcome the dynamicity challenge within underwater environment, resulting a secure and reliable UOWC system. In Shin, Baek & Song (2024), authors addressed the misalignment problem in point-to-point UOWC and focused on optimizing the communication between an underwater sensor and an unmanned surface vehicle (USV) that may irregularly shake above sea level. To this end, a two-step two-agent deep reinforcement learning algorithm was proposed. Furthermore, in Li *et al.* (2024b), a deep reinforcement learning (DRL)-based cooperative movement scheme for multiple autonomous underwater vehicle (AUV)s-based UOWC was proposed to overcome the misalignment and positional uncertainty problems typical of underwater environments. In Shin, Kim & Song (2023), the misalignment and power allocation problems in UOWC were evaluated, where authors jointly optimized the beam divergence and transmission power to maintain a seamless connection in P2P UOWC while minimizing the battery consumption.

While the literature birefly reviewed above demonstrates that RL can be a powerful tool for optimizing UOWC networks, most previous studies still exhibit certain research gaps. In particular, this body of work considers a NOMA-enabled multi-beam, multi-user underwater system. Due to the inherent energy constraints of optical transmitters in underwater environments, it is essential to maximize the system's EE. However, optimizing EE alone does not guarantee rate fairness among distributed nodes, which can degrade the overall quality of service. In this context, the following open challenges emerge:

1. Joint Optimization Challenge: How can beamforming and power allocation be jointly optimized in NOMA-enabled multi-user UOWC systems, while simultaneously maximizing EE and ensuring max-min fairness among users?
2. Scalable Learning-Based Optimization: How can network performance be improved through a scalable, learning-based framework that enables continuous selection of beam orientations and transmit power levels, thereby overcoming the sub-optimality introduced by discrete-action selection methods Palitharathna *et al.* (2022)?

In this work, we address these research gaps by designing a deep deterministic policy gradient(DDPG)-based multi-task DRL framework that jointly maximizes EE and fairness in NOMA-enabled multi-user UOWC networks.

Contributions: In this paper, our primary goal is to optimize the beamforming task, including beam steering and optimizing the power allocation among various receiving nodes, using NOMA in each transmitted beam. This optimization aims to maximize both the network's minimum rate and EE, while taking into consideration the quasi-stationarity of nodes underwater. The main contributions of this work can be summarized as follows:

- We propose a multi-beam UOWC network model, where each beam employs NOMA to simultaneously support multiple underwater user devices. A joint beamforming and power allocation problem is devised to jointly maximize the long-term EE and minimum rate of the network.
- A dynamic RL-based optimization framework is proposed to effectively solve the joint beamforming and power optimization problem. We propose both single-agent and sequential multi-agent DDPG models, with a comparison of their results in terms of the network's minimum rate, EE, and algorithm complexity. The proposed multi-agent model incorporates sequential decision making that enhances beam steering and power allocation tasks. Since this procedure enables training agents tailored to the specific task, these agents have a better decision making capability in dynamic environment. As shown in our simulations, compared to the single-agent approach, the multi-agent approach with sequential decision-making capability achieves better results with a slightly higher execution time.

The remainder of the paper is organized as follows. Section II outlines our system model. In section III, we formulate the problem. Section IV introduces the preliminaries of DDPG algorithm. In section V, we present the proposed DDPG-based beamforming solutions. Section VI details the simulation results. Finally, Section VII concludes. The notations used in this article are summarized in Table 5.1.

Table 5.1   Table of notations

Variable	Definition
ϵ	Uncertainty sphere radius
$\nu_{m,n}$	Thermal noise added to $y_{m,n}$
ψ	Concentrator field of view
$\psi_{n,m}$	Incident angle at the receiver
$\theta_{n,m}$	Irradiance angle at the transmitter
θ^{μ}	Online actor network
θ^{Q}	Online critic network
$\theta^{\mu'}$	Target actor network
$\theta^{Q'}$	Target critic network
$\theta_{1/2}$	Transmitter half power angle
τ	Update rate of the target networks
$\gamma_{n,i}(m)$	SINR for i^{th} user when decoding the m^{th} signal
γ	Discount factor
α_a, α_c	Learning rate
σ^2	Noise power
A_{o_n}	Continuous action set taken on orientation
A_{p_m}	Continuous action set taken on m^{th} node
A_r	Node aperture area
a_{o_n}	Orientation action
a_{p_i}	Node power action
c	Water attenuation coefficient
$c(\psi_{n,m})$	Gain of the optical concentrator
$d_{n,m}$	Distance from beam n to node m
d_{ϵ}	The random moving distance of nodes
EE	Energy efficiency
$h_{n,m}$	Underwater channel attenuation function from beam n to node m
$I_{n,i}(m)$	Sum off inter and intra-interference
$L(\theta^{Q})$	Loss function
M	Total number of nodes
M_n	Number of nodes covered by beam n
N	Number of clusters/beams
N_B	Batch size
N_0	Agent discovery end
$N_{0,o}$	Orientation agent discovery end
$N_{0,p}$	Power agents discovery end
N_{ep}	Number of episodes
n_{water}	Water refraction index
$o_n^{(t)}$	Beam orientation at time t

Table 5.2 Table of notations (continued)

P_f	Circuit and hardware power consumption
P_{max}	Maximum transmitted power
$p_{n,m}$	Power allocated to node m from cluster n
$p_n^{(t)}$	Power assigned to beam n
$p_{n,m}^{(t)}$	Power allocated to node m from cluster n at time t
Q	Q-function
$R_{n,i}(m)$	i^{th} user achievable rate when decoding m^{th} signal
$R_n(m)$	Achievable rate for m^{th} user from cluster n
R_{min}	Minimum rate for all NOMA users
r_t	Reward at time t
$r_o^{(t)}$	Orientation agent reward at time t
$r_p^{(t)}$	Beam power agent reward at time t
$r_{p,n}^{(t)}$	Node power agent reward at time t
S_i	Continuous state set
$S_{o,i}$	Continuous orientation state set
S_{p_m}	Continuous beam power state set
$S_{p_{n,m}}$	Continuous node power state set
$s_i^{(t)}$	Continuous state value at time t
$s_{n,m}$	Signal transmitted to node m in cluster n
t	Time slot
T	Number of steps per episode
x_n	Transmitted signal by beam n
$(x_i^{(t)}, y_i^{(t)})$	Node coordinates at time t
$(x_{0,n}, y_{0,n})$	Beam source coordinates
y_n	Received signal at node m in cluster n
$\hat{y}_{n,i}(m)$	Recovered m^{th} user signal at node m in cluster n

5.2 System Model

In this paper, we consider a multi-beam NOMA-aided UOWC where the transmitter is equipped with multi LEDs and can simultaneously forward multi beams into different directions Jiang, Li, Chen, Li & Wang (2023a); Li *et al.* (2023c); Shi, Ma, Tian, Guo & Ao (2024). The underwater receivers are clustered into groups based on their geographical locations, where each group, also referred to as cluster, is covered by one of the transmitted beams (see Figure 5.1). Due to underwater dynamicity, nodes are considered quasi-stationary, where their locations randomly

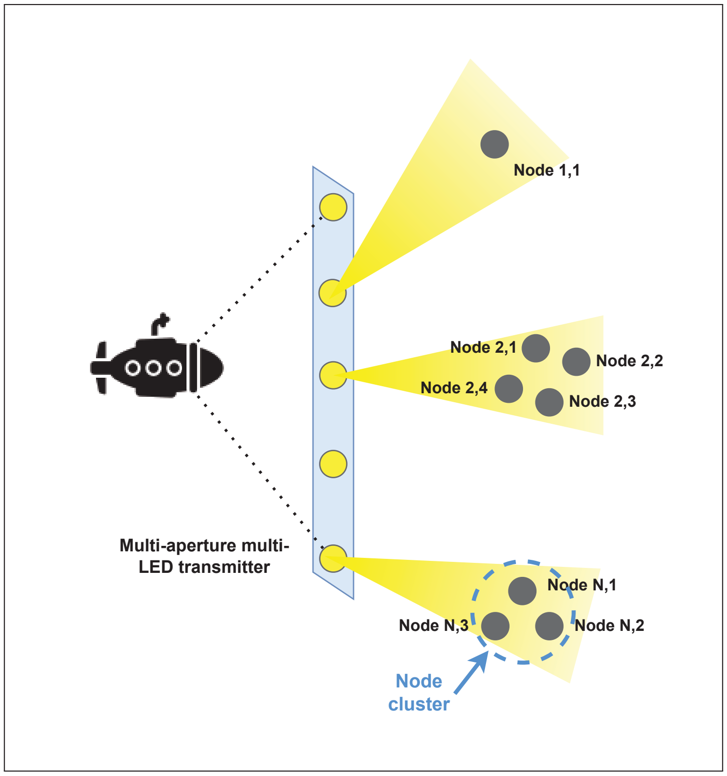


Figure 5.1 System model

shift after a certain period of time within a sphere of radius ϵ centered on the estimated location of the node. The total transmitted power cannot exceed a maximum value P_{max}. Our main goal is to maximize the minimum rate and the EE of the network by optimizing the beams orientation and power allocation among the different receiving nodes. NOMA is adopted at each transmitted beam to simultaneously serve different nodes within the same cluster. Let us assume N beams are transmitted simultaneously. The transmitted signal from each beam, n, is defined as shown in (5.1)

$$x_n = \sum_{m'=1}^{|M_n|} \sqrt{p_{n,m'}} s_{n,m'}, \tag{5.1}$$

where $|M_n|$ is the number of nodes covered by beam n, defined as cluster n, and $p_{n,m'}$ and $s_{n,m'}$ are the power allocated and the signal transmitted to node m' from cluster n, respectively. The nodes included in cluster n are defined based on their geographic locations. The receiving nodes are initially decomposed into multiple clusters using standard clustering algorOithms such as the K-means clustering[1]. Each cluster is supported by one beam. The received signal at each node m covered by beam n can be expressed as shown in (5.2)

$$y_{n,m} = h_{n,m}x_n + \sum_{\substack{n' = 1 \\ n' \neq n}}^{N} h_{n',m} \sum_{m'=1}^{|M_{n'}|} \sqrt{p_{n',m'}} s_{n',m'} + v_{m,n}, \tag{5.2}$$

where $v_{m,n} \sim \mathcal{N}(0, \sigma^2)$ is thermal noise and $h_{n,m}$ is underwater channel attenuation function from beam n to node m defined using the Beer-Lambert laww as shown in (5.3)

$$h^2_{n,m} = \begin{cases} \dfrac{(m+1)A_r}{2\pi d^2_{n,m}} \cos^m(\theta_{n,m}) \cos(\psi_{n,m}) c(\psi_{n,m}) e^{-cd_{n,m}}, & \theta_{n,m} \in \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right], \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

where A_r is the node m's aperture areas in m^2. $\theta_{n,m}$ is the irradiance angle at the transmitter, $\psi_{n,m}$ is the incident angle at the receiver. The distance from beam n to node m is denoted by $d_{n,m}$, c is the water attenuation coefficient, and $m = -ln(2)/ln(\cos(\theta_{1/2}))$ is the Lambertian order of the transmitter where $\theta_{1/2}$ is the half power angle of the transmitter Palitharathna *et al.* (2021). $c(\psi_{n,m})$ is the gain of the optical concentrator defined as specified in (5.4)

$$c(\psi_{n,m}) = \begin{cases} \frac{n^2_{water}}{\sin^2(\psi_{n,m})}), & 0 \leq \psi_{n,m} \leq \psi \\ 0, & \psi_{n,m} > \psi, \end{cases} \tag{5.4}$$

where n_{water} in the water refraction index and ψ is the concentrator field of view (FOV). The channel attenuation function captures unique characteristics of UOWC systems, where the attenuation is an exponential function of the distance, while the incident and irradiance angle directly impact the channel. This contrasts with RF communications, where path loss typically follows a polynomial decay and is less sensitive to nodes alignment. Different NOMA users within the same cluster are sorted in the ascending order of their channel attenuation function, i.e. $h_{n,1} > h_{n,2} > ... > h_{n,|M_n|}$. At the reception, each node retrieves its information using the

[1] The proposed resource allocation is also valid for other clustering schemes.

SIC technique. Said differently, each node m from cluster n will first decode the signals of the users from the same cluster with worse channel attenuation functions and will then remove them from the received signal until it detects its own information. Hence, when any node i from cluster n with worse attenuation function than m (i.e. $i < m$) is decoding the m^{th} user signal, the remaining signal $\hat{y}_{n,i}(m)$ can be expressed as shown in (5.5)Jiao *et al.* (2022b)

$$\begin{aligned}\hat{y}_{n,i}(m) = &\underbrace{h_{n,m}\sqrt{p_{n,m}}s_{n,m}}_{\text{Decoded signal}} \\ &+ \underbrace{h_{n,m}\sum_{m'=1}^{m-1}\sqrt{p_{n,m'}}s_{n,m'}}_{\text{Intra-beam interference}} \\ &+ \underbrace{\sum_{\substack{n'=1 \\ n' \neq n}}^{N} h_{n',m}\sum_{m'=1}^{|M_{n'}|}\sqrt{p_{n',m'}}s_{n',m'} + \nu_{m,i}}_{\text{Inter-beam interference}}.\end{aligned} \tag{5.5}$$

Consequently, the SINR for the i^{th} user when decoding the m^{th} signal can be written as shown in (5.6)

$$\gamma_{n,i}(m) = \frac{h_{n,m}^2 p_{n,m}}{I_{n,i}(m) + \sigma^2}, \tag{5.6}$$

where $I_{n,i}(m)$ is the sum of the inter and intra interference effects and can be computed as shown in (5.7)

$$I_{n,i}(m) = h_{n,m}^2 \sum_{m'=1}^{m-1} p_{n,m'} + \sum_{\substack{n'=1 \\ n' \neq n}}^{N} h_{n',m}^2 \sum_{m'=1}^{|M_{n'}|} p_{n',m'}. \tag{5.7}$$

The corresponding achievable rate can be deduced using $\gamma_{n,i}(m)$ (see (5.8))

$$R_{n,i}(m) = \log_2(1 + \gamma_{n,i}(m)). \tag{5.8}$$

Hence, the corresponding achievable rate is the minimum of these rates (see (5.9))

$$R_n(m) = \min_{i=1,..,n} R_{n,i}(m). \tag{5.9}$$

The minimum rate for all NOMA users is defined as shown in (5.10)

$$R_{min} = \min_{n,m} R_n(m), \tag{5.10}$$

and the EE is defined as shown in (5.11) Xu, Li, Ji & Du (2014)

$$EE = \frac{\sum_{n=1}^{N} \sum_{m=1}^{|M_n|} R_n(m)}{\sum_{n=1}^{N} \sum_{m=1}^{|M_n|} p_{n,m} + P_f}, \tag{5.11}$$

where P_f is a fixed power including the circuit power consumption and hardware power consumption. We assume that devices accurately perform SIC operations. In addition, we assume that each beam has a fixed width but variable orientation, which is optimized according to the served devices locations. Finally, a time-slotted optimization is considered and we assume that the channel gains remain fixed within a given time slot and independently vary at different time slots. The assumptions taken in this paper are listed as follows.

