2018

Intelligent and Secure Underwater Acoustic Communication Networks

Chaofeng Wang
Michigan Technological University, cwang8@mtu.edu

INTELLIGENT AND SECURE UNDERWATER ACOUSTIC COMMUNICATION NETWORKS

By

Chaofeng Wang

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Electrical Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Electrical Engineering.

Department of Electrical and Computer Engineering

Dissertation Advisor:	*Dr. Zhaohui Wang*
Committee Member:	*Dr. Daniel R. Fuhrmann*
Committee Member:	*Dr. Min Song*
Committee Member:	*Dr. Ossama Abdelkhalik*
Department Chair:	*Dr. Daniel R. Fuhrmann*

Dedication

I dedicate this dissertation to Xiangjun Wang, my father; Ying Huang, my mother; Qiao Xiao, my wife; and Huisha Wang, my elder sister.

Contents

List of Figures

List of Tables

Preface

This dissertation presents the research works during my Ph.D. study in Electrical Engineering at Michigan Technological University. This dissertation consists of three articles which are previously published or accepted. All my research works were under the supervision of my advisor Dr. Zhaohui Wang.

The work in Chapter 2 was published by *IEEE Access* as in [1]. As the first author of the paper, under the guidance of my advisor, I completed system modeling and algorithm design, implementation, and evaluation. As the third author, Wensheng Sun provided useful comments on channel characteristics and modeling. As the fourth author, Dr. Daniel R. Fuhrmann provided insightful advice to the paper. This article was completed by me and my advisor.

The work in Chapter 3 was accepted by the *2018 MTS/IEEE Oceans Conference* as in [2]. As the first author of the paper, under the guidance of my advisor, I proposed the algorithm and completed the implementation. As the second author, Li Wei provided advice for practical underwater systems. As the fourth author, Dr. Min Song provided constructive suggestions to the paper. As the fifth author, Dr. Nina Mahmoudian provided helpful comments on AUV trajectory modeling. This article was completed by me and my advisor.

The work in Chapter 4 was published in *IEEE Transactions on Signal Processing* as in [3]. As the first author, under the guidance of my advisor, I completed the algorithm design and implementation. This article was completed by me and my advisor.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Zhaohui Wang for her support and guidance during my Ph.D. study at Michigan Tech. From my advisor who is an excellent researcher, I keep learning how to do research with forward and rigorous thinking. This memorable learning experience will definitely continue to benefit my future career. It is my great honor to be her first Ph.D. student in her faculty career.

I would like to thank Professor Daniel R. Fuhrmann, Professor Min Song, and Professor Ossama Abdelkhalik to serve on my committee. I appreciate the guidance and support from Professor Daniel R. Fuhrmann at the beginning of my Ph.D. program. I would also like to thank Professor Saeid Nooshabadi and Professor Nina Mahmoudian for their helpful comments on my research work.

I would like to thank all my friends and colleagues, Wensheng Sun, Jing Liu, Yang Liu, Shuo Wang, Xu Zhang, Marco La Manna, Caoyang Jiang, Li Wei, Krishna C. Poduru, and Nina Huang for their invaluable support and help during my years at Michigan Tech.

Last but not least, I would like to thank my family. I am greatly indebted to my wife for her love, companion, and sacrifices, and to my parents and my sister for

their help to strengthen my belief and courage to pursue my dream with their endless encouragement and support. My parents-in-law inspire me with their understanding and trusts.

Abstract

Underwater acoustic (UWA) communication networks are promising techniques for medium- to long-range wireless information transfer in aquatic applications. The harsh and dynamic water environment poses grand challenges to the design of UWA networks. This dissertation leverages the advances in machine learning and signal processing to develop intelligent and secure UWA communication networks. Three research topics are studied: 1) reinforcement learning (RL)-based adaptive transmission in UWA channels; 2) reinforcement learning-based adaptive trajectory planning for autonomous underwater vehicles (AUVs) in under-ice environments; 3) signal alignment to secure underwater coordinated multipoint (CoMP) transmissions.

First, a RL-based algorithm is developed for adaptive transmission in long-term operating UWA point-to-point communication systems. The UWA channel dynamics are learned and exploited to trade off energy consumption with information delivery latency. The adaptive transmission problem is formulated as a partially observable Markov decision process (POMDP) which is solved by a Monte Carlo sampling-based approach, and an expectation-maximization-type of algorithm is developed to recursively estimate the channel model parameters. The experimental data processing reveals that the proposed algorithm achieves a good balance between energy efficiency and information delivery latency.

Secondly, an online learning-based algorithm is developed for adaptive trajectory planning of multiple AUVs in under-ice environments to reconstruct a water parameter field of interest. The field knowledge is learned online to guide the trajectories of AUVs for collection of informative water parameter samples in the near future. The trajectory planning problem is formulated as a Markov decision process (MDP) which is solved by an actor-critic algorithm, where the field knowledge is estimated online using the Gaussian process regression. The simulation results show that the proposed algorithm achieves the performance close to a benchmark method that assumes perfect field knowledge.

Thirdly, the dissertation presents a signal alignment method to secure underwater CoMP transmissions of geographically distributed antenna elements (DAEs) against eavesdropping. Exploiting the low sound speed in water and the spatial diversity of DAEs, the signal alignment method is developed such that useful signals will collide at the eavesdropper while stay collision-free at the legitimate user. The signal alignment mechanism is formulated as a mixed integer and nonlinear optimization problem which is solved through a combination of the simulated annealing method and the linear programming. Taking the orthogonal frequency-division multiplexing (OFDM) as the modulation technique, simulation and emulated experimental results demonstrate that the proposed method significantly degrades the eavesdropper's interception capability.

Chapter 1

Introduction

1.1 Background and Challenges

The earth is a planet where more than 70% of its surface is covered by water, such as lakes and oceans. Those water bodies not only itself have a significant impact on the nature but also provide tremendous resources for human beings. We harvest food in sea farms, extract metal ions from the sea water, and even take advantage of the cold temperature in water as cooling systems for infrastructures such as data centers. Furthermore, the mystery at the ocean bottom awaits to be discovered since about 95% of the ocean remains unexplored. As resource depletion has been a big concern in the new era, the underwater exploration becomes a potential remedy for

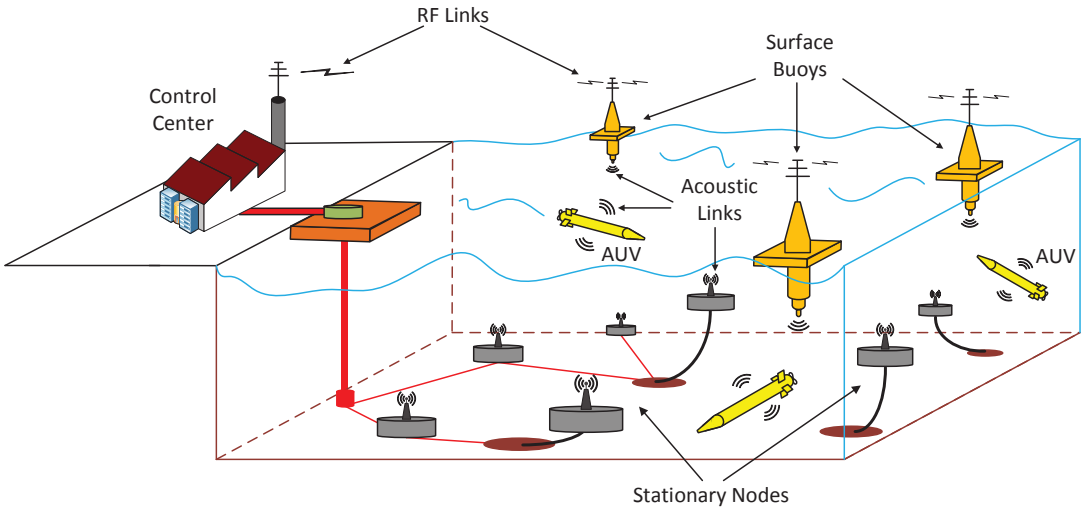


Figure 1.1: An example of an underwater acoustic communication network. The stationary sensor nodes, AUVs, and surface buoys can communicate with each other using acoustic links. Some stationary nodes and the buoys are connected to a control center via cables and high-rate radio links, respectively.

sustainable development. Underwater acoustic (UWA) communication networks are the promising techniques for wireless information transfer over medium and long ranges in various underwater applications.

A typical structure of UWA networks is presented in Fig. 1.1. The UWA communication networks are capable of different tasks with the help of data fusion and information transferring. Specifically, there exist stationary sensor nodes either mounted directly or moored at the water bottom. Those nodes usually live in water for a long period of time, from weeks to years. The mobile autonomous underwater vehicles (AUVs) can play various roles, such as to scan the water body at different depths and to relay the information for remotely located sensor nodes. There are surface buoys and some stationary nodes which are connected to a control center via high-rate radio

links and cables, respectively. Their major job is to relay the collected information from other sensor nodes via acoustic links to the control center for real-time analysis and decision-making while the cooperation of them can also provide better services to improve the system performance of the whole network.

The UWA communications networks are very different from terrestrial radio networks. The UWA channel has large sound propagation delay due to the low sound speed in water as 1500 m/s while radio waves travel at the speed of 3×10^8 m/s. The UWA channel has large Doppler effect and frequency-dependent signal attenuation. The signal attenuation is large at the higher frequency bands, which leads to limited communication bandwidth. About the multi-path effect, the channel delay spread is large in UWA channels, and different paths have different Doppler scaling factors. The investigation of UWA communication networks for practical applications is still at its early stage. Many existing solutions to UWA networks are modifications of techniques which are designed for terrestrial radio networks. Hence, many research problems and opportunities arise associated with the unique characteristics of UWA networks at different layers [4, 5, 6, 7, 8, 9, 10, 11]. In this dissertation, three critical issues of UWA communication networks are studied as follows.

- Many underwater nodes are powered by batteries, and the cost of battery replacement in the remote water area is high. The energy-efficient adaptive transmission in UWA networks should be specifically designed based on the characteristics of UWA channels.

- The capture of the water parameter field of interest such as the temperature map and underwater acoustic field is an effective method to study the underwater phenomenon. With a goal of providing an accurate field reconstruction, mobile sensor nodes such as AUVs are often employed to take field samples along their trajectories. How to design the optimal sampling trajectories for multiple AUVs is a key factor to accurately reconstruct the field of interest.

- The broadcasting nature of acoustic transmissions makes the UWA communication networks vulnerable to adversarial attacks. Due to the low sound speed in water, many security algorithms designed for terrestrial radio networks cannot be applied directly to UWA communication networks. The security in UWA communication networks has drawn considerable attention recently.

Nowadays, we enjoy the benefit brought by artificial intelligence in our daily life. The boom in artificial intelligence provides new techniques and perspectives to solve the arising research problems in UWA networks. The importance of the UWA networks also inspires us to care more about their security. We can foresee the rapid growth and development of intelligent and secure UWA networks in the near future.

1.2 Contributions

The dissertation focuses on developing intelligent and secure UWA communications networks. The solutions to the three research issues as discussed are presented in Chapters 2 to 4, respectively.

Chapter 2 studies adaptive transmission in an UWA point-to-point communication system that operates on an epoch-by-epoch basis for a long term. A fixed amount of information bits periodically arrive at the transmitter data queue, and wait for transmission via a number of packets within each epoch. To trade off energy consumption with transmission latency, the transmitter decides the transmission action at the beginning of each epoch, including to transmit or not, the transmission power and the modulation-and-coding parameters, based on the data queue status and the predicted channel conditions in the current and future epochs. To describe both the fast fading and the large-scale shadowing of UWA channels, the channel within each epoch is characterized by a compound Nakagami-lognormal distribution, and the evolution of the distribution parameters is modeled as an unknown Markov process. Given that the channel can only be observed during active transmissions, we formulate the adaptive transmission problem as a partially observable Markov decision process (POMDP), and develop an online algorithm in a model-based reinforcement learning (RL) framework. The algorithm recursively estimates the channel model

parameters, tracks the channel dynamics, and computes the optimal transmission action that minimizes a long-term system cost. Emulated results based on channel measurements from two field experiments demonstrate that the proposed algorithm achieves decent performance relative to a benchmark method that assumes perfect and non-causal channel knowledge.

Chapter 3 studies online learning-based trajectory planning for multiple autonomous underwater vehicles (AUVs) to estimate a water parameter field of interest in the under-ice environment. A centralized system is considered, where several fixed access points (APs) on the ice layer are introduced as gateways for communications between the AUVs and a remote data fusion center (FC). We model the water parameter field of interest as a Gaussian process (GP) with unknown hyper-parameters. The AUV trajectories for sampling are determined on an epoch-by-epoch basis. At the end of each epoch, the APs relay the observed field samples from all the AUVs to the FC which computes the posterior distribution of the field based on the Gaussian process regression (GPR) and estimates the field hyper-parameters. The optimal trajectories of all the AUVs in the next epoch are determined to minimize a long-term cost that is defined based on the field uncertainty reduction and the AUV mobility cost, subject to the kinematics constraint, the communication range constraint and the sensing area constraint. We formulate the adaptive trajectory planning problem as a Markov decision process (MDP). A reinforcement learning (RL)-based online

learning algorithm is designed to determine the optimal AUV trajectories in a constrained continuous space. Simulation results show that the proposed learning-based trajectory planning algorithm has performance similar to a benchmark method that assumes perfect knowledge of the field hyper-parameters.

In Chapter 4, we investigate countermeasures against eavesdropping attack in the coordinated multipoint (CoMP) transmission of geographically distributed antenna elements (DAEs) to an underwater legitimate user. Exploiting the low sound speed in water and the spatial diversity of DAEs, we propose *signal alignment* for transmission secrecy, where a transmission strategy will be judiciously designed such that useful signals will collide at the eavesdropper while stay collision-free at the legitimate user. Specifically, the transmit DAE set, and the transmission schedule and transmission power of each active DAE, are jointly optimized with a goal of minimizing the maximal received signal-to-interference-and-noise ratio (SINR) of useful signals at the eavesdropper, under a lower bound constraint of the received signal-to-noise ratio (SNR) at the legitimate user. Taking the orthogonal frequency-division multiplexing (OFDM) as the modulation technique, simulation and emulated experimental results demonstrate that the proposed method significantly degrades the eavesdropper's interception capability. We further investigate the secrecy capacity and the secure degrees-of-freedom (d.o.f.) of the signal alignment method from an information-theoretic perspective, which reveals that without external helpers, secure d.o.f. greater than $\frac{1}{2}$ can be achieved.

Contributions of the dissertation are summarized in Chapter 5.

Notation: Bold upper case letters and lower case letters are used to denote matrices and column vectors, respectively. $\mathbf{A}^{\mathrm{T}}$ denotes the transpose of matrix $\mathbf{A}$. $[\mathbf{a}]_m$ denotes the mth element of vector $\mathbf{a}$. $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. ∇_a denotes the derivative w.r.t. a. $[\mathbf{A}]_{m,k}$ denotes the (m,k)th element of matrix $\mathbf{A}$. $\mathbf{1}$ denotes a column vector with unity elements. $[\cdot]^+$ is defined as $\max\{\cdot, 0\}$. $\Re\{\cdot\}$ represents the real part of a complex variable.

Chapter 2

Reinforcement learning-based Adaptive Transmission in Underwater Acoustic Channels[1]

[1]The work in this chapter was published in "*IEEE Access*" ©2017 IEEE. See Appendix B.1 for the letter of permission from IEEE.

2.1 Introduction

2.1.1 Background

Due to the high deployment cost, the lifespan of underwater systems varies from months to years. For instance, underwater monitoring systems, such as scientific data collection systems, could be mounted at the water bottom for months to collect parameters of interest, and large-scale ocean observation systems, such as the NEPTUNE and VENUS ocean observatories [12] and the Ocean Observatory Initiative (OOI) [13], could have projected lifespans of more than 20 years. On the other hand, underwater nodes are often powered by batteries, and battery replacement and recharging are time-consuming and costly. Energy-efficient operation is critical for system longevity.

This chapter considers a long-term operating underwater system with deterministic data arrivals (e.g., periodic data collection systems), and studies energy-efficient acoustic transmission that adapts the transmission schedule and the transmission parameters, including the transmission power and the modulation-and-coding parameters, to the system state (e.g., the transmitter data queue length) and the current and future predicted channel conditions, with a goal of minimizing a long-term average

cost. The UWA channel exhibits both small-scale fast fading and long-term large-scale shadowing. Adapting transmission strategy to the channel dynamics could yield considerable energy saving.

2.1.2 Existing Works in Terrestrial Radio Networks

The channel-aware transmission to trade off energy consumption with information delivery latency has been extensively studied in terrestrial radio communications. Particularly for correlated fading channels, most of existing works model the channel as a finite-state Markov chain (FSMC) with known transitional probabilities, and formulate the problem as a Markov Decision Process (MDP) to determine the control variables, such as the transmission schedule, the transmission power, and the modulation-and-coding parameters, based on the channel state and the communication system state (e.g., the data queue length, the incoming traffic rate, and the packet delay constraint). Given that the MDP is generally computationally intractable to solve, special structures of the optimal policy are identified and exploited to find the optimal or near-optimal solution [14, 15, 16, 17]. However, the channel state transition probability and the traffic statistics could be hard to obtain in practice. Some works propose to solve the MDP *online* using reinforcement learning (RL) [18], where model-free RL methods (e.g., Q-learning, and the actor-critic algorithm) are used to learn from past experiences (namely, how to map "situations" to "actions")

without explicit modeling of the channel and/or traffic dynamics [19, 20, 21, 22, 23]. Recent applications of RL in radio-frequency networks include stochastic power control for energy harvesting systems [24, 25], data scheduling and admission control for backscatter sensor networks [26], and rate and mode adaptation for Wi-Fi/LTE-U coexistence [27].

2.1.3 Existing Works in Underwater Acoustic Networks

Compared to radio networks, studies on energy-efficient transmission in UWA networks have been limited. At the physical layer, relevant research includes adapting the transmission power, the frequency band, and the modulation-and-coding parameters to channel dynamics [28, 29, 30, 31]. At the link layer, assuming a two-state FSMC channel model with known transition probabilities and accounting the non-negligible cost of channel probing, energy-efficient transmission scheduling with partial and discontinuous channel state information (CSI) is studied in [32]. The transmission scheduling is formulated as a dynamic programming problem, and different ways of providing the CSI from the receiver are examined. The above work is extended in [33] when only partial data queue state information is available. In [34], the RL is introduced to optimize the parameters in a slotted Carrier Sensing Multiple Access (slotted CSMA) protocol. Assuming a binary symmetric channel (BSC) with unknown transition probabilities, the model-free RL (Q-learning augmented by virtual

experience and state-action aggregation) was introduced in [35] to adapt the link-layer transmission schedule and transmission parameters to the channel dynamics. Q-learning has also been used for designing routing protocols [36] with an aim to balance the workload among network nodes and to prolong the network lifetime.

For long-term operating underwater systems, the UWA channel exhibits both fast fading and large-scale shadowing; see field experiment observations in, e.g., [37, 38, 39]. Data analysis of different field experiments revealed that the fast fading could follow Rayleigh, Rician, Nakagami-m, or compound-K distributions; see [40] and references therein. Based on field measurements, a lognormal model was suggested for large-scale shadowing [41, 42]. Furthermore, the fading and shadowing statistics could change continuously over time; for instance, channel stationarity over an average of three-minute-long interval [43], nonstationarity and cyclostationarity [40] have been observed in different field experiments.

Existing solutions with the FSMC channel model assumption may not work well for adaptive transmission in long-term operating UWA systems. Specifically, the large channel dynamics require a sufficient number of discrete channel states for an adequate description of the channel behavior. Additionally, the FSMC parameters could change continuously over time. The high-dimensionality of the channel state space and the short-term channel stationarity could prevent model-free RL methods from convergence, which eventually leads to degraded performance.

2.1.4 Our Work

In this chapter, we introduce a *continuous* channel model to describe the temporal dynamics of UWA channels, and adopt a model-based RL framework to determine the transmission strategy with the aim of optimizing a long-term system performance measure. Specifically, to better capture the channel variation over a long term, we introduce a compound Nakagami-lognormal distribution to characterize the channel fast fading and the large-scale shadowing, and model the evolution of the distribution parameters as a first-order Markov process. Based on the above channel model, the model-based RL framework is employed for adaptive transmission. The framework has two components: channel model estimation and online planning. Following the maximum likelihood principle and the expectation-maximization concept [44], an algorithm is developed to recursively estimate the channel model parameters and predict the channel state based on newly obtained channel measurements. The online planning is then performed via a Monte Carlo sampling method which finds a near-optimal transmission strategy through constructing an online state-action tree.

The proposed algorithms are validated using data sets collected from two experiments, one held off the coast of Martha's Vineyard, Massachusetts, in 2008, and the other held in the ice-covered Keweenaw Waterway near Michigan Tech, Michigan, in 2014. The experimental results show that: 1) the recursive channel estimation method

yields decent performance on tracking the UWA channel dynamics; and 2) the model-based RL algorithm achieves performance close to a genie-aided method that assumes perfect and non-causal channel knowledge.

To the best of our knowledge, this is the first attempt that adopts the model-based RL framework for adaptive transmission in long-term operating UWA systems, where the channel statistical parameters in continuous spaces are explicitly learned from past transmissions.

The rest of the chapter is organized as follows. The system model is presented in Section 2.2. The model-based RL algorithm for adaptive transmission is developed in Section 2.3. The Monte Carlo sampling method for online planning is presented in Section 2.4. A recursive algorithm for channel model estimation and channel tracking is described in Section 2.5. Evaluation of the proposed algorithm is included in Section 2.6. Summary is presented in Section 2.7.

2.2 System Model and Problem Formulation

2.2.1 System Description

This work focuses on adaptive transmission in a long-term operating UWA point-to-point data transmission system. The time is divided into epochs as shown in Fig. 2.1. Each epoch consists of N time slots, and each time slot is used to transmit one data packet. At the end of the epoch, an acknowledgement packet is sent from the receiver through an error-free channel to the transmitter, which includes information of the packets that are successfully delivered and the received signal-to-noise ratio (SNR) of each packet. We further assume that at the transmitter, a fixed amount of information bits are generated at the application layer in each epoch and arrive at the data queue of the transmitter at the beginning of an epoch. The transmission schedule and the transmission parameters will be determined recursively epoch by epoch based on the data queue state and information about the channel state, with an ultimate goal of minimizing a long-term system cost.

For each time epoch, the transmission parameters include the transmission power, the modulation size and the channel coding rate. Note that the acoustic modem in practical systems only maintains a finite number of modulation and coding pairs

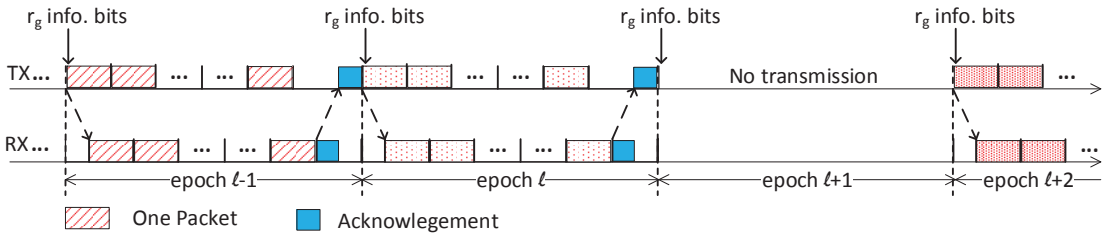


Figure 2.1: Epoch structure at the transmitter and the receiver. The transmission parameters, including the transmission power, the modulation size and the channel coding rate, could vary from epoch to epoch.

as well as a finite number of transmission power levels. We consider a finite set of discrete power levels $\mathcal{P} = \{P_0, P_1, P_2, \cdots\}$, with $P_0 = 0$ for no transmission, a finite set of discrete modulation sizes $\mathcal{M} = \{M_1, M_2, \cdots\}$, and a finite set of channel coding rates $\mathcal{R}_{\rm c} = \{r_{{\rm c},1}, r_{{\rm c},2}, \cdots\}$. A combination of the modulation size M_i and the coding rate $r_{{\rm c},j}$ yields a data rate of $r_{{\rm c},j} \cdot \log_2 M_i$. Stack the triplet of transmission parameters $\{P \in \mathcal{P}, M \in \mathcal{M}, r_{\rm c} \in \mathcal{R}_{\rm c}\}$ into a vector, $\mathbf{a} := [P, M, r_{\rm c}]^{\rm T}$. Denote $\mathbf{a}(\ell)$ as the transmission parameter vector in the ℓth epoch.

