Reinforcement Learning-based Adaptive Transmission for the Underwater Full-duplex Relay Network with Energy Harvesting

by

© **Ranning Wang**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Computer Engineering.

Faculty of Engineering and Applied Science

Memorial University

January 2020

St. John's, Newfoundland and Labrador, Canada

Abstract

The acoustic wave is the only known effective method for long-haul underwater wireless communication, as compared to radio-frequency waves and light waves. The demands for oceanic environment monitoring, disaster surveillance, and business applications have propelled the growth of the underwater acoustic communication market. However, underwater acoustic communication is still in its infancy due to the challenge characteristic–narrow effective bandwidth. To address this challenge, underwater cooperative communication, which introduces relay nodes to forward messages from the source node to the destination node, can increase the effective bandwidth.

The nature of long-term operational communication networks is dynamic in time scale. For example, energy arrivals in energy harvesting communication are stochastic and the channel conditions are time-varying in wireless communication. In this thesis, we focus on system optimization for the long-term operational communication network. To this end, the optimization problem is formulated as maximizing or minimizing the accumulated utility function from the current to a future time instant. Given that the causal information of the system is available, this type of problem is known as stochastic optimization problem in which some parameters are random variables, and thus, traditional optimization tools cannot directly be applied to solve the problem. Instead, the solution is provided by the reinforcement learning technique that describes how an agent interacts with the environment over time to maximize

the accumulative reward.

In this thesis, the long-term operational underwater relay network is investigated, which consists of one sensor node, one relay node, and one destination node. The relay node operates in full-duplex mode, and can transmit and receive signals at the same frequency and time. Also, the relay node relies on the harvested energy from the ambient environment, whereas the source and the destination nodes have fixed power supplies. We evaluate the network performance with respect to the end-to-end spectral efficiency and average energy efficiency and aim to improve these performance metrics in the long-term. Due to the stochastic characteristic of harvested energy and channel state information, we develop adaptive transmission policies for the considered system to optimize system performance. Considering that the practical condition in which the causal knowledge of the system is known, the problem is then formulated as an online sequential decision-making problem and the reinforcement learning technique is used to obtain the transmission policies. Two major benefits of the reinforcement learning framework are: 1) it obtains an optimal solution, and 2) it does not require the knowledge of future information. On the other hand, one can apply the conventional optimization approach; however, this focuses on maximizing only the current reward, not the future reward, and hence, is not optimal. Simulation results show that the proposed transmission policies improve the system performance when compared with the benchmark policy.

To my parents, Dongmei Liu and Jiancai Wang,

and to my grandparents, Yongqin Sang and Cheng Liu.

Acknowledgements

First and foremost, I express my sincere thanks and deepest gratitude to my supervisor Professor Octavia A. Dobre–a leading scientist in wireless communication academic community–for giving me an invaluable opportunity to study in her lab and providing continuous guidance, support, and encouragement throughout my master's study. She taught me not only how to do research, but also helped me to be a better man. This memorable learning experience will continue to benefit my future career, and undoubtedly, she is the best example for my future life and work.

Also, my thanks go to Dr. Animesh Yadav, who was previously a postdoctoral fellow in our lab, for the guidance in my research. He is very knowledgeable in the wireless communication field and has been always ready to help me. In addition, I would like to thank Esraa Makled who has provided constructive comments for my research.

I am also grateful to have my lab mates-Xiang Lin, Sunish Kumar, Ming Zeng, Quang Le, Svlvester Aboagye, Ahmad Al-habob, Dr. Phong Nguyen, Dr. Ahmed Ibrahim, Prof. Ruiqin Zhao, Muhammad Raza, Ashraf Fata, Ibrahim Alnahhal, Ali Esswie, Dr. Yahia Eldemerdash, Dr. Abdelkerim Amari, Dr. Oluyemi Omomukuyo, Dr. Mostafa Mohammadkarimi, Dr. Yi Zhang, Dr. Shu Zhang, etc.-for sharing their joy of everyday life and work.

As an experienced amateur radio enthusiast, I am honored and lucky to live and study in St. John's, NL, Canada, where Guglielmo Marconi, a pioneer of wireless communication, successfully received the first trans-Atlantic wireless signals and opened the historical door for wireless communication in December 1901. Also, from 2005 to 2013, I have received numerous awards from the Chinese youngster amateur radio contests regionally and nationally, and from worldwide amateur radio contests. It has motivated me to continually study wireless communication thereafter.

Last but not least, I would like to thank my parents, Dongmei Liu and Jiancai Wang, and my grandparents, Yongqin Sang and Cheng Liu, for their endless love, support, and encouragement, and for letting me pursue my dreams far from home.

September 2019, at St. John's, NL, Canada

Table of contents

List of tables

List of figures

List of abbreviations

AF	Amplify-and-forward
AMC	Adaptive modulation and coding
CSI	Channel state information
DP	Dynamic programming
EE	Energy efficiency
EH	Energy harvesting
ESI	Energy state information
FD	Full-duplex
HD	Half-duplex
MDP	Markov decision process
PSDs	Power spectral densities
P2P	Point-to-point communication
QoS	Quality-of-service
RL	Reinforcement learning
RSI	Residual self interference
RX	Receiver
SE	Spectral efficiency
SI	Self interference
SIC	Self interference cancellation
SINR	Signal-to-interference-plus-noise-ratio
SNR	Signal-to-noise-ratio
TD	Temporal-difference
TX	Transmitter
USD	U.S. dollar
UWAC	Underwater acoustic communication

Co-authorship Statement

I, Ranning Wang, have the principal authorship status for all the manuscripts included in this thesis. However, all the manuscripts resulting from this work are co-authored by my supervisor and collaborators, namely, Prof. Octavia A. Dobre, Dr. Animesh Yadav, Esraa A. Makled, Prof. Ruiqin Zhao, and Prof. Pramod K. Varshney. The list of manuscripts resulting from this thesis are cited below:

1. **Ranning Wang**, Animesh Yadav, Esraa A. Makled, Octavia A. Dobre, Ruiqin Zhao, and Pramod K. Varshney, "Optimal Power Allocation for Full-Duplex Underwater Relay Networks with Energy Harvesting: A Reinforcement Learning Approach", accepted by *IEEE Wireless Communications Letter,* Oct. 2019.

2. **Ranning Wang**, Esraa A. Makled, Animesh Yadav, Octavia A. Dobre, and Ruiqin Zhao, "Reinforcement Learning-based Energy-efficient power allocation for underwater relay-based network", submitted to *IEEE/MTS OCEANS* 2020, Singapore.

Chapter 1

Introduction

1.1 Introduction

About two-thirds of the planet's surface is covered by ocean. Most of the underwater places are waiting to be explored. Many species of fish and aquatic invertebrates use sound to communicate. For example, whales produce echoes of their own calls to hunt and navigate underwater. It was known that acoustic waves, in comparison with radio and light waves, are the only effective means for long-haul underwater wireless communication [1].

The underwater acoustic communication (UWAC) global market is expected to grow from the U.S. dollar (USD) 1.15 billion in 2016 to USD 2.86 billion by 2023. These increasing demands are from the oil and gas industry, naval defense, environmental monitoring, and academic research, thereby facilitating the development of UWAC [2]. However, it is challenging to provide a high quality-of-service (QoS) for UWAC. First, the UWAC is known for the low data rate due to the limited operational bandwidth. Second, the high transmission delay resulted from the slow propagation speed of acoustic waves (1500 m/s), e.g., the round-trip time is approximately 0.7 s

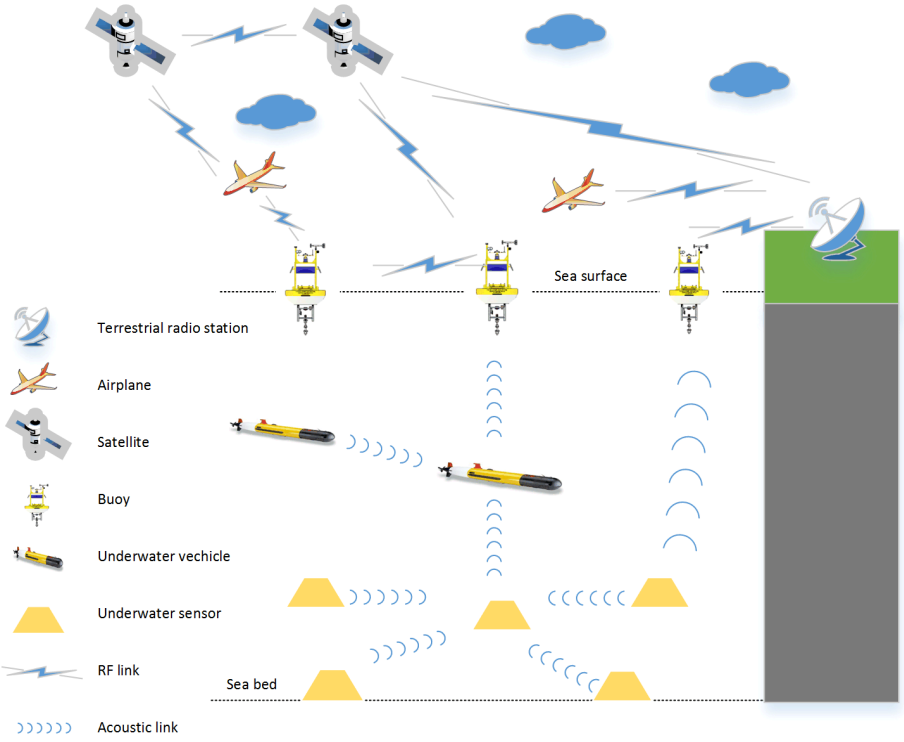

Figure 1.1: Future merged communication networks [3, 4].

for 1 km [3].

The future communication networks will merge underwater, space, air, and terrestrial communication networks [3, 4], as shown in Figure 1.1. The ubiquitous networks will enable us to be networked anywhere and anytime with anybody and anything to access desired information. This thesis focuses on UWAC networks, which consist of sensor nodes, buoy nodes, and underwater vehicles, and aim to build a high QoS, reliable, and intelligent network.

Underwater cooperative communication introduces the relay nodes to help forward the message from the source node to the destination node. Compared to the point-to-point (P2P) UWAC, the underwater relay network has several benefits: 1) the effective operation bandwidth is wider [5]; 2) it achieves higher throughput [5]; and 3) the energy consumption is lower [6].

1.1.1 Motivation

The research motivation of this thesis is to leverage the current popular machine learning techniques to build an intelligent underwater network so that it improves the QoS of the underwater relay networks. In this thesis, we focus on a long-term operational network where the communication system is dynamic in each time instant. Two basic performance metrics–throughput and energy efficiency (EE)–are used to evaluate the communication system. Based on the above, the motivations of the thesis are illustrated and detailed as follows.

Throughput

The maximum throughput (channel capacity) C describes how much information (in bits) can be transmitted over the system bandwidth B. The equation for calculating C is

$$C = \Lambda B \log_2 \left(1 + \frac{S}{N}\right), \tag{1.1}$$

where S and N are the average signal and noise powers over the bandwidth, respectively. Λ is 1 (0.5) for the single-relay network operating in the full-duplex (half-duplex) mode. It can be seen that the choices of duplex mode, the usage of bandwidth, and the signal-to-noise-ratio ($\frac{S}{N}$) affect the channel capacity C.

With increased throughput in cellular communication, more mobile phone services—from text, voice, to data—emerge to satisfy the demand. However, the throughput in UWAC is significantly lower than in cellular communication, and that leads to providing limited communication services. Therefore, it is crucial to improve the throughput in UWAC.

EE

EE is defined as the ratio of throughput over the consumed energy. The metric measures how much information (in bits) is transmitted for a unit consumed energy (in Joules). The consumed energy of the transceiver has a relationship with the transmit power. Thus, one important question is raised: "What is the optimal transmit power to achieve maximum EE?"

Energy consumption and EE for the underwater devices should be taken into considerable account in UWAC because most of the underwater devices are powered by batteries with limited capacity, and the replacement of the batteries is difficult due to the harsh oceanic environment and replacement costs.

Long-term operational network

The above two performance metrics will be improved throughout the long-term operational network. The optimization problem in the long-term operational network considers optimizing the current and future utility function, which is different from the conventional static network where the utility function is optimized only in the current time instant. Since the causal knowledge of the system is known and the long-term optimization problem involves random variables, convex optimization tools cannot directly be applied to solve the problem. Thus, reinforcement learning techniques are used instead to solve the stochastic optimization problem in a long-term operational network.

1.1.2 Thesis Outline

The outline for this thesis is as follows:

- Chapter 2 introduces the background knowledge in the areas of cooperative communication, energy harvesting communication, and reinforcement learning technique.

- Chapter 3 presents an overview of underwater acoustic propagation and underwater acoustic channels.

- Chapter 4 proposes an online transmission policy for self-sustainable underwater full-duplex single-relay networks which is powered by ambient harvested energy. The transmission policy is designed for the end-to-end sum rate maximization, and the problem is formulated as an online sequential decision-making problem and solved by reinforcement learning techniques.

- Chapter 5 designs an online stationary transmission policy for end-to-end average EE maximization under the same system model as in Chapter 4.

- Chapter 6 concludes the thesis and points out possible future research directions.

1.1.3 Research Contributions

Motivated by the need to improve the throughput and EE for the long-term operational underwater relay networks, the following research contributions are made:

- In Chapter 4, a three-node underwater network that consists of one source, one relay, and one destination is studied. In this scenario, the relay operates in the full-duplex (FD) mode, and can transmit and receive the signals simultaneously. Also, it is equipped with an energy harvesting (EH) unit to power the communication system. We formulate the end-to-end (source to destination) throughput

maximization problem over a finite time-slots (finite horizon) as a Markov decision process (MDP) framework. We obtain an optimal long-term operational transmission policy by using a reinforcement learning algorithm. Simulation results showed that the proposed policy achieves a higher throughput when compared to the greedy policy. Also, the FD performance is determined by the level of self-interference cancellation (SIC). The better the SIC level is, the higher the throughput is. Moreover, we investigate the system performance for different relay location and observe that the relay location affects the throughput performance.

- In Chapter 5, a single-relay network with one source, one relay, and one destination, is investigated. The direct link between the source and destination has been blocked due to obstacles; therefore, the relay forwards received data from the source to destination. Moreover, the relay uses the FD mode and has an EH unit. We aim to maximize the average energy efficiency of the single-relay network in the infinite horizon, and the problem is formulated as an infinite horizon average reward MDP problem. We develop a reinforcement learning-based energy-efficient transmission framework to obtain an optimal transmission policy. According to the relay's system state, such as the channel state information, battery level, and self-interference power level, it can select the optimal transmit power. Three transmission policies—optimal, greedy, and fixed—are compared and simulation results show that optimal policy outperforms others in terms of the average EE of the single-relay network. Further, the FD performance is better than the half-duplex (HD) one, because FD allows the transceiver to transmit and receive at the same time, whereas the transmit and receive operations in half-duplex are in two different time-slots.

Bibliography

[1] R. J. Urick, *Principles of Underwater Sound*. New York, US: McGraw-Hill Press, 1983.

[2] "Underwater Acoustic Communication Market," https://www.marketsandmarkets.com/PressReleases/underwater-acoustic-communication.asp, 2018.

[3] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad-hoc Networks*, vol. 3, no. 3, pp. 257–279, Mar. 2005.

[4] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 2th. Quart. 2018.

[5] R. Cao, L. Yang, and F. Qu, "On the capacity and system design of relay-aided underwater acoustic communications," in *Proc. IEEE Wireless Communication and Networking Conference*, Apr. 2010, pp. 1–6.

[6] Y. Li, Y. Zhang, H. Zhou, and T. Jiang, "To relay or not to relay: Open distance and optimal deployment for linear underwater acoustic networks," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3797–3808, Sep. 2018.

Chapter 2

Background

In this chapter, background knowledge on cooperative communications, energy harvesting communications, and reinforcement learning technique are introduced, respectively.

2.1 Cooperative Communications

When the P2P communications cannot meet the basic QoS requirement to the end-users, one or more relay nodes come into play with the role of forwarding the message to end-users, which is referred to as cooperative communications. Cooperative communications can not only extend the communication range, but also improve communication reliability[1].

2.2 Relay-aided Communications

In a communication system, the role of the relay is to forward the received signals from the source to the destination or another relay node. The benefits of relay are extending the communication range and increasing the spatial diversity of the communication system. For example, consider a three-node communication system with one source, one relay, and one destination. If there is a blockage between the source and the destination nodes, then the introduced relay node can extend the coverage of the communication system. If the channel between the source and the destination nodes is feasible, then the communication system gains spatial diversity by the added relay node.

There are different processing techniques on how to process the received signals at the relay. These techniques result in various relaying protocols, among which the most populalr are

- **amplify-and-forward**: the relay re-scales the received signals and transmits the amplified version of the signals to the destination.
- **decode-and-forward**: the relay decodes the received signals and sends the re-encoded information to the destination.

In addition to these commonly used protocols, there are other relaying techniques, such as compress-and-forward and coded cooperation [1].