Assumptions :

- Nodes are quasi-stationary. Their locations randomly change after a certain period of time within a sphere of radius ϵ centered on the estimated location of the node.
- The transmitter can transmit multiple optical beams simultaneously Jiang *et al.* (2023a); Li *et al.* (2023c); Shi *et al.* (2024).
- The transmission power is limited to P_{max}.
- Nodes are initially clustered using a clustering technique. Each cluster is assumed to be served by one beam. The proposed framework can be applied to any clustering technique.
- Devices accurately perform SIC operations.

- Each beam has a fixed width, but variable orientation. Beams orientations are optimized according to the predefined beam clusters while avoiding beam overlapping.
- A quasi-static oceanic turbulence fading channel is considered where the channel gains remain fixed within a given time slot and vary independently at different time slots.

Importantly, the considered nodes are not mobile. Instead, the node's movement is only due to underwater dynamicity.

5.3 Problem Formulation

In this section, we develop an optimization problem to conduct beamforming in a multi-beam NOMA technique, so as to jointly maximize the min rate and EE. Considering the exponential attenuation and directional nature of UOWC channels, the min-rate and EE largely depend on nodes positions and beam orientations. This further highlights the imbalance between receivers compared to RF systems, making minimum rate constraints particularly critical. In addition, power allocation, albeit essential, cannot fully compensate for poor channel conditions, which justifies the joint consideration of spatial and resource optimization in the proposed model. Hence, the beamforming task can be divided into the following two main sub-tasks: (1) The power allocation among different nodes and (2) the beams orientation optimization. For a multi-user network, maximizing the max-min fairness of resource allocation does not lead to optimal EE. Conversely, only maximizing EE can lead to unfair resource allocation among users. To strike a suitable balance, our ain in this study is to jointly maximize the system EE and max-min fairness of a multi-user UOWC network. To this end, we divide the overall duration into T non-overlapping time slots defined by t. Let $EE^{(t)}$ and $R_{min}^{(t)}$ be the EE and network min rate at the t-th time slot, respectively. The optimization problem can be written as shown in

(5.12)-(5.15)

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\max_{p_{n,m}^{(t)},o_n^{(t)}}\left\{R_{min}^{(t)},EE^{(t)}\right\} \tag{5.12}$$

$$\text{s.t } (C1): \sum_{n,m} p_{n,m}^{(t)} \leq P_{max}, \forall t \tag{5.13}$$

$$(C2): p_{n,m}^{(t)} > 0 \quad \forall n,m,t \tag{5.14}$$

$$(C3): -\tfrac{\pi}{2} + (n-1)\tfrac{\pi}{N} \leq o_n^{(t)} \leq -\tfrac{\pi}{2} + n\tfrac{\pi}{N} \quad \forall n,t, \tag{5.15}$$

where constraint (C1) implies that the sum of all node powers at time t, $p_{n,m}^{(t)}$, must be below the maximum power P_{max}; constraint (C2) implies that each node's power must be greater than zero; and constraint (C3) defines the possible orientation space for each beam orientation at time t, $o_n^{(t)}$, $n \in [1,N]$. To avoid beam overlapping and interference, it divides the half space $[\frac{-\pi}{2},\frac{\pi}{2}]$ into N equal sub-spaces. The considered problem is non-convex, as it aims to simultaneously maximize two fractions. Also, reducing the maximization problem to only min-max rate maximization over interference channel (i.e., by removing the EE from the objective function) makes it a classic NP-hard problem Luo & Zhang (2008). Hence, the considered problem is provably NP-hard as well. Solving the considered optimization problem using conventional techniques brings up several challenges that can be listed as follows,

- Conventional optimization methods may suffer from increased complexity due to the intricate algorithms and decision-making processes involved. The complexity of these approaches can lead to longer computation times and higher resource use, making them less efficient, particularly in real-time or resource-constrained scenarios.
- The computational resources needed to perform the allocation in conventional methods can be significant.
- Traditional approaches frequently rely on precise knowledge of the users' instantaneous SINRs/CSI to make allocation decisions. This requirement for accurate and up-to-date information can be challenging, especially in dynamic and unpredictable environments, where nodes are considered quasi-stationary and their locations, and hence the CSIs, is changing constantly in a random manner.

Therefore, considering the aforementioned challenges, RL is a good candidate to solve this problem in a dynamic manner.

Remark 1 : RL algorithms can learn the optimal policy directly from their own experiences. While traditional approaches heavily rely on predefined rules or heuristics and traditional machine learning techniques need ground truth, which is costly to obtain for large-scale systems, and lack adaptability in dynamic environment, RL methods can adapt and improve their performance over time through continuous interaction with the environment. This ability to learn from experience allows RL algorithms to refine their strategies and achieve higher levels of performance. Moreover, RL techniques are well-suited for dynamic environments, where the conditions or constraints may unpredictably change.

5.4 DDPG Preliminaries

DDPG is a a model-free, online, off-policy DRL algorithm that specifically deals with continuous states and action sets. The main elements of a DDPG and RL algorithm in general are outlined below:

- States S: It is a quantified definition of the environment encountered by the agent at certain time t. The environment represents anything that the agent can, directly or indirectly, interact with.
- Actions A: It is the decision taken by the agent considering the environment at time t.
- Policy $\pi(.)$: It is a state-action mapping. The RL agent aims to converge to the optimal policy that maximizes the cumulative reward.
- Reward: Maximizing the long-term reward function is the main objective of an RL problem. It is a quantitative feedback calculated following the action execution, as the environment moves from current state s_t to next state s_{t+1}.

At each time t, the agent at state $s_t \in S$ takes an action $a_t \in A$ following the policy $\pi(a_t/s_t)$. As a result, it receives a reward r_{t+1} and the state is updated to s_{t+1}.

The DDPG agent adopts an actor-critic approach to converge to an optimal policy π that maximizes the expected cumulative long-term reward $R_t = \sum_{i=t+1}^{\infty} \gamma r_{t+i+1}$, where $\gamma \in (0, 1]$ is the discount factor. The DDPG agent includes the following four main networks Mnih *et al.* (2015):

- An online actor network parameterized by θ^{μ}: This network takes as input the states and outputs the actions.
- An online critic network parameterized by θ^{Q}: This network takes as input the states and actions decided by the online actor network, and outputs the Q-function Q(s,a) that evaluates the given state-action tuple.
- A target actor network parameterized by $\theta^{\mu'}$: This network is a previous copy of the online actor network used for tracking the learned network to ensure stability.
- An target critic network parameterized by $\theta^{Q'}$: It is a previous copy of the online critic network.

The DDPG algorithm starts by an discovery phase where the agent discovers the environment by taking random actions and stores the corresponding resulting states transition and rewards in its memory. Afterwards, the algorithm learns the action using the actor network and adding certain exploration noise $\mathcal{N}_t$ to the action to cover more values of the continues action set. At each time t, the critic network is updated by minimizing the loss function, defined as shown in (5.16)

$$L(\theta^{Q}) = \frac{1}{N_B} \sum_{t} \left[Q(s_t, a_t|\theta^{Q}) - y_t \right]^2, \tag{5.16}$$

where N_B is the size of the batch sampled from the memory and y_t is calculated using the Bellman equation (see (5.17))

$$y_t = r_t + \gamma Q' \left(s_{t+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'} \right). \tag{5.17}$$

The actor network is updated using the sampled policy gradient (see (5.18))

$$\nabla_{\theta^{\mu}} \approx \frac{1}{N_B} \sum_{t} \left[\nabla_a Q(s, a|\theta^{Q})|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})|_{s=s_t} \right]. \tag{5.18}$$

Finally, the target networks are updated using (5.19a)-(5.19b)

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}, \tag{5.19a}$$

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}, \tag{5.19b}$$

where τ is the update rate of the target networks. The following theorem establishes the theoretical convergence of the DDPG RL algorithm.

Theorem 1:Fan, Wang, Xie & Yang (2020) Let $\mathcal{T}Q(s_t, a_t) = r_t + \gamma\mathbb{E}Q(s_{t+1}, a_{t+1})$ be the Bellmann optimality operator and $\mathcal{F}$ be the class of functions that the critic network can create. The difference between the estimated Q-value and $\mathcal{T}Q$ is defined as the one-step approximation error. Considering n independent experiences (s_t, a_t, s_{t+1}, r_t) and defining $\hat{Q}$ as the solution that minimizes the loss in (5.16), for any $0 \leq \epsilon \leq 1$ and $\delta \geq 0$, the one step approximation error is upper-bounded by (5.20)

$$\begin{aligned}\left\|\hat{Q} - \mathcal{T}Q\right\|^2 \leq (1+\epsilon)\sup_{g\in\mathcal{F}}\inf_{f\in\mathcal{F}}\left(\|f - \mathcal{T}g\|^2\right)\\ +C\cdot\frac{V_{\max}^2}{n\epsilon}\cdot\ln\mathcal{N}(\mathcal{F},\delta) + C'\cdot V_{\max}\cdot\delta\end{aligned} \tag{5.20}$$

where C and C' are two constants, $V_{max} = R_{max}/(1-\lambda)$ is the maximum value in the underlying Markov decision process (MDP), and $\mathcal{N}(\mathcal{F},\delta)$ is the δ-covering number associated with the function class $\mathcal{F}$, defined as the minimum number of euclidean balls of radius δ required to cover the space spanned by $\mathcal{F}$ applied on a fixed set of inputs.

Theorem 1 confirms that, under the assumptions of an MDP and the availability of i.i.d. experiences, the convergence of DDPG can be guaranteed within a bound. Building on this theorem, in Section V we introduce single-agent and multi-agent DDPG reinforcement learning algorithms, and then validate their convergence through numerical simulations presented in Figures 3-4 of Section VI.

5.5 DDPG-Based Beamforming Method

5.5.1 Single-Agent DRL Approach

MDP Formulation: The considered problem can be converted into an MDP problem. The actions, states, and reward are defined as follows:

- Actions: We define actions set $A = \{A_{o_1}, ..A_{o_N}, A_{p_1}, .., A_{p_M}\}$, where $A_{o,n}$ is the continuous action set taken on the orientation of the n^{th} beam, $i \in [1, N]$, and A_{p_m} is the continuous action set taken on the power of the m^{th} node, $m \in [1, M]$, where $M = \sum_n |M_n|$ is the total number of nodes in the network. Each action takes a value between 0 and 1. The beams orientations at each time t are extracted using the first N actions using (5.21)

$$o_n^{(t)} = \frac{\pi}{N} a_{o_n}^{(t)} - \frac{\pi}{2} + (n-1)\frac{\pi}{N}, n \in [1, N]. \tag{5.21}$$

  This divides the half space $[-\frac{\pi}{2}, \frac{\pi}{2}]$ into N equal sub-spaces, so assuming that beams cannot overlap to avoid interference, each beam orientation is learned in its corresponding sub-space. The power allocation is learned using the M following actions using (5.22), where $n \in [1, N], m \in [1, M_n]$

$$p_{n,m}^{(t)} = \frac{a_{p_i}^{(t)} P_{max}}{M}, i = \sum_{1}^{n-1} |M_n| + m. \tag{5.22}$$

- States: We define states set $S = \{S_1, .., S_M\}$, where each state is a continuous value $s_i^{(t)} \in S_i$ referring to the slope of the line between node i and its corresponding beam origin defined as shown in (5.23)

$$s_i^{(t)} = \frac{y_i^{(t)} - y_{0,n}}{x_i^{(t)} - x_{0,n}}, \tag{5.23}$$

  where $(x_i^{(t)}, y_i^{(t)})$ and $(y_{0,n}, x_{0,n})$ are the corresponding coordinates of node i at time t and its corresponding beam sources, respectively.
- Reward: The reward function is defined as the minimum rate in (5.10) for selected actions. Such a reward function is selected after exhaustively trying all possible options of the reward

design, i.e., considering the minimum rate, network EE, and the weighted sum of both functions. The minimum rate provides the most suitable outcome.

DRL Algorithm Development: Since the defined states and actions sets are continuous, in this study, we propose a DDPG-based solution to solve the considered optimization problem. The agent defined on the multi-beam transmitter level starts by initializing the actor critic networks parameters. For each episode, the states are reset to their initial values. Then, for the first N_0 time slots in the episode, the agent takes at each time slot t random actions, and after N_0 time slots, it takes actions using the actor network and adds exploration noise $\mathcal{N}_t$ to the chosen action. Then, the agent calculates the new orientation $o_n^{(t)}$ for each beam n and power $p_{n,m}^{(t)}$ for each node m in each cluster n using (5.21) and (5.22), respectively. It also calculates reward r_t using (5.10) and observes next state s_{t+1}. The resulting tuple (s_t, a_t, r_t, s_{t+1}) is stored in experience memory B and a random batch of size N_B is sampled from B. Finally, the agent updates the actor and critic networks. We assume that the nodes randomly move by a distance $d_\epsilon \leq \epsilon$ from their estimated location each 10 time slots. The proposed single-agent DDPG approach is summarized in Algorithm 1. The agent tests its training each N_{test} episodes. For the testing, an algorithm similar to the training algorithm is used. Except that during the test, the agent runs only one episode and always takes actions using the actor network and without adding and exploration noise. In addition, the resulting tuples in each time slot are not stored in memory, and the actor and critic networks are not updated. Theorem 1 confirms that, under the assumptions of an MDP and the availability of i.i.d. experiences, the convergence of DDPG can be guaranteed within a bound. Building on this theorem, we introduce single-agent and multi-agent DDPG reinforcement learning algorithms in Section V, and validate their convergence through numerical simulations presented in Figure 3 and Figure 4 of Section VI.

Complexity of the single-agent DDPG algorithm is $O(T.N_B.(C_{actor}+2C_{critic}))$, where C_{actor} and C_{critic} are the cost of forward/backward pass through the actor and critic networks, respectively Lillicrap *et al.* (2015). The cost of forward/backward pass largely depends on the architecture of the network and can be defined using the number of layers in each network and the number of input and output of each layer (see Eq. (5.24)).

$$C = \sum_{h=1}^{H} (n_{h,in}.n_{h,out}) + n_{h,out} \tag{5.24}$$

where H is the number of hidden layers in the network, $n_{h,in}$ is the number of input units in the layer, and $n_{h,out}$ is the number of output units. Consequently, computational complexity of the proposed single layer DDPG algorithm is $O(T.N_B.(M(6H+1)+N(3H+1)+3H^2+8H+2))$.