In the next, we will develop an UWA channel model and an evolution model of the transmitter data queue, and then formulate the adaptive transmission as an optimization problem.

2.2.2 Underwater Acoustic Channel Model

To model both the fast fading and the large-scale shadowing of UWA channels, the UWA channel within one epoch is statistically characterized via a compound

Nakagami-lognormal distribution. Accordingly, the received SNR follows a gamma-lognormal distribution [45]. Denote $\rho := P_{\text{tx}}/N_0$ as the transmission signal-power-to-noise ratio in an epoch, and denote x as the corresponding received SNR. The probability density function (PDF) of x can be expressed as

$$f_X(x; m, \mu, \sigma) = \int_0^\infty \frac{x^{m-1} \exp\left(-\frac{mx}{\rho y}\right)}{\Gamma(m)} \left(\frac{m}{\rho y}\right)^m \frac{\exp\left[-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right]}{\sqrt{2\pi}\sigma y} dy, \tag{2.1}$$

where $\Gamma(\cdot)$ is the gamma function, $m \in [1/2, \infty)$ is the fading parameter in the Nakagami-m fading, and μ and σ are the mean and the standard deviation of the log-normal shadowing, respectively [45]. Therefore, the UWA channel can be statistically parameterized by the triplet $\{m, \mu, \sigma\}$.

Define $\mathbf{s}_{\text{ch}} := [m, \mu, \sigma]^{\text{T}}$, and denote $\mathbf{s}_{\text{ch}}(\ell)$ as the channel state in the ℓth epoch. We model the long-term channel temporal variation as a first-order Markov process,

$$\mathbf{s}_{\text{ch}}(\ell) = \mathbf{A}\mathbf{s}_{\text{ch}}(\ell - 1) + \mathbf{w}_{\text{ch}}(\ell), \tag{2.2}$$

where $\mathbf{A}$ is a 3×3 unknown matrix, and $\mathbf{w}_{\text{ch}}(\ell)$ is the process noise vector for modeling inaccuracy, and is assumed following a zero-mean Gaussian distribution with an unknown covariance matrix $\mathbf{C}_{\mathbf{w}}$, namely, $\mathbf{w}_{\text{ch}}(\ell) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{w}})$.

The UWA channel in an epoch can be measured during packet transmissions. We assume that the receiver can measure the received SNR of each packet even if the

packet cannot be successfully decoded. The collected received SNR measurements are piggybacked on the acknowledgement packet sent from the receiver to the transmitter at the end of each active epoch. Denote $\{x_{\ell,1}, x_{\ell,2}, \cdots, x_{\ell,N}\}$ as the received SNRs of N packets in the ℓth epoch. Given the knowledge of the transmission SNR $\rho(\ell)$, the channel statistical parameters, $\{m(\ell), \mu(\ell), \sigma(\ell)\}$, can be estimated via the method of moments [44] according to (2.1).

We denote $\mathbf{z}_{\text{ch}}(\ell)$ as the vector stacked by the estimated parameters, $\{\hat{m}(\ell), \hat{\mu}(\ell), \hat{\sigma}(\ell)\}$, and take $\mathbf{z}_{\text{ch}}(\ell)$ as the observation vector of $\mathbf{s}_{\text{ch}}(\ell)$. Hence,

$$\mathbf{z}_{\text{ch}}(\ell) = \mathbf{s}_{\text{ch}}(\ell) + \mathbf{v}_{\text{ch}}(\ell), \tag{2.3}$$

where $\mathbf{v}_{\text{ch}}(\ell) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\text{v}})$ is the observation noise with an unknown covariance matrix $\mathbf{C}_{\text{v}}$, and is assumed independent from the process noise $\mathbf{w}_{\text{ch}}(\ell)$.

The channel model can then be uniquely represented by the unknown parameter set $\boldsymbol{\Theta} := \{\mathbf{A}, \mathbf{C}_{\text{w}}, \mathbf{C}_{\text{v}}\}$. Due to the water environment dynamics, the parameter set could be slowly time-varying.

Remark 1: For the epochs without active transmissions, a channel probing sequence could be transmitted to collect information about the channel dynamics. Although this work does not consider the probing sequence, the obtained theoretical results can be applied with slight modification to the scenario with channel probing sequences.

2.2.3 Evolution of the Data Queue

For each transmission parameter triplet $\mathbf{a} = [P, M, r_{\mathrm{c}}]^{\mathrm{T}}$, the packet error rate (PER) can be determined based on the compound distribution of the received SNR using an information-theoretic approach [46, 47] or an empirical formula estimated by real data [30]. For a channel state $\mathbf{s}_{\mathrm{ch}}$ and a transmission parameter vector $\mathbf{a}$, we denote the PER by function $\mathrm{PER}(\mathbf{s}_{\mathrm{ch}}, \mathbf{a})$.

At the beginning of epoch ℓ, the data queue length can be recursively represented as

$$q(\ell) = q(\ell-1) - r(\ell-1)N_{\mathrm{s}}(\ell-1) + r_{\mathrm{g}}, \tag{2.4}$$

where $r(\ell-1)$ is the amount of information bits carried by each packet according to the transmission parameter vector $\mathbf{a}(\ell-1)$, $N_{\mathrm{s}}(\ell-1)$ is the number of packets that are successfully delivered to the receiver in epoch $(\ell-1)$, and r_{g} is the amount of information bits from the application layer arriving at the beginning of epoch ℓ. Given $\mathrm{PER}(\mathbf{s}_{\mathrm{ch}}, \mathbf{a})$, the number of packets that can be successfully received follows a binomial distribution $\mathcal{B}(N, 1-\mathrm{PER}(\mathbf{s}_{\mathrm{ch}}, \mathbf{a}))$, namely,

$$\Pr(N_{\mathrm{s}} = k|\mathbf{s}_{\mathrm{ch}}, \mathbf{a}) = \binom{N}{k} \left(1-\mathrm{PER}(\mathbf{s}_{\mathrm{ch}}, \mathbf{a})\right)^{k} \left(\mathrm{PER}(\mathbf{s}_{\mathrm{ch}}, \mathbf{a})\right)^{N-k}. \tag{2.5}$$

Therefore, given the channel state $\mathbf{s}_{\mathrm{ch}}(\ell-1)$ and the transmission parameter vector

$\mathbf{a}(\ell-1)$, the probability distribution of $N_{\rm s}(\ell-1)$, and the transition probability from $q(\ell-1)$ to $q(\ell)$ can be determined.

2.2.4 Problem Formulation for Optimal Transmission

We define the *system state* of epoch ℓ as $\mathbf{s}(\ell) := \{\mathbf{s}_{\rm ch}(\ell), q(\ell)\} \in \mathcal{S},\ \forall \ell = 0, \cdots, \infty$. The transmission vector in each epoch, $\{\mathbf{a}(\ell) \in \mathcal{A}, \forall \ell\}$, can be determined to minimize the expected total discounted cost,

$$\min_{\{\mathbf{a}(\ell)\in\mathcal{A}\}_{\ell=0}^{\infty}} \mathbb{E}\left\{\sum_{\ell=0}^{\infty} \gamma^{\ell} C(\mathbf{s}(\ell), \mathbf{a}(\ell))\right\}, \tag{2.6}$$

where $\gamma \in (0, 1]$ is a discount factor, and the cost function $C(\mathbf{s}, \mathbf{a}) : \mathcal{S} \times \mathcal{A} \to \mathcal{R}$ is application-dependent, and can be defined by the system designer. In this work, we take the cost function as

$$C(\mathbf{s}(\ell), \mathbf{a}(\ell)) = f_{\rm p}\big(P(\ell)\big) + f_{\rm q}\big(q(\ell) - r(\ell)N_{\rm s}(\ell)\big), \forall\ \ell \tag{2.7}$$

where $f_{\rm p}(\cdot)$ and $f_{\rm q}(\cdot)$ are two generic functions that are related to the energy consumption and the queue length, respectively, $(q(\ell) - r(\ell)N_{\rm s}(\ell))$ is the queue length at the end of epoch ℓ, and the number of successfully delivered packets $N_{\rm s}(\ell)$ depends on the channel state $\mathbf{s}_{\rm ch}(\ell)$ and the action $\mathbf{a}(\ell)$. We note that the cost function $C(\mathbf{s}(\ell), \mathbf{a}(\ell))$ is a random variable due to the randomness of the channel state $\mathbf{s}_{\rm ch}(\ell)$

and the number of successfully delivered packets $N_{\rm s}(\ell)$.

2.3 Reinforcement Learning-based Adaptive Transmission

The optimization problem in (2.6) falls into the category of RL [48], where the transmitter (a.k.a. an agent in RL) interacts with the stochastic and dynamic UWA channel, with a goal of finding an optimal transmission strategy that minimizes the system long-term cost. In this section, we will reformulate the optimization problem in (2.6) in the model-based RL framework, and provide an overview of the proposed algorithm for online adaptive transmission. For notation convenience, we include the epoch index ℓ as a subscript.

2.3.1 Model-based RL for Adaptive transmission

Should the system state be completely observable, the optimal transmission strategy can be determined by solving the Bellman optimality equation (BOE),

$$V^*(\mathbf{s}) = \min_{\mathbf{a}\in\mathcal{A}} \left[C(\mathbf{s},\mathbf{a}) + \gamma \int_{\mathcal{S}} p(\mathbf{s}'|\mathbf{s},\mathbf{a}) V^*(\mathbf{s}') d\mathbf{s}' \right], \tag{2.8}$$

where $V^*(\mathbf{s})$ is referred to as the optimal value function of state $\mathbf{s}$, and $p(\mathbf{s}'|\mathbf{s},\mathbf{a})$ is the state transition probability after taking action $\mathbf{a}$. The minimand in (2.8) consists of two terms: one is the cost of taking action $\mathbf{a}$ at the current state $\mathbf{s}$, and the other is the expected cost in the successor states after taking action $\mathbf{a}$. In the problem under consideration, although the queue state can be completely observed, the UWA channel cannot be directly observed, especially in epochs with no transmissions. The interaction between the transmitter and the underwater channel can be modeled as a partially observable Markov Decision process (POMDP) [48].

We define $b(\mathbf{s}_{\mathrm{ch},\ell})$ as the belief of channel state $\mathbf{s}_{\mathrm{ch},\ell}$, which corresponds to *a priori* PDF of state $\mathbf{s}_{\mathrm{ch},\ell}$, and can be inferred based on past observations $\{\mathbf{z}_{\mathrm{ch},\ell'}; \ell' < \ell\}$. Consider $\mathbf{z}_{\mathrm{ch},\ell} \in \mathcal{Z}, \forall \ell$, with the empty set $\Phi \in \mathcal{Z}$ to represent the scenario without active transmissions, and $q_\ell \in \mathcal{Q}, \forall \ell$. To indicate the dependence of the value function on the channel model, we include the model parameter set $\mathbf{\Theta}$ in the value function representation. The BOE in (2.8) can be reformulated as in (2.9),

$$\begin{aligned} V^*(q_\ell, b(\mathbf{s}_{\mathrm{ch},\ell}); \mathbf{\Theta}) = \min_{\mathbf{a}\in\mathcal{A}} \Big[& \sum_{k=0}^{N} \int_{\mathcal{S}} C(\mathbf{s}_{\mathrm{ch},\ell}, q_\ell, \mathbf{a}) \Pr(N_{\mathrm{s},\ell} = k|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a}) b(\mathbf{s}_{\mathrm{ch},\ell}) d\mathbf{s}_{\mathrm{ch},\ell} \\ & + \gamma \sum_{k=0}^{N} \int_{\mathcal{Z}} \int_{\mathcal{S}} \Pr(N_{\mathrm{s},\ell} = k|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a}) f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a}) b(\mathbf{s}_{\mathrm{ch},\ell}) \\ & \times V^*(q_{\ell+1}, b(\mathbf{s}_{\mathrm{ch},\ell+1}); \mathbf{\Theta}) d\mathbf{s}_{\mathrm{ch},\ell} d\mathbf{z}_{\mathrm{ch},\ell} \Big], \end{aligned} \tag{2.9}$$

where q_ℓ, $q_{\ell+1}$, $N_{\mathrm{s},\ell}$ and $\mathbf{a}$ are related according to (2.4). Similar to (2.8), the minimand

in (2.9) has two terms: the first term is the expected cost in the current epoch based on the current channel belief state and action, and the second term is the expected cost in future epochs. The optimal action in the current epoch is the one that minimizes the total expected cost in the current and future epochs.

We next discuss the probability functions in (2.9) for two types of actions. For the actions leading to packet transmissions, namely, $[\mathbf{a}]_1 \neq 0$ (c.f. Section 2.2), the probability functions in (2.9) can be determined based on (2.2), (2.3) and (2.4). The channel state belief $b(\mathbf{s}_{\mathrm{ch},\ell+1})$ can be recursively updated as

$$b(\mathbf{s}_{\mathrm{ch},\ell+1}) \propto \int_{\mathcal{S}} f(\mathbf{s}_{\mathrm{ch},\ell+1}|\mathbf{s}_{\mathrm{ch},\ell}) f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a}) b(\mathbf{s}_{\mathrm{ch},\ell}) d\mathbf{s}_{\mathrm{ch},\ell}. \tag{2.10}$$

For the action of no transmission, namely, $[\mathbf{a}]_1 = 0$, we have $\mathbf{z}_{\mathrm{ch},\ell} \in \Phi$. The probability function $f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a})$ is non-informative and is independent of $\mathbf{s}_{\mathrm{ch},\ell}$, hence $\int_{\mathcal{Z}} f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{a}) d\mathbf{z}_{\mathrm{ch},\ell} = 1$. Therefore, the integral w.r.t. $\mathbf{z}_{\mathrm{ch},\ell}$ in the second summand of (2.9) can be separated from the double integral and be removed. Furthermore, since no transmission is scheduled, $q_{\ell+1}$ can be computed directly based on q_ℓ according to (2.4). The minimand in (2.9) is simplified as

$$C(\mathbf{s}_{\mathrm{ch},\ell}, q_\ell, \mathbf{a})|_{[\mathbf{a}]_1=0, N_{\mathrm{s},\ell}=0} + \gamma \int_{\mathcal{S}} b(\mathbf{s}_{\mathrm{ch},\ell}) V^*(q_{\ell+1}, b(\mathbf{s}_{\mathrm{ch},\ell+1})) d\mathbf{s}_{\mathrm{ch},\ell}. \tag{2.11}$$

The channel state belief $b(\mathbf{s}_{\text{ch},\ell+1})$ can be recursively updated as

$$b(\mathbf{s}_{\text{ch},\ell+1}) \propto \int_{\mathcal{S}} f(\mathbf{s}_{\text{ch},\ell+1}|\mathbf{s}_{\text{ch},\ell}) b(\mathbf{s}_{\text{ch},\ell}) d\mathbf{s}_{\text{ch},\ell}. \tag{2.12}$$

Given the Gaussian assumption in (2.2) and (2.3), the channel state belief in (2.10) and (2.12) can be computed through operating over the mean vectors and the covariance matrices of relevant random vectors using Kalman filtering [44]. Detailed discussions will be provided in Section 2.5.

2.3.2 An Overview of the Proposed Algorithm for Online Adaptive Transmission

Finding the optimal online transmission strategy requires estimation of channel model parameters and online planning at the beginning of each epoch. The model parameter estimation is performed based on channel measurements collected in the past epochs. A recursive estimator is desirable for online implementation, and especially in the presence of temporal variation of UWA channels. Given the model estimation, the optimal transmission strategy can be obtained by solving (2.9). Due to the mix of continuous and discrete random variables, the optimal solution to the BOE is not straightforward. In Section 2.4, we will develop a Monte Carlo sampling approach for online approximation of the optimal solution. In Section 2.5, an algorithm will be

designed to recursively estimate the unknown model parameter set $\mathbf{\Theta}$ and track the channel state.

At the outset, an overview of the proposed algorithm is in the following. At the beginning of epoch ℓ, the belief state $b(\mathbf{s}_{\text{ch},\ell})$ is computed recursively via (2.10) or (2.12), based on the parameter set estimation $\hat{\mathbf{\Theta}}_{\ell-1}$, the belief state $b(\mathbf{s}_{\text{ch},\ell-1})$, and the observation $\mathbf{z}_{\text{ch},\ell-1}$. The queue length q_ℓ can be observed. Based on the current knowledge of the system state and the channel model estimation, the optimal transmission strategy (i.e. action) can be obtained by solving (2.9). The transmitter applies the obtained transmission strategy. At the end of the epoch, the transmitter collects possible feedback from the receiver. Based on the observation $\mathbf{z}_{\text{ch},\ell}$ and the previous model estimation $\hat{\mathbf{\Theta}}_{\ell-1}$, the transmitter updates the channel model estimation, denoted by $\hat{\mathbf{\Theta}}_\ell$. The belief state $b(\mathbf{s}_{\text{ch},\ell})$, the observation $\mathbf{z}_{\text{ch},\ell}$, and the model estimation $\hat{\mathbf{\Theta}}_\ell$ will be used to compute the belief state $b(\mathbf{s}_{\text{ch},\ell+1})$ in the next epoch. The above process is repeated for each epoch.

2.4 Monte Carlo Sampling for Online Approximation

The mix of continuous and discrete random variables in the BOE (2.9) makes it intractable to solve. In this section, we develop a Monte Carlo sampling-based approach [49] to approximate the value function and to find a near-optimal solution. The approach is also known as Monte Carlo planning.

2.4.1 Value Function Approximation

The BOE in (2.9) has a recursive form. Given an estimation of the model parameters $\hat{\boldsymbol{\Theta}}$, sampling-based methods [50] can be applied to approximate the value function recursively through constructing a state-action tree (see Fig. 2.2 for an illustration, details provided later). The approximation accuracy increases as the number of samples in the state-action tree increases, which however, incurs higher computational complexity.

In this work, the idea of sparse sampling [49] is applied during the state-action tree construction. To guide the selection of "important" samples, a linear regression (LR) method [51] is introduced to approximate the value function of the system state based

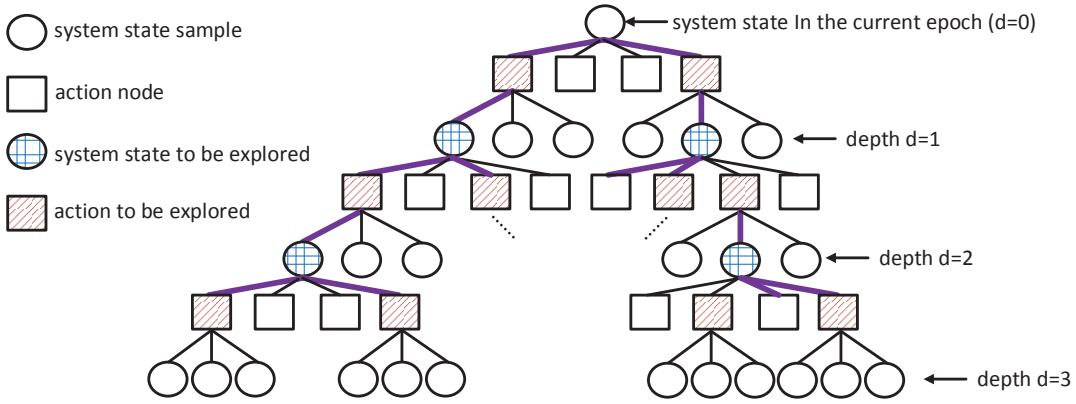


Figure 2.2: An illustration of the state-action tree for online planning, with the tree depth $D = 3$. There are 4 actions in the action space $\mathcal{A}$. At depth d, $N_{\mathrm{o}} = 3$ system state samples are drawn based on the action and the system state at depth $(d-1)$. $N_{\mathrm{a}} = 2$ actions and 1 child system state node are further explored at each depth.

on past value function approximations. Specifically, for the system state $\{q, b(\mathbf{s}_{\mathrm{ch}}), \hat{\mathbf{\Theta}}\}$, denote $\mathbf{x}$ as a vector stacked by q and the scalar elements in the mean vector and the covariance matrix of the channel belief state $b(\mathbf{s}_{\mathrm{ch}})$. The value function can be approximated as

$$V(\mathbf{x}; \boldsymbol{\phi}) = \phi_0 + \mathbf{x}^{\mathrm{T}} \boldsymbol{\phi}_1, \tag{2.13}$$

where $\boldsymbol{\phi}^{\mathrm{T}} := [\phi_0, \boldsymbol{\phi}_1^{\mathrm{T}}]$ is the LR coefficient vector[2]. The LR coefficient vector can be updated via the stochastic gradient decent method [51] based on past value function approximations.

The proposed Monte Carlo sampling approach has two steps. The first step is to construct a state-action planning tree, as depicted in Fig. 2.2. The second step is to approximate the value function recursively based on the state-action tree, as described

[2]For the elements in $\mathbf{x}$ which have relatively higher orders of magnitude, they can be multiplied by constants to reduce their orders of magnitude. For example, the values of μ and q are multiplied by 0.1 and $1/r_{\mathrm{g}}$, respectively, in Section 2.6 for the LR.

Algorithm II-1 Value function approximation: $\mathcal{V}(q, b, \hat{\mathbf{\Theta}}, d, \gamma, \alpha, \beta, \lambda)$

Input: Discount factor γ, temporal difference (TD) learning rate α, learning rate in the linear regression (LR) β, regularization parameter λ, current planning depth d, and system state $(q, b, \hat{\mathbf{\Theta}})$

Set the LR coefficient vector $\boldsymbol{\phi}^{\mathrm{T}} := [\phi_0, \boldsymbol{\phi}_1^{\mathrm{T}}]$ as a global parameter

Output: Approximated value function V_{opt}

Assign an integer value to D ($D > 0$) and set V_{opt}=$+\infty$
if $d = D$ **then**
    For system state $(q, b, \hat{\mathbf{\Theta}})$, set $\mathbf{x}$ as a vector consisting of q and all the scalar elements in the mean vector and the covariance matrix of the channel belief state b
    return $V(\mathbf{x}; \boldsymbol{\phi}) = \phi_0 + \mathbf{x}^{\mathrm{T}}\boldsymbol{\phi}_1$
for $i = 1$ to $|\mathcal{A}|$ **do**
    Select an action $\mathbf{a}$ from the action space $\mathcal{A}$ without replacement
    Compute the expected immediate cost c_i, and set $v_i = c_i$
    for $j = 1$ to N_{o} **do**
        Obtain a state sample $(q', b', \hat{\mathbf{\Theta}}')$ according to action $\mathbf{a}$ based on Algorithm II-2
        For $(q', b', \hat{\mathbf{\Theta}}')$, set $\mathbf{x}'_{ij}$ as a vector consisting of q' and all the scalar elements in the mean vector and the covariance matrix of the channel belief state b'
        $v_i \leftarrow v_i + \frac{\gamma}{N_{\mathrm{o}}} V(\mathbf{x}'_{ij}; \boldsymbol{\phi})$
Sort elements in $\{v_1, v_2, \cdots, v_{|\mathcal{A}|}\}$ in an increasing order as $\{v_{(1)}, v_{(2)}, \cdots, v_{(|\mathcal{A}|)}\}$
for $i = 1$ to N_{a} **do**
    Choose action $\mathbf{a}$ yielding $v_{(i)}$
    Randomly select a state sample $(q', b', \hat{\mathbf{\Theta}}')$ obtained after taking action $\mathbf{a}$
    Perform the TD learning:
    $v_{(i)} \leftarrow v_{(i)} + \alpha(c_{(i)} + \gamma \mathcal{V}(q', b', \hat{\mathbf{\Theta}}', d+1, \gamma, \alpha, \beta, \lambda) - v_{(i)})$
    Update the LR vector:
    $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} - \beta(V(\mathbf{x}; \boldsymbol{\phi}) - v_{(i)}) \nabla_{\boldsymbol{\phi}} V(\mathbf{x}; \boldsymbol{\phi}) - \beta\lambda\boldsymbol{\phi}$
    if $v_{(i)} < V_{\mathrm{opt}}$ **then**
        $V_{\mathrm{opt}} = v_{(i)}$
        $\mathbf{a}_{\mathrm{opt}} = \mathbf{a}$
return V_{opt}

in Algorithm II-1. Details about the two steps are in the following.