2.2.1 Half-duplex (HD)

In the conventional HD mode, the transmissions are divided into two orthogonal phases, either in time-division multiplexing or frequency-division multiplexing manner

[1]:

- In the first phase, the source sends the signals to both destination and relay.
- In the second phase, the relay re-transmits the signals to the destination.

The disadvantage is that the system rate reduces to half, because the re-transmission of the relay occupies an extra resource block.

2.2.2 Full-duplex (FD)

With the recent advance of self-interference cancellation (SIC) techniques, the FD mode becomes feasible and promising in communication systems, as it allows the transceiver to receive and transmit signals at the same frequency and time [2]. Thus, the throughput of the communication system would be double if the self-interference is fully suppressed, when compared to the conventional HD mode.

2.2.3 Relay-aided (UWAC)

When compared with point-to-point communication, the relay-aided underwater communication system can increase the throughput [3, 4] and reduce energy consumption [5]. The reason behind is that the effective operational bandwidth increases as the communication distance decrease, which is due to the effect of underwater path loss [3].

2.3 Energy Harvesting (EH) Communications

Green communication has attracted significant attention in recent years with the increasing carbon footprint in the Earth's environment. EH devices can harvest energy from the ambient environment, such as solar energy, wind, tidal waves, and radio-frequency waves, and store it in the rechargeable battery. These devices are sustainable and self-containable from the energy supply perspective, and do not depend on the conventional power grid. EH communication has numerous benefits: not only it saves energy consumption and cost, but also it reduces the carbon footprint [6].

2.3.1 EH Techniques in the Oceanic Environments

EH techniques have shown strong potential for powering the underwater devices by harvesting energy from the ambient environment, such as solar energy, microbe, and sea waves [7–9]. In [7], the authors investigated the use of solar energy to power autonomous jellyfish vehicle. The experimental results revealed that the degree of harvested energy decreases with the ocean depth and with increased turbidity. In [8], the authors designed an electronic circuit to harvest energy from benthic microbes in a littoral tidal basin. This is also called microbial fuel cell, which converts chemical energy to electrical energy by the action of microorganisms. In [9], the piezoelectric bimorphs elements are used to convert mechanical energy to electrical energy on the sea bottom.

2.4 Adaptive Communications

In a communication system, both transmitter and receiver should adopt the adaptive techniques against the stochastic characteristic of the channel and or any dynamic changes occurring in the system to achieve a specific system performance goal. For example, for a communication system in the time-varying wireless channels, the adaptive mechanism is the receiver estimates the channel information and feeds back to the transmitter. Therefore, the transmission scheme can adapt to the channel characteristic [10].

2.4.1 Adaptive Communications in UWAC

At the physical layer of UWAC, the authors in [11] investigated an adaptive modulation and coding (AMC) scheme with a finite number of transmission modes to combat the fast-varying underwater channel. They proposed the effective SNR as the indicator for the AMC scheme and showed that the performance outperforms the benchmark ones. In [12], an adaptive modulation and power scheme is explored to maximize the system throughput under a target average bit error rate. The results showed that the adaptive scheme achieved a higher throughput than the non-adaptive scheme that allocated uniform power and modulation. In [13], three power allocation strategies and their effects on the achievable rate for the underwater system are studied.

2.4.2 Scheduling Problems in Adaptive Communications

Depending on the availability of the causal knowledge of the communication system, there are two research approaches, *offline* and *online* settings, for the scheduling

problems in adaptive communications. Such knowledge could be the energy state information (ESI) in the energy harvesting communication systems and/or the channel state information (CSI) [6, 14].

- For the *offline* setting, the communication system has non-causal (past and future) knowledge of the CSI and/or ESI over a period of time at the beginning of the transmission. Although this setting is impractical as assuming the non-causal knowledge of the channel fading and the energy arrival, it provides the upper-bound on the system performance.

- For the *online* setting, the communication system has past and current (causal) knowledge of the CSI and/or ESI, which is realistic for most of the communication systems. Dynamic programming, which is one of reinforcement learning algorithms, is the typical mathematical tool to solve the online setting optimization problem [15, 16].

2.5 Reinforcement Learning

Reinforcement learning (RL) is a machine learning technique that describes an agent interacting over time with its environment with the goal of maximizing the cumulative rewards [16]. The Markov decision process (MDP) model [15] is used to formulate the RL problem in terms of states, actions, and rewards, as can be seen in Fig. 2.1. Specifically, an agent occupies a *state* in each time epoch, and it receives the corresponding *reward* when taking a certain *action*. Also, the policy is the set of state-action pairs, which is the mapping function between states and actions. RL is a mathematical tool that can be applied to solve a class of resource management and allocation problems [17].

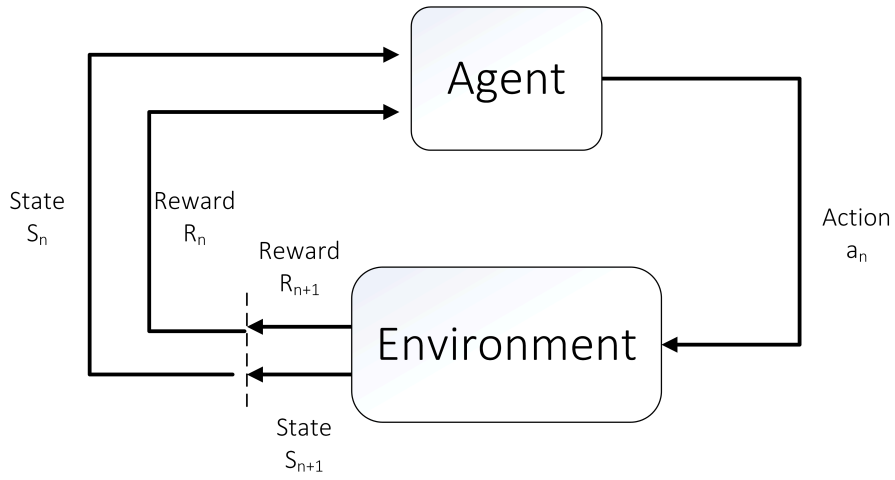

Figure 2.1: Diagram of reinforcement learning.

2.5.1 History of RL

The modern RL is developed in three threads. The first main thread is built upon the problem of optimal control and its well-known solution dynamic programming (DP). The second main thread concerns the trial-and-error process that started in the psychology of animal learning. The last and less distinct thread concerns the temporal-difference (TD) methods [16].

The term optimal control came into use in the late 1950s to describe the problem of designing a controller to minimize a measure of a dynamical system's behavior over time. Richard Bellman developed one of the approaches to this problem in the mid-1950s. This approach uses the concepts of a value function to define a functional equation, called the Bellman equation. The drawback is described by Bellman as the curse of dimensionality, which means the computational requirements grow exponentially with the number of state variables.

TD learning is a combination of Monte Carlo ideas and DP ideas. Like Monte Carlo methods, TD methods can learn directly from raw experience without a model

of the environment's dynamics. Like DP, TD methods update estimates based in part on other learned estimates, without waiting for a final outcome.

2.5.2 RL Algorithms

Learning an efficient strategy or policy is the primary objective of the RL framework. The function of RL algorithms is to obtain the strategy for the specifically defined goal. There are two types of RL modeling: model-based and mode-free frameworks, which rely on the availability of the prior knowledge of the environment [16]. The details are presented in the following:

- **Model-based RL framework** has the knowledge of the environment model. It takes a model as input and produces or improves a policy for interacting with the modeled environment. For example, algorithms in DP are the model-based framework. This framework is used to solve the problems in Chapters 4 and 5 because the environmental models are known to the system.

- **Model-free RL framework** does not have the knowledge of the environment model. It relies on the learning through trial-and-error experiences. For instance, Q-learning and Deep Q-learning are model-free frameworks.

Q-learning

Q-learning is a model-free RL algorithm that does not need the prior knowledge of the specific statistical or deterministic model of the environment, and it is different from the model-based RL framework. The Q-learning algorithm can solve the problem with the unknown environment. The policy is updated iteratively with the combination of the past (history) and the current learned values [16].

Q-learning works well if the state and action spaces of the problem are small, and a look-up table can be used to obtain the updated rule. However, it is impossible when the state-action space becomes very large. In this situation, many states may be rarely visited, thus the corresponding Q values are seldom updated, leading to a much longer time to converge. This can be addressed by deep RL, which will be introduced in the following paragraph.

Deep RL

In 2013, a group of researchers in an English startup called Deepmind proposed the deep RL algorithm to play the Atari games and achieved outstanding performance. The principle of deep RL is that the deep learning model is adopted for function approximation of parameters in reinforcement learning [16, 18, 19]. One of deep RL advantages is that it can use large state and action space and get a faster convergence rate to the optimal policy [20]. This framework could be implemented in our future work.

2.5.3 Two Steps for Solving RL Problems

The-state-of-the-art artificial intelligent (AI) is still in its early stage. The AI/ML business applications in the market are mapping problems, which refer to a set of data mapping to another set of data through a certain relationship. Examples are language translation and identification of images.

There are two basic steps to solve RL: training the model or learning the mapping function and deploying the model. The training aims to derive the policy, which is a set of state-action pairs. Once the policy is obtained, it can be deployed in the system to guide it how to make a decision under a certain system state.

Bibliography

[1] K. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking*. Cambridge University Press, 2009.

[2] L. J. Rodrguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex af relaying in the presence of residual self-interference," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[3] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[4] R. Cao, L. Yang, and F. Qu, "On the capacity and system design of relay-aided underwater acoustic communications," in *Proc. IEEE Wireless Communication and Networking Conference*, Apr. 2010, pp. 1–6.

[5] Y. Li, Y. Zhang, H. Zhou, and T. Jiang, "To relay or not to relay: Open distance and optimal deployment for linear underwater acoustic networks," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3797–3808, Sep. 2018.

[6] M. Ku, W. Li, Y. Chen, and K. J. R. Liu, "Advances in energy harvesting communications: Past, present, and future challenges," *IEEE Commun. Surveys Tutorials*, vol. 18, no. 2, pp. 1384–1412, 2nd. Quart. 2016.

[7] K. B. Joshi, J. H. Costello, and S. Priya, "Estimation of solar energy harvested for autonomous jellyfish vehicles (ajvs)," *IEEE Journal of Oceanic Engineering*, vol. 36, no. 4, pp. 539–551, Oct. 2011.

[8] P. R. Bandyopadhyay, D. P. Thivierge, F. M. McNeilly, and A. Fredette, "An

electronic circuit for trickle charge harvesting from littoral microbial fuel cells," *IEEE J. Oceanic Eng.*, vol. 38, no. 1, pp. 32–42, Jan. 2013.

[9] D. M. Toma, J. del Rio, M. Carbonell-Ventura, and J. M. Masalles, "Underwater energy harvesting system based on plucked-driven piezoelectrics," in *Proc. OCEANS 2015 - Genova*, May 2015, pp. 1–5.

[10] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[11] L. Wan, H. Zhou, X. Xu, Y. Huang, S. Zhou, Z. Shi, and J.-H. Cui, "Adaptive modulation and coding for underwater acoustic ofdm," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 2, pp. 327–336, 2014.

[12] A. Radosevic, R. Ahmed, T. M. Duman, J. G. Proakis, and M. Stojanovic, "Adaptive ofdm modulation for underwater acoustic communications: Design considerations and experimental results," *IEEE J. Oceanic Eng.*, vol. 39, no. 2, pp. 357–370, 2013.

[13] Y. M. Aval, S. K. Wilson, and M. Stojanovic, "On the achievable rate of a class of acoustic channels and practical power allocation strategies for ofdm systems," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 4, pp. 785–795, Oct 2015.

[14] M. Dong, W. Li, and F. Amirnavaei, "Online joint power control for two-hop wireless relay networks with energy harvesting," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 463–478, Jan 2018.

[15] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. Press, 1994.

[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[17] H. Ye, L. Liang, G. Y. Li, J. Kim, L. Lu, and M. Wu, "Machine learning for vehicular networks: Recent advances and application examples," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 94–101, June 2018.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[19] A. Géron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media, Inc., 2017.

[20] L. Xiao, Donghua, Jiang, X. Wan, W. Su, and Y. Tang, "Anti-jamming underwater transmission with mobility and learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 542–545, Mar. 2018.

Chapter 3

Underwater Acoustic Propagation and Channel

This chapter gives a comprehensive overview of underwater acoustic propagation and channel modeling, in terms of the physical acoustic properties, propagation phenomenon, and multi-path propagation. Next, the system designing and strategy are illustrated for the practical underwater communication system. Finally, the performance analysis for the underwater channel is presented.

3.1 Underwater Acoustic Propagation

This section compares the acoustic waves with other underwater information propagation ways and presents how the acoustic waves propagate at different underwater layers. Finally, the section shows how to design a practical underwater communication system.

3.1.1 Introduction

The sea covers about two-thirds of our planet's surface. There are promising and developing communication techniques and scenarios within the underwater acoustic communication field for commercial, navigation, and military use. However, how to characterize the underwater acoustic channel is a challenge. Knowledge of the channel model will help predict the overall performance of the communication system. As the transmission distance increases, the signal energy will inevitably decrease. Additionally, underwater ambient noise is an issue for the signal quality. From the communication model perspective, we ask the questions: what is the shape of transmission acoustic ray in different locations and environments? What is the underwater environments impact on transmitted signals, which are carefully addressed in this section.

3.1.2 Transmission Ways in Underwater Communications

Similar to terrestrial communications, underwater communications can utilize both wired and wireless ways to send information. The acoustic waves are the effective transmission carrier for a long-haul underwater wireless communication. The advantages and disadvantages of different transmission ways for underwater communications will be described in the following paragraphs. Also, Table 3.1 compared three wireless transmission ways with different evaluation metrics [1].

Cable (wired): The wired cables connect underwater nodes to communicate with each other. Although it is the most reliable way to provide communication services, the drawbacks are the high deployment and maintenance costs in the harsh oceanic environment.

Acoustic waves (wireless): Acoustic waves are an effective way for long-haul underwater wireless communications, and its effective communication range is on the orders of kilometers. Nonetheless, the disadvantages are the low data rate and high transmission latency because of the limited operational bandwidth and the slow transmission speed for acoustic waves, respectively.

Electromagnetic waves (wireless): Seawater has a strong absorption effect on the electromagnetic waves. Therefore, the electromagnetic waves can transmit the signals merely up to tens of meters.

Optical waves (wireless): Optical waves can transmit a very high data rate. However, the transmission distance is short due to the scattering effect and the turbidity of the seawater.

Table 3.1: Comparison of three wireless transmission ways in underwater communications [2].

	Acoustic	Electromagnetic	Optical
Nominal speed (m/s)	$\sim$ 1500	$\sim$ 33 333 333	$\sim$ 33 333 333
Power loss	relatively small	large	$\propto$ turbidity
Bandwidth	$\sim$ kHz	$\sim$ MHz	$\sim$ 10-150 MHz
Frequency band	$\sim$ kHz	$\sim$ MHz	$\sim 10^{14}$-10^{15} Hz
Antenna size	$\sim$ 0.1 m	$\sim$ 0.5 m	$\sim$ 0.1 m
Effective range	$\sim$ km	$\sim$ 10 m	$\sim$ 10-100 m

3.1.3 Underwater Acoustic Propagation

The propagation of underwater acoustic waves is more complicated than that of light in free-space, as the acoustic speed varies with time, geographical location, and depth of the seawater. It is known that the acoustic speed depends on temperature, salinity, and pressure. Illustrative plots of the three parameters as a function of depth

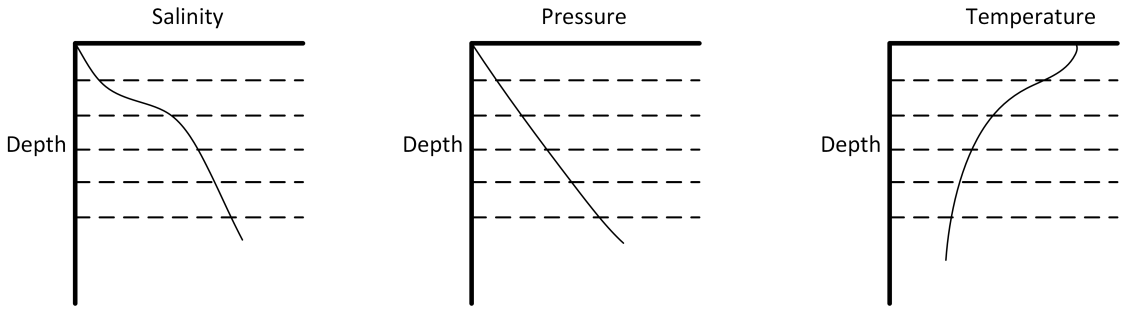

Figure 3.1: Salinity, pressure, and temperature as a function of depth [1].

are shown in Figure 3.1. The empirical acoustic speed c can be calculated by [3]

$$\begin{aligned} c =& 1448.96 + 4.591T - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 + 1.34(S-35) \\ &+ 1.63 \times 10^{-2}D + 1.675 \times 10^{-7}D^2 - 1.025 \times 10^{-2}T(S-35) - 7.139 \times 10^{-13}TD^3, \end{aligned} \tag{3.1}$$

where T is the temperature (in degrees Celsius), S is the salinity (in parts per thousand), and D is the depth (in meters). Note that this equation is valid for $0 \leq T \leq 30°$, $30 \leq S \leq 40$, and $0 \leq D \leq 8000$.