Algorithm 5.1 Single-agent DDPG beamforming algorithm

Algorithm: Single-agent DDPG beamforming algorithm
Input: $S, A, N_{ep}, T, N_0, \gamma$.
Initialize: Q, μ, Q', μ', B.
for $e = 0, 1, 2 \ldots N_{ep}$ **do**
  Reset $s_1 \in S$
  for $t \in [1, T]$ **do**
    if $t \leq N_0$ **then**
      Randomly choose $a^{(t)} = \{a_{o_1}^{(t)}, \ldots, a_{o_N}^{(t)}, a_{p_1}^{(t)}, \ldots, a_{p_M}^{(t)}\}$
    else
      Get $a^{(t)} = \mu(s^{(t)}|\theta^{\mu}) + \mathcal{N}^{(t)}$ according to the current policy and exploration noise
    end if
    Calculate $o_n^{(t)}, n \in [1, N]$ using (5.21)
    Calculate $p_{n,m}^{(t)}, n \in [1, N], m \in [1, M_n]$ using (5.22)
    Calculate reward $r^{(t)}$ using (5.10)
    Observe $s^{(t+1)}$
    Store $(s^{(t)}, a^{(t)}, r^{(t)}, s^{(t+1)})$ in experience memory B
    Sample random batch of size N_B from B
    Update Q, μ, Q', μ'
    if $mod(t, 10) = 0$ **then**
      Move randomly the nodes by a distance $d_\epsilon \leq \epsilon$ from their estimated location.
    end if
  end for
end for

This single-agent model outputs $N + M$ actions. Hence, by increasing the number of nodes or beams within the system, the actions size considerably increases, thereby slowing down the convergence of the system. To overcome this limitation, we develop a multi-agent RL approach by distributing the decision process among different agents that cooperate to reach the optimal beam orientation and power allocation in the network. Further detail on the multi-agent DRL approach is provided below.

5.5.2 Multi-Agent DRL Approach

In wireless network optimization, complex problems can be tackled effectively by breaking them down into smaller, manageable sub-problems and solving each one optimally. Inspired by this strategy, we propose a sequential multi-agent reinforcement learning approach that ensures a distributed beamforming task instead of a central one. The main task is divided into the following three main tasks: (1) optimizing beam orientation; (2) power allocation among beams; and (3) power allocation among nodes within each cluster. The different agents executing each of these sub-tasks are defined as follows:

- Beam orientation agent: This agent has M states defined as the slope values of the lines between the node i and its corresponding beam, as explained in the previous section, $S_o = \{S_{o_1}, .., S_{o_M}\}$. This variable offers an information of the elevation of node from the beam axis, which is sufficient to decide the optimal beam orientation to cover the corresponding nodes. It outputs N actions $A_o = \{A_{o_1}, ..A_{o_N}\}$. The beams orientation are calculated using (5.21). The reward, $r_o^{(t)}$, is defined to be the EE of the network, since the beam orientation affects the rates of all nodes, not only the minimum rate. Hence, focusing on the sum rate included in the EE equation would give a more balanced result than focusing on the min rate only.
- Beam power allocation: This agent has M states, corresponding to the channel information between each node and its corresponding beam, $S_p = \{h_{n,m}^2, n \in [1, N], m \in [1, M_n]\} = \{S_{p_1}, .., S_{p_M}\}$. It has N actions corresponding to the power allocated to each beam $A_p = \{p_n, n \in [1, N]\} = \{A_{p_1}, .., A_{p_N}\}$. The reward function, $r_p^{(t)}$, is the minimum rate of the network. This is so because maximizing the minimum rate guarantees that all node rates are above that minimum rate. By contrast, the EE function tends to maximize the total system capacity, which can be achieved by favoring nodes with good propagation

conditions—frequently at the expense of others with weaker channels. This inherently creates an imbalance among the nodes' achievable rates and may significantly degrade fairness by deteriorating the minimum rate to very low values. Hence, maximizing the minimum rate enforces the beam power agent and node power allocation agent to learn a policy that would increase data rate of all the nodes within a cluster while ensuring fairness among them.

- Node power allocation: There are N node power allocation agents, one per cluster. Each of these agents has M_n states, $n \in N$, defined as the channel information between the considered beam and its corresponding nodes. $S_{p_n} = \{h^2_{n,m}, m \in [1, M_n]\} = \{S_{p_{n,1}}, .., S_{p_{n,M_n}}\}$. It outputs M_n actions, $A_{p_n} = \{p_{n,m}, m \in [1, M_n]\} = \{A_{p_{n,1}}, .., A_{p_{n,M_n}}\}$. Following the same logic behind the choice of the beam power allocation agent's reward function, the reward function, $r^{(t)}_{p,n}$, for each agent n is the minimum rate of its corresponding cluster.

Unlike the single-agent model, this model alleviates the action size for each node, where the number of actions taken by the first and second agents is N, and the number of actions taken by the last agents equals the number of nodes in each agent. To strike a suitable balance between system EE and minimum rate, a multi-level reward function, described as follows, is selected for these agents. The orientation agent's reward is set to be the overall EE of the network, the beam power agent's reward is defined as the overall min rate of the network, while the reward of each node's power agent is the min rate of each corresponding cluster.

The proposed multi-agent DDPG-RL framework executes the following tasks sequentially.

1. The orientation agent decides the beam orientations based on the nodes elevation from the beam axis using (5.21).
2. Considering the channel gain based on the chosen orientation at time t, the beam power agent decides the power allocated for each beam using the (5.25)

$$p^{(t)}_n = \frac{a^{(t)}_{p_n} P_{max}}{N}, n \in [1, N], \tag{5.25}$$

   where $p^{(t)}_n$ is the power assigned for beam n and $a^{(t)}_{p_n}$ is the n^{th} action taken by the second agent.

3. Once the power allocated for each beam is decided by the beam power agent at time t, each node power agent decides on the power allocated for each node within its corresponding cluster, using (5.26)

$$p_{n,m}^{(t)} = \frac{a_{p_{n,m}}^{(t)} p_n^{(t)}}{M_n}, n \in [1, N], m \in [1, M_m], \tag{5.26}$$

where $a_{p_{n,m}}^{(t)}$ is the m^{th} action taken by agent n.

4. The agents receive their rewards for their selected actions.
5. The agents update their actor and critical networks.

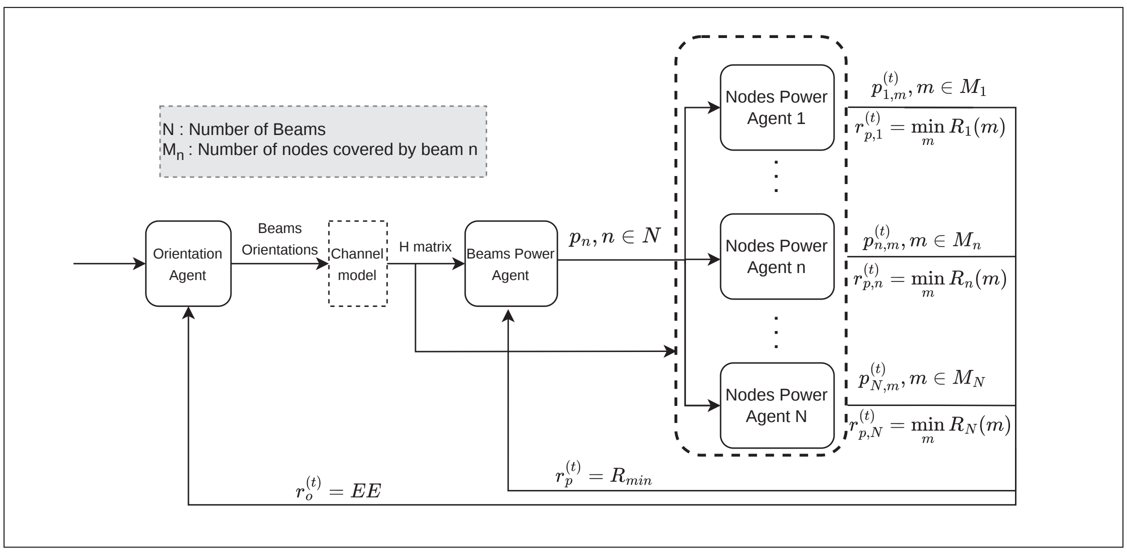


Figure 5.2 Structure of the proposed multi-agent DDPG model

The structure of the proposed multi-agent model is illustrated in Figure 5.2. To further improve the results, we propose a sequential DDPG approach where we introduce a delay between the discovery phase of the beam orientation agent and the discovery phase of the remaining agents. In fact, the discovery phase of the power agents starts simultaneously with the learning of the orientation agent. The reason behind choosing the sequential model is that the nodes are quasi-stationary. Hence, once converged, beam orientations do not considerably change. Consequently, the sequential model focuses first on learning the beam orientation, then learns the optimal power distribution using a close to optimal orientation-based data. Since a given agent focuses on learning a particular action optimally instead of learning all the possible actions, the sequential approach shows improved capability to learn optimal solution to intricate optimization problem (5.12) in the complex and dynamic UOWC environment. The proposed sequential multi-agent DDPG approach is summarized in Algorithm 1. Complexity of the proposed multi-agent DDPG algorithm is defined as the sum of complexity values of all networks, hence it is $O(T.N_B.(M(12H+1)+N(3H^2+12H+4)+6H^2+18H+4))$. Compared to the single-agent DDPG algorithm, the sequential multi-agent DDPG algorithm requires a higher computational complexity. This will be further demonstrated by simulations in the following section.

5.6 Results and Discussion

5.6.1 Simulation Setup

In this section, we evaluate the performance of the proposed algorithms. The simulations were performed using Python 3.7. We consider a 25×25 m^2 square area where nodes are randomly located in each cluster. We consider three clusters with one beam allocated to each cluster, i.e. $N = 3$. The transmitted beams are separated by a distance of 10 cm vertically. We consider pure sea water with extinction coefficient $c = 0.043$ for $\lambda = 514$ nm. The node's aperture area is set to $A_r = 19.6$ mm^2 and the noise power $\sigma^2 = 5 \times 10^{-12}$. The half power angle is $\theta_{1/2} = 7.5°$ and the concentrator field of view (FOV) is $\psi = 70°$. The water refractive index is $n_{water} = 1.33$Li *et al.* (2020a); Palitharathna *et al.* (2021). To replicate node's position uncertaintly due to underwater turbulence, we assume that each 10 steps, the nodes move from their estimated position by $\epsilon = 0.25$ m, defined as the uncertainty radius. The maximum power budget $P_{max} = 10$ W. The system parameters are summarized in Table 5.3.

The hyperparameters of the proposed DDPG algorithms are defined as follows. Each DDPG agent has four networks: online and target actor networks, denoted as μ and μ', respectively, and online and target critic networks, denoted as Q and Q', respectively. These networks have one input layer, one hidden layer, and one output layer each. The hidden layer has 100 neurons. The activation function for the input and hidden layer is the RELU function. For the actor network's

Algorithm 5.2 Sequential multi-agent DDPG beamforming algorithm

Algorithm: Sequential multi-agent DDPG beamforming algorithm

Input: $S_o, S_p, S_{p_n}, n \in N, A_o, A_p, A_{p_n}, n \in N, N_{ep}, N_{0,o}, N_{0,p}, \gamma$.

Initialize:

- $Q_o, \mu_o, Q'_o, \mu'_o, B_o$
- $Q_p, \mu_p, Q'_p, \mu'_p, B_p$
- $Q_{p,n}, \mu_{p,n}, Q'_{p,n}, \mu'_{p,n}, B_{p,n}, n \in [1, N]$.

for $e = 0, 1, 2 \ldots N_{ep}$ **do**

  Reset $s_o^{(1)}, \in S_o$, $s_p^{(1)} \in S_p$, and $s_{p_n}^{(1)} \in S_{p_n}, n \in [1, N]$

  for $t \in [1, T]$ **do**

    if $e \le N_{0,o}$ **then**

      Randomly choose $a_o^{(t)} = \{a_{o_1}^{(t)}, \ldots, a_{o_N}^{(t)}\}$

    else

      Get $a_o^{(t)} = \mu_o(s_o^{(t)}|\theta_o^{\mu_o}) + \mathcal{N}^{(t)}$ according to current policy and exploration noise

    end if

    Calculate $o_n^{(t)}, n \in [1, N]$ using (5.21)

    Update $s_p^{(t)}$ according to $o_n^{(t)}$

    if $e \le N_{0,p}$ **then**

      Randomly choose $a_p^{(t)} = \{a_{p_1}^{(t)}, \ldots, a_{p_N}^{(t)}\}$

    else

      Get $a_p^{(t)} = \mu_p(s_p^{(t)}|\theta_p^{\mu_p}) + \mathcal{N}_t$ according to current policy and exploration noise

    end if

    Calculate $p_n^{(t)}, n \in [1, N]$ using (5.25)

    Update $s_{p_n}^{(t)}$ according to $p_n^{(t)}$

    if $e \le N_{0,p}$ **then**

      Randomly choose $a_{p_n}^{(t)} = \{a_{p_{n,1}}^{(t)}, \ldots, a_{p_{n,N}}^{(t)}\}$

    else

      Get $a_{p_n}^{(t)} = \mu_{p_n}(s_{p_n}^{(t)}|\theta_{p_n}^{\mu_{p_n}}) + \mathcal{N}_t$ according to current policy and exploration noise

    end if

    Calculate $p_{n,m}^{(t)}, n \in [1, N], m \in M_n$ using (5.26)

    Calculate $r_o^{(t)}, r_p^{(t)}$, and $r_{p,n}^{(t)}$ using (5.11) and (5.10)

    Observe $s_o^{(t+1)}, s_p^{(t+1)}, s_{p,1}^{(t+1)}, .., s_{p,N}^{(t+1)}$

    Store $(s_o^{(t)}, a_o^{(t)}, r_o^{(t)}, s_o^{(t+1)})$ in B_o

    Sample random batch of size N_B from B_o and update Update Q_o, μ_o, Q'_o, μ'_o

    if $e \ge N_{0,o}$ **then**

      Store $(s_p^{(t)}, a_p^{(t)}, r_p^{(t)}, s_p^{(t+1)})$ in B_p

      Sample random batch of size N_B from B_p and update Q_p, μ_p, Q'_p, μ'_p

      Store $(s_{p,n}^{(t)}, a_{p,n}^{(t)}, r_{p,n}^{(t)}, s_{p,n}^{(t+1)})$ in $B_{p,n}, n \in [1, N]$

      Sample random batch of size N_B from $B_{p,n}, n \in [1, N]$ and update $Q_{p,n}, \mu_{p,n}, Q'_{p,n}, \mu'_{p,n}, n \in [1, N]$

    end if

    if $mod(t, 10) = 0$ **then**

      Move the nodes randomly by a distance $d_\epsilon \le \epsilon$ from their estimated location.

    end if

  end for

end for

Table 5.3 System parameters

Parameter	Value
Simulation area	$25 \times 25\ m^2$
Beam centers	$\{(0, -0.1); (0, 0); (0, 0.1)\}$
Number of beams	N=3
Extinction coefficient	$c = 0.043$
Wavelength	$\lambda = 514$ nm
Nodes aperture area	$A_r = 19.6\ mm^2$
Noise power	$\sigma^2 = 5 \times 10^{-12}$
Half power angle	$\theta_{1/2} = 7.5°$
FOV	$\psi = 70°$
Water refractive index	$n_{water} = 1.33$
Uncertainty radius	$\epsilon = 0.25$ m
Power budget	$P_{max} = 10$ W

output layer, we choose the tanh function to obtain an action in a defined set, i.e. between $[-1, 1]$. Conversely, for the critic network's output layer, we select the linear activation function. The experience memory capacity is set to 8000 with a batch size of $N_B = 128$. The actor networks learning rate is $\alpha_a = 10^{-4}$, while the critic networks learning rate is $\alpha_c = 10^{-3}$ and the discount factor is $\gamma = 0.99$. We use the Adam optimizer to update the target networks, and the update rate is set to $\tau = 0.01$. The number of episode is set to $N_{ep} = 80$ episodes, with each episode including $T = 50$ steps. The discovery phase of the orientation agent takes the first $N_{0,o} = 10$ episodes, while the discovery phase of the power agents takes the following 20 episodes and ends at $N_{0,p} = 30^{th}$ episode. The DDPG-related parameters are summarized in Table 5.4.