Algorithm II-2 Sample the queue state and the belief state in the next epoch

Input: Belief state b, queue length q, action $\mathbf{a}$, and model estimation $\hat{\mathbf{\Theta}}$
Output: Belief state b', and queue length q' in the next epoch, and updated estimated $\hat{\mathbf{\Theta}}'$

if $\mathbf{a}$ indicates transmissions **then**
    Sample the channel state $\mathbf{s}_{\text{ch}}$ from the belief state b
    Sample the observation noise $\mathbf{w}$ from $\mathcal{N}(0, \hat{\mathbf{C}}_{\mathbf{w}})$
    Compute the observation $\mathbf{z} = \mathbf{s}_{\text{ch}} + \mathbf{w}$
    Compute b' via Kalman filtering based on based on b and observation $\mathbf{z}$
    Sample the number of packets that are successfully decoded by the receiver, N_{s}, based on the channel state samples and the action $\mathbf{a}$, according to (2.5).
    Compute the queue length in the next epoch $q' = q + r_{\text{g}} - rN_{\text{s}}$
else
    Set $q' = q + r_{\text{g}}$
    Compute b' based on b via Kalman filtering without channel observation
Update $\hat{\mathbf{\Theta}}'$ as described in Section 2.5
return $(q', b', \hat{\mathbf{\Theta}})$

2.4.1.1 State-action Tree Construction

Given a root node which represents the current system state, the state-action tree is constructed by sequentially drawing samples of actions and samples of the system states up to a certain planning depth (denoted by D). Specifically,

- Let the current system state described by a triplet $(q, b, \hat{\mathbf{\Theta}})$ be the root state node of the state-action tree, where q is the queue length, and b is the channel belief state;

- For each system state node (including the root node) in the state-action tree, a small number (N_{a}) of actions which yield less approximated expected costs

will be selected to expand the tree. To do so, one first enumerates all the actions in the action space. For each enumerated action, a number (N_{o}) of child nodes describing the system states in the next epoch can be obtained through drawing samples of the channel state, the observations of the channel state, and the number of successfully delivered packets; see Algorithm II-2. The value of each child system state node can be approximated by the LR (c.f. Lines 8 to 11 in Algorithm II-1). The expected cost induced by each action can be approximated by summing up the expected immediate cost and the averaged value of its child system state nodes.

The expected immediate cost of each action can be computed by drawing a sufficient number of channel samples according to the belief state. The immediate cost corresponding to each channel sample can be obtained based on the packet error rate of the channel sample according to (2.5) and (2.7). The averaged immediate costs based on all samples yields the expected immediate cost.

- For each action to be further explored via the tree expansion, for computational efficiency, only one of its child nodes is randomly selected and serves as the system state to be explored in the next epoch.

The above process is repeated until the tree reaches the maximal planning depth, namely, the maximal number of future epochs to be evaluated. The values of D, N_{a}, and N_{o} can be determined to strike a balance between the approximation accuracy

and the computational complexity. Benefited from the LR-based value function approximation and the sparse sampling, the structure of the state-action tree can be much simplified compared to the case when all the actions or a large amount of child system state samples are explored to reach similar approximation accuracy.

2.4.1.2 Value Function Calculation

The value of the root state node (i.e., the current system state), can be calculated by propagating the values of all the child nodes in the state-action tree to the root node. Specifically,

- The value of a particular system state node at the planning depth d ($d < D$) is set as the minimal expected cost induced by the selected actions to be explored, and the action with the minimal expected cost is taken as the optimal action. For the system state nodes at the tree leaves, their values are approximated by the LR (c.f. Lines 2 to 4 in Algorithm II-1).

- For each action, we follow the concept of the temporal difference (TD) learning [18] to approximate its expected cost (as shown in Line 16 of Algorithm II-1), based on its expected immediate cost, the value of its child system state node in the state-action tree, and the approximated cost obtained via the LR method (c.f. Lines 8 to 11 in Algorithm II-1). Compared to the method that calculates

the expected cost as the summation of the expected immediate cost and the value of the child system state node, the above method can exploit the historical value function approximation results obtained via the LR to achieve higher approximation accuracy.

2.4.2 Computational Complexity

Denote K as the total number of channel state samples to calculate the expected immediate cost of each action and C_{PER} as the complexity of calculating the PER. The computational complexity of the expected immediate cost is $C_{\mathrm{cur}} = O(KC_{\mathrm{PER}})$. The computational complexity to sample the triplet in Algorithm II-2 in the worst case, namely, every action indicating packet transmissions, is $C_2 = O(C_{\mathrm{est}} + C_{\mathrm{PER}})$ where C_{est} is the computational complexity for the channel model estimation. The complexity of Algorithm II-1 in the worst case is $C_1 = O(|\mathcal{A}|N_{\mathrm{a}}^{D-1}C_{\mathrm{cur}}^{D} + |\mathcal{A}|N_{\mathrm{a}}^{D-1}N_{\mathrm{o}}C_2)$. Hence, the total complexity of the algorithm in the worst case is $O(C_1 + C_{\mathrm{est}})$.

2.5 Recursive Estimation of Unknown Channel Model Parameters

To facilitate online implementation, we will develop a low-complexity recursive algorithm to estimate the parameter set $\mathbf{\Theta}$ and the channel state vector based on the sequentially obtained observations $\{\mathbf{z}_{\text{ch},\ell}\}$. For notation convenience, we denote $\mathbf{z}_{\text{ch},\ell_1}^{\ell_2} := \{\mathbf{z}_{\text{ch},\ell_1}, \cdots, \mathbf{z}_{\text{ch},\ell_2}\}$ and $\mathbf{s}_{\text{ch},\ell_1}^{\ell_2} := \{\mathbf{s}_{\text{ch},\ell_1}, \cdots, \mathbf{s}_{\text{ch},\ell_2}\}$.

At time epoch ℓ, the unknown parameters can be estimated by maximizing the log-likelihood function with respect to the complete data set $L_\ell(\mathbf{\Theta}) := \ln f(\mathbf{z}_{\text{ch},0}^{\ell}, \mathbf{s}_{\text{ch},-1}, \mathbf{s}_{\text{ch},0}^{\ell}|\mathbf{\Theta})$. However, the channel state process $\{\mathbf{s}_{\text{ch},\ell'}\}$ is not observable. Instead, the expectation-maximization (EM) algorithm [51] can be used, which estimates the unknown parameters iteratively through an expectation step and a maximization step. Given a parameter set estimation $\hat{\mathbf{\Theta}}$, in the expectation step, the expectation of the log-likelihood function can be approximated as

$$\mathbb{E}\left[L_\ell(\mathbf{\Theta})|\hat{\mathbf{\Theta}}\right] = \int \left[\ln f(\mathbf{z}_{\text{ch},0}^{\ell}, \mathbf{s}_{\text{ch},-1}, \mathbf{s}_{\text{ch},0}^{\ell}|\mathbf{\Theta})\right] f\left(\mathbf{s}_{\text{ch},-1}^{\ell}|\mathbf{z}_{\text{ch},0}^{\ell}, \hat{\mathbf{\Theta}}\right) d\mathbf{s}_{\text{ch},-1}^{\ell}. \tag{2.14}$$

The parameter set estimation can be updated in the maximization step as $\hat{\mathbf{\Theta}}^{\text{new}} = \arg\max \mathbb{E}\big[L_\ell(\mathbf{\Theta})|\hat{\mathbf{\Theta}}\big]$.

The algorithm, however, requires processing within each iteration the data in the current and all the past epochs, hence is not amenable to online implementation. We next introduce several approximations, and then develop an EM-type and low-complexity recursive algorithm that estimates the parameter set $\mathbf{\Theta}$ in each epoch iteratively based on the new observation vector and the parameter estimation in the last epoch. We denote $\hat{\mathbf{\Theta}}_{\ell'}$ as the estimation at epoch ℓ'.

2.5.1 Approximation for Recursive Operation

Consider that

$$\ln f(\mathbf{z}_{\text{ch},0}^{\ell}, \mathbf{s}_{\text{ch},-1}^{\ell}|\mathbf{\Theta}) = \ln f(\mathbf{z}_{\text{ch},\ell}, \mathbf{s}_{\text{ch},\ell}|\mathbf{s}_{\text{ch},\ell-1}, \mathbf{\Theta}) + \ln f(\mathbf{z}_{\text{ch},0}^{\ell-1}, \mathbf{s}_{\text{ch},-1}^{\ell-1}|\mathbf{\Theta}). \tag{2.15}$$

The expectation in (2.14) can be decomposed as

$$\begin{aligned}\mathbb{E}\big[L_{\ell}(\mathbf{\Theta})|\hat{\mathbf{\Theta}}\big] = &\int [\ln f(\mathbf{s}_{\text{ch},-1}|\mathbf{\Theta})] f(\mathbf{s}_{\text{ch},-1}|\mathbf{z}_{\text{ch},0}^{\ell}, \hat{\mathbf{\Theta}}) d\mathbf{s}_{\text{ch},-1} \\ &+ \sum_{\ell'=0}^{\ell} \int [\ln f(\mathbf{s}_{\text{ch},\ell'}, \mathbf{z}_{\text{ch},\ell'}|\mathbf{s}_{\text{ch},\ell'-1}, \mathbf{\Theta})] \\ &\times f\left(\mathbf{s}_{\text{ch},\ell'}, \mathbf{s}_{\text{ch},\ell'-1}|\mathbf{z}_{\text{ch},0}^{\ell}, \hat{\mathbf{\Theta}}\right) d\mathbf{s}_{\text{ch},\ell'} d\mathbf{s}_{\text{ch},\ell'-1}. \end{aligned} \tag{2.16}$$

It can be approximated in two steps,

$$\begin{aligned}
\mathbb{E}\big[L_\ell(\boldsymbol{\Theta})|\hat{\boldsymbol{\Theta}}\big] \approx & \int [\ln f(\mathbf{s}_{\mathrm{ch},-1}|\boldsymbol{\Theta})] f(\mathbf{s}_{\mathrm{ch},-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}) d\mathbf{s}_{\mathrm{ch},-1} \\
& + \sum_{\ell'=0}^{\ell} \int [\ln f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{z}_{\mathrm{ch},\ell'}|\mathbf{s}_{\mathrm{ch},\ell'-1}, \boldsymbol{\Theta})] \\
& \times \underbrace{f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell'}, \hat{\boldsymbol{\Theta}})}_{\approx f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}) \text{ in Eq. (2.16)}} d\mathbf{s}_{\mathrm{ch},\ell'} d\mathbf{s}_{\mathrm{ch},\ell'-1}, \qquad (2.17a) \\
\approx & \int [\ln f(\mathbf{s}_{\mathrm{ch},-1}|\boldsymbol{\Theta})] f\left(\mathbf{s}_{\mathrm{ch},-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}_{-1}\right) d\mathbf{s}_{\mathrm{ch},-1} \\
& + \sum_{\ell'=0}^{\ell-1} \int [\ln f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{z}_{\mathrm{ch},\ell'}|\mathbf{s}_{\mathrm{ch},\ell'-1}, \boldsymbol{\Theta})] \\
& \times \underbrace{f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell'}, \hat{\boldsymbol{\Theta}}_{\ell'})}_{\approx f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell'}, \hat{\boldsymbol{\Theta}}) \text{ in Eq. (2.17a)}} d\mathbf{s}_{\mathrm{ch},\ell'} d\mathbf{s}_{\mathrm{ch},\ell'-1} \\
& + \int [\ln f(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \boldsymbol{\Theta})] \\
& \times f(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}) d\mathbf{s}_{\mathrm{ch},\ell} d\mathbf{s}_{\mathrm{ch},\ell-1}, \qquad (2.17b)
\end{aligned}$$

where the expectation of $\ln f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{z}_{\mathrm{ch},\ell'}|\mathbf{s}_{\mathrm{ch},\ell'-1}, \boldsymbol{\Theta})$ in (2.17a) is performed with respect to $f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell'}, \hat{\boldsymbol{\Theta}})$ instead of $f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}})$, and in (2.17b), the expectation of $[\ln f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{z}_{\mathrm{ch},\ell'}|\mathbf{s}_{\mathrm{ch},\ell'-1}, \boldsymbol{\Theta})]$ can be computed at epoch ℓ' based on $f(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}_{\mathrm{ch},0}^{\ell'}, \hat{\boldsymbol{\Theta}}_{\ell'})$. The above approximations enable recursive computation of the summation on the right side of (2.17b).

One more approximation is made for recursive computation of the PDF

$f(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}})$. Note that

$$f(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}}) = \frac{1}{c_0} f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \hat{\mathbf{\Theta}}) f(\mathbf{s}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \hat{\mathbf{\Theta}}) f(\mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell-1}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}}) \quad (2.18)$$

where c_0 is a normalization constant. We approximate the joint PDF by

$$\tilde{f}(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}}) := \frac{1}{c'_0} f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \hat{\mathbf{\Theta}}) f(\mathbf{s}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \hat{\mathbf{\Theta}}) \tilde{f}(\mathbf{s}_{\mathrm{ch},\ell-1}|\hat{\mathbf{\Theta}}_{\ell-1}), \quad (2.19)$$

through replacing $f(\mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell-1}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}})$ by $\tilde{f}(\mathbf{s}_{\mathrm{ch},\ell-1}|\hat{\mathbf{\Theta}}_{\ell-1})$ in (2.18), where$\tilde{f}(\mathbf{s}_{\mathrm{ch},\ell'}|\hat{\mathbf{\Theta}}_{\ell'})$ is defined as the marginalization of $\tilde{f}(\mathbf{s}_{\mathrm{ch},\ell'}, \mathbf{s}_{\mathrm{ch},\ell'-1}|\mathbf{z}^{\ell'}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}}_{\ell'})$ with respect to $\mathbf{s}_{\mathrm{ch},\ell'}$, and c'_0 is a normalization constant.

Finally, based on (2.17b) and (2.19), the expectation $\mathbb{E}\big[L_\ell(\mathbf{\Theta})|\hat{\mathbf{\Theta}}\big]$ is approximated by $Q_\ell(\mathbf{\Theta}|\hat{\mathbf{\Theta}})$ which is recursively defined as

$$\begin{aligned} Q_\ell(\mathbf{\Theta}|\hat{\mathbf{\Theta}}) = & \gamma_{\mathrm{ch}} Q_{\ell-1}(\mathbf{\Theta}|\hat{\mathbf{\Theta}}_{\ell-1}) + \\ & \int \left[\ln f(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \mathbf{\Theta})\right] \tilde{f}(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell-1}|\mathbf{z}^{\ell}_{\mathrm{ch},0}, \hat{\mathbf{\Theta}}) d\mathbf{s}_{\mathrm{ch},\ell} d\mathbf{s}_{\mathrm{ch},\ell-1}, \quad (2.20) \end{aligned}$$

where $\gamma_{\mathrm{ch}} \in (0, 1]$ is a forgetting factor that accounts for the temporal variation of unknown parameters. Based on (2.20), the expectation and maximization operations in the EM algorithm can be applied for recursive and iterative parameter estimation and channel tracking, as described in the next subsection.

2.5.2 Recursive Model and Channel State Estimation

Denote $\hat{\mathbf{\Theta}}_{\ell}^{(i)} = \{\hat{\mathbf{A}}_{\ell}^{(i)}, \hat{\mathbf{C}}_{\mathrm{w},\ell}^{(i)}, \hat{\mathbf{C}}_{\mathrm{v},\ell}^{(i)}\}$ as the estimation of the unknown parameters in the ith iteration at epoch ℓ. The parameter estimation can be updated via maximizing $Q_{\ell}(\mathbf{\Theta}|\hat{\mathbf{\Theta}}_{\ell}^{(i)})$. Note that $f(\mathbf{z}_{\mathrm{ch},\ell}, \mathbf{s}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \mathbf{\Theta}) = f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{\Theta}) f(\mathbf{s}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \mathbf{\Theta})$. Substitute

$$f(\mathbf{z}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{\Theta}) \sim \mathcal{N}(\mathbf{s}_{\mathrm{ch},\ell}, \mathbf{C}_{\mathrm{w}}),$$

$$f(\mathbf{s}_{\mathrm{ch},\ell}|\mathbf{s}_{\mathrm{ch},\ell-1}, \mathbf{\Theta}) \sim \mathcal{N}(\mathbf{A}\mathbf{s}_{\mathrm{ch},\ell-1}, \mathbf{C}_{\mathrm{v}})$$

into the log-likelihood function in (2.20). Set the partial derivative of $Q_{\ell}(\mathbf{\Theta}|\hat{\mathbf{\Theta}}_{\ell}^{(i)})$ with respect to each unknown parameter to zero. A set of recursive equations can be obtained,

$$\hat{\mathbf{A}}_{\ell}^{(i+1)} = \hat{\mathbf{A}}_{\ell-1} + \left(\mathbb{E}[\mathbf{s}_{\mathrm{ch},\ell}\mathbf{s}_{\mathrm{ch},\ell-1}^{\mathrm{T}}] - \hat{\mathbf{A}}_{\ell-1}\mathbb{E}[\mathbf{s}_{\mathrm{ch},\ell-1}\mathbf{s}_{\mathrm{ch},\ell-1}^{\mathrm{T}}]\right)\mathbf{M}_{\ell-1}^{-1}, \tag{2.21a}$$

$$\hat{\mathbf{C}}_{\mathrm{w},\ell}^{(i+1)} = \hat{\mathbf{C}}_{\mathrm{w},\ell-1} + \frac{1-\gamma_{\mathrm{ch}}}{1-\gamma_{\mathrm{ch}}^{\ell}} \times \left\{\mathbb{E}\left[(\mathbf{s}_{\mathrm{ch},\ell} - \hat{\mathbf{A}}_{\ell}^{(i+1)}\mathbf{s}_{\mathrm{ch},\ell-1})(\mathbf{s}_{\mathrm{ch},\ell} - \hat{\mathbf{A}}_{\ell}^{(i+1)}\mathbf{s}_{\mathrm{ch},\ell-1})^{\mathrm{T}}\right] - \hat{\mathbf{C}}_{\mathrm{w},\ell-1}\right\}, \tag{2.21b}$$

$$\hat{\mathbf{C}}_{\mathrm{v},\ell}^{(i+1)} = \hat{\mathbf{C}}_{\mathrm{v},\ell-1} + \frac{1-\gamma_{\mathrm{ch}}}{1-\gamma_{\mathrm{ch}}^{\ell+1}}\left\{\mathbb{E}\left[(\mathbf{z}_{\mathrm{ch},\ell} - \mathbf{s}_{\mathrm{ch},\ell})(\mathbf{z}_{\mathrm{ch},\ell} - \mathbf{s}_{\mathrm{ch},\ell})^{\mathrm{T}}\right] - \hat{\mathbf{C}}_{\mathrm{v},\ell-1}\right\}, \tag{2.21c}$$

where an auxiliary matrix is defined as

$$\mathbf{M}_{\ell-1} := \gamma_{\text{ch}}\mathbf{M}_{\ell-2} + \mathbb{E}[\mathbf{s}_{\text{ch},\ell-1}\mathbf{s}_{\text{ch},\ell-1}^{\text{T}}], \tag{2.22}$$

and the expectations are performed with respect to $\tilde{f}(\mathbf{s}_{\text{ch},\ell}, \mathbf{s}_{\text{ch},\ell-1}|\mathbf{z}_{\text{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}_{\ell}^{(i)})$.

The expectations in (2.21) and (2.22) can be computed via performing marginalization of $\tilde{f}(\mathbf{s}_{\text{ch},\ell}, \mathbf{s}_{\text{ch},\ell-1}|\mathbf{z}_{\text{ch},0}^{\ell}, \hat{\boldsymbol{\Theta}}_{\ell}^{(i)})$ (c.f. (2.19)). For convenience, denote $\tilde{f}(\mathbf{s}_{\text{ch},\ell-1}|\hat{\boldsymbol{\Theta}}_{\ell-1}) \sim \mathcal{N}(\boldsymbol{\mu}_{\ell-1}, \mathbf{C}_{\ell-1})$. It can be shown that [51]

$$\mathbb{E}[\mathbf{s}_{\text{ch},\ell}|\hat{\boldsymbol{\Theta}}_{\ell}^{(i)}] = \boldsymbol{\mu}_{\ell}^{(i)} = \hat{\mathbf{A}}_{\ell}^{(i)}\boldsymbol{\mu}_{\ell-1} + \mathbf{K}_{\ell}^{(i)}(\mathbf{z}_{\text{ch}} - \hat{\mathbf{A}}_{\ell}^{(i)}\boldsymbol{\mu}_{\ell-1}) \tag{2.23a}$$

$$\mathbb{E}[\mathbf{s}_{\text{ch},\ell-1}|\hat{\boldsymbol{\Theta}}_{\ell}^{(i)}] = \breve{\boldsymbol{\mu}}_{\ell-1}^{(i)} = \boldsymbol{\mu}_{\ell-1} + \mathbf{J}_{\ell-1}^{(i)}(\boldsymbol{\mu}_{\ell}^{(i)} - \hat{\mathbf{A}}_{\ell}^{(i)}\boldsymbol{\mu}_{\ell-1}) \tag{2.23b}$$

$$\mathbb{E}[\mathbf{s}_{\text{ch},\ell}\mathbf{s}_{\text{ch},\ell}^{\text{T}}|\hat{\boldsymbol{\Theta}}_{\ell}^{(i)}] = \mathbf{C}_{\ell}^{(i)} + \boldsymbol{\mu}_{\ell}^{(i)}\boldsymbol{\mu}_{\ell}^{(i),\text{T}} \tag{2.23c}$$

$$\mathbb{E}[\mathbf{s}_{\text{ch},\ell-1}\mathbf{s}_{\text{ch},\ell-1}^{\text{T}}|\hat{\boldsymbol{\Theta}}_{\ell}^{(i)}] = \breve{\mathbf{C}}_{\ell-1}^{(i)} + \breve{\boldsymbol{\mu}}_{\ell-1}^{(i)}\breve{\boldsymbol{\mu}}_{\ell-1}^{(i),\text{T}} \tag{2.23d}$$

$$\mathbb{E}[\mathbf{s}_{\text{ch},\ell}\mathbf{s}_{\text{ch},\ell-1}^{\text{T}}|\hat{\boldsymbol{\Theta}}_{\ell}^{(i)}] = \mathbf{C}_{\ell}^{(i)}\mathbf{J}_{\ell-1}^{(i),\text{T}} + \boldsymbol{\mu}_{\ell}^{(i)}\breve{\boldsymbol{\mu}}_{\ell-1}^{(i),\text{T}} \tag{2.23e}$$

where $\mathbf{K}_{\ell}^{(i)} = \mathbf{P}_{\ell}^{(i)}\big(\hat{\mathbf{C}}_{\text{v}}^{(i)} + \mathbf{P}_{\ell}^{(i)}\big)^{-1}$ with $\mathbf{P}_{\ell}^{(i)} = \hat{\mathbf{A}}_{\ell}^{(i)}\mathbf{C}_{\ell-1}\hat{\mathbf{A}}_{\ell}^{(i),\text{T}} + \hat{\mathbf{C}}_{\text{w}}^{(i)}$, $\mathbf{J}_{\ell-1}^{(i)} = \mathbf{C}_{\ell-1}\hat{\mathbf{A}}_{\ell}^{(i),\text{T}}\big(\mathbf{P}_{\ell}^{(i)}\big)^{-1}$, $\mathbf{C}_{\ell}^{(i)} = (\mathbf{I} - \mathbf{K}_{\ell}^{(i)})\mathbf{P}_{\ell}^{(i)}$, and $\breve{\mathbf{C}}_{\ell-1}^{(i)} = \mathbf{C}_{\ell-1} + \mathbf{J}_{\ell-1}^{(i)}(\mathbf{C}_{\ell}^{(i)} - \mathbf{P}_{\ell}^{(i)})\mathbf{J}_{\ell-1}^{(i),\text{T}}$.