Figure 3.2 shows the typical acoustic speed profile and the corresponding ray-tracing where both transmitter (TX) and receiver (RX) are located at the depth of 1300 m (acoustic channel axis), which the ray-tracing is generated from the Bellhop simulator.

According to Snell's law, the acoustic ray bends toward the direction of the minimum acoustic speed [4]. It can be seen in the right plot of Figure 3.2 that the rays bend toward the acoustic channel axis where there is a minimum acoustic speed. Therefore, the transmitted shape of the acoustic ray is a bend curve, as shown in the Figure 3.2.

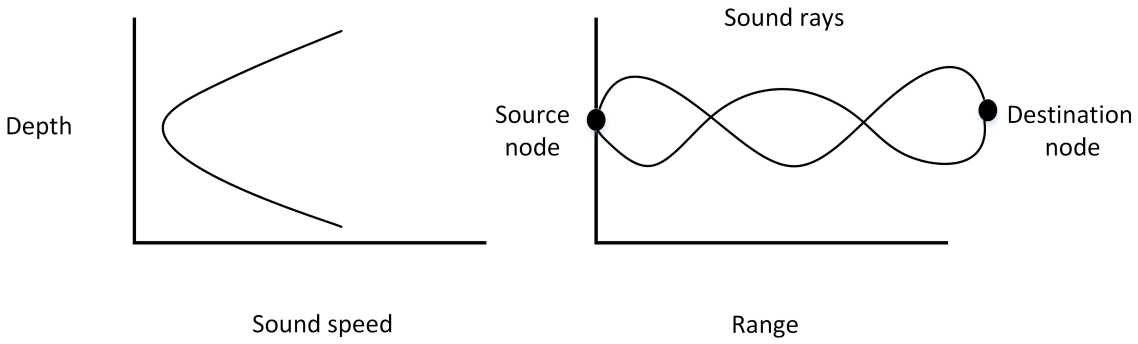

Figure 3.2: Typical acoustic speed profile and corresponding ray tracing between the source and destination nodes.

3.1.4 Profile of the Sea

As shown in Figure 3.3, the profile of the sea can be divided into 4 layers and is presented below [1].

- Surface layer (or mixed layer): the depth of the layer is a few tens of meters. Acoustic speed varies with local changes, such as heating, cooling, and wind action. Moreover, the acoustic speed is constant as both salinity and temperature tend to be homogeneous in the layer.

- Seasonal thermocline layer: the acoustic speed has a seasonal effect and is a negative gradient since the temperature decreases with depth.

- Main thermocline layer: the acoustic speed is decreasing because temperature decreases with depth. Also, the increased salinity and pressure cannot compensate for the decreased temperature.

- Deep isothermal layer: the temperature is nearly constant at around 4 degree Celsius. Therefore, the acoustic speed increases with pressure.

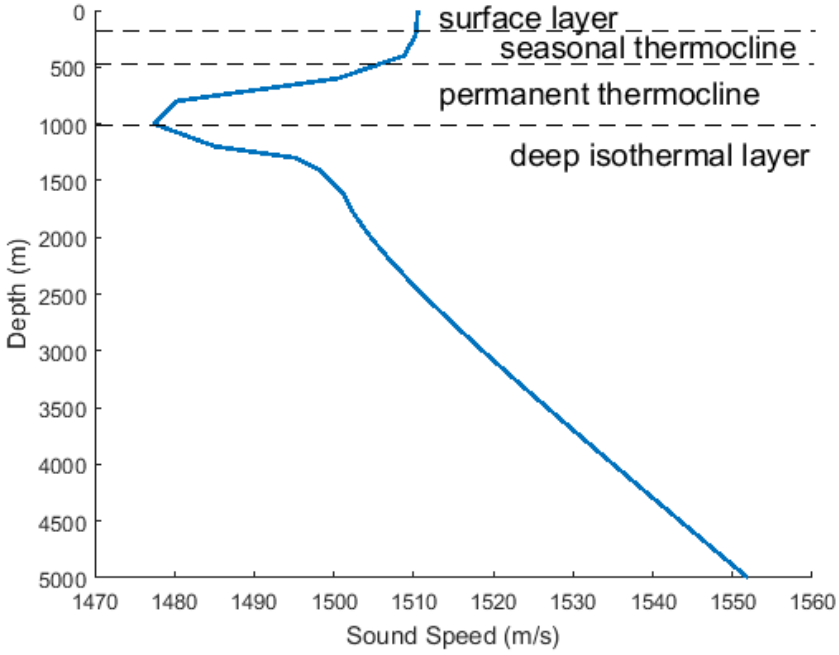

Figure 3.3: Typical acoustic speed profile of the sea [1].

3.1.5 The Propagation Path of Acoustic Ray

To better understand how the underwater acoustic ray propagates, the following section will introduce several basic propagation paths. These include the surface reflection, bottom bounce, surface duct, deep sound channel, convergence zone, and reliable acoustic path [4]. An illustration plot is shown in Figure 3.4.

Surface reflection: The acoustic ray is reflected by the sea surface. The smoothness of the sea surface affects the reflection performance. The transmission loss for surface reflection is determined by carrier frequency, wind speed, and grazing angle.

Bottom bounce: Similar to the surface reflection, the acoustic ray is reflected by the sea floor. The transmission loss is determined by the sediment type and grazing angle of the ray. When the grazing angle is less than the critical angle, the transmission energy will be lost.

Surface duct: In the surface layer, the acoustic speed has a positive gradient. If the layer is deep enough, the acoustic ray is channeled or confined in the layer and may bend toward the surface and reflect back into the layer.

Deep sound channel: There exists a minimum acoustic speed at a certain depth, called the acoustic channel axis, and the acoustic speed is increasing both above and below that depth. If the acoustic ray propagates near the acoustic channel axis, then it bends toward the channel axis back and forth. Thus, the ray is confined within that depth and no transmission losses are caused by reflecting from the surface or the bottom. The performance of this channel should be the best.

Convergence zone: The signal is transmitted from the shallow source and travels into the deep sea, and then travels back to shallow water. First, the ray bends downward since the acoustic speed has negative gradient due to the decreased temperature. Second, the ray bends upward because of the positive gradient of the acoustic speed. The depth of the sea should be deep enough to form a convergence zone.

Reliable acoustic path: If the acoustic source is located in a very deep sea and the receiver in the shallow water, then the propagation path forms as a reliable acoustic path since the wave is first refracted downward and refracted upward.

Shadow zone: The formation of the shadow zone is due to the acoustic ray bending. Typically, there are no signals in the zone. Therefore, the RX should not be placed inside the shadow zone.

3.1.6 Bellhop Simulator

The Bellhop is a ray-tracing simulator based on ray theory. It calculates the communication performance, such as ray-tracing, transmission loss, power delay profile,

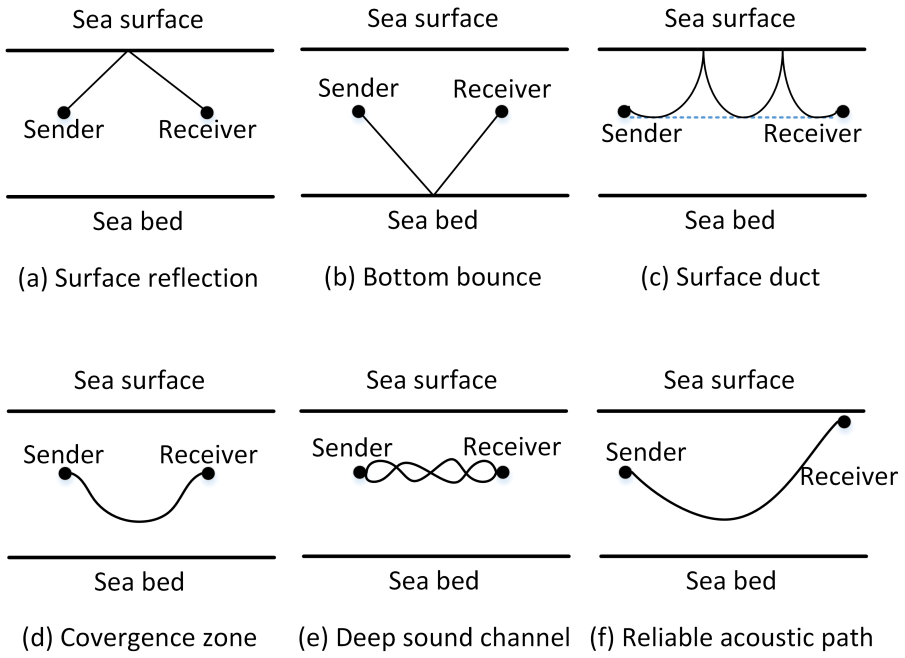

Figure 3.4: Acoustic propagation paths [4].

etc, from the input parameters [5]. Note that the underwater channel simulations in Chapters 4 and 5 are not generated from the Bellhop simulator.

Input Files

Input files are used for setting the communication scenario and system parameters. These include the operational bandwidth, center carrier frequency, location of the transceiver, the number of acoustic rays, inject direction of the acoustic ray, etc. For instance, the system environment parameters include the acoustic speed profile, the shape of the sea bed, the surface and bottom reflection coefficient, etc.

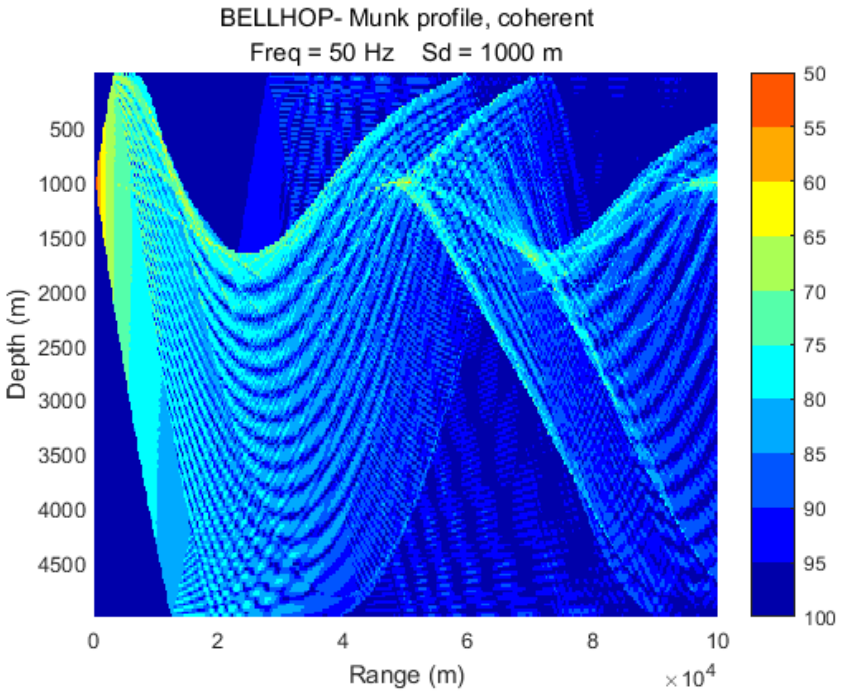

Figure 3.5: Transmission loss plot from the Bellhop simulator.

Output Files

The ray-tracing plot can get a sense of how the acoustic rays propagate in the channel. The eigenray represents the direct path from the TX to RX. Moreover, the transmission loss plot describes the signal intensity in both range and depth dimensions. Further, the signal amplitude and delay profile define the loudness and delay for every ray in the channel. For example, Figure 3.5 shows the transmission loss plot.

3.1.7 UWAC System Design

In order to build an efficient UWAC network, the network design should take into account the effects of the underwater acoustic propagation and channel [6].

Topology design

The locations of the TX and RX should be decided after the complete channel characterization, such as the acoustic propagation path and shadow zone, is known. Also, the physical property of the sea should be considered in the system design. Particularly, the deep sea communication performance is usually better than the shallow water counterpart. Further, introducing the relay nodes can improve the throughput and reduce the bit error rate. Thus, the optimal relay location should be studied through analysis and simulation.

Operating frequency

The selection of the operating frequency affects the intensity of the received signal because the path loss increases with the operating frequency. Also, there exists an optimal operating frequency under a given transmission range, for which the minimum bit error rate is achieved.

Environment-aware protocol design

The analysis revealed that the acoustic speed varies with seasons and sites. Therefore, the propagation path of the acoustic rays changes with different seasons and places. Thus, designing an environment-aware protocol which can adaptively adjust the transceiver parameters according to the seasons and deployment sites becomes essential.

3.2 Performance Analysis in Underwater Channel

In this section, underwater path loss and ambient noise are introduced and based on that, the signal-to-noise-ratio (SNR) is calculated for the performance analysis.

3.2.1 Underwater Path Loss Model

Path loss or transmission loss is a combination of *geometric spreading loss* and *attenuation loss*. *Geometric spreading loss* is a geometrical effect that represents the regular weakening of an acoustic signal as it spreads outward from the source. *Attenuation loss* includes the effects of absorption, scattering, and leakage out of acoustic channels.

The path loss experienced by a transmitting signal at frequency f in kHz over a distance l in km is given by [7]

$$A(l, f) = A_0 l^k a(f)^l, \tag{3.2}$$

where A_0 is a unit-normalizing constant, k is the spreading factor, and $a(f)$ is the absorption coefficient. Further, its expression in decibel (dB) is given by

$$10 \log \frac{A(l, f)}{A_0} = k \times 10 \log(l \times 1000) + l \times 10 \log a(f), \tag{3.3}$$

where the first term represents the spreading loss and the second term is the attenuation loss. The typical values for the spreading factor k are 1.5 for practical spreading, 1 for cylindrical spreading, and 2 for spherical spreading. $a(f)$ is the absorption

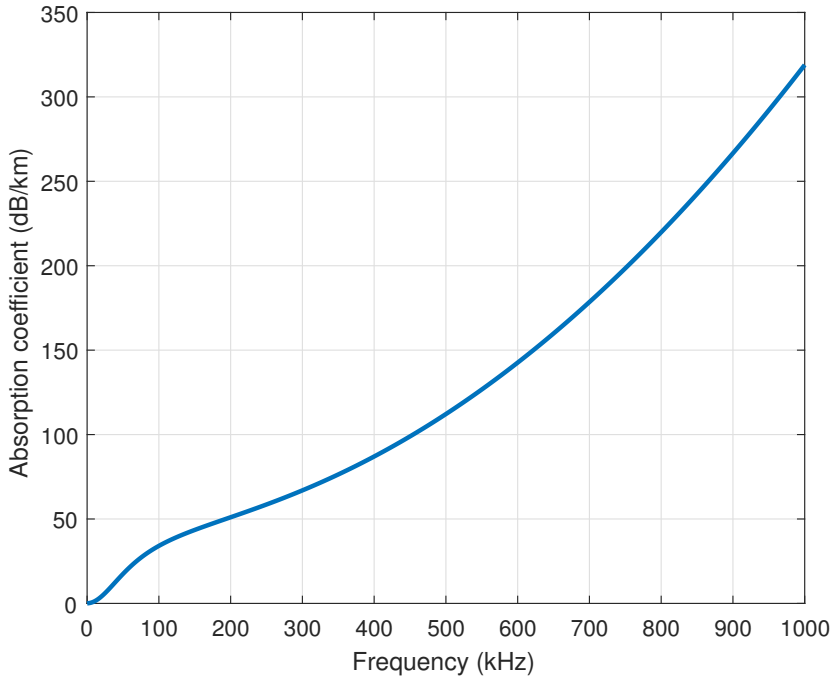

Figure 3.6: Absorption coefficient [7].

coefficient expressed using the Thorp's formula (in dB/km) as [7]

$$10 \log a(f) = \frac{0.11 f^2}{1 + f^2} + \frac{44 f^2}{4100 + f^2} + 2.75 \cdot 10^{-4} f^2 + 0.003. \tag{3.4}$$

Figure 3.6 shows the absorption coefficient $a(f)$ versus frequency. It increases with frequency, which illustrates that the absorption loss is higher when the operating frequency of a transmit signal is higher over a given distance.