5.6.2 Performance Assessment of the Proposed Models

We first evaluate the performance of the proposed single-agent and sequential multi-agent models. We also evaluate the effect of adding the sequential feature to the multi-agent DDPG solution by evaluating the performance of non-sequential multi-agent DDPG approach. Furthermore, we compare the results with the multi-agent DDPG approach where the reward for all agents is either set to the minimum rate or to EE. We consider the case of three beams where the first and third clusters include four nodes, while the second cluster includes three nodes, i.e. $M_1 = M_3 = 4$

Table 5.4 DDPG algorithm hyperparameters

Parameter	Value
Experience memory capacity	8000
Batch size	$N_B = 128$
Learning rate	$\alpha_a = 10^{-4}$, $\alpha_c = 10^{-3}$
Discount factor	$\gamma = 0.99$
Optimizer	Adam optimizer
Update rate	$\tau = 0.01$
Number of episodes	$N_{ep} = 80$
Number of steps per episode	$T = 50$
Orientation agent discovery end	$N_{0,o} = 10$
Power agents discovery end	$N_{0,p} = 30$

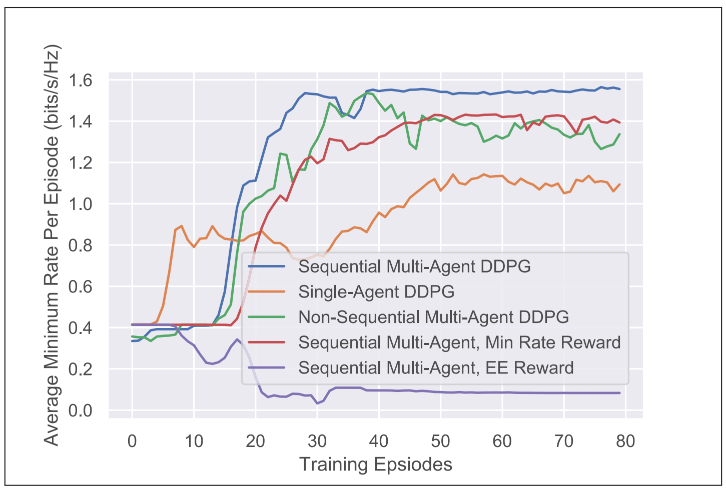


Figure 5.3 Average minimum rate per episode for $N = 3$, $M_1 = M_3 = 4$ and $M_2 = 3$, $T = 50$, $\epsilon = 0.25$, $N_0 = 30$, $N_{0,o} = 10$, and $N_{0,p} = 30$

and $M_2 = 3$. The results of the network's minimum rate and EE for these five schemes are shown in Figure 5.3 and Figure 5.4, respectively. These figures further justify the convergence of the proposed algorithms, proven theoretically above. The network's minimum rate for the single-agent DDPG beamforming model converges to around 1.1 bits/s/Hz after 50 episodes, while the network's minimum rate for the proposed sequential multi-agent beamforming model converges to around 1.5 bits/s/Hz after 30 episodes. Furthermore, the EE for the single-agent

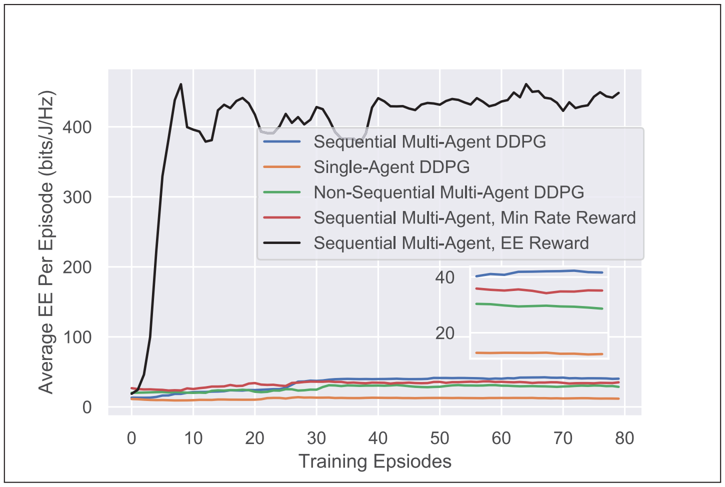


Figure 5.4 Average EE per episode for $N = 3$, $M_1 = M_3 = 4$ and $M_2 = 3$, $T = 50$, $\epsilon = 0.25$, $N_0 = 30$, $N_{0,o} = 10$, and $N_{0,p} = 30$

DDPG beamforming model reaches 15 bits/J/Hz, while the EE for the sequential multi-agent DDPG approach reaches 40 bits/J/Hz. This supports our claim that, for large number of nodes in the network, the single-agent's action size becomes considerably large, thereby affecting the performance of such networks. By contrast, adopting the multi-agent model where the decision is distributed among different agents alleviates the task and ensures better and faster convergence. Furthermore, introducing the sequential feature to the multi-agent DDPG approach ensures better minimum rate and EE and faster convergence. For instance, we find that, as compared to the non-sequential approach, the sequential multi-agent DDPG approach increases the network's minimum rate and EE by 0.1 bits/s/Hz and 10 bits/J/Hz, respectively, and converges 5 episodes earlier. Furthermore, the proposed sequential multi-agent approach gives better and faster convergence than the sequential multi-agent approach with all rewards set to R_{min}. Finally, the sequential multi-agent approach with all rewards set to EE maximized the EE compared to all the other approaches at the cost of a very poor network's minimum rate around 0.1 bits/s/Hz. This is well aligned with the analytical explanation of the choice of the reward functions presented above. Based on the evidence, we conclude that the proposed sequential multi-agent DDPG beamforming method effectively balances the network's minimum rate and EE trade-off.

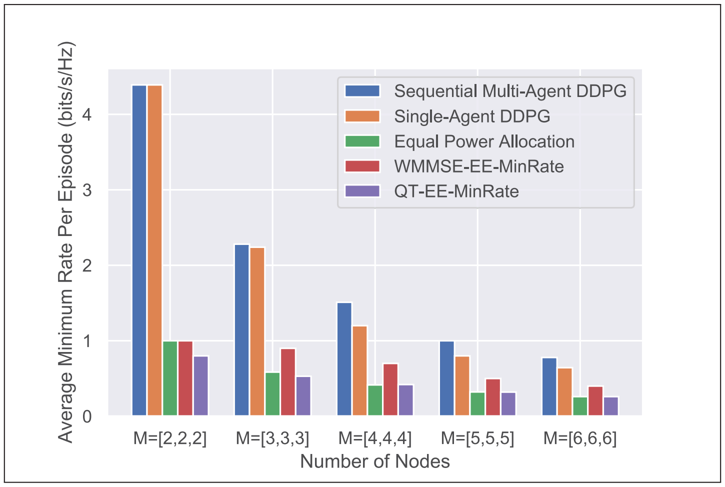


Figure 5.5 Average minimum rate per episode for different M_n values, $N = 3$, $T = 50$, $\epsilon = 0.25$, $N_0 = 30$, $N_{0,o} = 10$, and $N_{0,p} = 30$

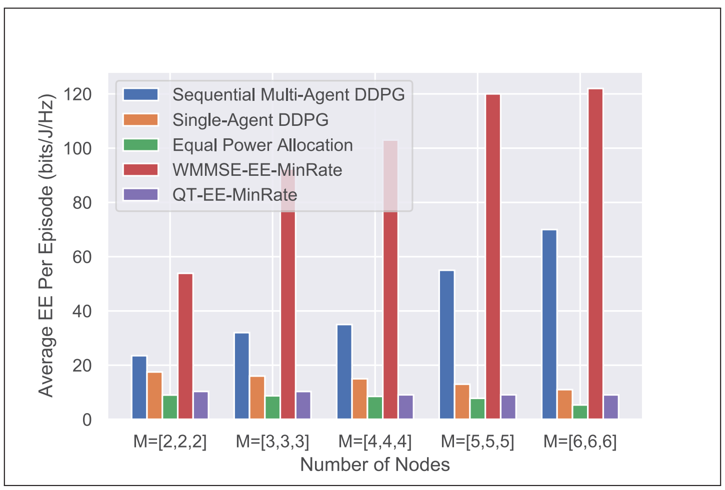


Figure 5.6 Average EE per episode for different M_n values, $N = 3$, $T = 50$, $\epsilon = 0.25$, $N_0 = 30$, $N_{0,o} = 10$, and $N_{0,p} = 30$

5.6.3 Performance Comparison with Benchmarks

Next, we evaluate the performance of the proposed single-agent and sequential multi-agent DDPG beamforming methods as a function of the number of nodes per cluster and water condition. For comparison, we use the following benchmarks.

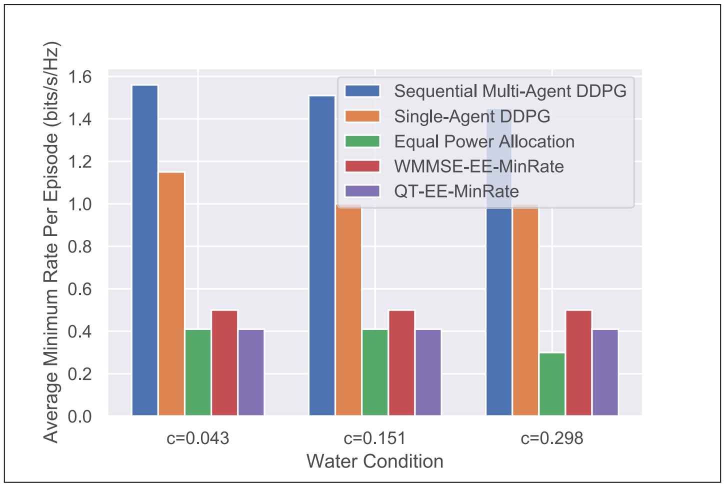


Figure 5.7 Average minimum rate per episode for different water conditions

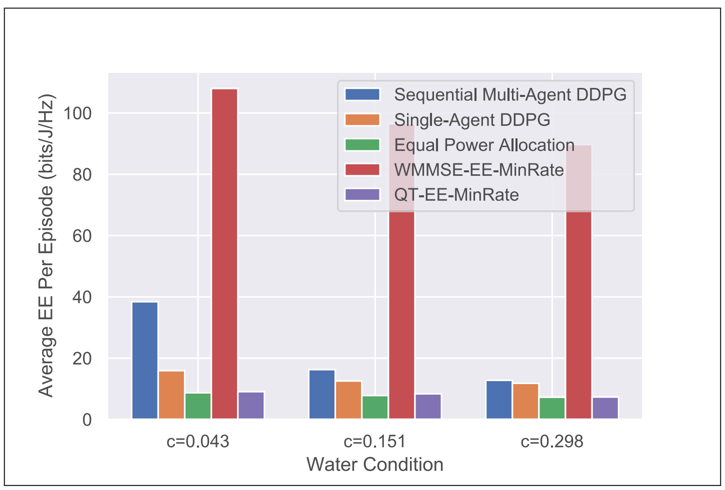


Figure 5.8 Average EE per episode for different water conditions

- Weighted minimum mean-square error (WMMSE)-based EE maximization with min-rate constraint (WMMSE-EE-MinRate): EE is defined as a fraction where the nominator is the sum of rates. Consequently, power allocation task is performed using fractional programming with Dinkelbach's method Dinkelbach (1967), combined with the WMMSE framework for handling the non-convex rate expressions Shi, Razaviyayn, Luo & He (2011). The beam orientations are defined according to the estimated locations of nodes.

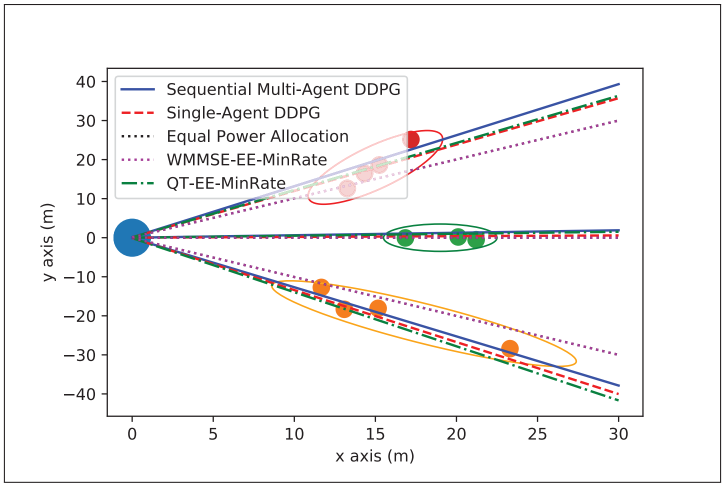
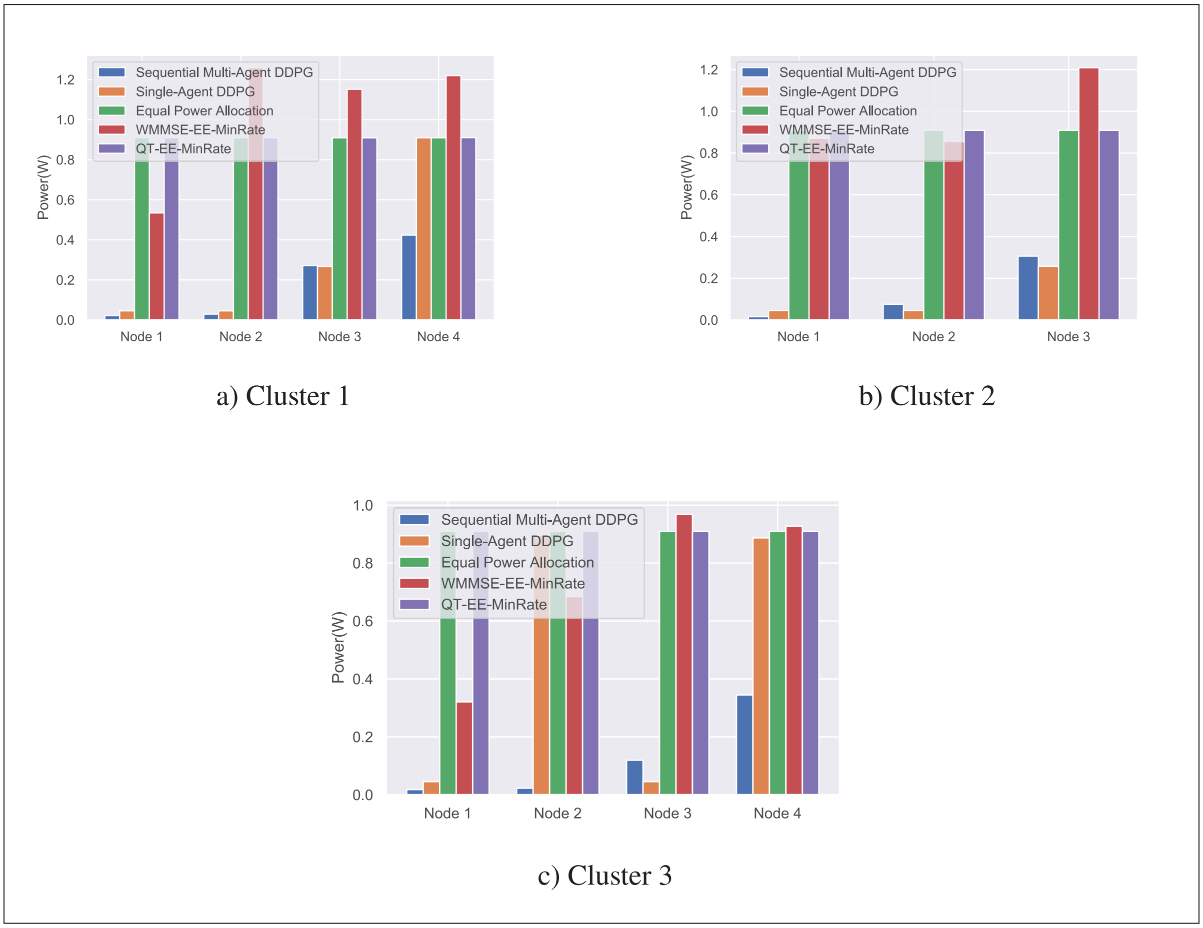


Figure 5.9 Comparison of the resulting beam axis orientations of the proposed methods and benchmarks

- Quadratic Transform (QT)-based EE maximization with min-rate constraint (QT-EE-MinRate): Similarly to the previous benchmark, this one uses fractional programming with Dinkelbach's method to solve the EE maximization problem, which is followed by using QT instead of WMMSE to solve the non-convex sum rate expression Shen & Yu (2018); Zappone & Jorswieck (2015). The beam orientations are defined according to the estimated locations of nodes.
- Equal power allocation: The beam orientation task is performed using a DDPG agent as defined in the proposed multi-agent DDPG approach, while the nodes are assigned equal power, i.e. $P_{n,m} = P_{max}/M$.