In summary, when $\mathbf{z}_{\text{ch},\ell}$ is available at the end of epoch ℓ, the iterative model parameter estimation can be initialized as $\hat{\boldsymbol{\Theta}}_{\ell}^{(0)} = \hat{\boldsymbol{\Theta}}_{\ell-1}$. The expectation and maximization operations are performed iteratively based on (2.23) and (2.21). Consider that the

operation terminates after a pre-determined number of iterations, denoted by N_{iter}. We set $\hat{\boldsymbol{\Theta}}_\ell = \hat{\boldsymbol{\Theta}}_\ell^{(N_{\text{iter}})}$ and $\tilde{f}(\mathbf{s}_{\text{ch},\ell}|\hat{\boldsymbol{\Theta}}_\ell) \simeq \mathcal{N}(\boldsymbol{\mu}_\ell, \mathbf{C}_\ell)$ with $\boldsymbol{\mu}_\ell = \boldsymbol{\mu}_\ell^{(N_{\text{iter}})}$ and $\mathbf{C}_\ell = \mathbf{C}_\ell^{(N_{\text{iter}})}$, which will be used for the operation in the next epoch. If no transmission is scheduled in epoch ℓ, namely, $\mathbf{z}_{\text{ch},\ell}$ is an empty set, no model parameter estimation is needed. One can set $\hat{\boldsymbol{\Theta}}_\ell = \hat{\boldsymbol{\Theta}}_{\ell-1}$, $\boldsymbol{\mu}_\ell = \hat{\mathbf{A}}_{\ell-1}\boldsymbol{\mu}_{\ell-1}$ and $\mathbf{C}_\ell = \hat{\mathbf{A}}_{\ell-1}\mathbf{C}_{\ell-1}\hat{\mathbf{A}}_{\ell-1}^{\text{T}} + \hat{\mathbf{C}}_{\text{w},\ell-1}$. In both cases, the *a posteriori* PDF $\tilde{f}(\mathbf{s}_{\text{ch},\ell}|\hat{\boldsymbol{\Theta}}_\ell)$ and the conditional PDF $\tilde{f}(\mathbf{s}_{\text{ch},\ell+1}|\mathbf{s}_{\text{ch},\ell}, \hat{\boldsymbol{\Theta}}_\ell)$ can be used to compute the belief state $b(\mathbf{s}_{\text{ch},\ell+1})$ according to (2.10) or (2.12).

Remark 2: The proposed algorithm does not guarantee that the Nakagami-fading parameter $m \geq 1/2$. In the Monte Carlo sampling method for online approximation, we only draw samples of m which are greater than $1/2$ based on the channel belief state.

2.6 Algorithm Evaluation

The proposed algorithm is evaluated using data sets collected from two experiments: one is the Surface Processes and Acoustic Communications Experiment (SPACE08), and the other is an experiment conducted in the Keweenaw Waterway near Michigan Tech in Nov. 2014 (KW-NOV14).

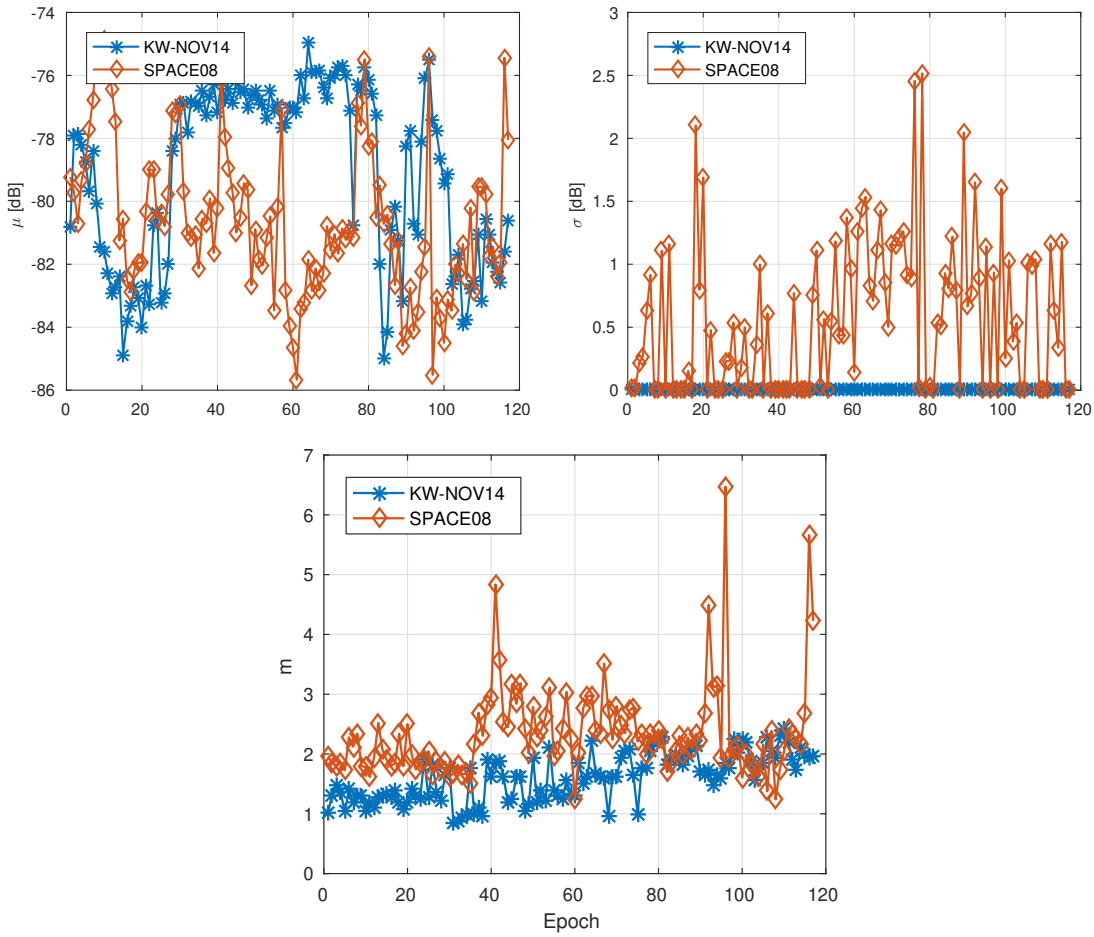


Figure 2.3: Estimated parameters $\{\mu, \sigma, m\}$ in two experiments. In KW-NOV14, the estimated σ's are on the order of 10^{-3}.

2.6.1 Experiment Description

The SPACE08 experiment was conducted near the coast of Martha's Vineyard, MA, from Oct. 14 to Nov. 1, 2008. We consider the data collected by a receiver which is 200 meters away from the transmitter, from Julian date 287 to Julian date 302. Due to the appearance of severe weather conditions during the experiment, some of the data files were damaged hence are excluded for algorithm evaluation. A waveform of

10 seconds was transmitted every two hours from the source to the receiver, leading to 12 transmissions per day. The waveform consists of 60 signaling blocks within the frequency band [8, 18] kHz, and each block has 672 symbols. In this work, we take each transmission as one epoch and take each signaling block as one packet. There are 117 epochs in total. The channel distribution parameters μ, σ and m within each epoch are estimated via the method of moments [44] based on the received SNR samples obtained within that epoch. The evolution of the distribution parameters is shown in Fig. 2.3.

The KW-NOV14 experiment was held in the Keweenaw Waterway adjacent to Michigan Tech from Nov. 22 to Nov. 28, 2014 when the water surface was covered by a thin layer of ice. The distance between the transmitter and the receiver is 312 m. A waveform of about 9 seconds was transmitted every 15 minutes. The waveform consists of 20 signaling blocks within the frequency band [14, 20] kHz, and each block has 672 symbols. Similar to SPACE08, we take each transmission as one epoch and take each signaling block as one packet. A total of 117 epochs are used for algorithm evaluation. Artificial Gaussian noise is added to the received signal in KW-NOV14 such that the two experiments have similar average channel losses over all the epochs. Evolution of the KW-NOV14 channel distribution parameters is shown in Fig. 2.3.

Comparing the channels in the two experiments, one can see that the channel in SPACE08 varies faster than that in KW-NOV14 due to a larger time interval between

Table 2.1
Transmission Modes.

Mode Index	Coding rate	Modulation	TSNR
1	N/A	N/A	0
2	1/2	BPSK	76 dB
3	1/2	BPSK	79 dB
4	1/2	QPSK	79 dB
5	1/2	BPSK	82 dB
6	1/2	QPSK	82 dB
7	3/4	QPSK	85 dB
8	3/4	QPSK	88 dB

two consecutive transmissions. Especially about KW-NOV14, the mean of the channel lognormal shadowing per epoch (μ) is quite stable from epoch 30 to 75, and the values of σ on the order of 10^{-3} reveals very slow variation.

2.6.2 Emulation Setup and Performance Metric

We consider 8 transmission modes as listed in Table 2.1. Mode 1 refers to no transmission. There are five non-zero discrete transmission power levels according to the listed transmission SNRs (TSNRs) (P_{tx}/N_0). We set the ambient noise level using an empirical formula N_0 [dB] $= 55 + 10\log_{10}(\text{bandwidth})$ *re* $1\mu\text{Pa}^2$ [52], which leads to 94.9 dB for SPACE08 and 92.8 dB for KW-NOV14. For a given transmission mode and a channel parameter triplet $\{\mu, \sigma, m\}$, the PER is computed using an information-theoretic method [46, Eq. (4)].

We define the cost function as

$$C(\mathbf{s}_\ell, \mathbf{a}_\ell) = \log_2 (1+P_\ell/P_{\max}) + \big(q_\ell - r_\ell N_{\mathrm{s},\ell}\big)/r_{\max}, \tag{2.24}$$

where P_ℓ is the transmission power in the ℓth epoch in Watts, $P_{\max}$ is the maximal transmission power in Watts, and $r_{\max}$ is the maximal amount of information bits that can be carried during one epoch. According to Table 2.1, $r_{\max}$ can be computed based on the mode with the highest data rate, namely, Mode 8, as $r_{\max} = 672 \times \frac{3}{4} \times \log_2 4 \times N_{\mathrm{pa}}$, where 672 is the number of symbols per packet, and N_{pa} is the total number of packets within one epoch.

The average observed cost is used as the performance metric,

$$\bar{C} = \frac{1}{N_{\mathrm{epoch}}} \sum_{\ell=1}^{N_{\mathrm{epoch}}} C(\mathbf{s}_\ell, \mathbf{a}_\ell), \tag{2.25}$$

where N_{epoch} is the total number of epochs in the algorithm evaluation.

To establish a performance upper bound, we consider a genie-aided transmission scheme with non-causal and perfect knowledge. It assumes that at the beginning of each epoch, the transmitter knows the number of successfully delivered packets corresponding to each transmission action in the current and all the future epochs. With the above knowledge, the system state only consists of the queue state. The

optimal action selection can be formulated as a dynamic programming (DP) problem,

$$V^*_{\text{Genie}}(q_\ell) = \min_{\mathbf{a}\in\mathcal{A}} \left[C(q_\ell, \mathbf{a}) + \gamma V^*_{\text{Genie}}(q_{\ell+1})\right], \tag{2.26}$$

where $C(q_\ell, \mathbf{a})$ is defined as in (2.24), and $q_{\ell+1}$ and q_ℓ are related as in (2.4) with perfect knowledge of $N_{\text{s},\ell}$ for a given $\mathbf{a}$. The optimization problem (2.26) is essentially a deterministic DP problem. However, the DP solver cannot be applied to (2.26) directly due to the curse of dimensionality [18] induced by the large total number of epochs and a large queue state space. To obtain a near-optimal solution, we modify Algorithm II-1 to approximate the value function in (2.26). Specifically, to approximate the expected cost induced by one action (c.f. Lines 8 to 11 in Algorithm II-1), the process of drawing system state samples is replaced by using the true system state directly. Correspondingly, the TD learning is performed based on the true system state instead of a system state sample in the next epoch (c.f. Line 16 in Algorithm II-1).

2.6.3 General Results

We set the data arrival rate $r_{\text{g}} = 20$ kilobits per epoch for SPACE08 and $r_{\text{g}} = 6$ kilobits per epoch for KW-NOV14. For an epoch with a small queue length, if the number of encoded data packets according to a chosen transmission mode is less than

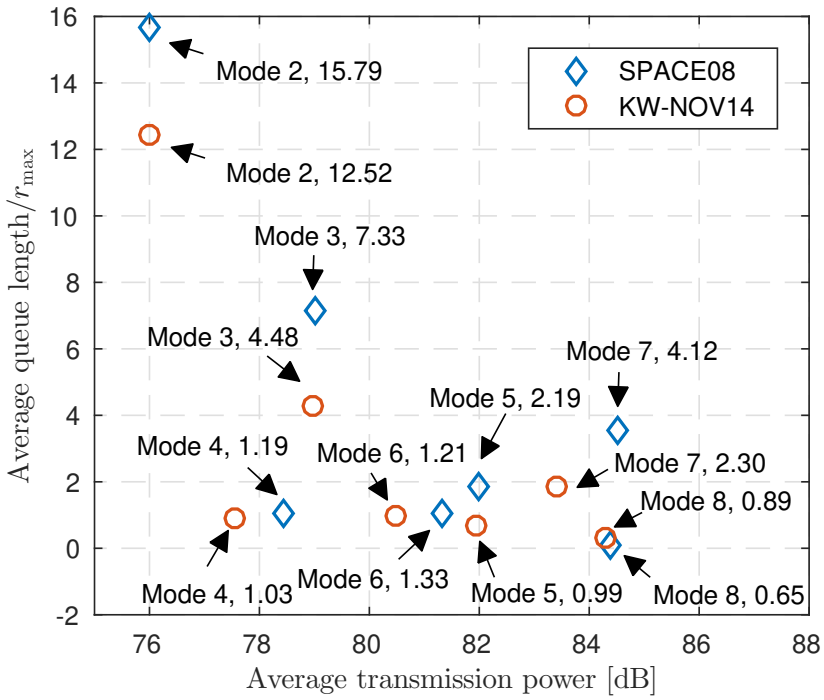


Figure 2.4: The performance of fixed-mode transmissions. The number next to each mode is the average cost calculated based on the cost function in (2.24).

the number of time slots within that epoch (see Fig. 2.1), the remaining time slots will be used to transmit dummy packets at a very low power level (with TSNR = 70 dB) for the purpose of channel probing. The average packet transmission power will be used to calculate the cost defined in (2.24).

To shed light on the tradeoff between energy consumption and information delivery latency, Fig. 2.4 depicts the performance of fixed-mode transmissions in both experiments. According to the cost function defined in (2.24), Mode 8 achieves the least average cost in both experiments.

We compare in details the performance of five schemes for the transmission action

selection.

- *Scheme 1*: The genie-aided transmission scheme;
- *Scheme 2*: The proposed online algorithm;
- *Scheme 3*: Randomly select a transmission action from the action space in each epoch;
- *Scheme 4*: Select the action with the least transmission power and rate, namely, Mode 2, in all epochs;
- *Scheme 5*: Select the action with the highest transmission power and rate, namely, Mode 8, in all epochs.

In the proposed algorithm, the number of the child system nodes for each action N_{o}, the number of the actions to be explored N_{a}, and the planning depth D are set to be 3, 3, and 5, respectively. We set the discount factor $\gamma = 0.8$ in both the genie-aided scheme and the proposed algorithm. The unknown channel model parameters in the first epoch $\hat{\mathbf{\Theta}}_0$ are initialized as $\hat{\mathbf{A}}_0 = \hat{\mathbf{C}}_{\mathrm{w},0} = \hat{\mathbf{C}}_{\mathrm{v},0} = \mathrm{diag}([1,1,1])$. We set the forgetting factor $\gamma_{\mathrm{ch}} = 0.8$ and the number of iterations $N_{\mathrm{iter}} = 20$. The learning rates for the TD learning and the LR, i.e., α and β, are set to be 0.01 and 0.01, respectively. The regularization parameter λ is set to be 1. The initial values of all the elements in $\boldsymbol{\phi}$ are set as 0. The number of channel state samples to calculate the expected immediate cost is set to be 100.

Table 2.2
Average Performance using the SPACE08 Data Set.

Scheme Index	1	2	3	4	5
Average Queue Length [kb]	10.9	13.1	254.8	949.6	7.1
Average Transmission Power [dB]	76.2	81.1	81.3	76.0	84.4
Average Cost	0.47	0.58	4.48	15.79	0.65

2.6.3.1 SPACE08

The performance of different schemes is shown in Table 2.2. It can be seen that the proposed algorithm has the least performance gap with the genie-aid method. Schemes 3 and 4 suffer from very large average queue lengths. Compared to the proposed algorithm, Scheme 5 has a smaller average queue length but requires more average transmission power.

The immediate costs per epoch of different schemes are shown in Fig. 2.5. One can see that the immediate cost of the proposed algorithm is close to that of the genie-aided method. With the immediate costs fluctuating with the mean of the channel lognormal shadowing, the proposed algorithm and the genie-aided method are able to maintain low costs when the average channel loss is small (i.e., when μ is large). When the average channel loss is large, the proposed algorithm still can maintain relatively low immediate costs. The immediate costs of Schemes 3 and 4 increase drastically due to the random selection of transmission actions in Scheme 3 and the adoption of the least transmission power and data rate in Scheme 4. Scheme 5 has

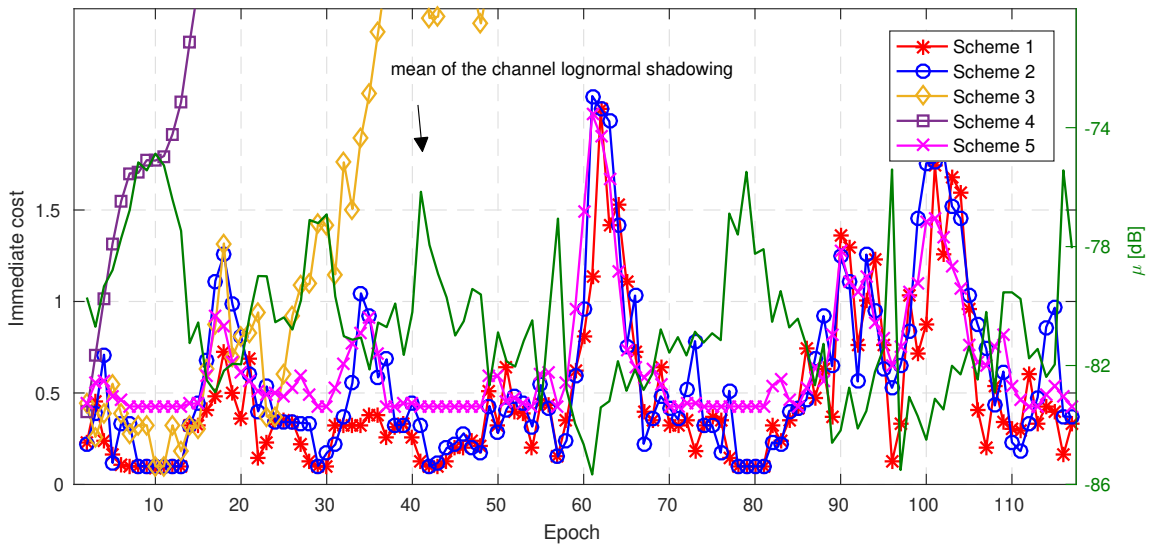


Figure 2.5: SPACE08: The mean of the channel lognormal shadowing and immediate collected costs by different schemes.

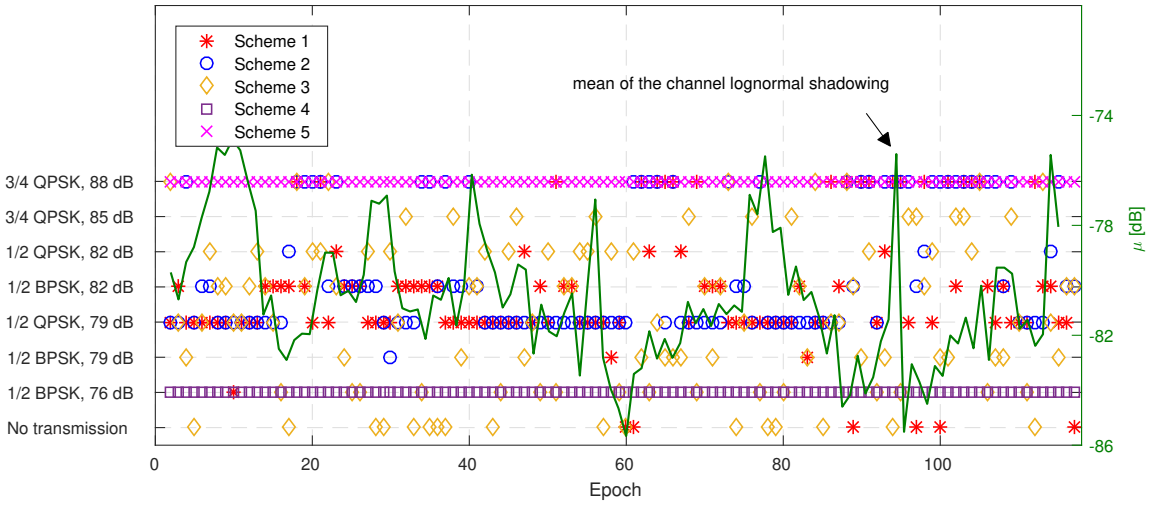


Figure 2.6: SPACE08: The mean of the channel lognormal shadowing and selected actions in different schemes.

larger immediate costs than the proposed algorithm and the genie-aided method in most epochs, due to its adoption of the largest transmission power.

The actions selected by different schemes are shown in Fig. 2.6. The proposed algorithm and the genie-aided method prefer in most epochs the transmission action with

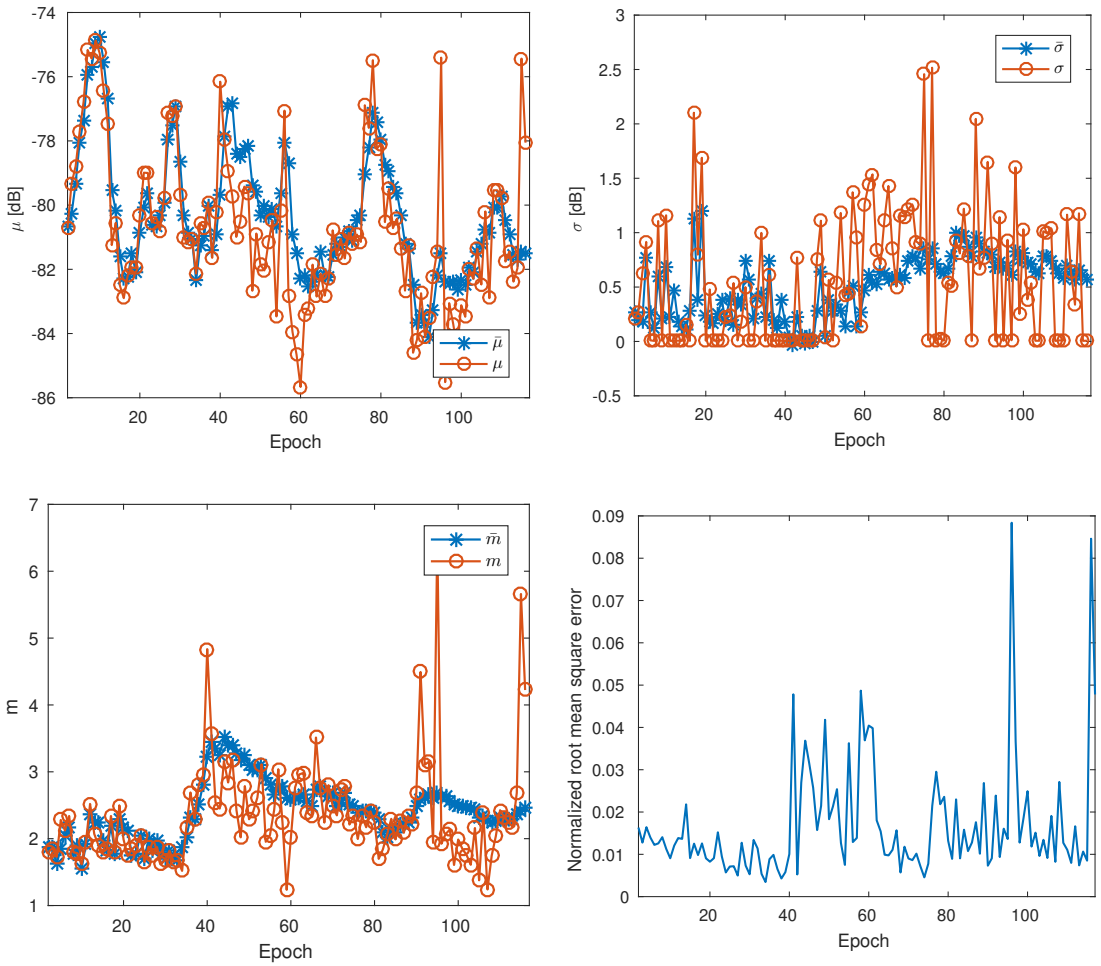


Figure 2.7: SPACE08: Comparison between the mean $(\bar{\mu}, \bar{\sigma}, \bar{m})$ of the estimated channel belief state and the true channel state (μ, σ, m), and the NRMSE.

a moderate transmission power level and a moderate data rate, i.e., 1/2 QPSK and 79 dB. In the epochs with large channel losses, the proposed algorithm opts for the transmission actions with larger transmission power levels to suppress the increase of the data queue length.