3.2.2 Underwater Ambient Noise

Typically, the underwater ambient noise is due to four sources: turbulence, shipping, waves, and thermal noise. The corresponding empirical power spectral densities

(PSDs) of these four components in μPa per kHz are given by [7]

$$\begin{aligned} N_t(f) &= 10^{((17-30\log_{10}(f))/10)}, \\ N_s(f) &= 10^{((40+20(s-0.5)+26\log_{10}(f)-60\log_{10}(f+0.03))/10)}, \\ N_w(f) &= 10^{((50+7.5w_s^{0.5}+20\log_{10}(f)-40\log_{10}(f+0.4))/10)}, \\ N_{th}(f) &= 10^{((-15+20\log_{10}(f))/10)}, \end{aligned} \tag{3.5}$$

where $N_t(f)$, $N_s(f)$, $N_w(f)$, and $N_{th}(f)$ are the turbulence, shipping, waves, and thermal noise PSDs, respectively. The shipping activity factor is denoted as s, while w_s is the wind speed. The overall PSDs of the ambient noise are calculated as

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f), \tag{3.6}$$

as shown in the Fig. 3.7. Also, the approximate PSDs is $N_{approx}(f) = 50 - 18\log f$. The underwater noise power in unit dB re μPa is calculated as

$$P_N = 10\log_{10}\left(\int_{f_o}^{f_o+B} N(f)df\right). \tag{3.7}$$

The turbulence noise affects the lower frequency as $f < 10$ kHz. Shipping noise dominates in the frequency range $10 < f < 100$ kHz, and the shipping activity factor s is between 0 and 1, which represents low to high. The motion of wind-driven surface waves influences the frequency region from 100 Hz to 100 kHz. Finally, the thermal noise affects the frequency over 100 kHz, due to the molecular motion.

Unlike the additive white Gaussian noise in the terrestrial wireless channel in which the PSD is flat across the band, the non-whiten nature of underwater noise should be considered for practical performance analysis.

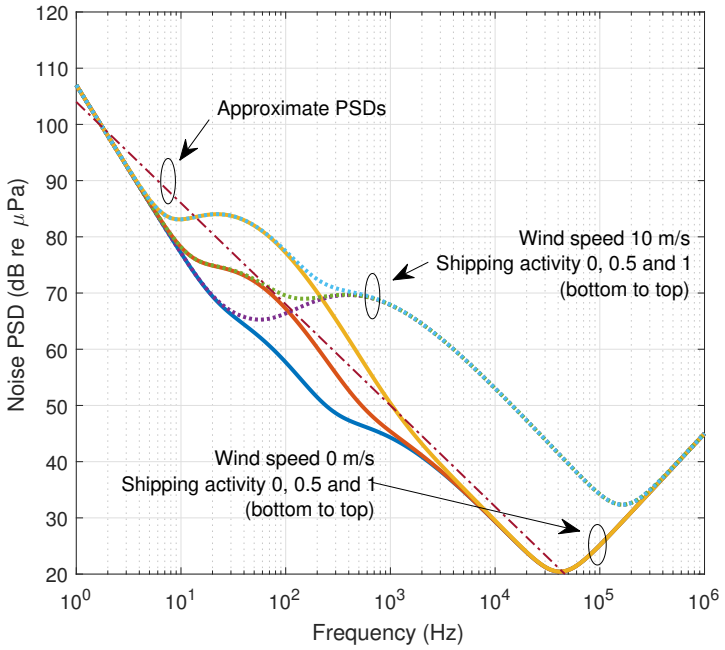

Figure 3.7: Underwater ambient noise [7].

3.2.3 Signal-to-Noise Ratio (SNR)

The SNR received at a frequency f over a distance l is defined as $SNR = \frac{S(l,f)}{A(l,f)N(f)}$, where $S(l, f)$ is the PSD of the transmitted signal.

Assume that the PSD of the narrow-band signal would be flat across the operational bandwidth, the narrow-band SNR over a frequency f and a distance l is calculated using (3.2) and (3.6) as

$$\gamma(l, f) = \frac{P}{A(l, f)N(f)\Delta f}, \tag{3.8}$$

where P is the power of the transmit signal and Δf is the narrow bandwidth around f. The factor $\frac{1}{A(l,f)N(f)}$ determines the SNR, and is plotted in Fig. 3.8 for different distances l. It can be seen that there exists an optimal frequency $f_o(l)$ that gives a maximum narrow-band SNR. Moreover, the optimal frequency $f_o(l)$ is shown in

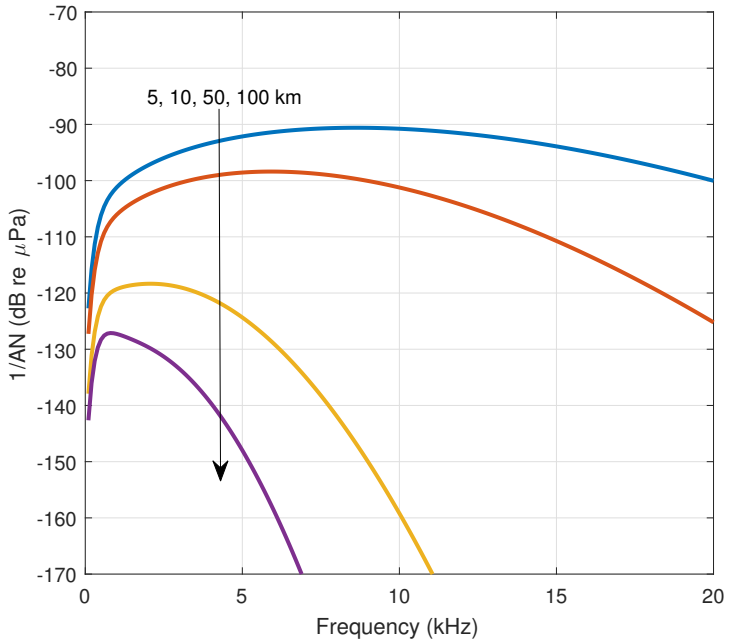

Figure 3.8: $\frac{1}{A(l,f)N(f)}$ [7].

Fig. 3.9. We observe that the optimal frequency $f_o(l)$ decreases with increasing the transmission distance. This suggests the importance of selecting the optimal operating frequency at a given distance.

3-dB Bandwidth

As plotted in Fig. 3.10, the 3-dB bandwidth $B_3(l)$ is defined as a frequency range around the optimal frequency $f_o(l)$, where the obtained narrow-band SNR is greater than half of the SNR achieved at the optimal frequency $f_o(l)$, i.e., $\gamma(l,f) > \frac{\gamma(l,f_o(l))}{2}$ or $A(l,f)N(f) < 2A(l,f_o(l))N(f_o(l))$.

Two calculation methods are used to obtain the $B_3(l)$: 1) exhaustive search is the easiest approach to find the 3-dB bandwidth by comparing the SNR of all frequencies with the optimal frequency, as illustrated in Algorithm 1 [7]; 2) the closed-form

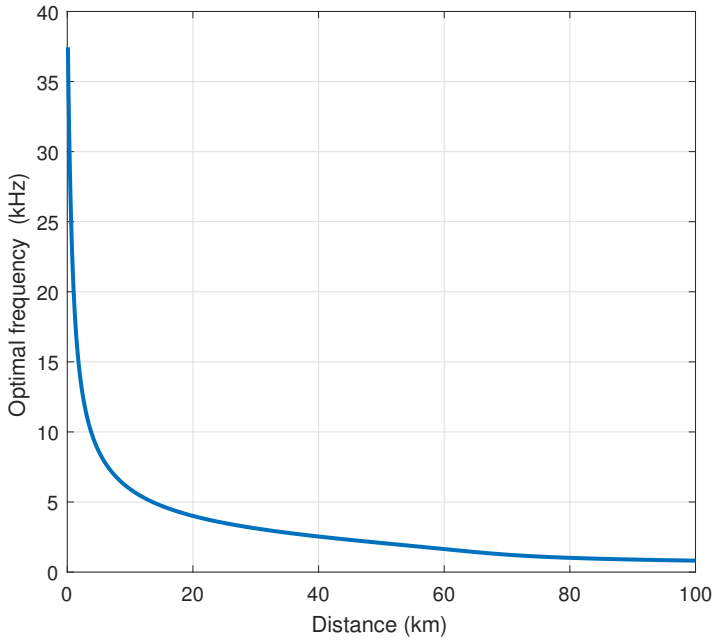

Figure 3.9: Optimal frequency $f_o(l)$ [7].

approximation approach uses the approximate equation to calculate 3-dB bandwidth, $B_3(l) = \omega l^{-\gamma}$, where $\omega = 10^{1.4291}$ and $\gamma = 0.5392$ [8]. This equation is proposed in [7] to illustrate the relationship between the system bandwidth and the transmission

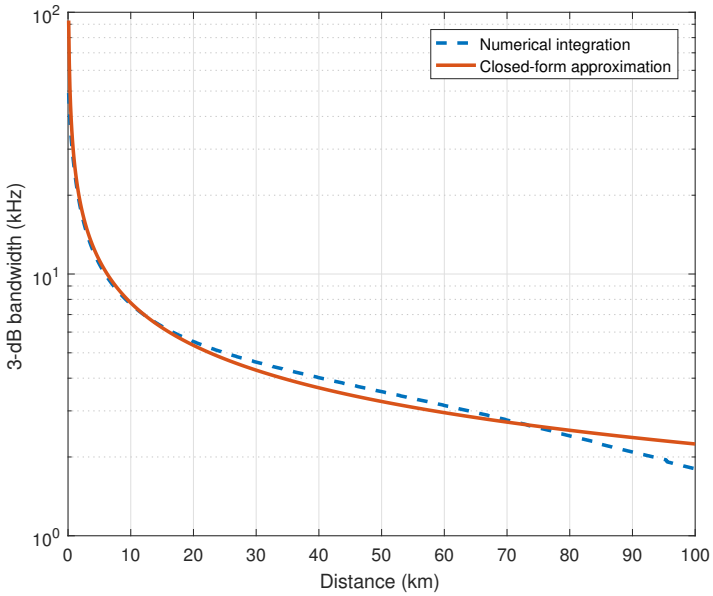

Figure 3.10: 3 dB bandwidth $B_3(l)$ [7].

distance, which comes from the examination of the numerical results.

Algorithm 1: Find the 3-dB bandwidth

Result: $B_3(l)$.

for l **do**
  while $A(l, f_{max})N(f_{max}) < 2A(l, f_o(l))N(f_o(l))$ **do**
    $f_{max} = f_{max} + 1$;
    Calculate $A(l, f_{max})N(f_{max})$;
  end
  while $A(l, f_{min})N(f_{min}) < 2A(l, f_o(l))N(f_o(l))$ **do**
    $f_{min} = f_{min} + 1$;
    Calculate $A(l, f_{min})N(f_{min})$;
  end
  $B_3(l) = f_{max} + f_{min}$;
end

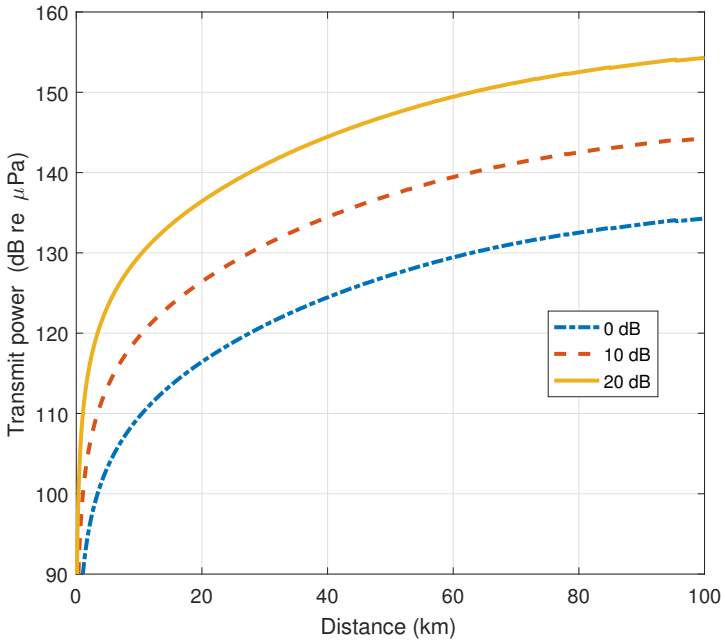

Figure 3.11: Minimum transmission power $P_{min}(l)$.

Minimum Transmission Power

In order to achieve a target SNR γ under an optimal frequency $f_o(l)$ and a 3-dB bandwidth $B_3(l)$, the minimum transmission power P_{min} can be calculated from (3.8) as $P_{min}(l) = \gamma \times A(l, f_o(l)) \times N(f_o(l)) \times B_3(l)$. The minimum transmission power versus transmission distance under different target SNRs is shown in Fig. 3.11.

The minimum transmission power for different frequencies and transmission distances is plotted in Fig. 3.12. It can be seen that a higher power is needed with the increasing frequency and distance. For long-haul communication, it is better to choose a suitable operating frequency to reduce energy consumption.

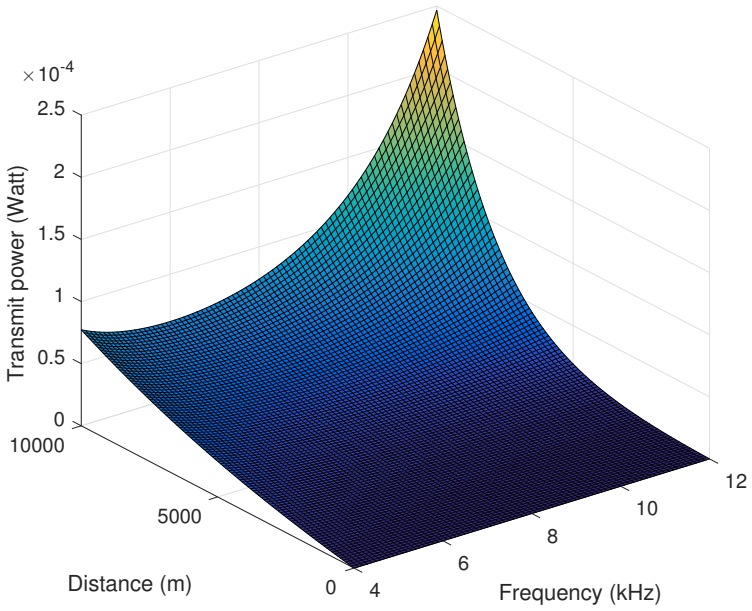

Figure 3.12: Minimum transmission power $P_{min}(l)$ for target SNR is equal to 20 dB.

3.2.4 Unit Conversion between Acoustic Power and Electric Power

The acoustic signal power is measured in Pascal (Pa) or micro Pascal (μPa). The TX will convert the electric energy in the circuit to acoustic energy and transmit acoustic waves outwards. The conversion from the electric power P^e in Watts to the acoustic power P^a in μPa and to the decibel form of the acoustic power in dB re μPa are given respectively by [9]

$$\begin{aligned} P_a &= P^e \cdot \phi \cdot 10^{17.15} \cdot DI, \\ 10\log_{10} P_a &= 10\log_{10} P^e + 10\log_{10}\phi + 171.5 + 10\log_{10} DI, \end{aligned} \tag{3.9}$$

where ϕ is the overall efficiency of the electric circuitry (power amplifier and transducer). Normally, $\phi < 1$ indicates that the electric power P^e fed into the projector

(transmit antenna) will have some degree of losses. DI is the transmitting directivity index of a projector and is the acoustic signal power difference from a non-directional projector which radiates the same amount power. The DI for a non-directional (omni-directional) projector is equal to 0.

Note that the units of the parameters in calculating the SNR should be used either in acoustic or in electric domain, i.e.,

$$\gamma = \frac{P_e G}{P_{N,e}} = \frac{P_a G}{P_{N,a}} = \frac{P_a \cdot \phi \cdot 10^{17.15} \cdot DI}{A(l,f) P_{N,a} \cdot \phi \cdot 10^{17.15} \cdot DI}, \tag{3.10}$$

where G is the underwater channel gain. $P_{N,e}$ and $P_{N,a}$ are the underwater noise powers in electric and acoustic, respectively.

3.2.5 Multi-Path Propagation

In this subsection, the time-invariant multi-path propagation and the time-varying multi-path propagation are introduced, respectively.

Time-Invariant Multi-Path Propagation

The TX sends the acoustical signals to the RX. These acoustic rays will experience reflection and refraction, and then, the RX will receive the superimposed *multi-path* acoustic rays, as shown in Fig. 3.13. In particular, reflections usually happen at the sea surface or bottom. In deep water, the acoustic rays will refract because of the non-homogeneous acoustic speed.

The delay spread D represents the maximum delay difference of the propagation

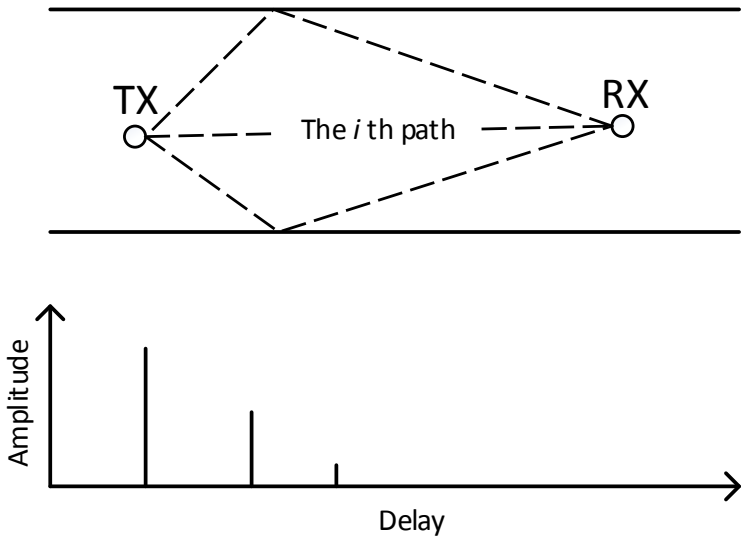

Figure 3.13: Time-invariant multi-path propagation and the delay profile.

paths in the time-of-arrival model,

$$D = \max\left\{\tau_i - \tau_j\right\}, \quad \forall\, i, j, \tag{3.11}$$

where $\tau_i = \frac{l_i}{c}$ is the propagation delay of the ith path, and c and l represent the acoustic speed and propagation length, respectively. The acoustic speed, e.g., 1500 m/s, causes a very large delay spread. For instance, the time difference would be 10 ms for two acoustic rays which differ 15 meters in path length.