The network's average minimum rate per episode is a decreasing function of the number of nodes per cluster (see Figure 5.5). This is because increasing the number of nodes increases the interference in the network, thereby degrading the minimum rate. For up to three nodes per beam, the proposed single and multi-agent DDPG approaches have the same performance in terms of the network's minimum rate. As the number of actions for the single-agent model is a function of the number of nodes, keeping a small number of nodes per cluster ensures a small action set size. Consequently, the performance of the two proposed approaches is similar. However, starting from four nodes per cluster, the proposed sequential multi-agent approach increases the



a) Cluster 1

b) Cluster 2

c) Cluster 3

Figure 5.10 Comparison of the resulting power distribution among nodes of the proposed methods and benchmarks

network's minimum rate by around 0.2 bits/s/Hz as compared to the single-agent approach. At five nodes per cluster, the proposed sequential multi-agent approach achieves the best minimum rate in the range of 1 bits/s/Hz, as compared to the other approaches. Based on this evidence, we conclude that not exceeding five nodes per cluster in the considered system will ensure tolerable rates for all nodes in the network. Regarding the considered benchmarks, the resulting average minimum rate per episode is considerably lower than that of the proposed models, reaching 0.3 bits/s/Hz for the equal power allocation and QT-EE-MinRate, and 0.4 bits/s/Hz for WMMSE-EE-MinRate. This is so because the equal power allocation benchmark maximizes the power for all nodes, thus maximizing the interference within the network, and the other two benchmarks aim to maximize the network's sum rate at the cost of the minimum rate.

The EE for different numbers of nodes is illustrated in Figure 5.6. The proposed sequential multi-agent DDPG approach has increasing EE values by increasing the number of nodes per cluster, exceeding 50 bits/J/Hz for the case of five nodes per cluster. Conversely, the proposed single-agent approach gives decreasing EE values by increasing the number of nodes per cluster, reaching 15 bits/J/Hz at five nodes per cluster. This is due to the fact that, in the single-agent decision, only the minimum rate is considered, i.e. without taking care of the EE results; by contrast, the multi-agent approach optimizes the trade-off between network's minimum rate and EE, and achieves a balanced results for both metrics. As for the equal power allocation and QT-EE-MinRate benchmarks, the EE values are almost constant for all number of nodes per cluster, ranging about 8 bits/J/Hz for the equal power allocation and around 10 bits/J/Hz for the QT-EE-MinRate. Concerning WMMSE-EE-MinRate, the EE is an increasing function reaching 120 bits/J/Hz for 6 nodes per cluster. This increase on EE is at the expense of min-rate performance as proven by Figure 5.5.

Table 5.5 Attenuation coefficient of various types of water for $\lambda = 514$ nm. Taken from: Mobley (1994)

Water type	$c(m^{-1})$
Pure sea water	0.043
Clean ocean	0.151
Coastal ocean	0.298
Turbid harbor	2.19

Furthermore, we evaluate the effect of water condition on the performance of the proposed models. To this end, we consider three different water conditions defined by variable water attenuation coefficient values, c , as shown in Table 5.5. The results displayed in Figures 5.7- 5.8 reveal that the average minimum data rate and EE of the proposed models deteriorate with increasing the attenuation coefficient. This can be explained by the fact that increasing the attenuation coefficient affects the communication channel, which has an impact on the data rate and EE of the nodes. However, the proposed models outperform the benchmarks in terms of minimum data rates for the different water conditions that are considered. For instance, for $c = 0.043$, the sequential multi-agent DDPG model achieved a 74% higher minimum rate per episode as compared to the equal power allocation and QT-EE-MinRate, and a 68% higher min-rate than WMMSE-EE-MinRate. Furthermore, the single-agent DDPG model achieved a 64% higher minimum rate than the equal power allocation and QT-EE-MinRate, and a 56% higher minimum rate than WMMSE-EE-MinRate. Regarding the EE performance, the sequential multi-agent and single agent models achieved 77% and 45% higher EE than equal power allocation and QT-EE-MinRate benchmarks, respectively. In its turn, WMMSE-EE-MinRate yielded a better EE than the proposed models, which can be explained by the fact that this benchmark is maximizing EE at the expense of the min-rate, despite the constraint set on min-rate. Finally, we can conclude that the proposed DDPG frameworks adapt well to the varying underwater environment.

To further evaluate the performance of the proposed methods in comparison with considered benchmarks, we consider a specific case study with three clusters of nodes, where the first (M_1) and third (M_3) clusters have fours nodes, while the second (M_2) cluster has three nodes. The resulting beam-axis orientations and power distribution among the nodes are displayed in Figures 5.9- 5.10, respectively. The resulting beam axis for the proposed methods and the equal power benchmark positions itself along the updated node positions, unlike WMMSE-EE-MinRate and QT-EE-MinRate that consider the estimated locations of nodes. This is so because the first three methods adopt the DDPG to converge to the optimal orientation according to the updated locations of the nodes. As for the power distribution among nodes, a certain balance is guaranteed among nodes' powers for the proposed methods, with higher power allocated to further nodes from the source (see Figure 10). Regarding the benchmarks, we observe an almost equal power allocation for the QT-EE-MinRate benchmark, which explains similarity in min-rate and EE results with the equal power allocation benchmark. In our case study, the proposed sequential multi-agent DDPG framework yields improved beam orientation and power distribution among nodes. As confirmed by the results presented in Figures 6-7, it achieves both higher EE and better rate fairness as compared to the benchmark schemes.

Comparison with Exhaustive Search and Branch and Bound Methods: The performance of the proposed sequential multi-agent DDPG and single-agent DDPG is also evaluated in

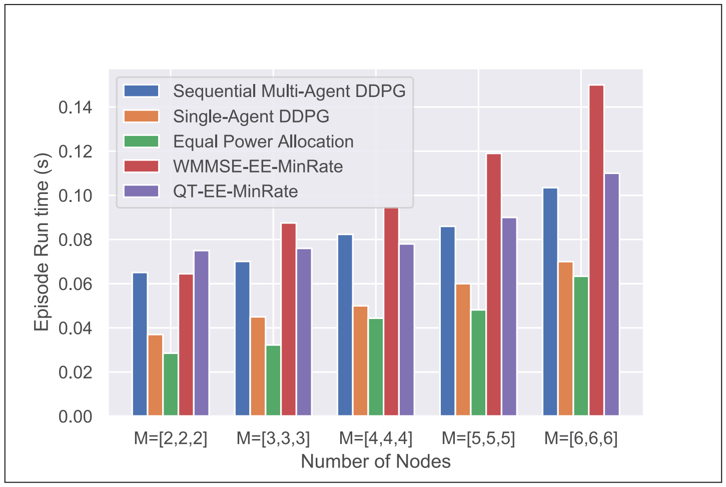


Figure 5.11 Episode run time for different M_n values, $N = 3$, $T = 50$, $\epsilon = 0.25$, $N_0 = 30$, $N_{0,o} = 10$, and $N_{0,p} = 30$

Table 5.6 Comparison with the exhaustive search method

	Minimum rate		EE	
	M=[2,2]	M=[3,3,3]	M=[2,2]	M=[3,3]
Sequential Multi-agent DDPG	4.4	2.24	26.16	14.87
Single-agent DDPG	4.38	2.22	21,43	14.26
Exhaustive search	0.42	0.29	14.32	14.57

Table 5.7 Comparison with the branch and bound method

	Minimum rate			EE		
M	[2,2,2]	[3,3,3]	[4,4,4]	[2,2,2]	[3,3,3]	[4,4,4]
Sequential Multi-agent DDPG	4.39	2.28	1.51	23.5	32	35
Single-agent DDPG	4.38	2.24	1.2	17.5	16	15
Branch and Bound	0.99	0.585	0.5	109	110	109

comparison to the exhaustive search and branch and bound method. Considering the high complexity of these methods, we limit the study to a small number of nodes. Specifically, we consider the following two cases for the exhaustive search method: (a) two clusters with two nodes each, and (b) two clusters with three nodes each. We also consider the three following cases for the branch and bound method: (a) three clusters with two nodes each, (b) three clusters with three nodes each, and (c) three clusters with four nodes each. To apply the exhaustive search, we discretize the continuous set of orientations and power to 10 equidistant values. The results are depicted in Table 5.6. We observe that both DDPG-based RL methods achieve a better minimum rate and EE as compared to the exhaustive search method. Of note, the computational complexity of an exhaustive search increases exponentially with the number of available beam orientations and transmit power levels. Therefore, in practice, exhaustive search is typically limited to selecting transmit power and beam orientations from a discrete set of possible actions. This limitation leads to sub-optimal solutions, despite the fact that, theoretically, exhaustive search can find the global optimum. By contrast, DDPG-based RL algorithms can select continuous action values with polynomial computational complexity. Consequently, they achieve superior performance in terms of minimum rate and EE as compared to the considered exhaustive search. Concerning the branch and bound method, we adopt the case of EE maximization with constraint on min-rate. The results show that the branch and bound method achieves higher EE results than the proposed DDPG-based RL methods, but with significantly smaller min-rate results. This outcome can be explained by the fact that, in this case, the branch and bound method tries to maximize the EE on the expense of min-rate. The branch and bound method is globally optimal for a single-objective EE maximization problem with a min-rate constraint. However, the proposed sequential multi-agent DDPG targets a fundamentally different joint objective that simultaneously maximizing both EE and min-rate fairness for which no classical global solver exists at tractable complexity. In this joint-objective sense, branch and bound achieves high EE at the cost of severely degraded min-rate (0.5˘0.99 bits/s/Hz), failing to balance both metrics. The proposed method, by contrast, achieves a superior trade-off across both objectives jointly. These results confirm that, in practice, the proposed sequential multi-agent DDPG-based RL algorithm not only converges, but also outperforms globally optimal solutions, which can be attributed to the latter's inherent implementation complexity.

5.6.4 Algorithm Complexity Assessment

Finally, we evaluate the complexity of the proposed algorithms. Figure 5.11 shows a comparison of the runtime per episode for the different considered algorithms. The single-agent algorithm has a lower runtime per episode than the sequential multi-agent allocation benchmark, reaching around 0.06 s for five nodes per cluster. Meanwhile, the sequential multi-agent algorithm's runtime per episode exceeds 0.08 s for five nodes per cluster. The proposed single-agent algorithm has one running agent, while the sequential multi-agent algorithm includes $2 + N$ agents, which makes it slower and more complex than the single-agent algorithm. Concerning the benchmarks, the equal power benchmark has a lower runtime than the proposed methods. This is so because it has one running agent that decides only on the beam orientations, which lightens the complexity of the algorithm. Meanwhile, QT-EE-MinRate has a similar run time as compared to the sequential multi-agent approach, and WMMSE-EE-MinRate exhibits the highest runtime for three or higher nodes per cluster. This behavior can be explained by the fact that these two benchmarks rely on repeated alternating updates and feasibility enforcement at every snapshot, whereas the trained DRL policy amortizes the optimization effort into an offline training stage, thereby enabling fast online decision-making. More specifically, the proposed framework's computational complexity is primarily dominated by neural network operations (i.e., multiplications and additions) and does not require any additional iterative optimization procedures.

Finally, we can conclude that the proposed single-agent and sequential multi-agent solutions achieve good results for the network's minimum rate/EE tradeoff. Compared to the single-agent model, the sequential multi-agent model enhances the network fairness and EE for higher number of nodes, with a slight increase in the run time. The sequential learning feature further improves the results of the multi-agent solution. Also, the proposed models considerably increases the network's fairness as compared to the considered benchmarks. Table 5.8 presents the results of a comparative analysis of the proposed models with the benchmarks.

Table 5.8 Comparative analysis of the proposed method with the considered benchmarks

Criteria	Equal Power	WMMSE-EE-MinRate	QT-EE-MinRate	Exhaustive search	Branch and bound	Single-agent DDPG	Sequential multi-agent DDPG
Complexity	Low	High	Moderate	Very high	Very high	Low	Moderate
Scalability	Poor	Poor	Very poor	Very poor	Very poor	Moderate	High
Optimality	Sub-optimal	Sub-optimal	Sub-optimal	Sub-optimal	Globally optimal (for given objective)	Near-optimal	Optimal

5.7 Conclusion

In this comprehensive exploration of the domain of UOWC, we aimed to optimize the beamforming task for a multi-beam network using the NOMA technique within each transmitted beam. A joint beam orientation and power control problem was formulated to concurrently optimize both system EE and minimum rate. To efficiently solve this problem, single-agent and multi-agent DDPG RL solutions were developed. A comparison of these solutions with the considered benchmarks- namely the equal power allocation, WMMSE-EE-MinRate and QT-EE-MinRate, proved that the proposed solution ensures more balanced results for the system's minimum rate and EE tradeoff. In addition, by distributing the task among different agents and hence alleviating the decision for each agent, the multi-agent solution improved the EE and fairness as compared to the single-agent solution for a higher number of nodes per cluster.