The channel state vector estimation and the normalized root mean squared error (NRMSE) of the estimation are depicted in Fig. 2.7. The results reveal that the proposed channel model can capture the channel dynamics reasonably well, and the

Table 2.3
Average Performance using the KW-NOV14 Data Set.

Scheme Index	1	2	3	4	5
Average Queue Length [kilobits]	4.9	11.4	31.3	25.1	7.0
Average Transmission Power [dB]	74.4	75.6	80.8	76.0	84.3
Average Cost	0.40	0.76	1.81	12.52	0.89

NRMSE less than 0.1 in each epoch shows the superior performance of the proposed recursive estimation algorithm.

2.6.3.2 KW-NOV14

The performance of different schemes is shown in Table 2.3. It can be seen that the proposed algorithm has the least performance gap with the genie-aid method. Schemes 3 and 4 suffer from large average queue lengths. Although Scheme 5 has a small average queue length, it requires the most average transmission power among all schemes.

The immediate costs and actions of different schemes are shown in Figs. 2.8 and 2.9, respectively. The immediate cost of Scheme 4 grows drastically due to its adoption of the least transmission power and data rate. Schemes 3 and 5 have larger immediate costs than the proposed algorithm and the genie-aided method in most epochs. The immediate cost of the proposed algorithm is close to that of the genie-aided method. A large performance gap between the proposed algorithm and the genie-aided method can be observed during epochs 8 to 32. Due to large channel dynamics and large

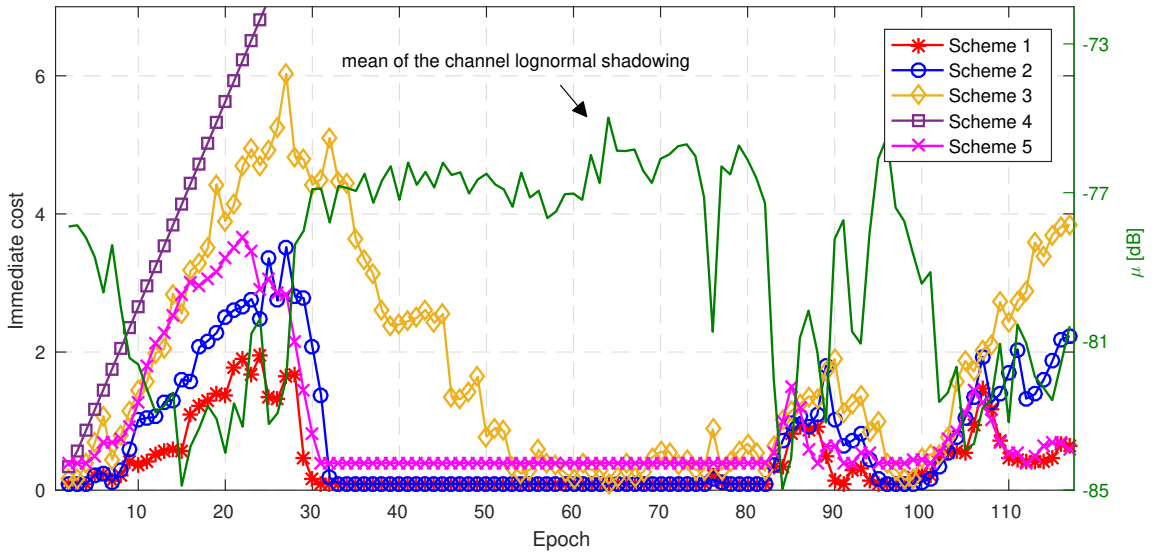


Figure 2.8: KW-NOV14: The mean of the channel lognormal shadowing and immediate collected costs by different schemes.

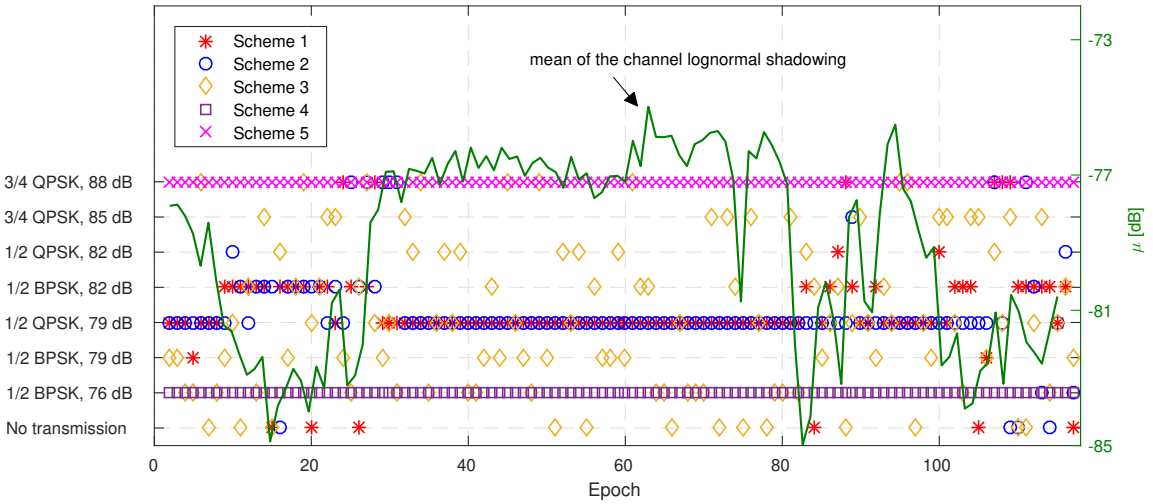


Figure 2.9: KW-NOV14: The mean of the channel lognormal shadowing and selected actions in different schemes.

channel losses in those epochs, the immediate cost of the proposed algorithm grows greater than that of the genie-aided method which can adapt the transmission mode more precisely. A little lag around epoch 30 can be observed between the changes of the immediate costs of those two schemes. During epochs 32 to 85, the immediate

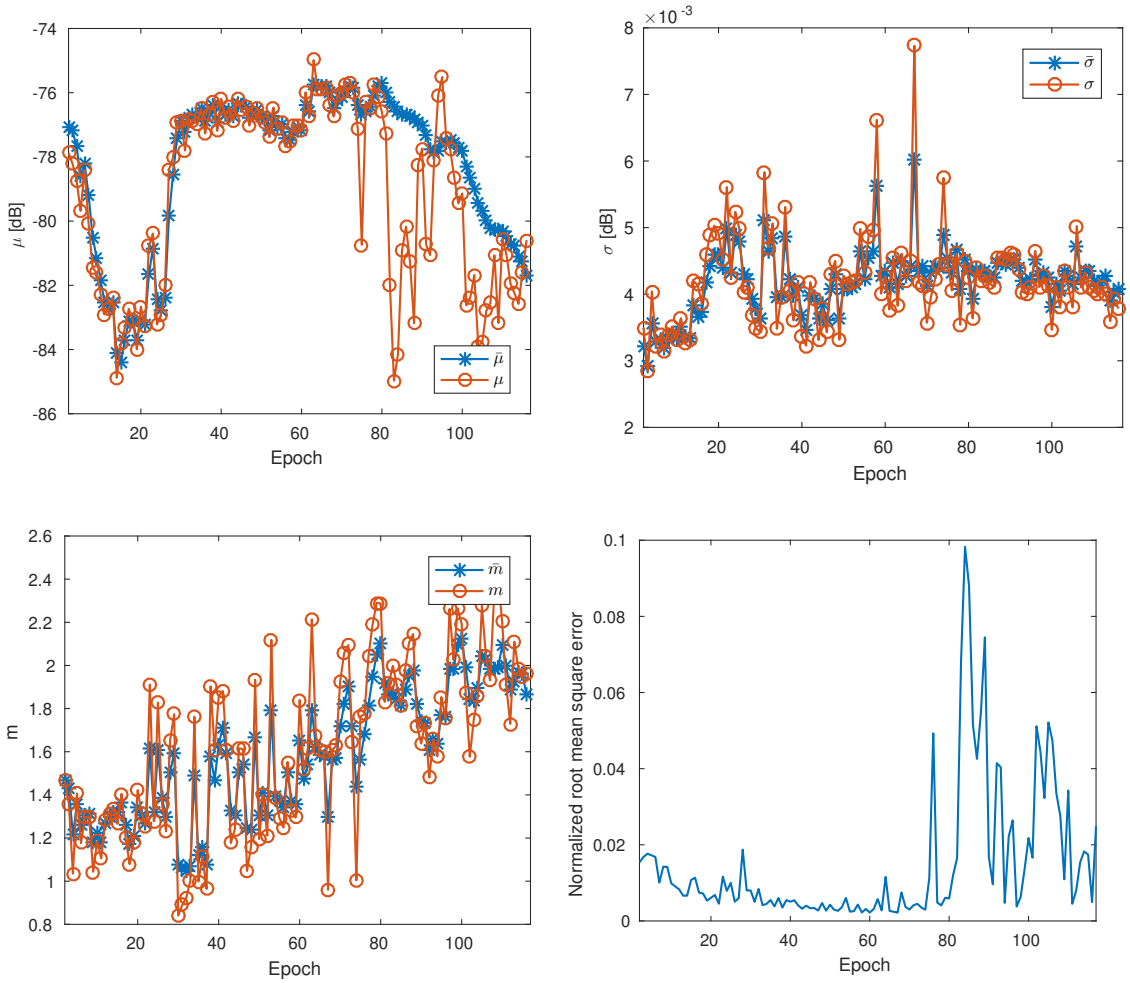


Figure 2.10: KW-NOV14: Comparison between the mean $(\bar{\mu}, \bar{\sigma}, \bar{m})$ of the estimated channel belief state and the true channel state (μ, σ, m), and the NRMSE.

costs obtained by the proposed algorithm and the genie-aided method are almost identical, as the transmitter in the proposed algorithm has learned adequate channel knowledge. Both schemes prefer the transmission action with a moderate transmission power level and a moderate data rate, i.e., 79 dB and 1/2 QPSK.

The channel state vector estimation and the NRMSE of the estimation are depicted in Fig. 2.10. Similar to the case in SPACE08, the difference between the mean values

of the channel belief state and the true channel states is small and the NRMSE is less than 0.1 in every epoch. The results validate the effectiveness of the proposed recursive estimation algorithm.

2.6.4 Performance of the Proposed Algorithm with Different System Setups

The performance of the proposed algorithm is examined in different system setups, including different data arrival rates from the application layer, different numbers of child system state samples in online approximation, different numbers of actions to be explored, and different depths of the state-action tree, in the Monte Carlo planning.

To quantify the performance of the proposed algorithm in different setups, we take the performance of the genie-aided scheme as a benchmark, and evaluate the *normalized difference* which is defined as $\left(\bar{C}-\bar{C}_{\text{Genie}}\right)/\bar{C}_{\text{Genie}}$, where $\bar{C}$ is the average cost defined in (2.25). For comparison purpose, $\bar{C}_{\text{Genie}}$ is obtained based on $N_{\text{o}}=3$, $N_{\text{a}}=3$, and $D=5$.

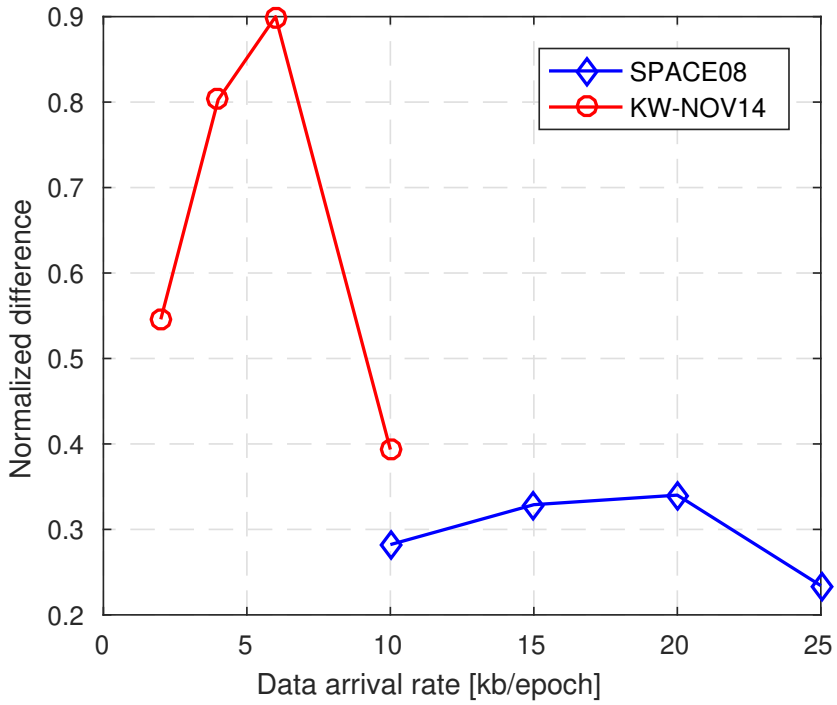
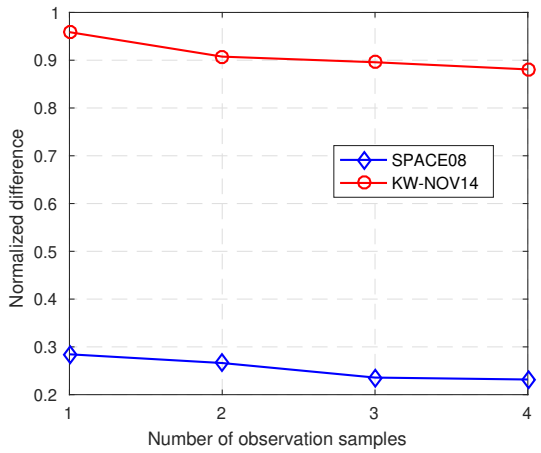


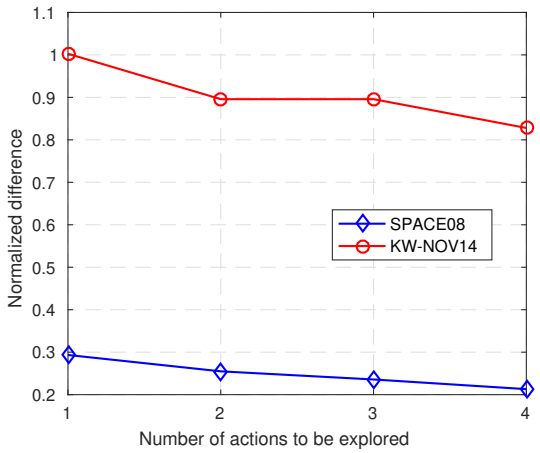
Figure 2.11: Normalized difference with respect to the genie-aided method with different data arrival rates with $N_{\mathrm{o}} = 3$, $N_{\mathrm{a}} = 3$, and $D = 5$.

2.6.4.1 Performance with Different Data Arrival Rates

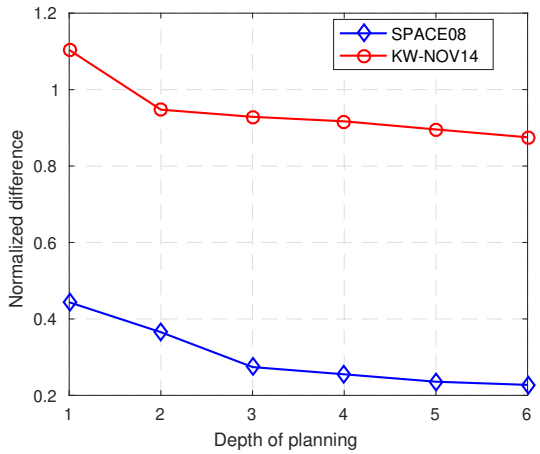
The data arrival rate will impact the performance of the proposed algorithm. As the data arrival rate increases, both the proposed algorithm and the genie-aided method prefer the transmission modes with high data rates to suppress the increase of the data queue length. Without precise channel knowledge, there are high chances that the proposed algorithm could schedule high-data-rate transmissions in epochs with bad channel conditions. Consequently, the proposed algorithm suffers an increased performance gap with the genie-aided method that determines the transmission actions based on non-causal and perfect knowledge. Fig. 2.11 shows the normalized



(a) With $N_{\mathrm{a}} = 3$ and $D = 5$



(b) With $N_{\mathrm{o}} = 3$ and $D = 5$



(c) With $N_{\mathrm{o}} = 3$ and $N_{\mathrm{a}} = 3$

Figure 2.12: Normalized difference with respect to the genie-aided method with different Monte Carlo planning parameters. $r_{\mathrm{g}} = 20$ kb/epoch in SPACE08, and $r_{\mathrm{g}} = 6$ kb/epoch in KW-NOV14.

performance difference of the proposed algorithm w.r.t. the genie-aided method with different data arrival rates. It can be seen that as the data arrival rate increases from a small value to a moderately large value, the normalized difference increases. However, with further increase of the data arrival rate, the normalized difference starts decreasing. This is caused by the large value of the average cost $\bar{C}_{\mathrm{Genie}}$ that increases monotonically with the data arrival rate.

2.6.4.2 Performance with Different Numbers of Child System State Samples and Actions To Be Explored in Online Approximation

The normalized performance difference of the proposed algorithm w.r.t. the genie-aided method with different numbers of child system state samples and different numbers of actions to be explored in online approximation are shown in Fig. 2.12(a) and Fig. 2.12(b), respectively. The performance improvement is minor with the increase of the numbers of child system state samples and actions to be explored. This indicates that with a small number of child system state samples and a small number of actions to be explored, the proposed algorithm can achieve good online approximation performance with a low computational complexity.

2.6.4.3 Performance with Different Depths of Monte Carlo Planning

The depth of Monte Carlo planning is a key factor in the tradeoff between the approximation accuracy and the computational complexity; see Section 2.4.2. Fig. 2.12(c) shows the normalized performance difference of the proposed algorithm w.r.t. the genie-aided method with different planning depths. It can be seen that considerable performance improvement is achieved when the depth of planning is increased from 1 to 2 in SPACE08 and from 1 to 3 in KW-NOV14. Further increase of the planning depth in both experiments leads to slight performance improvement, which, however,

is accompanied with exponentially increased computational cost. The results demonstrate that the proposed algorithm achieves decent performance with a small depth of planning since it stores and exploits the historical knowledge of the value function via the TD learning and the LR when evaluating the future expected costs.

2.7 Summary

This chapter focused on an UWA point-to-point transmission system which operates on an epoch-by-epoch basis over a long term, and developed an adaptive transmission algorithm which exploits the UWA channel dynamics to trade off energy consumption with information delivery latency. To describe both the short-term fading and the large-scale shadowing of UWA channels, the Nakagami-lognormal distribution was adopted for channel characterization. To account for the channel variation across epochs, the evolution of the channel distribution parameters was modeled as a Markov process with unknown parameters. Given that the channel can only be observed during active transmissions, we formulated the adaptive transmission problem as a POMDP to strike an optimal tradeoff between learning the channel dynamics via active transmissions and exploiting the learned channel knowledge for transmission efficiency. An algorithm in the model-based RL framework was developed, which recursively estimates the channel model parameters and computes the optimal transmission strategy that minimizes a long-term system cost. Thorough algorithm

evaluation was performed using channel measurements from two field experiments. The emulated results revealed that the proposed algorithm achieves decent performance relative to a benchmark method that assumes perfect and non-causal channel knowledge.

Chapter 3

Reinforcement Learning-based Adaptive Trajectory Planning for AUVs in Under-ice Environments[1]

[1]The work in this chapter was accepted by the "*2018 MTS/IEEE Oceans Conference*".

3.1 Introduction

The reconstruction of water parameter fields of interest such as water temperature and underwater acoustic field has drawn considerable attention in recent years [53, 54, 55]. To provide an accurate field reconstruction with stationary sensor networks, a large number of sensors at different locations are required. Besides, the deployment cost and the maintenance effort of stationary sensors are even higher in the under-ice environment as demanding activities involved in under-ice water exploration [56, 57, 58]. Autonomous underwater vehicles (AUVs) can detect and track the water parameter field of interest in a more flexible way since they can sense the field through either predetermined or adaptively determined trajectories. In addition, thanks to the spatial correlation of the underwater parameter field [59], it is not necessary to traverse all the area of interest to reduce the uncertainty of the field to a certain level. Instead, by visiting a subset of the area, an accurate field reconstruction can be achieved. However, the knowledge of field spatial correlation is often unavailable, especially in under-ice environments. In this chapter, we develop an online learning-based trajectory planning mechanism which adaptively determines the sampling locations of multiple AUVs to reduce the field uncertainty as more as possible without any field knowledge of the spatial correlation.

3.1.1 Existing Studies in Terrestrial Robotic Networks

The adaptive trajectory planning has been extensively studied in terrestrial robotic networks. Many works determine trajectories greedily to optimize some performance metrics such as field estimation error reduction and information entropy at each time step. For one robot, [60] studies trajectory planning to explore the maxima of the field of interest. The field is modeled as a Gaussian process (GP) where the mean function includes the information of the field maxima while the variance indicates the uncertainty of the field. The Bayesian optimization (BO) is leveraged which adaptively determines the optimal trajectory of a robot in each time step to strike a balance on tracking the field maxima and exploring the field with high uncertainty. In [61], trajectory planning to monitor physical phenomenon for multiple AUVs is investigated. The field is modeled as a GP and the variances at different locations can be updated through the Gaussian process regression (GPR) based on field samples at the sampling locations. At each time step, the next sampling locations for all the mobile sensors are determined to minimize the variances of the field over a set of locations (which could be a grid on the field or interested areas) based on the GPR. Both centralized and decentralized algorithms with coordination via wireless communications are presented. To monitor spatial phenomena by multiple wireless mobile sensors, an adaptive sampling strategy is proposed in [62]. At each time slot, the next sampling locations of all the sensors are determined to minimize the conditional

entropy based on the previous noisy measurements. It assumes that at each time slot, the robots could communicate to a based station via single hop or multi-hop links. The greedily optimized trajectory planning could results in myopic performance of the field reconstruction since it only cares about the immediate uncertainty reduction, and it often assumes that the field knowledge of spatial correlation is known *a priori*.

There exist some works on non-myopic adaptive trajectory planning which consider the long-term performance of field reconstruction. In [63], the trajectory planning of one robot for localization and mapping is proposed where the trajectory is determined to minimize the overall uncertainty about its pose (location and heading) and the locations of environmental landmarks. A partially observable Markov decision process (POMDP) is formulated where the system states are the robots' pose and the landmarks' locations, and the action includes the waypoints within the trajectory. The proposed POMDP is solved by direct policy search which uses BO to adjust the policy parameters. [64] studies adaptive trajectory planning method to observe environments by visiting a subset of sampling locations. For one robot, it selects the next sampling locations which maximize the expected utility related to "informativeness" within fixed timesteps based on the posterior belief of the field. For multiple robots, a greedy algorithm is proposed based on the sequential allocation. However, the objective function needs to be local-submodular to achieve the near-optimality. As extensions to [60] where the trajectory planning is myopic, the trajectory planning problem is reformulated as sequential BO (SBO) problems which consider BO

for future trajectories to seek non-myopic performance [65, 66]. It shows that the proposed SBO is essentially a POMDP where the system state consists of the field and the positions of the robots, and the action is the parameterized trajectory. It assumes that the transition probability of the state is known and deterministic, and a Monte Carlo tree-based method is used to solve the POMDP. The existing solutions to non-myopic adaptive trajectory planning usually formulate it as an MDP or a POMDP with discrete action spaces and one mobile robot, and the tree-based methods are often used to solve the problem. For GP-based field modeling for trajectory planning, the field hyper-parameters of the covariance function in GP which indicates the field spatial correlation is often considered to be known *a priori*. For non-myopic adaptive trajectory planning, one robot is often considered due to the low computational complexity.