Time-Varying Multi-Path Propagation

The time-variability of the propagated paths is one of the most challenging characteristics for UWAC. This is due to the relative motion between TX and RX or the motion of the surface waves generated by the wind.

The Doppler rate v_i is defined as the change rate of the propagation length for the ith path, and the Doppler spread D_d is the maximum Doppler rate difference among the propagated paths,

$$D_d = \max\left\{\frac{v_i - v_j}{c}\right\}, \quad \forall\, i, j. \tag{3.12}$$

The Doppler spread results in a frequency shift in the received signal, and thus, causes interference among different components of the signal. The Doppler frequency shift at the system center frequency f_c is $f_d = \frac{v}{c} f_c$.

Statistical Channel Model

There is no consensus on a standardized statistical underwater channel model for fading, and a number of experiments estimated the statistical channel performance based on particular experiment locations [6].

There are a number of underwater channel experiments from which a statistical fading model was obtained [10–12]. Chitre et. al. [10] modeled the fading of each acoustic ray through the Rayleigh distribution. K-distribution fading was justified in [11] for the envelope amplitude statistics. The authors in [12] measured the short-term path gains, and findings indicated that they indicate a conditional Ricean distribution with Bessel-type autocorrelation.

Bibliography

[1] S. Zhou and Z. Wang, *OFDM for Underwater Acoustic Communications*. John Wiley & Sons, 2014.

[2] L. Lanbo, Z. Shengli, and C. Jun-Hong, "Prospects and problems of wireless communication for underwater sensor networks," *Wireless Communications and Mobile Computing*, vol. 8, no. 8, pp. 977–994, Jul. 2008.

[3] K. V. Mackenzie, "Nine-term equation for sound speed in the oceans," *The Journal of the Acoustical Society of America*, vol. 70, no. 3, pp. 807–812, Sep. 1981.

[4] M. C. Domingo, "Overview of channel models for underwater wireless communication networks," *Physical Communication*, vol. 1, no. 3, pp. 163–182, Sep. 2008.

[5] M. B. Porter, "The bellhop manual and users guide: Preliminary draft," 2011.

[6] S. Milica and P. James, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[7] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[8] Y. Li, Y. Zhang, H. Zhou, and T. Jiang, "To relay or not to relay: Open distance and optimal deployment for linear underwater acoustic networks," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3797–3808, Sep. 2018.

[9] R. J. Urick, *Principles of Underwater Sound*. New York, US: McGraw-Hill Press, 1983.

[10] M. Chitre, "A high-frequency warm shallow water acoustic communications channel model and measurements," *J. Acoust. Soc. Amer.*, vol. 122, no. 5, pp. 2580–2586, Nov. 2007.

[11] W. B. Yang and T. C. Yang, "High-frequency channel characterization for m-ary frequency-shift-keying underwater acoustic communications," *J. Acoust. Soc. Amer.*, vol. 120, no. 5, pp. 2615–2626, Oct. 2006.

[12] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Oceanic Eng.*, vol. 38, no. 4, pp. 701–717, Oct. 2013.

Chapter 4

Optimal Power Allocation for Full-Duplex Underwater Relay Networks with Energy Harvesting: A Reinforcement Learning Approach

4.1 Abstract

In this chapter,[1] we study the optimal power allocation problem where the goal is to maximize the long-term end-to-end sum rate of an underwater full-duplex energy harvesting relay network. The problem is formulated as an online sequential decision-making problem, and a reinforcement learning algorithm is used to solve it. Simulation

[1]Part of this chapter has been published in *IEEE Wireless Communications Letter* [1].

results show that the optimal online power allocation policy achieves a higher sum rate than the computationally-efficient sub-optimal online greedy power allocation policy, especially under insufficient harvested energy. Besides, we also investigate the system performance for different relay positions in a single-relay network and observe that the highest sum rate is obtained when the relay is placed at the mid-point of the link.

4.2 Introduction

The demand for oceanic environment monitoring, disaster surveillance, and business applications has propelled the growth of the underwater acoustic communication (UWAC) market. The use of acoustic waves, in comparison with radio and light waves, is the only known effective means for long-haul underwater communication. Due to the limited operational bandwidth and the slow propagation speed of acoustic waves, communication suffers from low data rates and high transmission delays, making it challenging to provide high quality-of-service (QoS). Moreover, underwater devices are usually powered by batteries. Owing to the high maintenance costs and the harsh oceanic environment, it is infeasible to replace these devices regularly, so that, they are not sustainable and reliable for long-term underwater applications [2]. To support sustainability and reliability, the emerging energy harvesting (EH) devices, which use harvested energy to power the communication system, have become promising for future UWAC [3].

For a long-term communication system, it is essential to design a transmission policy to achieve a specific system performance goal. Such an adaptive policy can be derived using reinforcement learning (RL) techniques. One can develop an optimal

policy with the long-term goal of maximizing the cumulative reward [4]. For instance, a Q-learning based distributed routing protocol was proposed in [5] to prolong the lifetime of underwater sensor networks. RL was also applied to build an anti-jamming transmission framework that controls the transmit power and uses transducer mobility [6].

Recent work has been done for RL-based long-term adaptive transmission at the physical layer of UWAC [7–9]. In [7], power allocation was investigated to maximize the transmission throughput for point-to-point communication. In [8], data forwarding was studied through a model-based RL approach, aiming to minimize the system cost. Further, in [9], the Dyna-Q algorithm was explored to achieve maximum communication link throughput by adapting the modulation order. However, none of these works studied the performance of relay networks in UWAC.

In this chapter, we investigate an underwater relay network where the relay operates in full-duplex (FD) amplify-and-forward (AF) mode and harvests energy from the oceanic environment, such as benthic sources [10]. We formulate a long-term end-to-end sum rate maximization problem and solve it through RL. The problem is described as an agent interacting with its environment to maximize the cumulative reward on the long-run. Optimal and sub-optimal online power allocation policies are introduced, showing the importance of proper allocation of limited resources to achieve the desired goal of maximizing the sum rate. In addition, it is revealed that a highest sum rate is obtained when the relay is placed at the mid-point of the single-relay network.

The rest of the chapter is organized as follows. Section 4.3 describes the system model. Section 4.4 introduces the problem formulation and solutions. Section 4.5 presents the numerical results and discussion. Finally, Section 4.6 concludes the

chapter.

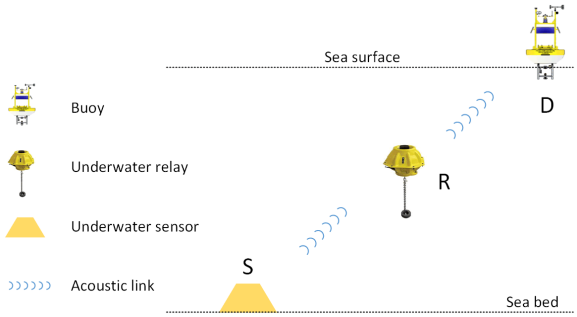

Figure 4.1: Underwater relay network.

4.3 System Model

We consider an underwater relay network consisting of a sensor (S), an EH FD relay (R), and a buoy (D), as shown in Fig. 4.1. The sensor sends the information to the buoy via the relay. The sensor and the buoy have a fixed power supply, whereas the relay node relies on harvested energy from the ambient environment to communicate with the buoy. The FD relay is equipped with a self-interference cancellation (SIC) unit. Since SIC is not perfect, we assume that there is residual self-interference (RSI) in the system [11, 12]. Also, we consider a discrete-time data transmission model, where data is transmitted in a slot of duration T.

4.3.1 Signal Model

The signal received at the relay in the nth time slot is given by

$$y_{\mathrm{R},n} = \sqrt{G_{\mathrm{SR},n}}\sqrt{P_{\mathrm{S}}}x_{\mathrm{S},n} + i_n + w_{\mathrm{R}}, \tag{4.1}$$

where $G_{\mathrm{SR},n}$ is the underwater channel gain for the source-to-relay (SR) link, P_{S} and $x_{\mathrm{S},n}$ are the transmit power and the information symbol with average unit energy at the source, respectively, i_n is the RSI, and w_{R} denotes the additive underwater noise at the relay. The underwater channel gain and noise are discussed in the subsequent subsections.

Hence, the signal-to-interference-plus-noise-ratio (SINR) at the relay can be expressed as

$$\gamma_{\mathrm{SR},n} = \frac{G_{\mathrm{SR},n} P_{\mathrm{S}}}{\sigma_{i,n}^2 + \sigma_{\mathrm{R}}^2}, \tag{4.2}$$

where σ_{R}^2 is the noise power and $\sigma_{i,n}^2 = \beta P_{\mathrm{R},n}^{\lambda}$ is the RSI power with parameters β and λ [11]. The smaller the values of β and λ are, the better the SIC performance is.

The relay node forwards the received signal to the buoy according to the AF protocol [11]. The signal received at the destination buoy in the nth time slot is given by

$$y_{\mathrm{D},n} = \sqrt{G_{\mathrm{RD},n}}\sqrt{P_{\mathrm{R},n}} x_{\mathrm{R},n} + w_{\mathrm{D}}, \tag{4.3}$$

where $G_{\mathrm{RD},n}$ is the underwater channel gain for the relay-to-destination (RD) link, $P_{\mathrm{R},n}$ and $x_{\mathrm{R},n} = \alpha y_{\mathrm{R},n-1}$ are the transmit power and the information symbol with average unit energy at the relay, respectively, with α being the amplification coefficient, and w_{D} is the additive underwater noise at the buoy. The corresponding signal-to-noise-ratio (SNR) is expressed as

$$\gamma_{\mathrm{RD},n} = \frac{G_{\mathrm{RD},n} P_{\mathrm{R},n}}{\sigma_{\mathrm{D}}^2}, \tag{4.4}$$

where σ_{D}^2 is the noise power at the buoy.

To obtain the end-to-end SINR expression, we substitute $x_{\mathrm{R},n} = \alpha y_{\mathrm{R},n-1}$ in (4.3),

which yields

$$\begin{aligned} y_{\mathrm{D},n} =& \sqrt{G_{\mathrm{RD},n}}\sqrt{P_{\mathrm{R},n}}\alpha \\ & \times (\sqrt{G_{\mathrm{SR},n-1}}\sqrt{P_{\mathrm{S}}}x_{\mathrm{S},n-1} + i_{n-1} + w_{\mathrm{R}}) + w_{\mathrm{D}}. \end{aligned} \tag{4.5}$$

Therefore, the end-to-end SINR in the nth time slot can be calculated as

$$\begin{aligned} \gamma_n &= \frac{G_{\mathrm{SR},n-1}P_{\mathrm{R},n}\alpha^2 G_{\mathrm{RD},n}P_{\mathrm{S}}}{G_{\mathrm{RD},n}P_{\mathrm{R},n}\alpha^2(\sigma_{i,n-1}^2 + \sigma_{\mathrm{R}}^2) + \sigma_{\mathrm{D}}^2} \\ &= \frac{\gamma_{\mathrm{SR},n-1}\gamma_{\mathrm{RD},n}}{\gamma_{\mathrm{SR},n-1} + \gamma_{\mathrm{RD},n} + 1}. \end{aligned} \tag{4.6}$$

Further, the throughput C_n (in bits/sec) in the nth time slot is

$$C_n = B\log_2\left(1+\gamma_n\right), \tag{4.7}$$

where B is the system bandwidth.

4.3.2 Channel Model

There is no standardized underwater channel model for fading [13]. In this chapter, we choose the model proposed in [14], which captures both large and small scale fading. Accordingly, the instantaneous channel gain is expressed as

$$G = \frac{1}{B}\int_{f_o}^{f_o+B} |\bar{H}_0(f)\sum_l h_l\tilde{\gamma}_l(f,t)e^{-j2\pi f\tau_l}|^2 df, \tag{4.8}$$

where f_o is the minimum operational frequency, $\bar{H}_0$ represents the channel filtering effect, and h_l and τ_l are large-scale parameters of the lth path. The small-scale fading effect is represented by $\tilde{\gamma}_l(f,t)e^{-j2\pi a_l f\tau_l}$, with $\tilde{\gamma}_l$ being the small-scale fading coefficient

and a_l being the Doppler scaling factor on the lth path.

4.3.3 Underwater Ambient Noise

Typically, the underwater ambient noise is due to four sources: turbulence, shipping, waves, and thermal noise. The corresponding empirical power spectral densities (PSDs) of these four components in μPa per kHz are given respectively by [15]

$$\begin{aligned} N_t(f) &= 10^{((17-30\log_{10}(f))/10)}, \\ N_s(f) &= 10^{((40+20(s-0.5)+26\log_{10}(f)-60\log_{10}(f+0.03))/10)}, \\ N_w(f) &= 10^{((50+7.5w_s^{0.5}+20\log_{10}(f)-40\log_{10}(f+0.4))/10)}, \\ N_{th}(f) &= 10^{((-15+20\log_{10}(f))/10)}, \end{aligned} \tag{4.9}$$

where s is the shipping activity factor and w_s is the wind speed. The overall PSD of the ambient noise is $N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f)$, which is the underwater noise power P_N in dB re μPa [1] and is calculated as

$$P_N = 10\log_{10}\left(\int_{f_o}^{f_o+B} N(f)df\right). \tag{4.10}$$

The conversion from acoustic power P^a in dB re μPa to electrical power P^e in Watts is given by [16][2]

$$P^e = 10^{\frac{P_a - 10\log_{10}\phi - 171.5}{10}}, \tag{4.11}$$

where ϕ is the overall efficiency of the electric circuitry (power amplifier and transducer).

[1] Unit for sound pressure level (SPL), which is the logarithmic ratio of acoustic pressure A to a reference pressure $A_{\text{ref}} = 1$ μPa, multiplied by 10, i.e., SPL$= 10\log_{10}\frac{A}{A_{\text{ref}}}$.

[2] Here DI in equation (3.9) is considered 1.

4.3.4 Energy Harvesting Model

The relay node harvests energy periodically and stores it in the rechargeable battery with capacity $B_{\max}$. We model the EH process as a stationary, temporally independent and identically distributed Bernoulli process [17]. That is, a node harvests the energy $E^h_{R,n}$ in the nth time slot with probability (w.p.) p and does not with probability $1-p$. Moreover, the battery's energy level in the nth time slot is represented by B_n, which is calculated as follows

$$B_n = \begin{cases} \min\left\{B_{n-1} - E_{\mathrm{R},n-1} + E^h_{\mathrm{R},n-1}, B_{\max}\right\}, \text{ w.p. } p, \\ B_{n-1} - E_{\mathrm{R},n-1}, \qquad\qquad\qquad\qquad \text{w.p. } 1-p, \end{cases} \tag{4.12}$$

where $E_{\mathrm{R},n} = P_{\mathrm{R},n}T$ is the energy consumption of the relay for transmitting in the nth time slot. Further, we assume that the energy consumption for signal processing and receiving at the relay is negligible.

4.4 Problem Formulation and Solutions

We aim to maximize the throughput over N time slots, and formulate the optimization problem as follows:

$$\underset{P_{\mathrm{R},n}}{\text{maximize}} \quad \sum_{n=1}^{N} C_n \tag{4.13a}$$

$$\text{subject to} \quad 0 \le P_{\mathrm{R},n} \le P_{\mathrm{R,max}}\,, \forall\, n = 1, \ldots, N\,, \tag{4.13b}$$

where $P_{\mathrm{R,max}}$ is the maximum transmit power of the relay.

The solution to (4.13) is an optimal power allocation policy that maximizes the

sum rate at the end of the time slot N. Thus, we cast the problem as an online sequential decision-making problem and apply the RL approach to obtain the optimal policy. RL, a machine learning technique, is essentially a Markov decision process (MDP) that defines the interaction between an agent and its environment in terms of states, actions, and rewards [18]. Further, based on observation, problem (4.13) is a finite-horizon discrete time MDP problem. The MDP model corresponding to problem (4.13) is discussed in the following.

4.4.1 Markov Decision Process (MDP) Model

The MDP model consists of decision epochs, states, actions, transition probabilities, and rewards [18]. Each of these elements is presented below.