The proposed methods can be deployed in real-world scenarios by adding a control unit to the multi-beam transmitter. In the case of AUVs, the control unit is usually already in place, since the AUVs are remotely monitored. Importantly, the training of the DDPG agent is also performed offline. Once the model converges, only the trained model is used to perform the beamforming task. Importantly, the computationally intensive training phase is limited to only a short time (e.g., a couple of seconds as depicted by the presented results), where the convergence is achieved within 50 episodes and the runtime per each episode is less than 0.1 seconds. Thus, our proposed DDPG framework is feasible for practical UOWC systems.

The proposed methods are designed for quasi-stationary nodes, perfect SIC and CSI conditions. In our future work, we intend to investigate the case of mobile nodes where the receiving sensors are mobile AUVs traveling underwater. Under such conditions, the clustering task should be conducted dynamically. Furthermore, to alleviate the power consumption on the underwater system and realize training using more powerful equipment, the integration of digital twin that replicates the underwater propagation scenario will be investigated. In this context, we will address the case of imperfect CSI estimations and develop an appropriate RL-guided robust optimization algorithm. Finally, a prospective experimental evaluation of the proposed models will be conducted to further demonstrate their effectiveness in real-world underwater scenarios.

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

As UOWC is gradually getting more scholarly attention, consistent evidence has been accumulated to indicate that it can serve as a potential means of communication in 5G. This remarkable interest can be explained by the advantages afforded by UOWC, such as high data rates over short distances (in Gbps), fast propagation, and reduced energy consumption of used light sources. However, despite these advantages, this type of communications still faces several challenges, mostly related to the harsh conditions in marine environments, including absorption and scattering of light, oceanic turbulence, background noises, and inter-node and multipath interference. Further challenges include the light source, specifically PEs and misalignment problems, the range/beamwidth trade off, and the limited energy budget. Although a substantial amount of previous research has sought to find solutions to these problems, it remains difficult to solve them using traditional techniques due to the dynamic nature of underwater space and its dependency on multiple factors. For that reason, RL algorithms can be a promising solution for UOWC as it learns by trial and error without requiring prior knowledge of the environment.

In this thesis, we aimed to construct an autonomous SON for UOWC robust against challenging underwater environment. Specifically, we focused on proposing novel solutions to maintain the connectivity between nodes, extending the network lifetime, and solving the fairness problem in UOWC networks. We also evaluated the effect of oceanic turbulence and PEs on frequently implemented underwater links -namely, the MRC-based direct link and the OIRS-based indirect link.

In Chapter 3, we proposed novel RL-based beam adaptation solutions for P2P UOWC systems. The beamwidth, beam orientation, or both were optimized using Q-learning and SARSA algorithms. The results revealed show that the proposed solutions yielded optimal beamwidth and orientation of the light source and improved the success rate of the optical link. We also found that our solutions guaranteed better SNR as compared to literature for four different underwater environments.

Furthermore in Chapter 4, we assessed the outage probability for UOWC systems over oceanic turbulence with PEs. A combined UOWC system including an MRC-based direct link and OIRS-assisted indirect link was considered, where the oceanic turbulence was modeled using the EGG distribution. The results showed that the derived outage probability expressions perfectly aligned with the MC simulation results. Based on this evidence, we conclude that the proposed system can enhance the connectivity of high-speed UOWC under different underwater channel conditions.

Next, in Chapter 5, we focused on link adaptation for P2P direct air-to-underwater optical communications, using DDPG and OFDM technique under noisy CSI conditions. The proposed solution was found to dynamically adapt the beamwidth, beam orientation, and power levels, treating this entities as continuous variables to account for the inherent randomness in real-world scenarios. The results demonstrated that the proposed solution outperforms the Q-learning-based link adaptation solution in terms of success rate and consumed energy.

Finally, in Chapter 6, we treated the resource allocation problem to optimize fairness and EE in NOMA-aided UOWC networks. A dynamic beamforming solution was proposed for a multi-beam UOWC network. NOMA was applied within each transmitted beam to support multi-user communications. We aimed to maximize the system EE of a multi-beam UOWC network while guaranteeing the max-min fairness. In this respect, two DDPG-based beamforming solutions were proposed to optimize beam orientations and power allocation, with due considation of the quasi-stationarity of underwater nodes. The results showed that the proposed solutions achieve improved fairness and EE as compared to WMMSE-EE-MinRate, QT-EE-MinRate, and equal power allocation solutions.

6.2 Author's Publications

The present author's Ph.D. research contributed to the following published research articles. Journal publications and conference proceedings are denoted by "J" and "C," respectively.

J1: Romdhane, I., & Kaddoum, G. (2022). A reinforcement-learning-based beam adaptation for underwater optical wireless communications. IEEE Internet of Things Journal, 9(20), 20270-20281.

J2. Celik, A., Romdhane, I., Kaddoum, G., & Eltawil, A. M. (2022). A top-down survey on optical wireless communications for the internet of things. IEEE Communications Surveys & Tutorials, 25(1), 1-45.

J3. Romdhane, I., Rahman, Z., Kaddoum, G., Bansal, A., & Zafaruddin, S. M. (2023). Outage probability analysis for UOWC system over oceanic turbulence with pointing errors. IEEE Communications Letters.

J4. Rahman, Z., Romdhane, I., & Kaddoum, G. (2023). Performance of RSMA-Based UOWC systems over oceanic turbulence channel with pointing errors. IEEE Open Journal of the Communications Society.

C1. Romdhane, I., Rahman, Z., & Kaddoum, G. (2024, May). DDPG-Based Link Adaptation for Direct Air-to-Underwater Optical Communication Under Noisy CSI Conditions. In 2024 20th International Conference on the Design of Reliable Communication Networks (DRCN) (pp. 1-6). IEEE.

C2. Rahman, Z., Romdhane, I., Hassan, M. Z., & Kaddoum, G. (2024). Effective Capacity Analysis for Multi-Aperture UOWC over Oceanic Turbulence with Pointing Errors. In 2024 IEEE Middle East Conference on Communications and Networking (MECOM) (pp. 333-338).

J5. Romdhane, I., Rahman, Z., Mohamed, N. B., Hassan, M.Z., & Kaddoum, G. (2026). Dynamic Resource Optimization for a Joint Max-Min Fairness and Energy-Efficiency Problem in NOMA-Aided Underwater Optical Wireless Systems. IEEE Open Journal of the Communications Society.

6.3 Future Work

Based on the literature review and research outcomes of this Ph.D. thesis, in what follows, we present future research directions.

6.3.1 Resource Optimization in AUV-based Multi-Beam UOWC Networks

Nowadays, UOWC networks are constantly expanding geographically. In many remote sensing applications, sensors are wide spread into the ocean and can be remotely controlled by a multi-beam central node. Considering the short range of optical communications, such communication can be achieved using multi-relay links. However, this solution will introduce transmission delays, which can be critical for real-time applications. Alternatively, mobile nodes, also known as AUVs, can be a good alternative to efficiently collect such information. This highlights the importance of AUVs that collect information from sensors to transfer it to central nodes, and vice versa. Accordingly, it is necessary to consider the case of UOWC networks that include AUVs. On the central node level, the number of receiving AUVs and their locations are dynamic. Therefore, the communicating AUVs at each time slot are clustered according to their current location. Then, the central node decides on the beams orientation and power allocation among them. To achieve this task, in future work, we will use DRL techniques to perform dynamic decision at each time.

6.3.2 Joint Routing and Link adaptation for OIRS-based UOWC Networks

The problem of connectivity is fundamental in UOWC networks. Communication between different nodes in the system can undergo light occlusion or blockage due to exterior obstacles. Accordingly, OIRS can come handy in such cases to overcome these obstacles. By functioning as a relay between interconnecting nodes, it can also extend the link range. However, in order to ensure a good communication between nodes, a good beam adaptation is needed. In addition, in such cases, challenges like inter-node and inter-beam interference must be taken into consideration. To date, however, beam adaptation in OIRS-based UOWC networks has been poorly addressed in literature. Although the OIRS-based indirect link can significantly improve connectivity within UOWC networks, in some cases, it is better to opt for the direct link to avoid interference problems and additional delays. Hence, a decision must be made between the direct/indirect link according to the current nodes and network state. In this context, in future work, we will consider an OIRS-based UOWC network where each node in the network will decide at each transmission whether or not to choose the direct or indirect link and perform the link adaptation task according to the receiving node's location and the network state. In this prospective study, we will use multi-agent RL techniques to perform a dynamic and distributed decision at each time.

6.3.3 Adaptive Positioning, Beam Steering, and SWIPT for AUV-Based UOWC Networks

Usually, underwater sensors are fed with batteries that are difficult to change or recharge. This makes it challenging to keep these sensors functioning for longer periods of time. One solution previously proposed in the literature was the simultaneous wireless information and power transfer (SWIPT) technology. Since AUVs can be fed by stations on the water surface, this technique can charge the underwater sensors by the AUVs during their communication. Said differently, while the AUV is transferring information to underwater sensors, it simultaneously recharges them by dividing the emitted power between information transfer and feeding sensors with power. However, in order to guarantee an optimal communication with underwater sensors, the AUV needs to find its best positioning and light steering. In addition, an optimal power division between information and power transfer is required to ensure the needed power charge while maintaining an acceptable information data rate, considering the data size and the residual power in each sensor. Hence, before launching the communication, the AUV needs to dynamically decide on these variables. Considering the high dimensions of states and actions in this case, conventional DRL techniques become slow to converge, which may be critical in time-sensitive scenarios. Hence, in future work, we will opt to use quantum DRL to alleviate the algorithm complexity and fasten convergence.

BIBLIOGRAPHY

D. Zwillinger. (2014). *Table of integrals, series, and products*. Elsevier.

Abo Rahama, Y., Ismail, M. H. & Hassan, M. S. (2018). On the Sum of Independent Fox's H -Function Variates With Applications. *IEEE Trans. Veh. Technol.*, 67(8), 6752-6760. doi: 10.1109/TVT.2018.2827180.

Agarwal, A. & Krikidis, I. (2025). Fairness-Driven Optimization for NOMA-UWOC Systems With Energy Harvesting Requirements. *IEEE J. Ocean. Eng.*, 50(1), 403–418. doi: 10.1109/JOE.2024.3455560.

Ahmad, R., Soltani, M. D., Safari, M. & Srivastava, A. (2020). Load Balancing of Hybrid LiFi WiFi Networks Using Reinforcement learning. *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-6. doi: 10.1109/PIMRC48278.2020.9217382.

Alenezi, A. M. & Hamdi, K. A. (2020). Reinforcement Learning Approach for Hybrid WiFi-VLC Networks. *2020 IEEE 91st Veh. Technol. Conf. (VTC2020-Spring)*, pp. 1–5.

Alghamdi, R., Saeed, N., Dahrouj, H., Alouini, M.-S. & Al-Naffouri, T. Y. (2019). Towards Ultra-Reliable Low-Latency Underwater Optical Wireless Communications. *2019 IEEE 90th Veh. Techol. Conf. (VTC2019-Fall)*, pp. 1-6. doi: 10.1109/VTCFall.2019.8891506.

Aman, W., Al-Kuwari, S., Muzzammil, M., Rahman, M. M. U. & Kumar, A. (2023). Security of underwater and air–water wireless communication: State-of-the-art, challenges and outlook. *Ad Hoc Netw.*, 142, 103114.

Ata, Y. & Kiasaleh, K. (2023). Analysis of Optical Wireless Communication Links in Turbulent Underwater Channels with Wide Range of Water Parameters. *IEEE Transactions on Vehicular Technology*, 72(5), 6363–6374. doi: 10.1109/TVT.2022.3224512.

Ata, Y., Abumarshoud, H., Bariah, L., Muhaidat, S. & Imran, M. A. (2023). Intelligent Reflecting Surfaces for Underwater Visible Light Communications. *IEEE Photon. J.*, 15(1), 1-10. doi: 10.1109/JPHOT.2023.3235916.

Bansal, A., Rahman, Z. & Zafaruddin, S. M. (Early Access, May 2023). Diversity Analysis of Multi-Aperture UWOC System over EGG Channel with Pointing Errors. *IEEE Commun. Lett.*, 1-1. doi: 10.1109/LCOMM.2023.3281613.

Baykal, Y., Ata, Y. & Gökçe, M. C. (2022). Underwater turbulence, its effects on optical wireless communication and imaging: A review. *Optics & Laser Technology*, 156, 108624.

Busoniu, L., Babuska, R. & De Schutter, B. (2008). A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Trans. Syst., Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2), 156-172.

Carver, C. J., Tian, Z., Zhang, H., Odame, K. M., Quattrini Li, A. & Zhou, X. (2021). Amphilight: Direct air-water communication with laser light. *GetMobile: Mobile Computing Commun.*, 24(3), 26–29.

Celik, A., Saeed, N., Shihada, B., Al-Naffouri, T. Y. & Alouini, M.-S. (2019). End-to-end performance analysis of underwater optical wireless relaying and routing techniques under location uncertainty. *IEEE Trans. Wireless Commun.*, 19(2), 1167–1181.

Celik, A., Romdhane, I., Kaddoum, G. & Eltawil, A. M. (2022). A Top-Down Survey on Optical Wireless Communications for the Internet of Things. *IEEE Commun. Surveys Tuts.*, 1-1. doi: 10.1109/COMST.2022.3220504.

Che, X., Wells, I., Dickers, G., Kear, P. & Gong, X. (2010). Re-evaluation of RF electromagnetic communication in underwater sensor networks. *IEEE Commun. Mag.*, 48(12), 143–151.

Chen, D., Fan, K., Wang, J., Lu, H., Jin, J. & Liu, C. (2022). Two-dimensional power allocation scheme for NOMA-based underwater visible light communication systems. *Applied Optics*, 62(1), 211–216.

Chitre, M., Shahabudeen, S. & Stojanovic, M. (2008). Underwater Acoustic Communications and Networking: Recent advances and future challenges. *Marine Technol. Soc. J.*, 42(1), 103–116.

Cox, C. & Munk, W. (1954). Measurement of the roughness of the sea surface from photographs of the sun's glitter. *Josa*, 44(11), 838–850.

Dani, M. N., So, D. K. C., Tang, J. & Ding, Z. (2022). Resource Allocation for Layered Multicast Video Streaming in NOMA Systems. *IEEE Trans. Veh. Technol.*, 71(11), 11379-11394. doi: 10.1109/TVT.2022.3193122.

Derakhshandeh, A., Pachnicke, S. & Hoeher, P. A. (2023). Underwater Wireless Laser-Based Communications Using Optical Phased Array Antennas. *IEEE Photonics Journal*.

Dinkelbach, W. (1967). On nonlinear fractional programming. *Manage. Sci.*, 13(7), 492–498.

Elamassie, M., Miramirkhani, F. & Uysal, M. (2018). Performance characterization of underwater visible light communication. *IEEE Trans. Commun.*, 67(1), 543–552.

Elamassie, M., Bariah, L., Uysal, M., Muhaidat, S. & Sofotasios, P. C. (2021). Capacity analysis of NOMA-enabled underwater VLC networks. *IEEE access*, 9, 153305–153315.

Elsayed, M., Samir, A., El-Banna, A. A. A., Khan, W. U., Chatzinotas, S. & ElHalawany, B. M. (2022). Mixed RIS-Relay NOMA-Based RF-UOWC Systems. *2022 IEEE 95th Veh. Technol. Conf.:(VTC2022-Spring)*, pp. 1–6.