3.1.2 Existing Studies in Underwater AUV Networks

The research on adaptive trajectory planning in underwater AUV networks has been limited. Some works first construct the uncertainty map of the interested area such as the estimation error over the area based on historical observation data and then determine trajectories of AUVs to globally maximize the uncertainty reduction. In [67], a trajectory planning method for one AUV is proposed to inspect underwater structures. The view locations are selected to minimize the uncertainty of the structure

surface based on known uncertainty maps. The optimization problem is solved by a sampling-based redundant roadmap method. In [68], adaptive trajectory planning of one AUV is studied to reduce the field uncertainty. The field spatial correlation is first estimated based on historical data. Then, the sampling locations are selected among pre-determined waypoints to minimize the mutual information between the sampled and unsampled locations. In [69], trajectories of multiple AUVs are determined to maximize the line integral of the uncertainty of water field estimates subject to constraints of primary motion, anticurling, vicinity, communications, and obstacle avoidance. It assumes the uncertainty of field estimates is known *a priori*, and a mixed integer optimization problem is formulated. It does not consider the coordination of multiple AUVs since the trajectories are determined offline. Those works consider the prior field knowledge, however, which is difficult and time-consuming to obtain in practice. An adaptive trajectory planning of multiple AUVs strategy which considers coordination of wireless communications is studied in [70]. In each time step, the sampling locations of all the AUVs are determined to minimize the total uncertainty of the field where the field is described by a GP. A decentralized algorithm is proposed where the Voronoi tessellation is used to distribute the optimization objective among the AUVs. The non-myopic adaptive trajectory planning of multiple AUVs for the long-term field reduction is rarely studied.

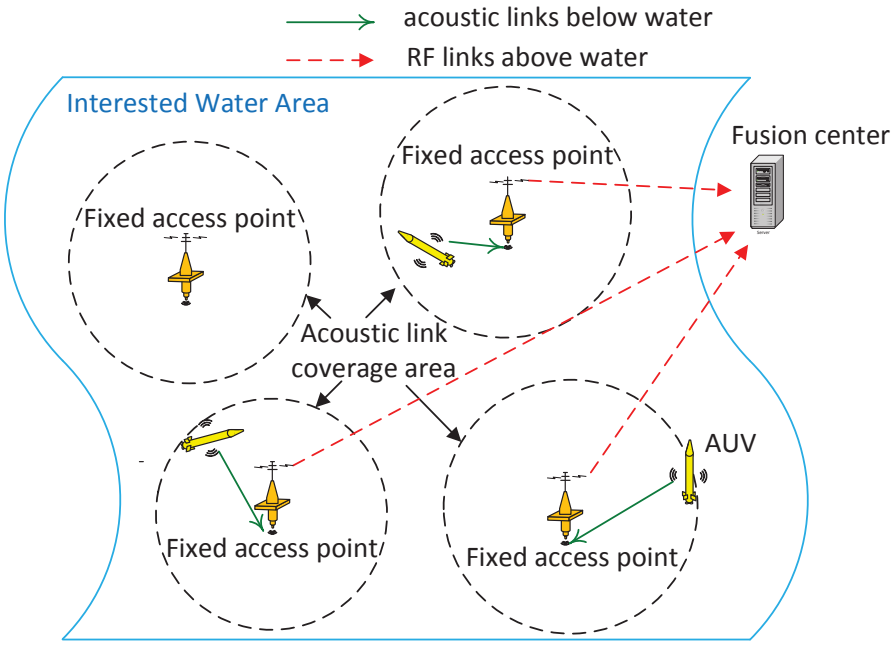


Figure 3.1: An illustration of a system layout with 3 AUVs and 4 APs.

3.1.3 Our Work

This chapter studies the adaptive trajectory planning of multiple AUVs in the under-ice environment for estimation of a water parameter field of interest. Particularly, we consider a centralized system as illustrated in Fig. 3.1, where the fixed access points (APs) on the ice layer serve as gateways for communications between the AUVs and a remote data fusion center (FC). The AUV trajectories are determined by the FC on a time epoch-by-epoch basis based on the samples collected in the past epochs.

In this work, the water parameter field of interest is modeled as a Gaussian process (GP) with unknown hyper-parameters [71]. At the end of each epoch, the APs relay

the field samples collected by the AUVs to the FC where the field hyper-parameters are estimated via the maximum likelihood method [71], and the posterior field distribution and the field uncertainty are computed via the Gaussian process regression (GPR) [72]. The AUV trajectories in the next epoch will then be determined based on the current system state including the current positions of all the AUVs and the field knowledge, with an aim of minimizing a long-term system cost that is defined based on the field uncertainty reduction and the AUV mobility cost. The AUV trajectories are expected to satisfy several practical constraints, including the kinematics constraint, the constraint on the communication range, and the constraint of being within the area of interest.

The adaptive trajectory planning problem is formulated as a Markov decision process (MDP) [73] with a constrained continuous action space. A reinforcement learning (RL)-based method is designed for online learning of the optimal action, i.e., the trajectories of all the AUVs, which satisfies the constraints. The knowledge for determining the optimal trajectories in each epoch is first obtained by transferring the historical knowledge to determine the trajectories in the previous epoch and then is further adjusted based on the newly collected system cost. The proposed RL-based trajectory planning algorithm is validated using simulated 2-dimensional (2D) fields. The simulation results show that the proposed algorithm achieves performance similar to a benchmark method that assumes perfect knowledge of the field hyper-parameters.

The main contributions of this work are in the following.

- The developed algorithm is non-myopic and for multiple AUVs, while most existing works on non-myopic planning consider only a single vehicle [63, 64, 65, 66].
- This work performs the online learning of the field hyper-parameters, while many existing works assume known *a priori* of the field knowledge [61, 68, 69, 70].
- The developed algorithm considers a continuous action space, while many existing works consider either a discrete action space or a finite number of predetermined trajectory patterns [62, 64, 65, 66].

The rest of the chapter is organized as follows. The system model is presented in Section 3.2. The RL-based adaptive trajectory planning algorithm is developed in Section 3.3. Evaluation of the proposed algorithm is included in Section 3.4. Summary is presented in Section 3.5.

3.2 System Model and Problem Formulation

In this section, we describe the system in details and build a mathematical model for the field estimation. The trajectory planning for multiple AUVs is then formulated

as an optimization problem under constraints.

3.2.1 System Description

The system under consideration consists of multiple AUVs, several fixed APs and a remote FC. Denote the set of the AUVs as $\mathcal{M} = \{1, 2, ..., |\mathcal{M}|\}$. The AUVs are equipped with sensors and acoustic communication devices. They take field measurements at different sampling locations as navigate along their trajectories. A total number of N_{AP} APs are placed at fixed locations which collect data from all the AUVs via acoustic links. The APs send the observation data and location information of all the AUVs to a data FC via high data rate radio links where the FC performs further data processing. An illustration of the system layout with 3 AUVs and 4 APs is shown in Fig. 3.1. The underwater area of interest can be described by a continuous location set $\mathcal{X}_{\text{area}} \subset \mathbb{R}^{\text{D}}$ with D = 2 or D = 3. The field can be described as $f(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X}_{\text{area}}$ represents a location in the area of interest.

The system operates on an epoch-by-epoch basis. The proposed trajectory planning mechanism for AUVs in each epoch is described as in Fig. 3.2. The planned trajectory of each AUV in the ℓth epoch consists of K waypoints in K time slots and is determined at the end of the $(\ell-1)$th epoch, i.e., $\tilde{\mathbf{y}}_i(\ell) := [\mathbf{y}_{i1}(\ell); \mathbf{y}_{i2}(\ell); \cdots; \mathbf{y}_{iK}(\ell)]$. Each AUV takes field measurements around the waypoints, and after reaching the last

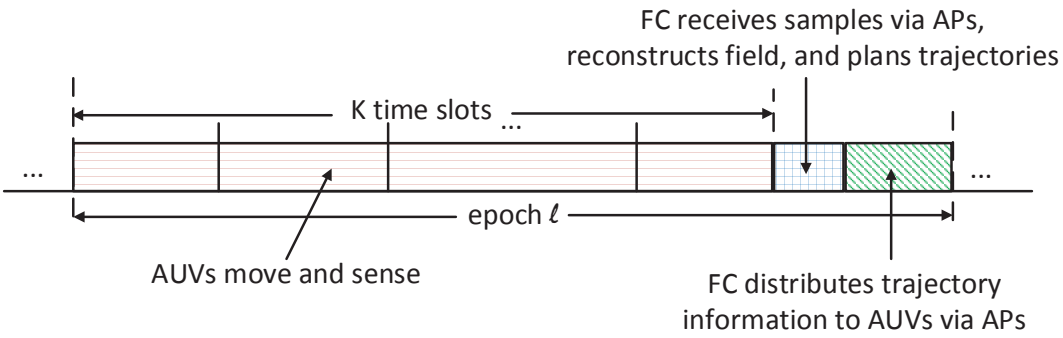


Figure 3.2: Epoch structure for water parameter field reconstruction using AUVs.

waypoint in the current epoch, it transmits the observed data and the corresponding sampling locations to the nearest AP via acoustic links in water. The APs then relay all the information to the FC via radio links above water. The FC estimates the field based on all the observation data, estimates the field knowledge, determines the trajectories $\{\tilde{\mathbf{y}}_i(\ell+1), i \in \mathcal{M}\}$ for all the AUVs in the next epoch, and transmits via APs the planned trajectories to all the AUVs. At the end of the ℓth epoch, all the AUVs receive their planned trajectories in the next epoch.

3.2.2 Constraints on Sampling Trajectories

The planned trajectories must satisfy practical constraints. In this chapter, we consider three constraints related to kinematics, the communication range, and the sensing area.

3.2.2.1 Kinematics Constraint

Due to the limited travel speed of an AUV, in each epoch, the distance between any two consecutive waypoints for each AUV is constrained as

$$\sqrt{||\mathbf{y}_{ij}(\ell)-\mathbf{y}_{i,j+1}(\ell)||^2} \leq \kappa_{\text{up}}, \forall i \in \mathcal{M} \tag{3.1}$$

with $1 \leq j \leq K-1$, and

$$\sqrt{||\mathbf{y}_{iK}(\ell) - \mathbf{y}_{i1}(\ell+1)||^2} \leq \kappa_{\text{up}}, \forall i \in \mathcal{M} \tag{3.2}$$

where κ_{up} is the maximal distance that an AUV can travel within one time slot.

3.2.2.2 Communication Range Constraint

Since the field samples of each AUV must be sent to an AP in the last time slot in each epoch, we must ensure that in the Kth time slot of each epoch, each AUV must be within the communication range of at least one of the N_{AP} APs, i.e.,

$$\sqrt{||\mathbf{y}_{iK}(\ell) - \mathbf{y}_{\text{AP}}^{(j)}||^2} < \kappa_{\text{comm}}, \ \exists j \in \mathcal{I}_{\text{AP}}, \ \forall i \in \mathcal{M} \tag{3.3}$$

where $\mathcal{I}_{\text{AP}} := \{1, 2, \cdots, N_{\text{AP}}\}$ is the AP index set, $\mathbf{y}_{\text{AP}}^{(j)}$ is the location of the jth AP, and κ_{comm} is the communication range that ensures error-free transmission between an AP and an AUV.

3.2.2.3 Sensing Area Constraint

We assume that all the AUVs should stay within the area of interest, i.e.,

$$\mathbf{y}_{ij}(\ell) \in \mathcal{X}_{\text{area}}, \forall i \in \mathcal{M}, i \geq 0, 0 \leq j \leq K, \ell \geq 0. \tag{3.4}$$

3.2.3 Modeling Real Trajectories of AUVs

Denote $\mathcal{Y}(\ell) := \left\{\tilde{\mathbf{y}}_1(\ell), \tilde{\mathbf{y}}_2(\ell), \cdots, \tilde{\mathbf{y}}_{|\mathcal{M}|}(\ell)\right\}$ as the planned trajectories consisting of waypoints for all the AUVs in the ℓth epoch. Due to the complex underwater environment, the AUVs may not arrive at each planned waypoint exactly. We model the true sampling location of the ith AUV in the kth time slot within the ℓth epoch as

$$\mathbf{x}_{ik}(\ell) = \mathbf{y}_{ik}(\ell) + \mathbf{e}_{ik}(\ell), \tag{3.5}$$

where $\mathbf{e}_{ik}(\ell) \in \mathbb{R}^{\text{D}}$ is a noise vector which describes the location inaccuracy, and is assumed following a uniform distribution $\mathcal{U}(-\epsilon, \epsilon)$ [74] with $\epsilon \ll \kappa_{\text{comm}}$ and $\epsilon \ll \kappa_{\text{up}}$

being the navigation error.

The exact sampling locations of the ith AUV in the ℓth epoch are described by $\tilde{\mathbf{x}}_i(\ell) = [\mathbf{x}_{i1}(\ell); \mathbf{x}_{i2}(\ell); \cdots; \mathbf{x}_{iK}(\ell)]$. Denote $\mathcal{X}_{\text{samp}}(\ell) := \left\{\tilde{\mathbf{x}}_1(\ell), \tilde{\mathbf{x}}_2(\ell), ..., \tilde{\mathbf{x}}_{|M|}(\ell)\right\}$ as the sampling locations of all the AUVs in the ℓth epoch, $\mathcal{Z}(\ell)$ as all the sampling location from epoch 0 to epoch ℓ, and $\tilde{\mathbf{p}}(\ell) := [\mathbf{x}_{1K}(\ell-1); \mathbf{x}_{2K}(\ell-1); \cdots; \mathbf{x}_{|\mathcal{M}|K}(\ell-1)]$ as the locations of all the AUVs at the beginning of the ℓth epoch.

3.2.4 Gaussian Process Regression for Field Estimation

In this work, we exploit the GPR to reconstruct the field of interest and estimate the field knowledge of spatial correlation. We first obtain a discrete set of target points $\mathcal{X}$ by discretizing the area $\mathcal{X}_{\text{area}}$. We intend to minimize the field uncertainty over the target points rather than the whole area of interest to reduce the computational complexity. The set $\mathcal{X}$ can be selected based on different application requirements or to balance the field reconstruction accuracy and the computational complexity. We assume that the total number of elements in $\mathcal{X}$ is $N_{\mathcal{X}}$. The field of interest is then modeled as a GP with zero mean,

$$f(\mathbf{x}) \sim \mathcal{GP}(0, K(\mathbf{x}, \mathbf{x}')), \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}_{\text{area}}, \tag{3.6}$$

where $K(\mathbf{x}, \mathbf{x}')$ is the value of covariance function at locations $\mathbf{x}$ and $\mathbf{x}'$ which describes the spatial correlation between locations $\mathbf{x}$ and $\mathbf{x}'$.

There are various types of covariance functions that can be employed [71]. In this work, we consider the squared exponential covariance function,

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left\{-(\mathbf{x} - \mathbf{x}')^{\mathrm{T}} \mathbf{\Lambda}^{-2} (\mathbf{x} - \mathbf{x}')\right\}, \tag{3.7}$$

where $\mathbf{\Lambda} = \mathrm{diag}([d_1, \cdots, d_{\mathrm{D}}])$ with $\mathrm{D} = 2$ or $\mathrm{D} = 3$ being the dimension of the water area and d_i being the distance scale that determines the spatial correlation of two locations, and σ_f^2 is the signal variance.

In the ℓth epoch, a set of field observations can be obtained,

$$\boldsymbol{\psi}(\ell) = f\left(\mathcal{X}_{\mathrm{samp}}(\ell)\right) + \mathbf{n}(\ell), \tag{3.8}$$

where $f(\mathcal{X}_{\mathrm{samp}}(\ell))$ are the field values at the locations in $\mathcal{X}_{\mathrm{samp}}(\ell)$, and $\mathbf{n}(\ell)$ is the observation noise with each of its elements assumed following a Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$.

Denote $\mathbf{\Psi}(\ell) = \{\boldsymbol{\psi}(\ell')\}_{\ell'=0}^{\ell}$ as available field observations. Denote $\mathbf{C}(\mathcal{A}, \mathcal{B})$ as a matrix whose the (i, j)th element is calculated as $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$, with $\mathbf{x}_i \in \mathcal{A}$ and $\mathbf{x}_j \in \mathcal{B}$. The posterior distribution of the field in the ℓth epoch over the target point set $\mathcal{X}$ can be

obtained as

$$f(\mathcal{X}) \sim \mathcal{N}(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell), \tag{3.9}$$

with $\boldsymbol{\mu}_\ell = \mathbf{C}(\mathcal{X}, \mathcal{Z}(\ell))\mathbf{C}_{\mathcal{Z}}^{-1}\boldsymbol{\Psi}(\ell)$, $\boldsymbol{\Sigma}_\ell = \mathbf{C}(\mathcal{X}, \mathcal{X}) - \mathbf{C}(\mathcal{X}, \mathcal{Z}(\ell))\mathbf{C}_{\mathcal{Z}}^{-1}\mathbf{C}(\mathcal{Z}(\ell), \mathcal{X})$, and $\mathbf{C}_{\mathcal{Z}} = \mathbf{C}(\mathcal{Z}(\ell), \mathcal{Z}(\ell)) + \sigma_n^2\mathbf{I}$, according to [71].

Based on the available observation $\boldsymbol{\Psi}(\ell)$ at the end of the ℓth epoch, the field hyper-parameters $\boldsymbol{\theta}_{\text{hyper}} := \{\sigma_f^2, \boldsymbol{\Lambda}\}$ can be estimated by maximizing the log marginal likelihood,

$$\hat{\boldsymbol{\theta}}_{\text{hyper}} = \max_{\boldsymbol{\theta}_{\text{hyper}}} \left\{ -\frac{1}{2}\boldsymbol{\Psi}(\ell)^{\mathrm{T}}\mathbf{C}_{\mathcal{Z}}^{-1}\boldsymbol{\Psi}(\ell) - \frac{1}{2}\log|\mathbf{C}_{\mathcal{Z}}| \right\}. \tag{3.10}$$

The hyper-parameters fully characterize the field spatial correlation, which are unknown *a priori* and estimated on the fly. The optimization problem (3.10) can be solved using a quasi-Newton method, i.e., the L-BFGS-B method [75].

3.2.5 Problem Formulation for Optimal Trajectory Planning

The field uncertainty can be obtained based on the field posterior distribution which is updated through the GPR. Specifically, we denote $\mathbf{u}_\ell := \text{diag}(\boldsymbol{\Sigma}_{\ell-1})$, to describe the uncertainty of all the target points in $\mathcal{X}$ based on the observations up to the $(\ell-1)$th epoch.

Denote $\mathbf{s}(\ell) = \{\tilde{\mathbf{p}}(\ell), \mathbf{u}_\ell\}$ as the system state at the beginning of the ℓth epoch.

Denote $\mathbf{a}(\ell)$ as the action in the ℓth epoch which consists of the planned waypoints for all the AUVs in the ℓth epoch. The desired trajectories for all the AUVs in the ℓth epoch can be determined to minimize the expected total discounted cost,

$$\min_{\{\mathbf{a}(\ell)\}_{\ell=0}^{\infty}} \mathbb{E}\left\{\sum_{\ell=0}^{\infty} \gamma^{\ell} C(\mathbf{s}(\ell), \mathbf{a}(\ell))\right\}, \tag{3.11}$$

where $\gamma \in (0, 1]$ is a discount factor, and $C(\mathbf{s}(\ell), \mathbf{a}(\ell))$ is an application-dependent cost function. In this work, the cost function considers the field uncertainty reduction, the AUV mobility cost based on the planned trajectories, and the constraints from (3.1) to (3.4). Next we present the formulation of the cost function used in this work.

3.2.5.1 Cost Function

Denote the current state $\mathbf{s} = \{\tilde{\mathbf{p}}, \mathbf{u}\}$ and the planned trajectories as $\mathbf{a}$. Denote the next state $\mathbf{s}' = \{\tilde{\mathbf{p}}', \mathbf{u}'\}$. The costs, reward, and penalties induced by action $\mathbf{a}$ under the current state $\mathbf{s}$ and the next state $\mathbf{s}'$ are as follows.

- *Uncertainty reduction reward:* The sampling reward to reduce the field uncertainty by performing the action $\mathbf{a}$ at the system state $\mathbf{s}$ is defined as

$$R(\mathbf{s}, \mathbf{a}) := \frac{\alpha_R}{N_{\mathcal{X}}} \left(||\mathbf{u}||_1 - ||\mathbf{u}'||_1 \right), \tag{3.12}$$

where α_R is a weighting factor, and $||\mathbf{u}'||_1$ is the summation of all the elements in $\mathbf{u}'$ which describes the total estimation error of the field. We intend to minimize the field uncertainty over the target set.

- *Trajectory cost:* The mobility cost is defined as

$$C_{\mathrm{T}}(\mathbf{a}) := \alpha_L L(\mathbf{a}) + \alpha_A A(\mathbf{a}), \tag{3.13}$$

where $L(\mathbf{a})$ is the total distance of the planned trajectories based on $\mathbf{a}$, $A(\mathbf{a})$ is the total angle that the AUVs travel along the planned trajectories based on $\mathbf{a}$, and α_L and α_{A} are weighting factors. Less energy will be consumed if an AUV travels less distance and makes less turns.

- *Trajectory constraint penalty:* We define a penalty term for the case if the planned trajectories do not satisfy constraints (3.3) to (3.4). The penalty is defined as

$$C_{\mathrm{P}}(\mathbf{a}) := \alpha_{\mathrm{p1}} I_1 + \alpha_{\mathrm{p2}} I_2, \tag{3.14}$$

where α_{p1} and α_{p2} are positive values, and I_1 and I_2 are indication functions for constraints (3.3) and (3.4), respectively, which equal 1 if the constraints are not satisfied and 0 otherwise.

Hence, the cost function in (3.11) used in this work can be described as

$$C(\mathbf{s}, \mathbf{a}) = -R(\mathbf{s}, \mathbf{a}) + C_{\mathrm{T}}(\mathbf{a}) + C_{\mathrm{P}}(\mathbf{a}). \tag{3.15}$$

3.3 Reinforcement Learning-based Adaptive Trajectory Planning

The proposed optimization problem (3.11) is essentially an MDP if the field hyper-parameters are known *a priori*. It has a continuous action space and a continuous state space, which generally is difficult to solve. In this work, we adopt one type of RL mechanism, i.e., the actor-critic method to solve the proposed MDP. Classic RL algorithms can be categorized into two types. One type is the actor-based method where an actor is trained to generate optimal actions while the other type is the critic-based method where a critic is trained to evaluate actions. The actor-critic method combines the two classic types of RL methods which can achieve higher learning performance. Specifically, in actor-critic-based algorithms, the actor is trained to generate the optimal actions while the critic is trained to provide action evaluation which helps the actor to improve its action generation strategy. Among various actor-critic-based algorithms, we employed the deep deterministic policy gradient (DDPG) algorithm [76] which deals with continuous action spaces and has high learning efficiency.

3.3.1 DDPG Basics and Design

In the DDPG algorithm, an actor is represented by a neural network which takes the system state $\mathbf{s}$ as the input and takes the optimal action $\mathbf{a}$ under the system state $\mathbf{s}$ as the output. A critic is also represented by a neural network which takes the system state $\mathbf{s}$ and the action $\mathbf{a}$ as the inputs and takes a Q-value function $Q(\mathbf{s}, \mathbf{a})$ as the outputs. The Q-value $Q(\mathbf{s}, \mathbf{a})$ indicates the expected cost after taking the $\mathbf{a}$ under the system state $\mathbf{s}$. In the learning process, the actor network is leveraged to provide the action $\mathbf{a}$ to be executed under the state $\mathbf{s}$. After performing the action $\mathbf{a}$, the corresponding cost $C(\mathbf{s}, \mathbf{a})$ can be obtained. Based on the obtained cost, the weights of the critic network are adjusted to better approximate the Q-value function $Q(\mathbf{s}, \mathbf{a})$. Then, the weights of the actor are adjusted using the policy gradient method such that the action obtained by the actor could result in lower expected cost, i.e., lower output of the critic network which takes the output of the actor network as the input. For more details about the DDPG method, please refer to [76].