Decision Epochs: We consider discrete and finite decision epochs (finite-horizon), in which the decision is made at the beginning of the time slot. Let $\mathcal{T} = \{1, 2, \ldots, N\}$ be the set of decision epochs.

States: The relay is characterized by a state during each decision epoch. The state space of the relay, $\mathcal{S}$, is given by

$$\mathcal{S} = \mathcal{B}_{\mathrm{R}} \times \mathcal{G}_{\mathrm{SR}} \times \mathcal{G}_{\mathrm{RD}} \times \mathcal{P}_{\mathrm{R}}, \tag{4.14}$$

where $\mathcal{B}_{\mathrm{R}} = \{0, \frac{B_{\max}}{l}, \ldots, B_{\max}\}$ is the set of battery levels, with $l+1$ being the number of battery levels. $\mathcal{G}_{\mathrm{SR}} = \{g^1_{\mathrm{SR}}, g^2_{\mathrm{SR}}, \ldots, g^m_{\mathrm{SR}}\}$ and $\mathcal{G}_{\mathrm{RD}} = \{g^1_{\mathrm{RD}}, g^2_{\mathrm{RD}}, \ldots, g^m_{\mathrm{RD}}\}$ are the sets of channel states of the SR and RD links, respectively, with m being the number of channel states. $\mathcal{P}_{\mathrm{R}} = \{0, P_1, \ldots, P_k\}$ is the set of transmit power levels, with $k+1$ being the number of transmit power levels.

In the nth time slot, the state of the relay $\mathbf{s}_n \in \mathcal{S}$ can be expressed as

$$\mathbf{s}_n = \{B_{\mathrm{R}}(n), G_{\mathrm{SR}}(n-1), G_{\mathrm{RD}}(n), P_{\mathrm{R}}(n-1)\}, \tag{4.15}$$

where $B_{\mathrm{R}}(n)$ and $P_{\mathrm{R}}(n-1)$ are the battery level and transmit power of the relay in the nth and $(n-1)$th time slot, respectively, while $G_{\mathrm{SR}}(n-1)$ and $G_{\mathrm{RD}}(n)$ represent the channel states of the SR and RD links, respectively.

Actions: For a given state, the relay selects an action from the action set, which is described as $\mathcal{A} = \{0, P_1, \ldots, P_k\}$. Moreover, $\mathbf{a}_n \in \mathcal{A}$ stands for the action in the nth time slot.

Transition Probabilities: The transition probability $\mathbf{P}(\mathbf{s}_{n+1} \mid \mathbf{s}_n, \mathbf{a}_n)$ is expressed in (4.16) below. This represents the probability of going to state $\mathbf{s}_{n+1}$ from $\mathbf{s}_n$ after taking an action $\mathbf{a}_n$. $\mathbf{P}(B_{\mathrm{R}}(n+1) \mid B_{\mathrm{R}}(n), P_{\mathrm{R}}(n))$ is the relay's battery transition probability. $\mathbf{P}(G_{\mathrm{SR}}(n) \mid G_{\mathrm{SR}}(n-1))$ and $\mathbf{P}(G_{\mathrm{RD}}(n+1) \mid G_{\mathrm{RD}}(n))$ are the transition probabilities of the channels SR and RD, respectively.

$$\begin{aligned}\mathbf{P}(\mathbf{s}_{n+1} \mid \mathbf{s}_n, \mathbf{a}_n) &= \mathbf{P}(B_{\mathrm{R}}(n+1), G_{\mathrm{SR}}(n), G_{\mathrm{RD}}(n+1), P_{\mathrm{R}}(n) \mid B_{\mathrm{R}}(n), G_{\mathrm{SR}}(n-1), G_{\mathrm{RD}}(n), P_{\mathrm{R}}(n-1), P_{\mathrm{R}}(n)) \\ &= \mathbf{P}(B_{\mathrm{R}}(n+1) \mid B_{\mathrm{R}}(n), P_{\mathrm{R}}(n)) \times \mathbf{P}(G_{\mathrm{SR}}(n) \mid G_{\mathrm{SR}}(n-1)) \times \mathbf{P}(G_{\mathrm{RD}}(n+1) \mid G_{\mathrm{RD}}(n))\end{aligned} \tag{4.16}$$

Rewards: After taking an action $\mathbf{a}_n$ in state $\mathbf{s}_n$, the relay receives a reward $\mathbf{R}_n(\mathbf{s}_n, \mathbf{a}_n)$, which is the same as (4.7)

$$\mathbf{R}_n(\mathbf{s}_n, \mathbf{a}_n) = B \log_2(1 + \gamma_n), \quad \forall \quad n = 1, 2, \ldots, N. \tag{4.17}$$

The decision rule is a function $\mathbf{d}_n(\mathbf{s}_n)$: $\mathcal{S} \to \mathcal{A}$, which specifies the action selection

when the system state is $\mathbf{s}_n$. Moreover, a policy $\pi = \{\mathbf{d}_1(\mathbf{s}_1), \mathbf{d}_2(\mathbf{s}_2), \ldots, \mathbf{d}_N(\mathbf{s}_N)\}$ is a sequence of decision rules. The set of all policies is denoted by Π.

Let $\mathbf{v}_N^{\pi}(\mathbf{s}_1)$ denote the expected total reward over N decision epochs, if the policy π is adopted and the beginning state of the relay is $\mathbf{s}_1$. The expected total reward $\mathbf{v}_N^{\pi}(\mathbf{s}_1)$ is

$$\mathbf{v}_N^{\pi}(\mathbf{s}_1) = \mathbb{E}^{\pi}\left\{\sum_{i=1}^{N} \mathbf{R}_i(\mathbf{s}_i, \mathbf{a}_i)\right\}, \tag{4.18}$$

where $\mathbb{E}^{\pi}\{\cdot\}$ denotes the statistical expectation, given that policy π is used. Equation (4.18) can be solved by the backward induction algorithm,[3] as shown in Algorithm 2 [18]. In this algorithm, Equations (4.19) and (4.21) provide the maximum expected reward from the ith decision epoch under state $\mathbf{s}_i$ to the last decision epoch.

Our goal is to seek an optimal policy $\pi^* = \{\mathbf{d}_1^*(\mathbf{s}_1), \mathbf{d}_2^*(\mathbf{s}_2), \ldots, \mathbf{d}_N^*(\mathbf{s}_N)\}$, which can be obtained through (4.20) and (4.22), that maximizes the expected cumulative reward.

4.4.2 Proposed Solutions

The optimal policy is introduced first and followed by a computationally-efficient sub-optimal policy.

Optimal Online Power Allocation Policy: In this policy, according to the system state $\mathbf{s}_n$, the relay chooses the transmit power $P_{\mathrm{R},n}$ by applying the optimal policy π^* at the beginning of each time slot. Therefore, the power allocation is

$$P_{\mathrm{R},n} = \mathbf{d}_n^*(\mathbf{s}_n), \qquad \forall \quad n = 1, 2, \ldots, N. \tag{4.23}$$

[3]The backward induction algorithm provides an efficient method for solving finite-horizon discrete time MDPs [18].

Algorithm 2: The Backward Induction Algorithm

1 Set $i = N$ and compute $\mathbf{u}_i^*(\mathbf{s}_i)$ and $\mathbf{d}_i^*(\mathbf{s}_i)$, for $\forall \mathbf{s}_i \in \mathcal{S}$ by

$$\mathbf{u}_i^*(\mathbf{s}_i) = \max_{\mathbf{a}_i \in \mathcal{A}} \left[\mathbf{R}_i(\mathbf{s}_i, \mathbf{a}_i)\right], \tag{4.19}$$

Set

$$\mathbf{d}_i^*(\mathbf{s}_i) = \arg\max_{\mathbf{a}_i \in \mathcal{A}} \left[\mathbf{R}_i(\mathbf{s}_i, \mathbf{a}_i)\right] \tag{4.20}$$

2 Set $i = i - 1$ and compute $\mathbf{u}_i^*(\mathbf{s}_i)$ and $\mathbf{d}_i^*(\mathbf{s}_i)$, for $\forall \mathbf{s}_i \in \mathcal{S}$ by

$$\mathbf{u}_i^*(\mathbf{s}_i) = \max_{\mathbf{a}_i \in \mathcal{A}} \left[\mathbf{R}_i(\mathbf{s}_i, \mathbf{a}_i) + \sum_{\mathbf{s}_{i+1} \in \mathcal{S}} \mathbf{P}(\mathbf{s}_{i+1} \mid \mathbf{s}_i, \mathbf{a}_i)\, \mathbf{u}_{i+1}^*(\mathbf{s}_{i+1})\right] \tag{4.21}$$

Set

$$\mathbf{d}_i^*(\mathbf{s}_i) = \arg\max_{\mathbf{a}_i \in \mathcal{A}} \left[\mathbf{R}_i(\mathbf{s}_i, \mathbf{a}_i) + \sum_{\mathbf{s}_{i+1} \in \mathcal{S}} \mathbf{P}(\mathbf{s}_{i+1} \mid \mathbf{s}_i, \mathbf{a}_i)\, \mathbf{u}_i^*(\mathbf{s}_{i+1})\right] \tag{4.22}$$

3 If $i = 1$, stop. Otherwise return to step 2.

The optimal policy π^* is stored at the relay prior to transmission, and the complexity of Algorithm 1 is $O(N|\mathcal{S}||\mathcal{A}|)$, where $|\cdot|$ denotes the cardinality of the set.

Sub-optimal Online Greedy Power Allocation Policy: At the beginning of each time slot, the relay chooses the transmit power $P_{\mathrm{R},n}$ to maximize the current reward. Thus, we turn to a greedy power allocation from (4.17) as

$$P_{\mathrm{R},n} = \arg\max_{\mathbf{a}_n \in \mathcal{A}} \mathbf{R}_n(\mathbf{s}_n, \mathbf{a}_n), \qquad \forall \quad n = 1, 2, \ldots, N. \tag{4.24}$$

As compared to the optimal policy, this policy has a lower computational complexity by avoiding the computation of the expected future rewards; therefore, the complexity of this policy is $O(N|\mathcal{A}|)$.

4.5 The Procedure to Solve the Problem (4.13)

As mentioned in Chapter 2.5.3, there are two steps to solve the RL problem: training the model and deploying the model. In the following, we illustrate the procedure to solve the problem in equation (4.13).

- Training phase: In this phase, we firstly build the MDP model according to the formulated problem (4.13), as presented in Section 4.4.1. Secondly, we run the backward induction algorithm based on the formulated MDP model, as illustrated in Algorithm 2. Finally, we obtain the policy π^* from the algorithm.
- Deploying phase: In this phase, the relay (agent) deploys the obtained policy and then uses the policy to allocate the transmit power according to the relay's system state.

Specifically, we explain how to define the channel states and calculate the transition probabilities as follows:

Define channel states: we set the number of channel states to two. Accordingly, a single threshold is applied to classify the channel gains into two states. The threshold is set to the mean value of channel gains, which are generated from the acoustic channel simulator made by the researchers at Northeastern University in U.S.A. [14, 19].

Calculate transition probabilities of channel states: we calculate the transition probabilities based on references [20, 21]. The simulation parameters for calculating the transition probabilities of channel states are listed in Table 4.1. Let there be K channel states in the Markov model, and assume that a transition happens between adjacent states only. The transition probability $t_{i,j}$ from state i to state j in the finite-state Markov chain can be approximated as

$$t_{k,k+1} \approx \frac{N(A_{k+1})}{R_t^{(k)}}, \quad \forall\, k = 1, 2, \ldots, K-1, \tag{4.25}$$

Table 4.1: Simulation parameters.

Parameters	**Value**
R_t	$\frac{1}{104.86\times10^{-3}}$
f_m	0.01
Transition probability of channel SR and RD	$\begin{bmatrix} 0.9998 & 0.0002 \\ 0.0002 & 0.9998 \end{bmatrix}$

$$t_{k,k-1} \approx \frac{N(A_k)}{R_t^{(k)}}, \quad \forall\, k = 2, 3, \ldots, K, \tag{4.26}$$

where $N(A_k)$ is the level crossing rate at a specific received instantaneous SNR level A_k. $R_t^{(k)} = R_t \times p_k$ is the average number of symbols transmitted per second during which the received SNR is in state k for a symbol rate R_t and a steady-state probability of state k p_k. Further, the level crossing rate is calculated as

$$N(A_k) = f_m \int_0^\infty \dot{y} p(A_k, \dot{y}) d\dot{y}, \tag{4.27}$$

where the dot indicates the time derivative of the received SNR, $p(A_k, \dot{y})$ is the joint density function of A_k and $\dot{y}$, and f_m is the Doppler frequency. Moreover, the probability density function of the received SNR is given by

$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}, \tag{4.28}$$

where μ and σ are the mean and variance of the normal distribution [14].

4.6 Numerical Results and Discussion

In this section, we evaluate the performance of the optimal online and sub-optimal online greedy power allocation policies for the considered underwater system. The

values of simulation parameters are listed in Table 4.2.

Sum rate versus EH rate

Figure 4.2 shows the sum rate versus the EH rate for different values of λ. We observe that the sum rate increases with the EH rate. This can be easily explained, as more energy can be used at the relay to forward data to the buoy. In addition, as expected, we can see that the sum rate of the optimal online policy outperforms that of the sub-optimal online greedy policy. The gap between the solutions of the two policies decreases in the higher EH rate region, where there are abundant energy resources at the relay in all time slots. Thus, the local optimum approaches the global one. Meanwhile, the sum rate of the FD mode at $\lambda = 0$ is better than that at $\lambda = 0.35$. This is because the RSI is directly proportional to λ. Moreover, by comparing the FD and half-duplex (HD) operating modes, we can see that FD is significantly better than HD, especially when λ is low.

Sum rate versus relay position

Figure 4.3 illustrates the sum rate versus the relay position in the single-relay network. We can see that the location of the relay is crucial in an underwater network, as it determines the throughput performance. Moreover, as expected, the highest sum rate is achieved when the relay is placed at the mid-point of the link. Also, the optimal online policy outperforms the corresponding sub-optimal online greedy policy. Furthermore, the FD performance is better than that of HD.

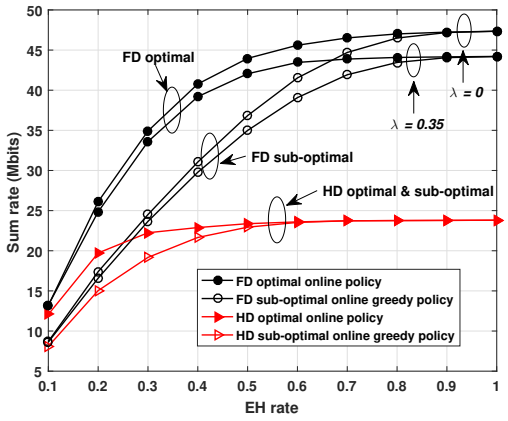

Figure 4.2: Sum rate vs. EH rate for SR distance equal to 5 km.

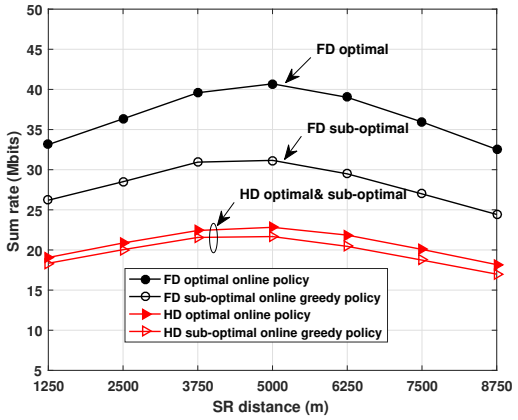

Figure 4.3: Sum rate vs. relay position for $p = 0.4$ and $\lambda = 0$.

Table 4.2: Simulation parameters.

Parameters	**Value**
Height of S, R, D and sea surface (m)	0, 50, 100, and 100
SD distance (km)	10
f_o and B (kHz)	15.5 and 15
P_{S} and P_{R} (W)	1 and $\{0, 1, 2\}$
ϕ, β and λ [11]	1, 1 and $\{0, 0.35\}$
p, $B_{\max}$ and E_R^h (Joule)	$[0, 1]$, 5 and 3
l, m, and k	5, 2, and 2
B_{R} (1), P_{R} (0)	0
w_s (m/s) and s [15]	10 and 0.5
T (s) and N	1 and 180

4.7 Conclusion

In this chapter, we investigated the optimal transmission policy of a long-term sum rate maximization problem for an underwater FD EH relay network. We formulated and solved the optimization problem through the RL approach. Simulation results revealed that the optimal online power allocation policy outperforms the computationally-efficient sub-optimal online greedy one, especially when the harvested energy is scarce. Moreover, as expected, the result showed that the highest sum rate is achieved when the relay is placed at the mid-point of the single-relay network.

Bibliography

[1] R. Wang, A. Yadav, E. A. Makled, O. A. Dobre, R. Zhao, and P. K. Varshney, "Optimal power allocation for full-duplex underwater relay networks with energy harvesting: A reinforcement learning approach," *Accepted by IEEE Wireless Communications Letters*, Oct. 2019.