Erol-Kantarci, M., Mouftah, H. T. & Oktug, S. (2011). A Survey of Architectures and Localization Techniques for Underwater Acoustic Sensor Networks. *IEEE Commun. Surveys Tuts.*, 13(3), 487-502. doi: 10.1109/SURV.2011.020211.00035.

Fan, J., Wang, Z., Xie, Y. & Yang, Z. (2020). A theoretical analysis of deep Q-learning. *Learning for dynamics and control*, pp. 486–489.

Farid, A. A. & Hranilovic, S. (2007). Outage Capacity Optimization for Free-Space Optical Links With Pointing Errors. *IEEE J. Lightw. Technol.*, 25(7), 1702-1710.

Frater, M. R., Ryan, M. J. & Dunbar, R. M. (2006). Electromagnetic communications within swarms of autonomous underwater vehicles. *Proc. of the 1st ACM Int. Workshop on Underwater Networks*, pp. 64–70.

Fu, Q. & Song, A. (2018). Adaptive modulation for underwater acoustic communications based on reinforcement learning. *OCEANS 2018 MTS/IEEE Charleston*, pp. 1–8.

Gao, J., Wang, J., Gu, J. & Shi, W. (2024). Q-learning-based routing optimization algorithm for underwater sensor networks. *IEEE Internet of Things Journal*, 11(22), 36350–36357.

Ge, J., Liang, Y.-C., Joung, J. & Sun, S. (2020). Deep reinforcement learning for distributed dynamic MISO downlink-beamforming coordination. *IEEE Trans. Commun.*, 68(10), 6070–6085.

Geng, X. & Zhang, B. (2023). Deep Q-Network-Based Intelligent Routing Protocol for Underwater Acoustic Sensor Network. *IEEE Sensors J.*, 23(4), 3936-3943. doi: 10.1109/JSEN.2023.3234112.

Geng, X. & Zheng, Y. R. (2023). Exploiting Propagation Delay in Underwater Acoustic Communication Networks via Deep Reinforcement Learning. *IEEE Trans. Neural Netw. Learning Syst.*, 34(12), 10626-10637. doi: 10.1109/TNNLS.2022.3170050.

Ghazy, A. S., Kaddoum, G., Talhi, C., Iqbal, N. & Muqaibel, A. H. (2025). Reinforcement learning-assisted secure reliable underwater wireless acoustic communications. *IEEE Access*.

Gou, Y., Zhang, T., Yang, T., Liu, J., Song, S. & Cui, J.-H. (2023). A Deep MARL-Based Power-Management Strategy for Improving the Fair Reuse of UWSNs. *IEEE Internet Things J.*, 10(7), 6507-6522. doi: 10.1109/JIOT.2022.3226953.

Gou, Y., Zhang, T., Liu, J., Yang, T., Song, S. & Cui, J.-H. (2024). Achieving Fair-Effective Communications and Robustness in Underwater Acoustic Sensor Networks: A Semi-Cooperative Approach. *IEEE Trans. Mobile Comput.*, 23(5), 5722–5739. doi: 10.1109/TMC.2023.3310476.

Guo, Y., Xie, C., Huang, P., Li, J., Zhang, L., Huang, D. & Zeng, G. (2018). Channel-parameter estimation for satellite-to-submarine continuous-variable quantum key distribution. *Phys. Rev. A*, 97(5), 052326.

Han, G., Gong, A., Wang, H., Martínez-García, M. & Peng, Y. (2021). Multi-AUV Collaborative Data Collection Algorithm Based on Q-Learning in Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.*, 70(9), 9294-9305. doi: 10.1109/TVT.2021.3097084.

Han, L., Liu, X., Wang, Y. & Li, B. (2022). Joint Impact of Channel Estimation Errors and Pointing Errors on FSO Communication Systems Over $\mathcal{F}$ Turbulence Channel. *J. Light. Technol.*, 40(14), 4555–4561.

Hu, T. & Fei, Y. (2010). QELAR: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks. *IEEE Trans. Mobile Comput.*, 9(6), 796–809.

Huang, A., Tao, L., Wang, C. & Zhang, L. (2018). Error performance of underwater wireless optical communications with spatial diversity under turbulence channels. *Applied optics*, 57(26), 7600–7608.

Hunt, K. P., Niemeier, J. J. & Kruger, A. (2010). RF communications in underwater wireless sensor networks. *2010 IEEE Int. Conf. on Electro/Inf. Technol.*, pp. 1-6. doi: 10.1109/EIT.2010.5612087.

Ijeh, I. C., Khalighi, M. A., Elamassie, M., Hranilovic, S. & Uysal, M. (2022). Outage probability analysis of a vertical underwater wireless optical link subject to oceanic turbulence and pointing errors. *Journal of Optical Communications and Networking*, 14(6), 439–453.

J. Zheng *et al.* (2019). Sum of Squared Fluctuating Two-Ray Random Variables With Wireless Applications. *IEEE Trans. Veh. Technol.*, 68(8), 8173-8177. doi: 10.1109/TVT.2019.2920533.

Jamali, M. V., Chizari, A. & Salehi, J. A. (2017a). Performance analysis of multi-hop underwater wireless optical communication systems. *IEEE Photon. Technol. Lett.*, 29(5), 462–465.

Jamali, M. V., Salehi, J. A. & Akhoundi, F. (2017b). Performance Studies of Underwater Wireless Optical Communication Systems With Spatial Diversity: MIMO Scheme. *IEEE Trans. Commun.*, 65(3), 1176-1192. doi: 10.1109/TCOMM.2016.2642943.

Jamali, M. V., Mirani, A., Parsay, A., Abolhassani, B., Nabavi, P., Chizari, A., Khorramshahi, P., Abdollahramezani, S. & Salehi, J. A. (2018). Statistical Studies of Fading in Underwater Wireless Optical Channels in the Presence of Air Bubble, Temperature, and Salinity Random Variations. *IEEE Trans. Commun.*, 66(10), 4706-4723. doi: 10.1109/TCOMM.2018.2842212.

Jiang, H., Qiu, H. et al. (2022). Ergodic Capacity and Error Performance of Spatial Diversity UWOC Systems over Generalized Gamma Turbulence Channels. *Optics Communications*, 505, 127476.

Jiang, Y., Varanasi, M. K. & Li, J. (2011). Performance Analysis of ZF and MMSE Equalizers for MIMO Systems: An In-Depth Study of the High SNR Regime. *IEEE Trans. Inf. Theory*, 57(4), 2008-2026. doi: 10.1109/TIT.2011.2112070.

Jiang, Y., Li, Y., Chen, X., Li, S. & Wang, T. (2023a). A Strategy for Reducing Interference in Underwater Optical Multi-beam Systems Based on NOMA. *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 11, 722-727. doi: 10.1109/ITAIC58329.2023.10408966.

Jiang, Y., Li, Y., Zhang, X., Ju, M. & Huang, P. (2023b). Outage performance analysis for relay-assisted UWOC systems over GGD weak turbulence with nonzero boresight pointing errors. *Physical Commun.*, 58, 102017.

Jiao, R., Dai, L., Wang, W., Lyu, F., Cheng, N. & Shen, X. (2022a). Max-Min Fairness for Beamspace MIMO-NOMA: From Single-Beam to Multi-Beam. *IEEE Trans. Wireless Commun.*, 21(2), 739-752. doi: 10.1109/TWC.2021.3098900.

Jiao, R., Dai, L., Wang, W., Lyu, F., Cheng, N. & Shen, X. (2022b). Max-Min Fairness for Beamspace MIMO-NOMA: From Single-Beam to Multi-Beam. *IEEE Trans. Wireless Commun.*, 21(2), 739-752. doi: 10.1109/TWC.2021.3098900.

Kahn, J. & Barry, J. (1997). Wireless infrared communications. *Proc. IEEE*, 85(2), 265-298. doi: 10.1109/5.554222.

Kaushal, H. & Kaddoum, G. (2016). Underwater optical wireless communication. *IEEE Access*, 4, 1518–1547.

Kim, I. I., McArthur, B. & Korevaar, E. J. (2001). Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications. *Optical wireless commun. III*, 4214, 26–37.

Komine, T. & Nakagawa, M. (2004). Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans. Consum. Electron.*, 50(1), 100–107.

Kong, J., Wu, Z.-Y., Ismail, M., Serpedin, E. & Qaraqe, K. A. (2019). Q-Learning Based Two-Timescale Power Allocation for Multi-Homing Hybrid RF/VLC Networks. *IEEE Wireless Commun. Lett.*, 9(4), 443–447.

Kulhandjian, H. (2014). Inside out: Underwater communications. *J. Ocean Technol.*, 9(2), 104–105.

Li, G., Shang, T., Kong, W. & Tang, T. (2024a). Energy Efficiency Optimization in a Parallel Relay-Assisted Underwater Wireless Optical Communication System with Simultaneous Lightwave Information and Power Transfer. *Applied Optics*, 63(4), 999–1010.

Li, J., Yang, B., Ye, D., Wang, L., Fu, K., Piao, J. & Wang, Y. (2020a). A Real-Time, Full-Duplex System for Underwater Wireless Optical Communication: Hardware Structure and Optical Link Model. *IEEE Access*, 8, 109372-109387. doi: 10.1109/ACCESS.2020.3001213.

Li, M., Luo, H., Li, X., Ruby, R., Tao, H. & Wu, K. (2023a). A Cooperative Multi-AUV Mobile Scheme with Optical Communication via DDPG Approach. *2023 IEEE 29th Int. Conf. Parallel and Distributed Syst. (ICPADS)*, pp. 1476-1483. doi: 10.1109/ICPADS60453.2023.00209.

Li, M., Luo, H., Tao, H., Li, X., Dong, P. & Wu, K. (2024b). Collaborative multi-AUV optical communication via deep reinforcement learning. *IEEE Sensors Journal*, 25(1), 1627–1640.

Li, S., Yang, L., da Costa, D. B. & Yu, S. (2021). Performance Analysis of UAV-Based Mixed RF-UWOC Transmission Systems. *IEEE Trans. Commun.*, 69(8), 5559-5572. doi: 10.1109/TCOMM.2021.3076790.

Li, X., Gui, L., Xia, Y., Yang, X., Li, Y., Li, H. & Lang, L. (2023b). Experimental demonstration of a real-time multi-user uplink UWOC system based on SIC-free NOMA. *Opt. Express*, 31(19), 30146–30159.

Li, X., Hu, X., Li, W. & Hu, H. (2019). A multi-agent reinforcement learning routing protocol for underwater optical sensor networks. *ICC 2019-2019 IEEE Int. Conf. Commun. (ICC)*, pp. 1–7.

Li, X., Hu, X., Zhang, R. & Yang, L. (2020b). Routing protocol design for underwater optical wireless sensor networks: A multiagent reinforcement learning approach. *IEEE Internet Things J.*, 7(10), 9805–9818.

Li, Y., Jiang, Y. & Wang, J. (2022). Research on Power Allocation of MIMO-NOMA in Multi-Beam Space Division Multiple Access for Underwater Optical Communication. *2022 14th Int. Conf. Signal Process. Syst. (ICSPS)*, pp. 645-650. doi: 10.1109/ICSPS58776.2022.00119.

Li, Y., Mohsan, S. A. H., Chen, X., Tehseen, R., Li, S. & Wang, J. (2023c). Research on power allocation in multiple-beam space division access based on NOMA for underwater optical communication. *Sensors*, 23(3), 1746.

Liang, Y., Yin, H., Jing, L., Ji, X. & Wang, J. (2022). Performance analysis of relay-aided NOMA optical wireless communication system in underwater turbulence environment. *Remote Sensing*, 14(16), 3894.

Liang, Y., Yin, H., Jing, L., Ji, X. & Wang, J. (2023). Solution for Self-Interference of NOMA-Based Wireless Optical Communication System in Underwater Turbulence Environment. *IEEE Access*, 11, 30223-30236. doi: 10.1109/ACCESS.2023.3260221.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Lin, C., Han, G., Zhang, T., Shah, S. B. H. & Peng, Y. (2023). Smart Underwater Pollution Detection Based on Graph-Based Multi-Agent Reinforcement Learning Towards AUV-Based Network ITS. *IEEE Trans. Intel. Transportation Syst.*, 24(7), 7494-7505. doi: 10.1109/TITS.2022.3162850.

Liu, J., Mao, T., He, D., Yang, Y., Gao, Z., Zheng, D. & Zhang, J. (2024). Reinforcement-learning-enabled beam alignment for water-air direct optical wireless communications. *2024 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 138–143.

Luo, Z.-Q. & Zhang, S. (2008). Dynamic Spectrum Management: Complexity and Duality. *IEEE J. Sel. Topics. Signal Process.*, 2(1), 57-73. doi: 10.1109/JSTSP.2007.914876.

Ma, S., Li, M., Deng, T., Wang, J., Ruby, R. & Luo, H. (2023). UAV-Assisted Maritime Data Collection via Optical Communications using Deep Reinforcement Learning. *2023 9th Int. Conf. Big Data Computing Commun. (BigCom)*, pp. 164-171. doi: 10.1109/BIGCOM61073.2023.00030.

Magbool, A., Kumar, V. & Flanagan, M. F. (2023). On Energy Efficiency and Fairness Maximization in RIS-Assisted MU-MISO mmWave Communications. *ICC 2023 - IEEE Int. Conf. Commun.*, pp. 5364-5369. doi: 10.1109/ICC45041.2023.10279200.

Mathai, A., Saxena, R. & Haubold, H. (2009). *The H-Function: Theory and Applications*. Springer New York.

McCluney, W. R. (2014). *Introduction to radiometry and photometry*. Artech House.

Miramirkhani, F. & Uysal, M. (2017). Visible light communication channel modeling for underwater environments with blocking and shadowing. *IEEE Access*, 6, 1082–1090.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G. et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540), 529–533.

Mobley, C. D. (1994). Light and water: radiative transfer in natural waters. *(No Title)*.

Naik, R. P. & Chung, W.-Y. (2022). Evaluation of Reconfigurable Intelligent Surface-Assisted Underwater Wireless Optical Communication System. *IEEE J. Lightw. Technol.*, 40(13), 4257-4267. doi: 10.1109/JLT.2022.3162627.

O. Korotkova, N. F. & Shchepakina, E. (2012). Light scintillation in oceanic turbulence. *Waves in Random and Complex Media*, 22(2), 260-266. doi: 10.1080/17455030.2012.656731.

Omeke, K. G., Mollel, M., Shah, S. T., Zhang, L., Abbasi, Q. H. & Imran, M. A. (2024). Toward a Sustainable Internet of Underwater Things Based on AUVs, SWIPT, and Reinforcement Learning. *IEEE Internet Things J.*, 11(5), 7640-7651. doi: 10.1109/JIOT.2023.3319250.

Otnes, R., Jenserud, T., Voldhaug, J. E. & Soldberg, C. (2009). A roadmap to ubiquitous underwater acoustic communications and networking. *Proc. Underwater Acoustic Measurement: Technol. and Results*, 1–8.