A critical issue of the DDPG method is how to design the actor and critic neural networks. The structural design of the actor and the critic is presented as follows. For the actor, as illustrated in Fig. 3.3, the uncertainty map and the current locations of all the AUVs go through two fully connected layers with rectified linear units (ReLU) as the activation functions. The output layer takes the summation of the outputs of

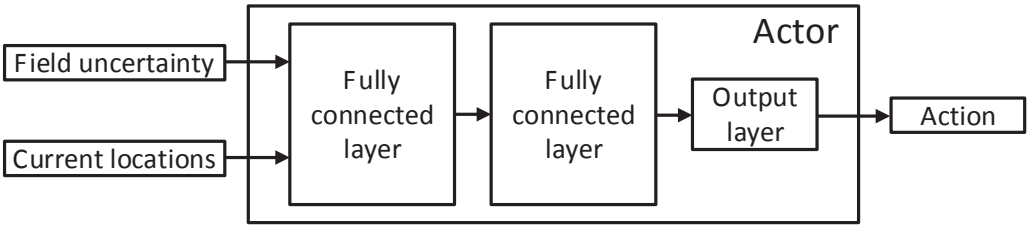


Figure 3.3: An example of the forward structure of actor network.

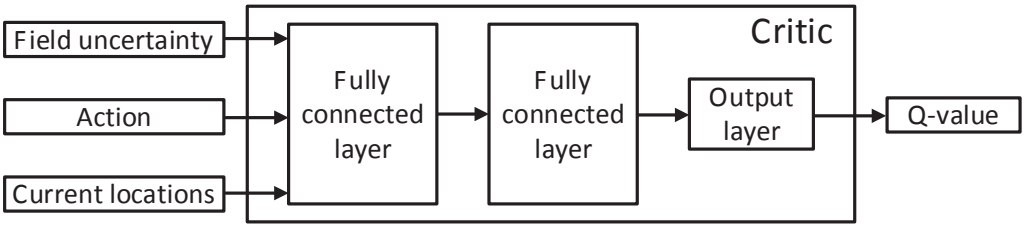


Figure 3.4: An example of the forward structure of critic network.

the second fully connected layer and uses the tanh activation function to bound the elements of the action to be within $[-1, 1]$. For the critic, as shown in Fig. 3.4, the uncertainty map, and the current locations and actions of all the AUVs go through two fully connected layers with rectified linear units (ReLU) as the activation functions. The output layer of the critic is just the summation of the outputs of the second fully connected layer. Considering the online application in this work, the structural design of the actor and critic networks should achieve learning efficiency to balance the system performance and the computational complexity.

In each training iteration, the parameters of the actor and the critic networks are updated based on one iteration of the backpropagation algorithm [51].

3.3.2 Training for Actions Under Constraints (3.1) to (3.4)

For the kinematics constraints (3.1) and (3.2), the tanh activation function is first exploited in the output layer of the actor as mentioned in the subsection above. In this way, the output of the actor is constrained to be from -1 to 1. Furthermore, the outputs of the actor network is multiplied by $\kappa_{\text{up}}/\sqrt{2}$ such that the distance that an AUV travels in each time slot is guaranteed to be not greater than κ_{up}, considering that the action for each AUV in each time slot describes the moving distance along each dimension of the area.

For the constraints (3.3) to (3.4), we first briefly introduce a technique called experience replay used in the DDPG algorithm [76]. Experience replay is a technique to train the agent with transition samples drawn from a buffer which consists of historical transitions from the previous training experience. Denote the replay buffer as $\mathcal{B}$. Denote a quadruple as (s, a, s', c) which consists of the current state s, the action performed under the state s, the next state s' by performing a based on s, the cost c collected by performing a based on s. The quadruple describes the system transition from one epoch to the next epoch. In each epoch, the transition is stored into the buffer $\mathcal{B}$. In each training iteration, the parameters in the actor and critic networks are adjusted by a mini-batch of samples of transitions which are randomly

taken from $\mathcal{B}$. By mini-batch samples, the neural networks can be trained more efficiently compared to the case with one sample per learning iteration. By training based on samples from the relay buffer rather than samples obtained sequentially, the correlation of training samples can be removed, which improves the convergence performance of the neural networks.

To better learning the actions which satisfy the constraints (3.3) to (3.4), we propose a modified DDPG algorithm where two replay buffers are exploited for training. Denote $\mathcal{B}_1$ and $\mathcal{B}_2$ as two buffers where one consists of transitions whose actions satisfy the constraints (3.3) and (3.4) and the other one consists of transitions whose actions do not satisfy the constraints (3.3) and (3.4). In the training process, it is found that most of the transition samples are from $\mathcal{B}_1$, which makes it difficult for the actor and the critic to learn from "bad" samples in $\mathcal{B}_2$. To learn the optimal actions which satisfy the constraints, we should ensure that the actor and the critic learn equally sufficient samples from $\mathcal{B}_1$ and $\mathcal{B}_2$ in the training process. In this way, the actor will generate actions which have less cost and satisfy the constraints (3.3) to (3.4) while the critic could evaluate the actions and states without a bias.

Denote the actor network as μ and the critic network as Q. The modified algorithm to obtain the optimal trajectories for all the epochs under the known field hyperparameters is described in Algorithm III-1. In the training process, one training episode refers to a process which begins from the initial state when all the AUVs are

at their initial positions (the first epoch) and ends at the final state when the whole sampling task is completed (the last epoch). In each epoch, an action $\mathbf{a}$ is randomly adjusted based on the output of the actor with an exploration noise. The purpose of the random action adjust is to introduce actions which potentially achieve less cost. The random adjust function considered in this chapter is described in Algorithm III-2. After performing the action $\mathbf{a}$, the immediate cost C and the next state $\mathbf{s}'$ can be obtained based on (3.15) and the GPR. Instead of training the transition quadruple $\{\mathbf{s}, \mathbf{a}, C, \mathbf{s}'\}$ immediately, the quadruple is stored into the replay buffers $\mathcal{B}_1$ or $\mathcal{B}_2$ based on the situation that whether $\mathbf{a}$ satisfies the communication and sense area constraints (3.3) and (3.4) or not. We will train the actor and the critic by a minibatch of transitions drawn from the buffers $\mathcal{B}_1$ or $\mathcal{B}_2$. To ensure that the actor and the critic learn sufficient samples in both $\mathcal{B}_1$ and $\mathcal{B}_2$, the transition samples from $\mathcal{B}_1$ and $\mathcal{B}_2$ are learned in consecutive learning iterations. Using the transition samples, the weights of the actor and critic networks are updated to minimize the prediction error of the Q-value function and to maximize the Q-value, respectively, using the stochastic gradient descent (SGD) method based on the critic and actor networks and the target networks. At the end of the training iteration, the target critic and actor networks are updated. The exploitation of the target networks improves the stability of learning [76].

If the field hyper-parameters are known *a priori*, the modified DDPG algorithm can be used to learn the optimal actions offline, which could provide the performance

Algorithm III-1 Modified DDPG algorithm:
MDDPG($N_{\text{initial}}, N_{\text{episode}}, N_{\text{epoch}}, N_{\mathcal{B}}, \gamma, \tau, Q, \theta^Q, \mu, \theta^\mu, Q', \theta^{Q'}, \theta^{\mu'}, \hat{\sigma_f}, \hat{\mathbf{D}}, \mathbf{s}$)

Input: Initial epoch N_{initial}, total training episode N_{episode}, total epochs in an episode N_{epoch}, minibatch size $N_{\mathcal{B}}$, discount factor γ, learning rate of target network τ, the critic network Q with its weight θ^Q, the actor network μ with its weights θ^μ, the target critic network Q' with its weight $\theta^{Q'}$, the target actor network μ' with its weights $\theta^{\mu'}$, the estimated hyper-parameters $\hat{\sigma_f}$ and $\hat{\mathbf{D}}$, and current system state $\mathbf{s}$

Output: Optimal action $\mathcal{A}_{\text{opt}}$ for future epochs, the critic and actor networks Q and μ, weights θ^Q and θ^μ, the target critic and actor networks Q' and μ', weights $\theta^{Q'}$ and $\theta^{\mu'}$

Initialize replay buffers $\mathcal{B}_1$ and $\mathcal{B}_2$. Set $i_{\text{iter}} = 0$ and $C_{\text{opt}} = \infty$
for episode = 1 to N_{episode} **do**
    Set $C_{\text{tot}} = 0$
    Receive the initial state $\mathbf{s}$
    for epoch = N_{initial} to N_{epoch} **do**
        Perform action $\mathbf{a}_{\text{epoch}} = \text{RandomAdjust}(\mu(\mathbf{s}) + \mathcal{N}(0, \epsilon))$ according to Algorithm III-2, obtain the immediate cost C based on (3.15), and observe the next state $\mathbf{s}'$.
        if $\mathbf{a}$ does not satisfy the communication and sensing area constraints (3.3) and (3.4) **then**
            Store the transition quadruple $\{\mathbf{s}, \mathbf{a}_{\text{epoch}}, C, \mathbf{s}'\}$ into the buffer $\mathcal{B}_2$
        else
            Store the transition quadruple $\{\mathbf{s}, \mathbf{a}_{\text{epoch}}, C, \mathbf{s}'\}$ into the buffer $\mathcal{B}_1$
        if i_{iter} mod 2 **then**
            Sample a random minibatch of $N_{\mathcal{B}}$ transitions $\{\mathbf{s}_i, \mathbf{a}_i, C_i, \mathbf{s}_{i+1}\}$ from $\mathcal{B}_1$
        else
            Sample a random minibatch of $N_{\mathcal{B}}$ transitions $\{\mathbf{s}_i, \mathbf{a}_i, C_i, \mathbf{s}_{i+1}\}$ from $\mathcal{B}_2$
        $\xi_i \leftarrow C_i + \gamma Q'(\mathbf{s}_{i+1}, \mu'(\mathbf{s}_{i+1}))$, update the critic by minimizing the error: $L = \frac{1}{N_{\mathcal{B}}} \sum_i (\xi_i - Q(\mathbf{s}_i, \mathbf{a}_i))$, and set $i_{\text{iter}} \leftarrow i_{\text{iter}} + 1$
        Update the actor by deterministic policy gradient theorem to maximize $Q(\mathbf{s}_i, \mu(\mathbf{s}_i))$
        Update the target networks: $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$ and $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
        $C_{\text{tot}} \leftarrow C_{\text{tot}} + C$
    if $C_{\text{opt}} > C_{\text{tot}}$ **then**
        $\mathcal{A}_{\text{opt}} = \{\mathbf{a}_{\text{epoch}}\}_{\text{epoch}=N_{\text{initial}}}^{N_{\text{epoch}}}$
Return $(\mathcal{A}_{\text{opt}}, Q, \mu, \theta^Q, \theta^\mu, Q', \mu', \theta^{Q'}, \theta^{\mu'})$

Algorithm III-2 Random action adjust: RandomAdjust($\mathbf{a}$)

Input: Action $\mathbf{a}$

Output: Adjusted action $\mathbf{a}$

1: Assign $\delta \in [0, 1]$ and drawn u from a uniform distribution $\mathcal{U}[0, 1]$
2: **if** $u < \delta$ **then**
3: Set the latitude travel distance in $\mathbf{a}$ to 0
4: **else**
5: Set the longitude travel distance in $\mathbf{a}$ to 0
6: Clip the elements in $\mathbf{a}$ to be within $[-\kappa_{\text{up}}/\sqrt{2}, \kappa_{\text{up}}/\sqrt{2}]$ to satisfy the kinematics constraints (3.1) and (3.2)
7: **Return a**

Algorithm III-3 Online trajectory planning algorithm in each epoch

Input: Current epoch N_{curr}, total episode N_{episode}, total epochs in each episode N_{epoch}, minibatch size $N_{\mathcal{B}}$, discount factor γ, learning rate of target network τ, the critic network Q with its weight θ^Q, the actor network $\mu(\mathbf{s})$ with its weights θ^μ, the target critic network Q' with its weight $\theta^{Q'}$, and the target actor network μ' with its weights $\theta^{\mu'}$

1: All AUVs take samples of the field via their planned trajectories
2: The FC receives the observations of field samples from all the AUVs
3: Hyper-parameters of field $\hat{\sigma_f}$ and $\hat{\mathbf{D}}$ are estimated based on (3.10)
4: The FC obtained the updated system state $\mathbf{s}$ of all the AUVs based on $\hat{\sigma_f}$ and $\hat{\mathbf{D}}$
5: $\mathcal{A}_{\text{opt}}, \tau, Q, \theta^Q, \mu, \theta^\mu, Q', \theta^{Q'}, \theta^{\mu'}$
$\leftarrow$ MDDPG($N_{\text{curr}}, N_{\text{episode}}, N_{\text{epoch}}, N_{\mathcal{B}}, \gamma, \tau, Q, \theta^Q, \mu, \theta^\mu, Q', \theta^{Q'}, \theta^{\mu'}, \hat{\sigma_f}, \hat{\mathbf{D}}, \mathbf{s}$)
6: Start to perform the action for the next epoch only according to $\mathcal{A}_{\text{opt}}$

upper bound for the proposed online learning strategy.

3.3.3 Online Learning for Trajectories Planning with Unknown Field Hyper-parameters

In practice, the perfect knowledge of the field hyper-parameters is often unavailable.

It is generally the case that those hyper-parameters should be estimated online during

the sampling process. We propose an online learning algorithm, i.e., Algorithm III-3, which incorporates the modified DDPG algorithm to online determine the optimal trajectories of all the AUVs at each epoch, where the field hyper-parameters online estimated. Specifically, after the moving and sampling process at each epoch, the unknown field hyper-parameters in the covariance function (3.7) can be estimated by solving the optimization problem (3.10) based on all the historical observations. After obtaining the estimated hyper-parameters, the previous learned knowledge, i.e., the critic network $Q(\mathbf{s}, \mathbf{a})$ with its weight θ^Q, the actor network $\mu(\mathbf{s})$ with its weights θ^μ, the target critic network $Q'(\mathbf{s}, \mathbf{a})$ with its weight $\theta^{Q'}$, and the target actor network $\mu'(\mathbf{s})$ with its weights $\theta^{\mu'}$ in the previous epoch are transferred to the current epoch. The modified DDPG algorithm then takes the available knowledge of the actors and the critics and the estimated field hyper-parameters as inputs to learn what will be the optimal trajectories for future epochs. In this way, the optimal trajectories for each epoch can be learned online according to the online estimated field hyper-parameters.

3.4 Algorithm Evaluation

We consider an under-ice field of interest in a 2D 15 km $\times$ 15 km square area, and the target set $\mathcal{X}$ consists of 16 $\times$ 16 grid points where the latitude and longitude distance between any two consecutive locations is 1 km. The 2D water parameter

field is generated based on the circulant embedding method [77] with the field hyper-parameters as $\sigma_f^2 = 1$ and $\mathbf{\Lambda} = \text{diag}([0.3, 0.3])$.

The duration of one time slot is 1,000 seconds (16.7 minutes), and one epoch consists of 3 time slots, leading to an epoch duration of 50 minutes. We consider a total of 9 epochs in the sampling process, which yields a deployment time duration of 7.5 hours in total. The simulated system consists of 4 AUVs and 4 APs. The 4 APs are located at $(4 \text{ km}, 4 \text{ km})$, $(4 \text{ km}, 11 \text{ km})$, $(11 \text{ km}, 4 \text{ km})$, and $(11 \text{ km}, 11 \text{ km})$, respectively. Those four locations are also the initial deployment sites of the 4 AUVs. The maximal navigation error is $\epsilon = 5$ m. The maximal speed of each AUV is 1 m/s, and the maximal distance of an AUV can travel within one time slot is therefore $\kappa_{\text{up}} = 1$ km. The communication range for underwater acoustic links between an AUV and an AP is $\kappa_{\text{comm}} = 3.5$ km. The discounted factor is $\gamma = 0.99$. The weights in the total cost function (3.15) are $\alpha_R = -10$, $\alpha_L = 1 \times 10^{-3}$, $\alpha_A = 5 \times 10^{-2}$, $\alpha_{p1} = 2$, and $\alpha_{p1} = 4$.

The hyper-parameters for the modified DDPG algorithm are as follows. The numbers of units in the first and the second hidden layers for both of the actor and critic networks are 400 and 300, respectively. The activation functions of the hidden layers are rectified linear units (ReLU). The batch normalization is exploited in the actor network. The learning rates for the actor and the critic networks are 1×10^{-3} and 1×10^{-4}, respectively. The L_2 weight decay for the actor and the critic networks are

1×10^{-1} and 1×10^{-2}, respectively. The mini-batch size for training is 10.

We evaluate the field estimation performance of three schemes.

- *Scheme 1:* The clairvoyant method which determines the sampling trajectories through the offline modified DDPG algorithm based on the perfect knowledge of the field hyper-parameters, according to Algorithm III-1;
- *Scheme 2:* The proposed online RL algorithm which determines the sampling trajectories epoch-by-epoch through the modified DDPG algorithm where the field hyper-parameters are online estimated in each epoch based on the collected samples, according to Algorithm III-3.
- *Scheme 3:* All the AUVs sample the water parameter field via random walk. Here, we present the simulation result that is selected among 10000 Monte Carlo runs which yields the minimal total cost.

We take the normalized mean square error (NMSE) as a performance metric for the field estimation, which describes the normalized difference between the true field and the estimated field,

$$\text{NMSE} := \frac{\int_{\mathcal{X}_{\text{area}}} ||f(\mathbf{x}) - \hat{f}(\mathbf{x})||^2 d\mathbf{x}}{\int_{\mathcal{X}_{\text{area}}} ||f(\mathbf{x})||^2 d\mathbf{x}}, \tag{3.16}$$

where f is the true field and $\hat{f}$ is the estimated field based on the mean of the GPR.

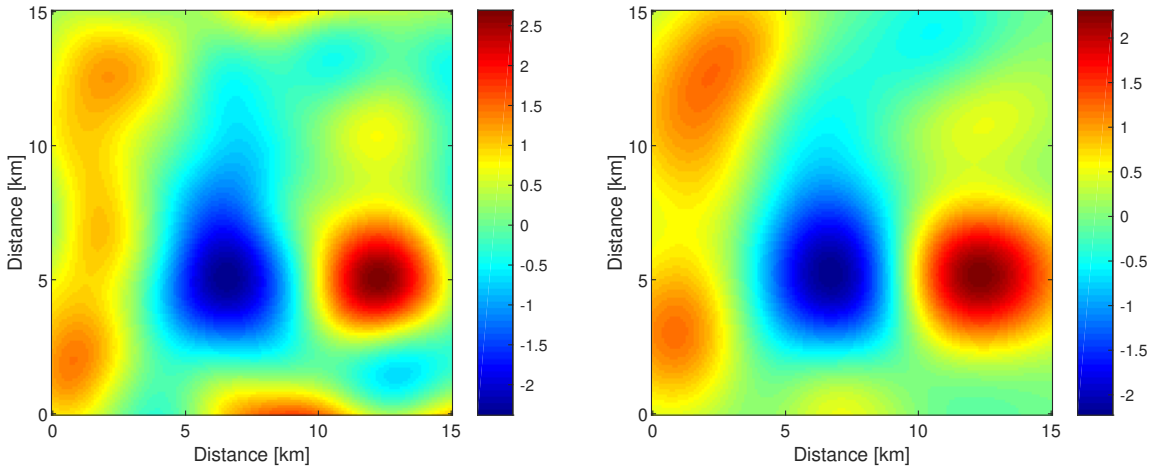
The three schemes are first examined from the perspectives of the total traveled

Table 3.1
Performance Comparison of the Three Schemes.

	Scheme 1	Scheme 2	Scheme 3
Total traveled distance [km]	74.4	77.9	78.1
Total traveled angle [rad]	76.6	117.4	131.5
NMSE	0.11	0.15	1.07

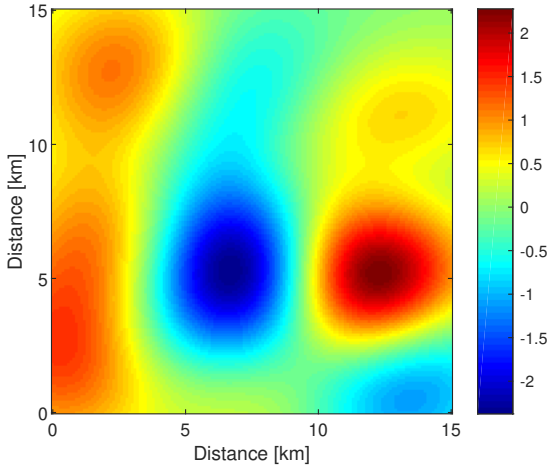
distance, the total traveled angle and the NMSE, as shown in Table 3.1. Scheme 1 achieves the least total traveled distance and the least total traveled angle, while Scheme 2 has a similar total traveled distance but greater total traveled angle. The performance gap is due to the fact that Scheme 2 estimates the field hyper-parameters and determines the actions online. The total traveled distance and the total traveled angle obtained by Scheme 3 are similar to those of Scheme 2. However, the NMSEs obtained by Schemes 1 and 2 are significantly smaller than that of Scheme 3, where a marginal difference of the NMSEs obtained by Schemes 1 and 2 can be observed.

The simulated true field and the estimated fields by the three schemes are presented in Fig. 3.5. One can see that Schemes 1 and 2 can capture important features of the true field, and the estimated field by Scheme 3 is significantly different from the true field. The planned trajectories obtained by the three schemes are shown in Fig. 3.6. To explore the area with high uncertainty, the trajectories determined by Scheme 1 spread out more than those of Schemes 2 and 3, which results in the largest sensed area. The sensed area based on the trajectories obtained by Scheme 2 at the early epochs is small due to the inaccurate field hyper-parameter estimation based on limited field samples at the early stage. With more field samples collected, the

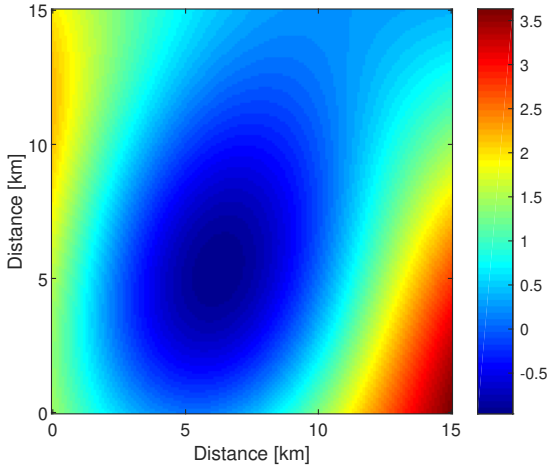


(a) True field

(b) Estimated field by Scheme 1



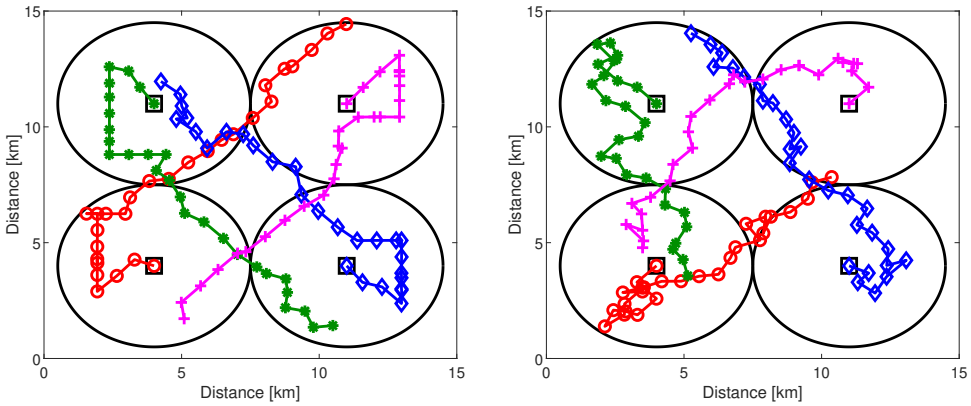
(c) Estimated field by Scheme 2



(d) Estimated field by Scheme 3

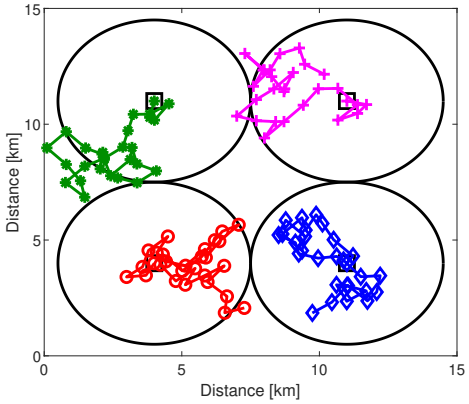
Figure 3.5: The true field and the estimated fields obtained by the three schemes.

trajectory pattern obtained by Scheme 2 is similar to the pattern obtained by Scheme 1 which tends to explore the area with high uncertainty.