[2] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad-hoc Networks*, vol. 3, no. 3, pp. 257–279, Mar. 2005.

[3] M. Ku, W. Li, Y. Chen, and K. J. R. Liu, "Advances in energy harvesting communications: Past, present, and future challenges," *IEEE Commun. Surveys Tutorials*, vol. 18, no. 2, pp. 1384–1412, 2nd. Quart. 2016.

[4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[5] T. Hu and Y. Fei, "Qelar: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 6, pp. 796–809, Jun. 2010.

[6] L. Xiao, Donghua, Jiang, X. Wan, W. Su, and Y. Tang, "Anti-jamming underwater transmission with mobility and learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 542–545, Mar. 2018.

[7] L. Jing, C. He, J. Huang, and Z. Ding, "Energy management and power allocation for underwater acoustic sensor network," *IEEE Sensors J.*, vol. 17, no. 19, pp. 6451–6462, Oct. 2017.

[8] C. Wang, Z. Wang, W. Sun, and D. R. Fuhrmann, "Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels," *IEEE Access*, vol. 6, pp. 2541–2558, Dec. 2017.

[9] Q. Fu and A. Song, "Adaptive modulation for underwater acoustic communications based on reinforcement learning," in *Proc. IEEE/MTS OCEANS*, Charleston, NC, USA, Oct. 2018, pp. 1–8.

[10] P. R. Bandyopadhyay, D. P. Thivierge, F. M. McNeilly, and A. Fredette, "An electronic circuit for trickle charge harvesting from littoral microbial fuel cells," *IEEE J. Oceanic Eng.*, vol. 38, no. 1, pp. 32–42, Jan. 2013.

[11] L. J. Rodrguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex af relaying in the presence of residual self-interference," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[12] E. A. Makled, A. Yadav, O. A. Dobre, and R. D. Haynes, "Hierarchical full-duplex underwater acoustic network: A noma approach," in *Proc. IEEE/MTS OCEANS*, Charleston, NC, USA, Oct. 2018, pp. 1–6.

[13] S. Milica and P. James, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[14] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Oceanic Eng.*, vol. 38, no. 4, pp. 701–717, Oct. 2013.

[15] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[16] R. J. Urick, *Principles of Underwater Sound*. New York, US: McGraw-Hill Press, 1983.

[17] A. Yadav, M. Goonewardena, W. Ajib, O. A. Dobre, and H. Elbiaze, "Energy management for energy harvesting wireless sensors with adaptive retransmission," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5487–5498, Dec. 2017.

[18] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. Press, 1994.

[19] "Acoustic Channel Simulator," https://http://millitsa.coe.neu.edu/projects/chann_sim/chann_sim_info.pdf.

[20] Qinqing Zhang and S. A. Kassam, "Finite-state markov model for rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[21] V. Bhaskar, "Finite-state markov model for lognormal, chi-square (central), chi-square (non-central), and k -distributions," *International Journal of Wireless Information Networks*, vol. 14, pp. 237–250, Dec. 2007.

Chapter 5

Reinforcement Learning-based Energy-Efficient Transmission Policy for Full-Duplex Underwater Relay Networks with Energy Harvesting

5.1 Introduction

Oceanic applications,[1] such as environment monitoring and offshore oil and gas extraction, drive the development of underwater acoustic communication (UWAC).

[1]Part of this chapter has been submitted to *IEEE/MTS OCEANS* 2020 Singapore.

A major challenge of UWAC is the low throughput due to the severely limited operational bandwidth [1]. This limitation can be partially overcome by deploying full-duplex (FD) relays, which transmit and receive signals at the same frequency and time. Specifically, the FD relay networks can achieve higher throughput than the half-duplex (HD) ones if the self-interference (SI) power is reduced to the noise level by applying SI cancellation techniques [2, 3]. Moreover, underwater devices are power-limited as they are usually powered by batteries [1]. Research on underwater energy harvesting (EH) devices has shown that energy from the ambient environment can be harvested, which becomes more sustainable and reliable in long-term applications [4].

Recent works for long-term adaptive communication have been limited to point-to-point UWAC [5, 6]. Here, we propose a reinforcement learning (RL)-based transmission policy to maximize the long-term energy efficiency (EE) of a single-relay underwater network, where the relay has an EH unit and operates in FD mode.

The rest of the chapter is organized as follows: Section 5.2 presents the system model. Section 5.3 illustrates the problem formulation and solution, and Section 5.4 provides the preliminary numerical results. Finally, Section 5.5 concludes the chapter.

Notations: SR, RD, and SD stand for sensor-to-relay, relay-to-buoy, and sensor-to-buoy. $(\cdot)_i$ represents the i-th time slot. log is the base-10 logarithm operation.

5.2 System Model

We consider a single-relay underwater network, where the sensor (S) sends information to the buoy (D) via the FD EH relay (R), as shown in Figure 5.1. The underwater channel is characterized by both path-loss (PL) and small-scale Rayleigh fading, and is modeled as a finite-state Markov chain. Also, a time-slotted transmission model is

considered, with each time slot of duration T.

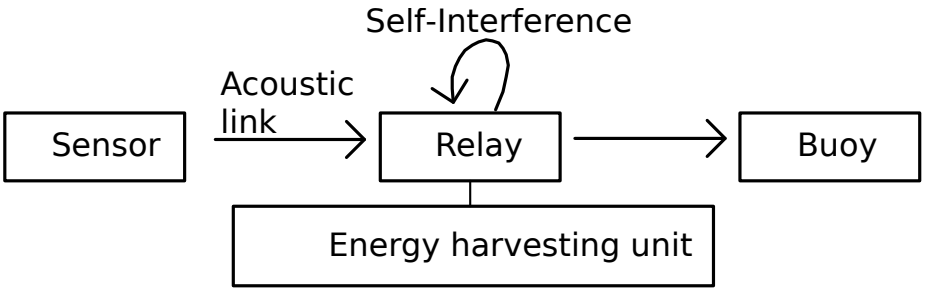

Figure 5.1: Underwater single-relay network.

In the following, underwater PL, ambient noise, signal model, and EH process are discussed.

Underwater PL: $A(l,f)$ is the PL at frequency f (in kHz) over a distance l (in km), which is given by $10\log A(l,f) = k10\log(1000l) + l10\log a(f)$, where k is the spreading factor, and $a(f)$ is the absorption coefficient expressed using the Thorp's formula (in dB/km) as $10\log a(f) = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 2.75\cdot 10^{-4}f^2 + 0.003$ [7].

Underwater Ambient Noise:[2] It is modeled as a stationary, temporally independent and identically distributed Bernoulli process[8]. That is, the battery with capacity $E^{\max}$ updates as $E_n = \min\left\{E_{n-1} - E^R_{n-1} + E^H, E^{\max}\right\}$ with probability p, and updates as $E_n = E_{n-1} - E^R_{n-1}$ with probability $1-p$, where $E^R_{n-1} = P^{\mathrm{R}}_{n-1}T$ and E^H are the consumed and the harvested energy of the relay, respectively.

Signal Model: The achievable signal-to-noise-plus-interference-ratio (SINR) at the relay at frequency f in the n-th time slot is

$$\gamma^{\mathrm{SR}}_n(f) = \frac{\left|h^{\mathrm{SR}}_n\right|^2 \left[A(l^{\mathrm{SR}},f)\right]^{-1} S_n(f)}{N(f) + I_n(f)}, \tag{5.1}$$

where h^{SR}_n is the small-scale Rayleigh fading coefficient, $S_n(f)$ denotes the PSD of

[2]The underwater ambient noise is the same as in Chapter 4.

the transmitted signal of the sensor, and $I_n(f) = b\left[R_n(f)\right]^{\lambda} B^{\lambda-1}$ is the PSD of the residual SI (RSI), with b and λ as the RSI parameters [3]. B and $R_n(f)$ are the system bandwidth and the PSD of the transmitted signal of the relay, respectively. Next, the achievable signal-to-noise-ratio (SNR) at the buoy is expressed as

$$\gamma_n^{\mathrm{RD}}(f) = \frac{\left|h_n^{\mathrm{RD}}\right|^2 \left[A(l^{\mathrm{RD}}, f)\right]^{-1} R_n(f)}{N(f)}. \tag{5.2}$$

Finally, the end-to-end SINR at the buoy when the relay applies the amplify-and-forward protocols [9] is

$$\gamma_n^{\mathrm{SD}}(f) = \frac{\gamma_{n-1}^{\mathrm{SR}}(f)\,\gamma_n^{\mathrm{RD}}(f)}{\gamma_{n-1}^{\mathrm{SR}}(f) + \gamma_n^{\mathrm{RD}}(f) + 1}. \tag{5.3}$$

Then, the EE in bits/J/Hz of this network can be calculated as

$$\rho_n = \frac{\int_{f^{\min}}^{f^{\min}+B} \log_2\left[1 + \gamma_n^{\mathrm{SD}}(f)\right]\, df}{B\left(P_{n-1}^{\mathrm{S}} + P_n^{\mathrm{R}}\right)}, \tag{5.4}$$

where $f^{\min}$ is the minimum operational frequency. P_{n-1}^{S} and P_n^{R} are the transmit power of the sensor and the relay, respectively.

EH Process:[3] It is modeled as a stationary, temporally independent and identically distributed Bernoulli process [8]. That is, the battery with capacity $E^{\max}$ updates as $E_n = \min\left\{E_{n-1} - E_{n-1}^R + E^H, E^{\max}\right\}$ with probability p, and updates as $E_n = E_{n-1} - E_{n-1}^R$ with probability $1-p$, where $E_{n-1}^R = P_{n-1}^{\mathrm{R}} T$ and E^H are the consumed and the harvested energy of the relay, respectively.

[3]The EH process is the same as in Chapter 4.

5.3 Problem Formulation and Solution

Our goal is to maximize the long-term EE of the single-relay network by optimizing the power allocation at the relay. We apply the RL framework to solve it. In this framework, the learning takes place as a result of the interaction between an agent and the environment. The whole process can be described by a Markov decision process (MDP). Next, we will build an MDP model of the problem, which consists of states, actions, rewards, and transitions probabilities [10].

The relay (i.e., the agent) holds a state $s_n = \left\{g_{n-1}^{\mathrm{SR}}, g_n^{\mathrm{RD}}, P_n^{\mathrm{R,RSI}}, B_n\right\}$, where g_{n-1}^{SR} and g_n^{RD} represent the channel states of the SR and the RD links, respectively. $P_n^{\mathrm{R,RSI}}$ is the RSI power and B_n is the battery level. Thus, the state space is $\mathcal{S} = \mathcal{G}^{\mathrm{SR}} \times \mathcal{G}^{\mathrm{RD}} \times \mathcal{P}^{\mathrm{R,RSI}} \times \mathcal{B}$, where $\mathcal{G}^{\mathrm{SR}}$ and $\mathcal{G}^{\mathrm{RD}}$ are the sets of channel states for the SR and the RD links, respectively. $\mathcal{P}^{\mathrm{R,RSI}}$ is the set of RSI powers and $\mathcal{B}$ is the set of battery levels. The relay chooses an action $a_n \in \mathcal{P}^{\mathrm{R}}$ under state s_n, where $\mathcal{P}^{\mathrm{R}}$ is the set of transmit powers. $r_n\left(s_n, a_n\right) = \rho_n$ denotes the reward under state s_n after action a_n is chosen. The state transition probability is $p\left(s_{n+1} \mid s_n, a_n\right) = p\left(g_n^{\mathrm{SR}} \mid g_{n-1}^{\mathrm{SR}}\right) p\left(g_{n+1}^{\mathrm{RD}} \mid g_n^{\mathrm{RD}}\right) p\left(B_{n+1} \mid B_n, P_n^{\mathrm{R}}\right)$, which represents the probability that state s_{n+1} will be occupied, when action a_n is chosen under state s_n.

The policy $\pi : \mathcal{S} \rightarrow \mathcal{P}^{\mathrm{R}}$ is a mapping from state s_n to action a_n.

We cast the problem as an infinite-horizon average reward MDP problem and the goal is to find an optimal policy π^* that maximizes the expected long-term average reward given a start state s_1,

$$J^{\pi^*}(s_1) = \lim_{h \rightarrow \infty} \frac{1}{h} \mathbb{E}^{\pi^*} \left\{ \sum_{i=1}^{h} r_i\left(s_i, a_i\right) \mid s_1 \right\}. \tag{5.5}$$

The stationary optimal policy π^* can be derived using the value iteration algorithm[4], as shown in Algorithm 3 [11]. The span semi-norm of a vector $\overrightarrow{x}$ denoted by $sp(\overrightarrow{x})$ in row 4 of the Algorithm 3, is defined as $sp(\overrightarrow{x}) = \max x - \min x$, where x is the element in $\overrightarrow{x}$.

The value-state function $v(s_n)$ defines how good it is for the agent to be in terms of the expected future rewards for a given state s_n,

$$v(s_n)=\max_{a_n\in\mathcal{A}}\left\{r_n(s_n,a_n)+\sum_{s_{n+1}\in\mathcal{S}}p(s_{n+1}\mid s_n,a_n)v(s_{n+1})\right\}. \tag{5.6}$$

Algorithm 3: The value iteration algorithm

Result: Optimal policy π^*.

1 **Input:** $v^0 = 0$, $i = 0$, $\epsilon = 0.01$;
2 **for** $\forall s_n \in \mathcal{S}$ **do**
3 Compute value-state function $v^{i+1}(s_n)$;
4 **if** $sp(v^{i+1} - v^i) < \epsilon$ **then**
5 **for** $\forall s_n \in \mathcal{S}$ **do**
6 $\pi^* = \arg\max_{a_n\in\mathcal{A}} v^{i+1}(s_n)$;
7 **end**
8 **else**
9 $i = i + 1$;
10 **end**
11 **end**

5.4 The Procedure to Solve the Problem (5.5)

[4]The value iteration algorithm provides a method for solving infinite-horizon average reward MDPs [11]. The backward induction algorithm cannot be applied to this problem, since it is the solution for finite-horizon MDPs.

The procedure to solve problem (5.5) is the same as explained in Section 4.5: training the model and deploying the model. In the following, we illustrate the procedure to solve the problem (5.5).

- Training phase: In this phase, we firstly build the MDP model according to the formulated problem (5.5), as presented in Section 5.3. Secondly, we run the value iteration algorithm based on the formulated MDP model, as shown in Algorithm 3. Finally, we obtain the policy π^* from the algorithm.
- Deploying phase: In this phase, the relay (agent) deploys the obtained policy and then uses the policy to allocate the transmit power according to the relay's system state.

Next, we explain how to define the channel states and calculate transition probabilities:

Define channel gain and channel states: The channel gain follows Rayleigh fading and the corresponding probability density function is given by

$$p_Y(y) = \frac{1}{\gamma_0} e^{\frac{\gamma}{\gamma_0}}, \tag{5.7}$$

where γ_0 is the average SNR.

We set the number of channel states to three. Accordingly, there exist two thresholds to divide the channel gains into three states, or equivalently, three intervals. The thresholds T_1 and T_2 are set respectively as

$$\int_0^{T_1} p_Y(y)dy = 0.2, \tag{5.8}$$

$$\int_{T_1}^{T_2} p_Y(y)dy = \int_{T_2}^{\infty} p_Y(y)dy = 0.4. \tag{5.9}$$

Additionally, we assume that three channel states correspond to three channel gains, ch_i for $i = 1, 2, 3$. The channel gains are derived as follows

$$\int_{0}^{ch_1} p_Y(y)dy = 0.1, \tag{5.10}$$

$$\int_{0}^{ch_2} p_Y(y)dy = 0.6, \tag{5.11}$$

$$\int_{0}^{ch_3} p_Y(y)dy = 0.8. \tag{5.12}$$

Calculate transition probabilities of channel states: we calculate the transition probabilities based on reference [12]. Let us consider K states in the Markov model, and that a transition occurs between adjacent states only. The transition probabilities in the finite-state Markov chain can be approximated respectively as [12]

$$t_{k,k+1} \approx \frac{N(A_{k+1})}{R_t^{(k)}}, \ \ \forall\, k = 1, 2, \ldots, K-1, \tag{5.13}$$

$$t_{k,k-1} \approx \frac{N(A_k)}{R_t^{(k)}}, \ \ \forall\, k = 2, 3, \ldots, K, \tag{5.14}$$

where $N(A_k)$ is the level crossing rate at a specific received instantaneous SNR level A_k, $R_t^{(k)} = R_t \times p_k$ is the average number of symbols transmitted per second during which the received SNR is in state k for the symbol rate R_t and the steady-state

Table 5.1: Simulation parameters.

Parameters	**Value**
R_t	$\frac{1}{104.86\times10^{-3}}$
f_m	10
γ_0	1
$\{ch_1, ch_2, ch_3\}$	$\{0.106, 0.511, 1.61\}$
Transition probability of channel SR and RD	$\begin{bmatrix} 0.82 & 0.18 & 0 \\ 0.09 & 0.81 & 0.1 \\ 0 & 0.09 & 0.91 \end{bmatrix}$

probability of state k p_k. The level crossing rate is calculated as

$$N(A_k) = \sqrt{\frac{2\pi A_k}{\gamma_0}} f_m e^{-\frac{A_k}{\gamma_0}}, \tag{5.15}$$

where f_m is the Doppler frequency.