Palitharathna, K. W. S., Suraweera, H. A., Godaliyadda, R. I., Herath, V. R. & Thompson, J. S. (2021). Average Rate Analysis of Cooperative NOMA Aided Underwater Optical Wireless Systems. *IEEE Open J. Commun. Society*, 2, 2292-2310. doi: 10.1109/OJCOMS.2021.3116146.

Palitharathna, K. W. S., Suraweera, H. A., Godaliyadda, R. I., Herath, V. R. & Ding, Z. (2022). Lightwave Power Transfer in Full-Duplex NOMA Underwater Optical Wireless Communication Systems. *IEEE Commun. Lett.*, 26(3), 622-626. doi: 10.1109/LCOMM.2021.3140117.

Partan, J., Kurose, J. & Levine, B. N. (2007). A survey of practical issues in underwater networks. *ACM SIGMOBILE Mobile Computing and Commun. Review*, 11(4), 23–33.

Pompili, D. & Akyildiz, I. F. (2009). Overview of networking protocols for underwater wireless communications. *IEEE Commun. Mag.*, 47(1), 97–102.

Qasem, Z. A. H., Ali, A., Deng, B., Li, Q. & Fu, H. Y. (2023). Unipolar X-Transform OFDM With Index Modulation for Underwater Optical Wireless Communications. *IEEE Photon. Technol. Lett.*, 35(11), 581-584. doi: 10.1109/LPT.2023.3264561.

Qian, Y., Chen, C., Du, P. & Liu, M. (2023). Hybrid space division multiple access and quasi-orthogonal multiple access for multi-user underwater visible light communication. *IEEE Photon. J.*, 15(4), 1–7.

Rahman, Z., Tailor, N. V., Zafaruddin, S. M. & Chaubey, V. K. (2022). Unified Performance Assessment of Optical Wireless Communication Over Multi-Layer Underwater Channels. *IEEE Photon. J.*, 14(5), 1-14. doi: 10.1109/JPHOT.2022.3201081.

Rahman, Z., Zafaruddin, S. M. & Chaubey, V. K. (2023). Direct Air-to-Underwater Optical Wireless Communication: Statistical Characterization and Outage Performance. *IEEE Trans. Veh. Technol.*, 72(2), 2655-2660. doi: 10.1109/TVT.2022.3211186.

Rahman, Z., Hassan, M. Z. & Kaddoum, G. (2024a). Ergodic Capacity Optimization for RSMA-Based UOWC Systems Over EGG Turbulence Channel. *IEEE Commun. Lett.*, 28(3), 587-591. doi: 10.1109/LCOMM.2024.3356413.

Rahman, Z., Romdhane, I. & Kaddoum, G. (2024b). Performance of RSMA-Based UOWC Systems Over Oceanic Turbulence Channel With Pointing Errors. *IEEE Open J. Commun. Society*, 5, 127-138. doi: 10.1109/OJCOMS.2023.3339790.

Rahmani, A. M., Tanveer, J., Mutairi, A., Altulyan, M., Gemeay, E., Alam, M., Yousefpoor, M. S., Yousefpoor, E. & Hosseinzadeh, M. (2025). A Q-learning-based hierarchical routing protocol in underwater acoustic sensor networks. *Computers and Electrical Engineering*, 123, 110211.

Rahmani, M., Bashar, M., Dehghani, M. J., Akbari, A., Xiao, P., Tafazolli, R. & Debbah, M. (2023). Deep Reinforcement Learning-Based Sum Rate Fairness Trade-Off for Cell-Free mMIMO. *IEEE Trans. Veh. Technol.*, 72(5), 6039-6055. doi: 10.1109/TVT.2022.3230041.

Riksfjord, H., Haug, O. T. & Hovem, J. M. (2009). Underwater acoustic networks-survey on communication challenges with transmission simulations. *2009 3rd Int. Conf. on Sensor Technol. and Appl.*, pp. 300–305.

Romdhane, I. & Kaddoum, G. (2022). A Reinforcement Learning based Beam Adaptation for Underwater Optical Wireless Communications. *IEEE Internet Things J.*, 1-1. doi: 10.1109/JIOT.2022.3173211.

Saeed, N., Celik, A., Al-Naffouri, T. Y. & Alouini, M.-S. (2019a). Underwater optical wireless communications, networking, and localization: A survey. *Ad Hoc Netw.*, 94, 101935.

Saeed, N., Celik, A., Alouini, M.-S. & Al-Naffouri, T. Y. (2019b). Performance Analysis of Connectivity and Localization in Multi-Hop Underwater Optical Wireless Sensor Networks. *IEEE Trans. Mobile Comput.*, 18(11), 2604-2615. doi: 10.1109/TMC.2018.2878672.

Salam, R., Srivastava, A., Bohara, V. A. & Ashok, A. (2023). An optical intelligent reflecting surface-assisted underwater wireless communication system. *IEEE Open Journal of the Communications Society*, 4, 1774–1786.

Salama, W. M., Aly, M. H. & Amer, E. S. (2023). Underwater optical wireless communication system: Deep learning CNN with NOMA-based performance analysis. *Opt. Quantum Electron.*, 55(5), 436.

Salcedo-Serrano, P., Boluda-Ruiz, R., Garrido-Balsells, J. M. & García-Zambrana, A. (2022). Ergodic capacity analysis of underwater FSO systems over scattering-induced fading channels in the presence of Weibull oceanic turbulence. *ICC 2022 - IEEE Int. Conf. on Commun.*, pp. 3814-3819. doi: 10.1109/ICC45855.2022.9838728.

Seah, W. K., Lee, C.-H., Lin, Y.-D. & Lai, Y.-C. (2022). Combined Communication and Computing Resource Scheduling in Sliced 5G Multi-Access Edge Computing Systems. *IEEE Trans. Veh. Technol.*, 71(3), 3144-3154. doi: 10.1109/TVT.2021.3139026.

Shen, K. & Yu, W. (2018). Fractional Programming for Communication Systems—Part I: Power Control and Beamforming. *IEEE Trans. Signal Process.*, 66(10), 2616–2630. doi: 10.1109/TSP.2018.2812733.

Shen, Z., Yin, H., Jing, L., Liang, Y. & Wang, J. (2022). A Cooperative Routing Protocol Based on Q-Learning for Underwater Optical-Acoustic Hybrid Wireless Sensor Networks. *IEEE Sensors J.*, 22(1), 1041-1050. doi: 10.1109/JSEN.2021.3128594.

Shen, Z., Yin, H., Jing, L., Ji, X., Liang, Y. & Wang, J. (2023). A Power Control-Aided Q-Learning-Based Routing Protocol for Optical-Acoustic Hybrid Underwater Sensor Networks. *IEEE Trans. Green Commun. Netw.*, 7(4), 2117-2129. doi: 10.1109/TGCN.2023.3270867.

Shetty, C. S. S., Naik, R. P., Acharya, U. S. & Chung, W.-Y. (2023). Performance Analysis of MIMO-EGC System for the Underwater Vertical Wireless Optical Communication Link. *IEEE Access*, 11, 99253–99267. doi: 10.1109/ACCESS.2023.3313931.

SHI, H., Prasad, R. V., Onur, E. & Niemegeers, I. (2014). Fairness in Wireless Networks: Issues, Measures and Challenges. *IEEE Commun. Surveys Tuts.*, 16(1), 5-24. doi: 10.1109/SURV.2013.050113.00015.

Shi, J., Zhang, S. & Yang, C. (2012). High frequency RF based non-contact underwater communication. *Proc. IEEE OCEANS, Yeosu, South Korea*, pp. 1-6. doi: 10.1109/OCEANS-Yeosu.2012.6263403.

Shi, J., Ma, C., Tian, X., Guo, H. & Ao, J. (2024). Optimization and modeling of optical emission spatial coverage from underwater multi-faceted optical base stations. *Photonics*, 12(1), 4.

Shi, Q., Razaviyayn, M., Luo, Z.-Q. & He, C. (2011). An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans. Signal Process.*, 59(9), 4331–4340.

Shi, Y., Yang, Q., Huang, X., Li, D. & Huang, X. (2023). An SDN-Enabled Framework for a Load-Balanced and QoS-Aware Internet of Underwater Things. *IEEE Internet Things J.*, 10(9), 7824-7834. doi: 10.1109/JIOT.2022.3231329.

Shin, H., Kim, S. M. & Song, Y. (2023). Learning-Aided Joint Beam Divergence Angle and Power Optimization for Seamless and Energy-Efficient Underwater Optical Communication. *IEEE Internet of Things J.*, 10(24), 22726-22739. doi: 10.1109/JIOT.2023.3304655.

Shin, H., Baek, S. & Song, Y. (2024). Multidimensional Beam Optimization in Underwater Optical Wireless Communication Based on Deep Reinforcement Learning. *IEEE Internet of Things J.*, 11(17), 28623-28634. doi: 10.1109/JIOT.2024.3404476.

Shirley, J. W. (1951). An early experimental determination of Snell's law. *Am. J. Phys.*, 19(9), 507–508.

Sticklus, J., Hieronymi, M. & Hoeher, P. A. (2018). Effects and constraints of optical filtering on ambient light suppression in LED-based underwater communications. *Sensors*, 18(11), 3710.

Sun, X., Kang, C. H., Kong, M., Alkhazragi, O., Guo, Y., Ouhssain, M., Weng, Y., Jones, B. H., Ng, T. K. & Ooi, B. S. (2020). A review on practical considerations and solutions in underwater wireless optical communication. *Journal of Lightwave Technology*, 38(2), 421–431.

Sutton, R. S. & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Thai, T.-H. H., Nguyen, C. T., Le, H. D. & Pham, A. T. (2025). Reliable downlink transmission for underwater optical wireless systems with energy harvesting and HARQ. *Applied Optics*, 64(26), 7768–7778.

Wang, C., Shen, X., Wang, H., Xie, W., Zhang, H. & Mei, H. (2024a). Multi-Agent Reinforcement Learning-Based Routing Protocol for Underwater Wireless Sensor Networks With Value of Information. *IEEE Sensors J.*, 24(5), 7042-7054. doi: 10.1109/JSEN.2023.3345947.

Wang, C., Wang, Z., Sun, W. & Fuhrmann, D. R. (2017). Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels. *IEEE Access*, 6, 2541–2558.

Wang, H., Li, Y. & Qian, J. (2019a). Self-Adaptive Resource Allocation in Underwater Acoustic Interference Channel: A Reinforcement Learning Approach. *IEEE Internet Things J.*, 7(4), 2816–2827.

Wang, J., Wang, J., Li, Y., Jiang, Y. & Chen, X. (2022). Power Allocation Algorithm Considering QoS and Max-Min Fairness in Single-Beam UVLC System. *2022 14th Int. Conf. Signal Process. Syst. (ICSPS)*, pp. 628-633. doi: 10.1109/ICSPS58776.2022.00116.

Wang, J., Luo, H., Ruby, R., Liu, J., Wang, S. & Wu, K. (2024b). Enabling Reliable Water-Air Direct Optical Wireless Communication for Uncrewed Vehicular Networks: A Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.*, 1-16. doi: 10.1109/TVT.2024.3376521.

Wang, R., Yadav, A., Makled, E. A., Dobre, O. A., Zhao, R. & Varshney, P. K. (2019b). Optimal power allocation for full-duplex underwater relay networks with energy harvesting: A reinforcement learning approach. *IEEE Wireless Commun. Lett.*, 9(2), 223–227.

Wang, W., Li, Y., Hao, B., Lv, Y. & Ju, M. (2025). Error performance and power allocation for MIMO-UWOC systems in composite oceanic fading scenario. *Opt. Commun.*, 132301.

Weng, Y., Matsuda, T., Sekimori, Y., Pajarinen, J., Peters, J. & Maki, T. (2022a). Pointing error control of underwater wireless optical communication on mobile platform. *IEEE Photonics Technology Letters*, 34(13), 699–702.

Weng, Y., Pajarinen, J., Akrour, R., Matsuda, T., Peters, J. & Maki, T. (2022b). Reinforcement Learning Based Underwater Wireless Optical Communication Alignment for Autonomous Underwater Vehicles. *IEEE J. Ocean. Eng*, 47(4), 1231-1245. doi: 10.1109/JOE.2022.3165805.

Wolfram function. https://functions.wolfram.com/Hypergeometric-Functions/MeijerG. *Accessed Feb. 10, 2022*.

Xu, Q., Li, X., Ji, H. & Du, X. (2014). Energy-Efficient Resource Allocation for Heterogeneous Services in OFDMA Downlink Networks: Systematic Perspective. *IEEE Trans. Veh. Technol.*, 63(5), 2071-2082. doi: 10.1109/TVT.2014.2312288.

Xu, X., Hu, D. & Lu, X. (2007). Kernel-based least squares policy iteration for reinforcement learning. *IEEE Trans. Neural Netw.*, 18(4), 973–992.

Yang, H., Du, P., Zhong, W.-D., Chen, C., Alphones, A. & Zhang, S. (2019). Reinforcement learning-based intelligent resource allocation for integrated VLCP systems. *IEEE Wireless Commun. Lett.*, 8(4), 1204–1207.

Yang, L., Zhu, Q., Li, S., Ansari, I. S. & Yu, S. (2021). On the Performance of Mixed FSO-UWOC Dual-Hop Transmission Systems. *IEEE Wirel. Commun. Lett.*, 10(9), 2041-2045. doi: 10.1109/LWC.2021.3091429.

Yue, P., Wang, X., Xu, D. & Xu, S. (2025). Non-line-of-sight scattering channel modeling of MIMO links for underwater wireless optical communication. *Opt. Commun.*, 578, 131468.

Zappone, A. & Jorswieck, E. (2015). Energy Efficiency in Wireless Networks via Fractional Programming Theory. *Found. Trends Commun. Inf. Theory*, 11(3–4), 185–396. doi: 10.1561/0100000088.

Zedini, E., Oubei, H. M., Kammoun, A., Hamdi, M., Ooi, B. S. & Alouini, M.-S. (2019). Unified Statistical Channel Model for Turbulence-Induced Fading in Underwater Wireless Optical Communication Systems. *IEEE Trans. Commun.*, 67(4), 2893-2907. doi: 10.1109/TCOMM.2019.2891542.

Zedini, E., Kammoun, A., Soury, H., Hamdi, M. & Alouini, M.-S. (2020). Performance Analysis of Dual-Hop Underwater Wireless Optical Communication Systems Over Mixture Exponential-Generalized Gamma Turbulence Channels. *IEEE Trans. Commun.*, 68(9), 5718-5731. doi: 10.1109/TCOMM.2020.3006146.

Zeng, Z., Fu, S., Zhang, H., Dong, Y. & Cheng, J. (2017). A Survey of Underwater Optical Wireless Communications. *IEEE Commun. Surveys Tuts.*, 19(1), 204-238.

Zhang, S., Chen, H. & Xie, L. (2025). Service-aware Q-learning-based routing protocol in the underwater acoustic sensor network. *Computer Networks*, 257, 110986.

Zhang, Z., Zhang, L. & Chen, Z. (2021). Multi-Agent Reinforcement Learning Based Channel Access Scheme for Underwater Optical Wireless Communication Networks. *2021 15th Int. Symp. on Med. Inf. and Commun. Technology (ISMICT)*, pp. 65-69. doi: 10.1109/ISMICT51748.2021.9434918.