5.5 Numerical Results

In the simulation, we set T as 1 s. $f^{\min}$ and B are 9.5 kHz and 5 kHz, respectively. l^{SR} and l^{RD} are 5 km. k, w_s, and s are 2, 0, and 0.5, respectively. P^{S} and $\mathcal{P}^{\mathrm{R}}$ are 130 and $\{0, 110, 120, 130, 140, 150\}$ dB re μPa. We assume that the channel gain $|h_n|^2 \in \{0.106, 0.511, 1.61\}$ with the channel states transition matrix for the SR and the RD are $\begin{bmatrix} 0.82 & 0.18 & 0 \\ 0.09 & 0.81 & 0.1 \\ 0 & 0.09 & 0.91 \end{bmatrix}$. Further, assume that the battery has 11 levels. $E^{\max} = 10E^{\min}$ and $E^H = 3E^{\min}$, where E^{min} is the minimum energy. The values of $\mathcal{P}^{\mathrm{R}}$ correspond to the first six battery levels. The number of time slots is 50. At the start state, the values for B_1 and $P_1^{\mathrm{R,RSI}}$ are 0.

Three policies are compared in the problem: the optimal policy, which chooses the

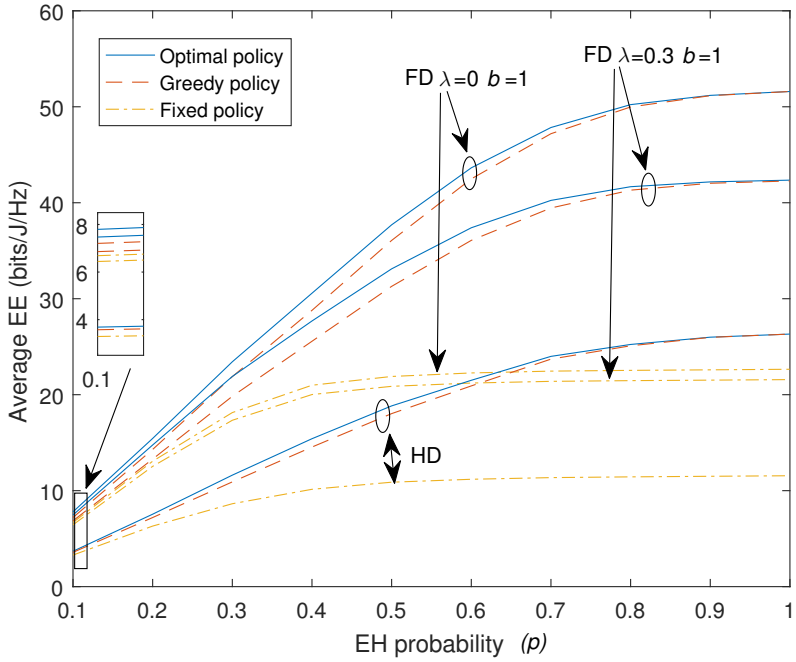

Figure 5.2: Average EE vs. EH probability.

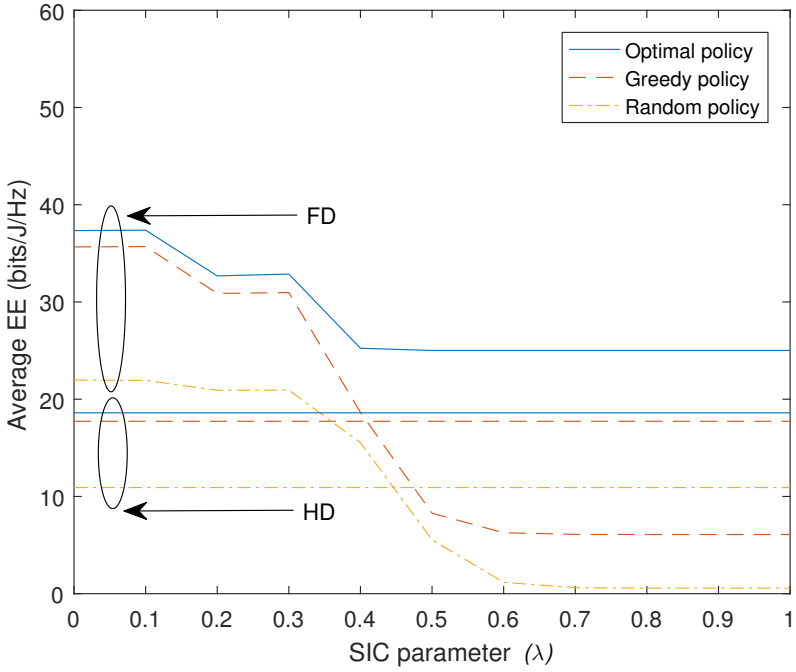

Figure 5.3: Average EE vs. SIC parameter (λ) for $p = 0.5$ and $b = 1$.

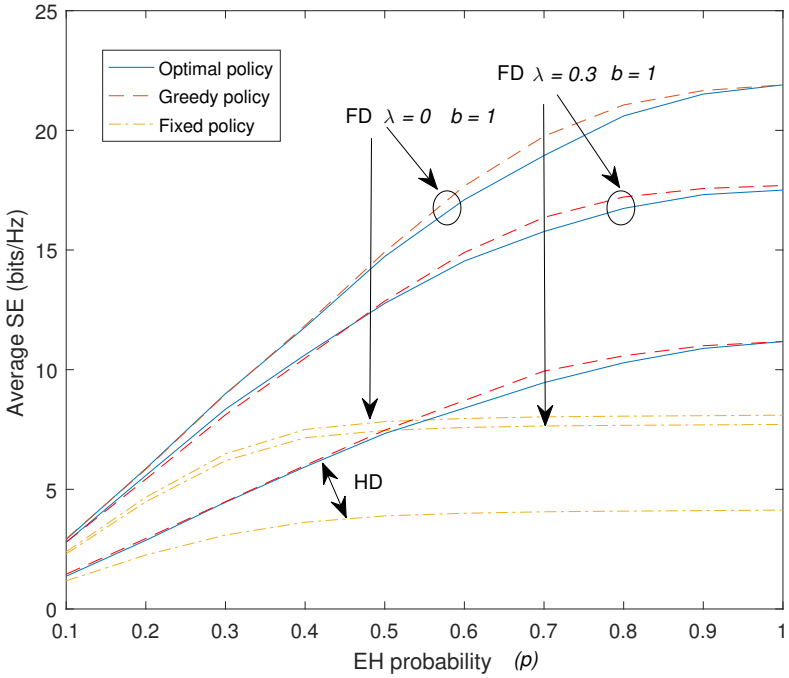

Figure 5.4: Average SE vs. EH probability.

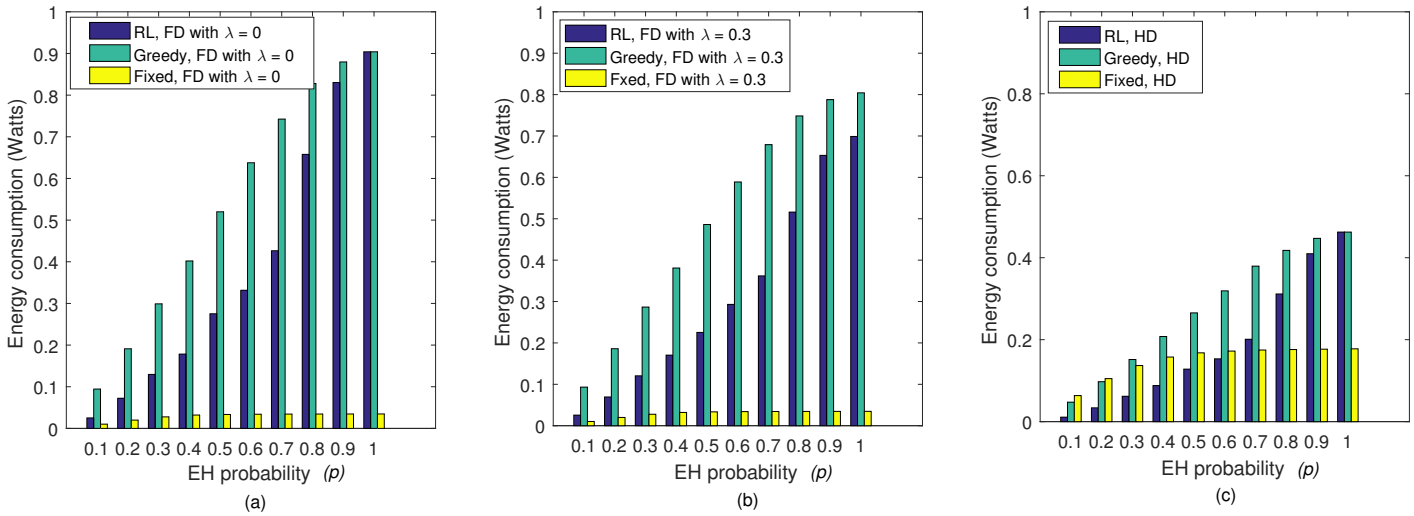

Figure 5.5: Energy consumption vs. EH probability.

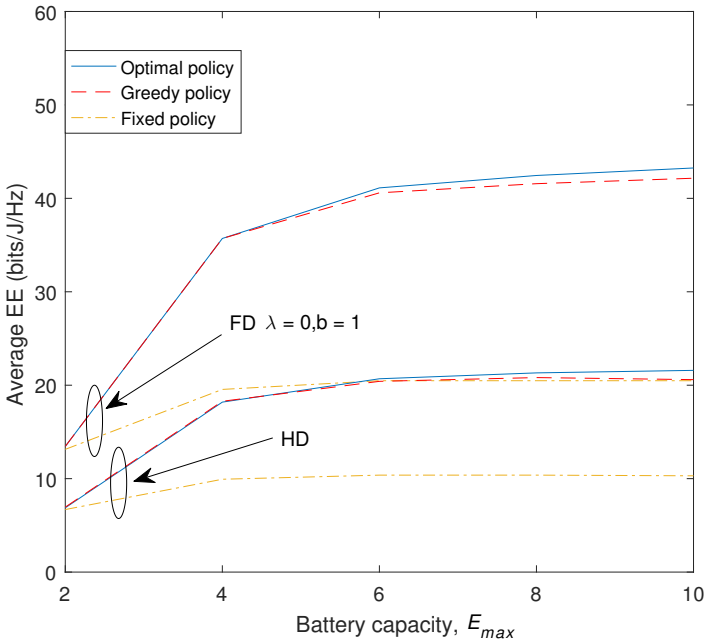

Figure 5.6: Average EE vs. Battery capacity for $p = 0.6$.

transmit power based on π^*; the greedy policy, which maximizes the current reward; and the fixed policy, which chooses the fixed transmit power equal to 110 dB re μPa. Figure 5.2 shows the long-term average EE results. It can be seen that results from the optimal policy outperform those from the greedy and the fixed ones. Also, the performance of the long-term average EE is increasing with the EH probability. Moreover, the FD mode achieves a higher average EE than the HD one, especially as the SI cancellation improves (perfect SI cancellation is when $\lambda = 0$). Figure 5.3 illustrates that the average EE in the FD mode is decreasing as the SIC parameter λ increases. This indicates that a better SIC cancellation technique should be applied in the FD operation for improving the average EE. Moreover, since the SI does not exist in the HD mode, the average EE in the HD mode remains constant as λ varies. Figure 5.4 presents the average spectral efficiency (SE) performance versus different EH probability values under the formulated problem that maximizes average EE.

The greedy policy has a slightly higher average SE in the high EH region. Further, Figure 5.5 shows the energy consumption versus the EH probability with respect to different transmission policies and operation modes. The energy consumption for the greedy policy is larger than the optimal and fixed policies. The fixed transmission policy consumed the least energy, since the transmit power sets to the smallest power level (i.e., 110 dB re μPa) in every time slot. Also, the energy consumption for the HD mode is less than that for the FD mode, because the FD mode can transmit the signals in every time slot, whereas the HD mode operates in a time-multiplexing manner that the transmission and reception operate in two different time slots. Thus, the FD mode consumed more energy than the HD mode. Figure 5.6 shows the average EE performance versus the battery capacity. The average EE increases as the battery capacity increases due to more energy available for data transmission.

For backward induction algorithm, the size of the optimal policy table is $N \times |\mathcal{S}| \times |\mathcal{A}|$, where $|\cdot|$ denotes the cardinality of the set. However, the size is $|\mathcal{S}| \times |\mathcal{A}|$ for value iteration algorithm, which has a lower size. This is because the value iteration algorithm can derive a stationary policy that does not vary with time [11].

5.6 Conclusion

In this chapter, the problem of maximizing the long-term EE of an FD EH single-relay underwater network is investigated. The RL framework is applied to obtain an optimal energy-efficient transmission policy, which provides improved results when compared with the greedy and fixed policies. Further, the FD performance is better than the HD one, especially under good SI cancellation.

Bibliography

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

[2] E. A. Makled, A. Yadav, O. A. Dobre, and R. D. Haynes, "Hierarchical full-duplex underwater acoustic network: A noma approach," in *Proc. IEEE/MTS OCEANS*, Charleston, NC, USA, Oct. 2018, pp. 1–6.

[3] L. J. Rodriguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex af relaying in the presence of residual self-interference," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[4] P. R. Bandyopadhyay, D. P. Thivierge, F. M. McNeilly, and A. Fredette, "An electronic circuit for trickle charge harvesting from littoral microbial fuel cells," *IEEE J. Oceanic Eng.*, vol. 38, no. 1, pp. 32–42, Jan. 2013.

[5] C. Wang, Z. Wang, W. Sun, and D. R. Fuhrmann, "Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels," *IEEE Access*, vol. 6, pp. 2541–2558, Dec. 2017.

[6] Q. Fu and A. Song, "Adaptive modulation for underwater acoustic communications based on reinforcement learning," in *Proc. IEEE/MTS OCEANS*, Charleston, NC, USA, Oct. 2018, pp. 1–8.

[7] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[8] A. Yadav, M. Goonewardena, W. Ajib, O. A. Dobre, and H. Elbiaze, "Energy management for energy harvesting wireless sensors with adaptive retransmission," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5487–5498, Dec. 2017.

[9] M. O. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1126–1131, Nov. 2003.

[10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[11] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. Press, 1994.

[12] Qinqing Zhang and S. A. Kassam, "Finite-state markov model for rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

Chapter 6

Conclusions and Future Works

In this chapter, conclusions are drawn by summarizing the content for each chapter, and some possible research directions and additional works are mentioned from the preliminary results of the thesis.

6.1 Conclusions

- Chapter 1 presented motivation, thesis outline, and contributions of the thesis.
- Chapter 2 reviewed the background knowledge, including cooperative communication, energy harvesting communication, and reinforcement learning technique.
- Chapter 3 overviewed the underwater acoustic propagation and channel. The propagation paths of acoustic waves vary due to the non-homogeneous underwater sound speed profile. Moreover, the path loss is transmission distance- and operating frequency-dependent. Additionally, the standard model for underwater channel fading is still an open issue in the community. Moreover, the UWAC system designing should consider the propagation paths of acoustic rays.

- Chapter 4 proposed the RL-based transmission policy for the end-to-end sum rate maximization in an FD EH single-relay underwater network over a finite horizon. First, the signal transmission model is analyzed, and the end-to-end SNR is given under the AF protocol of the relay node. Second, the optimization problem is defined as maximizing the end-to-end sum rate over a finite N time slot. Third, the RL framework is proposed to derive an optimal transmission policy. Finally, the performance of the proposed transmission policy is compared with the benchmark greedy policy, which shows the outstanding performance of the proposed policy.

- Chapter 5 proposed the RL-based stationary transmission policy for the long-term end-to-end average EE maximization in an FD EH single-relay underwater networks. The MDP model is formulated, and the RL algorithm is used to derive an optimal transmission policy. The policy is stored in the relay before the start of the transmission. During the transmission, the relay selects the optimal transmit power according to the harvested energy amount, battery level, channel state information, and interference level.

In sum, this thesis developed the adaptive transmission policies for the long-term operational three-node underwater relay networks. Assuming that the causal knowledge of the considered system is known, the online sequential decision-making problem is formulated. The RL technique is adopted to solve the problem and derive transmission policies. The transmission policies obtained are optimal under the formulated model and setting, obtained through the RL framework and achieve better performance compared with the benchmark ones.

6.2 Future Directions of Research

According to the results of the thesis, the possible research directions and additional future work may be as follows:

- The system model in Chapters 4 and 5 is a three-node underwater relay network. This model may be extended to multiple nodes in future work. The coordination of the multiple underwater nodes in the long-term operational networks brings up interesting problems.
- The system performance in Chapters 4 and 5 could further be improved by increasing the number of states and actions. To tackle this issue, powerful deep RL can be investigated.