(a) Scheme 1

(b) Scheme 2



(c) Scheme 3

Figure 3.6: Trajectories of 4 AUVs obtained by the three schemes, where the black squares and the black circles indicate the positions of 4 APs and the communication ranges of the APs, respectively. The black circles also are the initial deployment locations of the 4 AUVs.

3.5 Summary

In this chapter, we investigated the adaptive trajectory planning of multiple AUVs for the water parameter field estimation in the under-ice environment. An online learning-based trajectory planning algorithm was proposed to adaptively determine the trajectories of AUVs. The field of interest was modeled as a GP with unknown

hyper-parameters. The field hyper-parameters and the field posterior distribution were estimated online based on the collected samples. The adaptive trajectory planning problem was formulated as an MDP with a goal of minimizing a long-term cost that is defined based on the field uncertainty reduction and the AUV mobility cost, subject to the kinematics constraint, the communication range constraint, and the sensing area constraint. A RL-based method was designed to solve the above MDP with a constrained action space. The simulation results showed that the proposed RL-based adaptive trajectory planning algorithm achieved the performance close to a benchmark method that assumes perfect knowledge of the field hyper-parameters.

Chapter 4

Signal Alignment for Secure Underwater Coordinated Multipoint Transmissions[1]

[1]The work in this chapter was published in "*IEEE Transactions on Signal Processing*"
See Appendix B.2 for the letter of permission from IEEE.

4.1 Introduction

Coordinated multipoint (CoMP) has recently attracted considerable attention in the radio community [78, 79], and been taken as one of the core techniques for the LTE-Advanced system, the 5th generation (5G) cellular networks and the recently proposed C-RAN architecture [80, 81, 82]. Compared to centralized operations, CoMP promises a larger coverage and a higher throughput for the network with nomadic users [78, 79, 83, 84, 85]. The system architecture with interconnected and geographically distributed antenna elements (DAEs) has also been widely used by engineers for underwater acoustic (UWA) system development[2]. Two examples are shown in Fig. 4.1, where one is formed by distributed nodes which are anchored at the water bottom and connected via cables, and the other is formed by gateways which can communicate instantaneously with a control center via high-rate radio links. A large-scale testbed with 96 DAEs distributed over an area of 30 km $\times$ 50 km has been deployed in early days by the Atlantic Underwater Test and Evaluation Center (AUTEC) [86], where the DAEs are separated by more than 4 km. Another example is the Ocean-TUNE testbed [87] which has four distributed stationary nodes with distances among nodes varying from 500 m to 12 km.

[2]Due to the large sound propagation latency in water, the distributed antenna system and the coordination among DAEs (CoMP) are naturally coupled in underwater acoustic environment, which slightly differ from relevant research efforts in terrestrial radio networks.

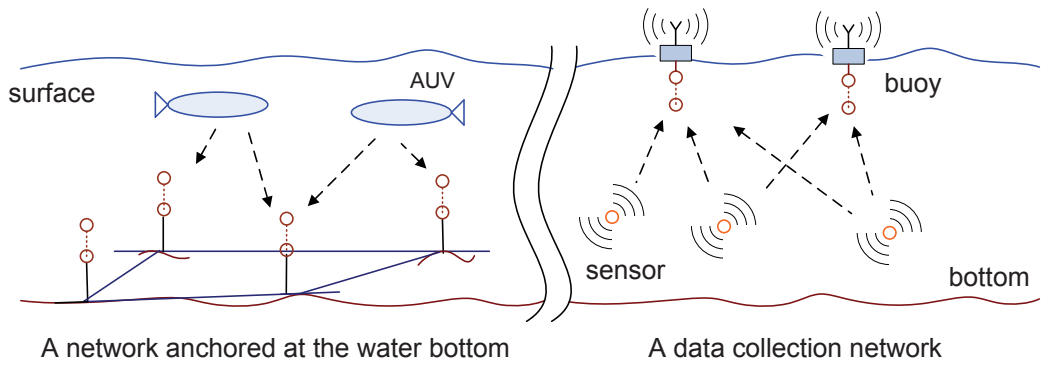


Figure 4.1: Two examples of underwater distributed antenna systems. In (a), the nodes anchored at water bottom are interconnected and also connected to a control center via fibers and cables. In (b), the gateways can communicate with each other and also with a control center via high-rate radio links.

Similar to terrestrial radio networks, UWA networks are prone to adversarial attacks, especially in critical missions, such as tactical surveillance, underwater asset protection, and commercial offshore oil and gas exploration. The attacks could be passive, such as eavesdropping and traffic analysis, or active, such as denial-of-service (DoS) attacks, resource consumption, reply attacks, and message modification [88]. Among all these attacks, the eavesdropping attack often precedes other types of attacks, due to the low cost and low probability of detection. In this chapter, we investigate the physical-layer security for underwater acoustic CoMP transmissions in the presence of eavesdropping attacks. Despite considerable progress on UWA communications and networking in the last decade [59, 89, 90, 91], research on UWA communication and network security has been very limited.

4.1.1 Physical-layer Security in Terrestrial Radio Networks

Due to the nature of broadcast transmission, wireless communication is vulnerable to security attacks. Besides classical security mechanisms such as cryptography, it has recently been shown that communication security can be largely enhanced by exploiting the physical-layer randomness [92, 93, 94]. Typical physical-layer transmission secrecy mechanisms include: (1) secrecy beamforming or precoding with a multi-antenna source and a single-antenna or multi-antenna destination [95, 96, 97]; (2) artificial noise-assisted friendly jamming where the jamming signal can be transmitted by helpers or embedded in the transmitted signal from the source node [93]; (3) full-duplex transceiving where the destination can simultaneously receive the desired signal from the source and transmit jamming signal to eavesdroppers [98, 99]; and (4) security key generation based on physical-layer randomness [100, 101].

Particularly about cooperative jamming in the Gaussian wiretap channel, interference alignment is investigated in [102] based on both the legitimate user's and the eavesdropper's channel state information (CSI), where the message from the legitimate transmitter is divided into M submessages, and the cooperative jamming signals from M helpers are specifically designed so that at the legitimate receiver, they are aligned in the same dimension and occupy the smallest signal space, while at the

eavesdropper, each jamming signal is aligned with a submessage for transmission secrecy. This approach achieves secure degrees-of-freedom (d.o.f.) of $\frac{M}{M+1}$. The above result is further extended in [103] to the scenario without the eavesdropper's CSI. Through transmitting an extra cooperative jamming signal after M submessages from the legitimate transmitter, the same secure d.o.f. of $\frac{M}{M+1}$ can be achieved. In [104], interference alignment and secrecy precoding are applied to a K-user Gaussian interference channel with confidential messages and a K-user interference channel with an external eavesdropper. It is shown that each user can achieve secure d.o.f. of $\frac{K-2}{2K-2}$ in the former scenario with probability one and secure d.o.f. of $\frac{K-2}{2K}$ in the latter scenario in the ergodic setting. For an overview on interference alignment for physical-layer security, please refer to [105, 106] and references therein.

4.1.2 Underwater Acoustic Network Security

In contrast to terrestrial radio networks, there has been very limited study on communication and network security in the UWA environment [107]. A literature survey of recent research about secure communication protocols over the network stack is presented in [108], which calls for more investigation on UWA network security to meet their rapidly growing applications. The DoS attacks, especially jamming attacks on the physical layer and the network layer are investigated in [109, 110, 111]. Particularly about transmission secrecy, a cooperative jamming method is investigated in

[112] for point-to-point transmissions with a friendly jammer. The jamming signal is known *a priori* only to the legitimate users. Therefore, the legitimate receiver can retrieve the useful information by joint channel estimation and interference suppression, while the eavesdropper cannot perform the above operation due to the lack of knowledge about the jamming signal. When multiple friendly jammers are available, jammer selection and power allocation strategies are investigated to minimize the signal-to-interference-and-noise ratio (SINR) at the eavesdropper while maintain a sufficiently large SINR at the legitimate receiver. A game-theory based anti-jamming system is proposed in [113] to secure transmissions from underwater nodes to a surface sink in the presence of a reactive jammer. The static game and the dynamic game between the underwater nodes and the jammer are studied to maximize their individual utilities based on the received SINR of the useful signal and the transmission cost. A closed form of the optimal allocation strategy is derived for the static game with known channel gains, and the reinforcement learning technique is adopted for the dynamic game where the channel gain is not available. The challenges of applying the received signal strength (RSS) for key generation in UWA networks are reviewed in [114]. It points out that, the RSS-based key generation approaches for terrestrial radio networks cannot be directly used in UWA networks since the long transmission delay of probes leads to a low key generation rate and the asymmetric RSS measurements decrease the key agreement probability. To increase the RSS-based key generation rate in UWA networks, it suggests to dividing the communication band

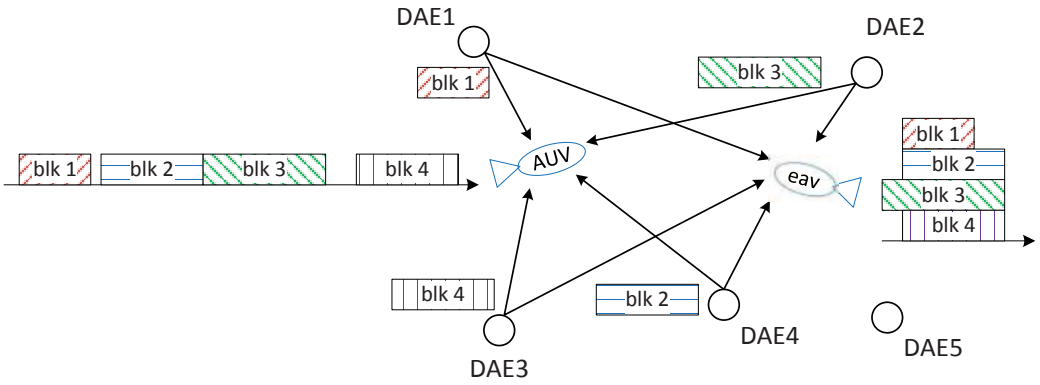


Figure 4.2: Illustration of signal alignment for secure coordinated multipoint transmissions

into multiple independent subchannels and performing multi-channel key generation (see details in [114]). To improve the key agreement probability, a smoothing filter is proposed to reduce the RSS random fluctuation. The proposed methods are validated via data collected from sea trials. In [115], a secret key generation method based on the UWA channel randomness is investigated. A predefined linear block code (e.g., a BCH code) is used for key bits extraction based on the observed channel frequency response at each user. To mitigate the channel observation difference at the two users due to noise or channel asymmetry, the syndrome information at one user is sent to and used at the other user for key reconciliation. The developed approach is validated via field experiments.

4.1.3 Our Work

The underwater acoustic environment features large sound propagation latency. For instance, at a nominal sound speed of 1500 m/s, the signal propagation latency for

a transmission distance of 3 km is two seconds. In this work, we exploit the low sound speed in water and the spatial diversity of system entities, including the DAEs, the legitimate user and the eavesdropper, and develop *signal alignment* strategies to secure underwater CoMP transmissions, where the transmission strategy is judiciously designed such that useful signals will collide at the eavesdropper while stay collision-free at the legitimate user, as illustrated in Fig. 4.2.

The main contributions of this work are the following.

- We develop practical signal alignment strategies for secure CoMP transmissions. Assuming *a priori* knowledge of the eavesdropper's location, the transmit DAE set and the transmission schedule and power level of each active DAE, are optimized with a goal of minimizing the maximal value of the eavesdropper's received SINRs under the constraint that the signals are well-separated at the legitimate user and the received signal-to-nose ratio (SNR) is lower bounded by a predetermined threshold. Taking the orthogonal frequency-division multiplexing (OFDM) as the physical-layer modulation technique, simulation and emulated experimental results show that the proposed method achieves much higher confidentiality than a benchmark method.

  When the eavesdropper's location is not available, a simplified version of the proposed method is developed to exploit the spatial diversity of DAEs and the legitimate user for transmission secrecy. Assuming a randomly located

eavesdropper, simulation results reveal that the proposed method still achieves a decent transmission secrecy performance.

- We further derive the lower and upper bounds of the secrecy capacity and the secure d.o.f. of the signal alignment method from an information-theoretic perspective. The derivation reveals that without the assistance of external helpers, a secure d.o.f. greater than $\frac{1}{2}$ can be achieved, and the lower bound of the secure d.o.f. increases with the total number of DAEs.

The eavesdropper considered in this work could be a legitimate user who is curious about the message transmitted to other system users, or an adversary who passively intercepts data transmissions under water. The honest-but-curious eavesdropper has been considered in many existing works, e.g., [93, 116, 117, 118, 119, 120], where *a priori* knowledge of the eavesdropper's location and/or CSI is often assumed. The underwater localization techniques developed in e.g., [121], can be used for system user positioning. When the eavesdropper is a passive adversary, techniques for passive target localization, such as range and bearing estimation [122, 123, 124] and target tracking [125] in passive sonar applications, can be applied to estimate the eavesdropper's location.

The proposed signal alignment concept for secure underwater CoMP transmission falls into the general category of interference alignment for physical-layer security. Different from existing research in radio networks that requires accurate CSI and

receiver front-end characteristics [102, 103, 126], the low sound speed in water provides grand opportunities to align signal and/or interference in the time domain. Other benefits of exploiting the low sound speed in transmission scheduling, such as UWA network throughput improvement, have been demonstrated in existing works [5, 8].

The rest of the chapter is organized as follows. The system model is presented in Section 4.2. The receiver processing method at the eavesdropper is developed in Section 4.3. The signal alignment method for secure CoMP transmissions is developed in Section 4.4. The secrecy capacity and secure d.o.f. are investigated in Section 4.5. Simulation and emulated experimental results are presented in Sections 4.6 and 4.7, respectively. Summary is presented in Section 4.8.

4.2 System Model for Coordinated Multipoint Transmissions

We consider an underwater system with N_{DAE} DAEs that are connected via cables or high-rate radio links. The information message from DAEs to a legitimate user is encoded into multiple blocks using an identical parameter set, with the center frequency, the frequency band, and the time duration of each block denoted by f_{c}, $\mathcal{B}$, and T_{bl}, respectively. Among a set of DAEs for transmission (denoted by $\mathcal{T}_{\mathrm{DAE}}$),

each DAE is assigned to transmit an individual block. When the total number of blocks is larger than the number of active DAEs, i.e., $M := |\mathcal{T}_{\text{DAE}}|$, a round-based transmission can be performed.

Denote d_μ as the distance between the μth DAE and the legitimate user, and denote t_μ as the transmission starting time of the μth DAE. The time-of-arrival of the signal from the μth DAE at the legitimate user is

$$\xi_\mu := t_\mu + d_\mu / c, \tag{4.1}$$

where c is the sound speed in water. Denote $N_{\text{pa},\mu}$ as the number of channel paths between the μth DAE and the legitimate user. We assume: (1) the amplitude of the pth path $A_{p,\mu}(t)$, is constant within one block, i.e., $A_{p,\mu}(t) = A_{p,\mu}$, and (2) the time varying delay of the pth path $\tau_{p,\mu}(t)$ relative to the delay in the line-of-sight transmission can be approximated as $\tau_{p,\mu}(t) \approx \tau_{p,\mu}$. The channel impulse response is

$$h_\mu(\tau) = \sum_{p=1}^{N_{\text{pa},\mu}} A_{p,\mu} \delta(\tau - \tau_{p,\mu} - \xi_\mu). \tag{4.2}$$

Denote $\tilde{s}_\mu(t)$ as the transmitted signal of the μth DAE. The signal arrived at the legitimate user can be formulated as

$$\tilde{y}_\mu(t) = \sum_{p=1}^{N_{\text{pa},\mu}} A_{p,\mu} \tilde{s}_\mu \left(t - \tau_{p,\mu} - \xi_\mu\right). \tag{4.3}$$

Corresponding to the active DAE set $\mathcal{T}_{\text{DAE}}$, the overall received signal at the legitimated user can be cast as

$$\tilde{y}(t) = \sum_{\mu \in \mathcal{T}_{\text{DAE}}} \sum_{p=1}^{N_{\text{pa},\mu}} A_{p,\mu} \tilde{s}_{\mu} \left(t - \tau_{p,\mu} - \xi_{\mu}\right) + \tilde{n}(t), \tag{4.4}$$

where $\tilde{n}(t)$ is the ambient noise.

According to the pth signal propagation path, denote $A(l_p(t), f) = A_0 l_p^k(t) \alpha(f)^{l_p(t)}$ as the power loss experienced by a signal of frequency f travelling over the distance $l_p(t)$ where k is the spreading factor, $\alpha(f)$ is the absorbtion coefficient, and A_0 is a scaling constant. Based on [127], the transmission power loss from the μth DAE to the legitimate user can be calculated as

$$P_{\text{loss}}(f_{\text{c}}, d_{\mu}) = \frac{\sum_{p=1}^{N_{\text{pa},\nu}} |A_{p,\nu}|^2}{B} \int_{f_{\text{c}}-B/2}^{f_{\text{c}}+B/2} Q^2(f) df \tag{4.5}$$

where $Q(f)$ is the path transfer function can be expressed in terms of a reference path's ($p = 0$) nominal transfer function as

$$Q(f) = \frac{1}{\sqrt{A\left(\bar{l}_0, f\right)}} \tag{4.6}$$

where $\bar{l}_0$ is the nominal length of a reference path.

Denote $P_{\text{tx},\mu}$ as the transmission power of the μth DAE, and denote $P_{\text{loss}}(f, d)$ as

the transmission power loss of an acoustic signal at frequency f for a distance d. Corresponding to the signal from the μth DAE, the received signal power at the legitimate user can be approximated as

$$P_{\mathrm{rx},\mu} = P_{\mathrm{tx},\mu} \cdot P_{\mathrm{loss}}(f_{\mathrm{c}}, d_{\mu}). \tag{4.7}$$

A similar system model can be developed for the eavesdropper. Denote $d_{\mu}^{(\mathrm{e})}$ as the distance between the μth DAE and the eavesdropper. The time-of-arrival of the signal from the μth DAE at the eavesdropper is

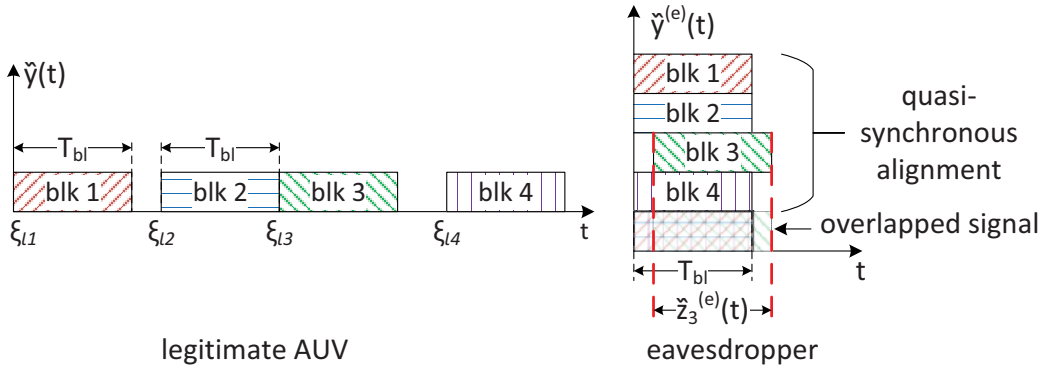
$$\xi_{\mu}^{(\mathrm{e})} := t_{\mu} + d_{\mu}^{(\mathrm{e})}/c. \tag{4.8}$$

The received signal at the eavesdropper can be formulated as

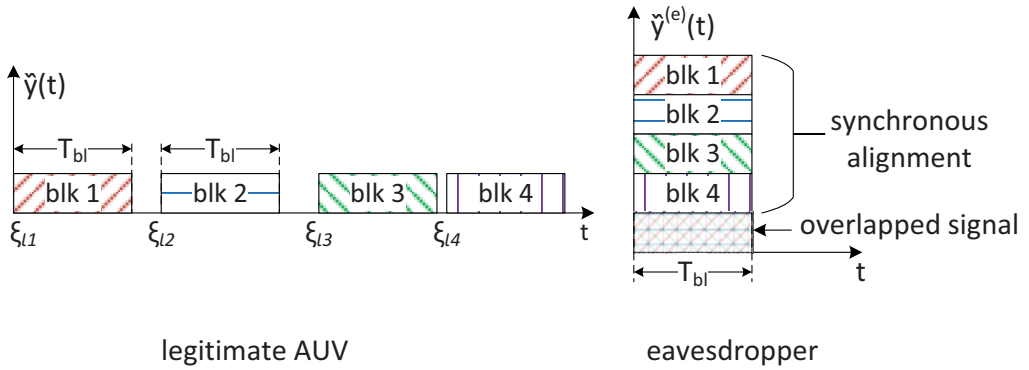
$$\tilde{y}^{(\mathrm{e})}(t) = \sum_{\mu \in \mathcal{T}_{\mathrm{DAE}}} \sum_{p=1}^{N_{\mathrm{pa},\mu}^{(\mathrm{e})}} A_{p,\mu}^{(\mathrm{e})} \tilde{s}_{\mu}^{(\mathrm{e})} \left(t - \tau_{p,\mu}^{(\mathrm{e})} - \xi_{\mu}^{(\mathrm{e})}\right) + \tilde{n}^{(\mathrm{e})}(t), \tag{4.9}$$

where $\{A_{p,\mu}^{(\mathrm{e})}, \tau_{p,\mu}^{(\mathrm{e})}\}$ denote the amplitude and delay of the pth path between the μth DAE and the eavesdropper, respectively, and $\tilde{n}^{(\mathrm{e})}(t)$ is the ambient noise. The received signal power corresponding to the μth DAE's transmission can be similarly obtained as

$$P_{\mathrm{rx},\mu}^{(\mathrm{e})} := P_{\mathrm{tx},\mu} \cdot P_{\mathrm{loss}}(f_{\mathrm{c}}, d_{\mu}^{(\mathrm{e})}). \tag{4.10}$$



(a) Quasi-synchronous alignment



(b) Synchronous alignment

Figure 4.3: Illustration of quasi-synchronous alignment and synchronous alignment of transmit blocks at the eavesdropper. As an illustration of the receiver processing at the eavesdropper, $\tilde{z}_3^{(e)}(t)$ in (a) is truncated from the received signal to recover the information symbols in block 3.

4.3 Receiver Processing at the Eavesdropper

To reduce the eavesdropper's interception capability, one can exploit the low sound speed in water to create signal collision at the eavesdropper while keep the signals well-separated at the legitimate user.

We consider that the eavesdropper has one receive antenna and performs the single-block decoding in the frequency domain by truncating each desired block from the received signal based on its time-of-arrival [128, 129, 130, 131], as illustrated in Fig. 4.3 (a). Taking the block transmitted by the μth DAE as the desired signal, the truncated signal can be expressed as

$$\tilde{z}_{\mu}^{(\mathrm{e})}(t) = \tilde{y}^{(\mathrm{e})}(t + \xi_{\mu}^{(\mathrm{e})}) \tag{4.11}$$

$$= \tilde{y}_{\mu}^{(\mathrm{e})}(t) + \underbrace{\sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} \tilde{y}_{\nu}^{(\mathrm{e})} \left(t - \xi_{\nu}^{(\mathrm{e})} + \xi_{\mu}^{(\mathrm{e})}\right)}_{:= \tilde{I}_{\mu}^{(\mathrm{e})}(t)} + \tilde{n}_{\mu}^{(\mathrm{e})}(t) \tag{4.12}$$

for $t \in [0,\ T_{\mathrm{bl}}]$, where $\tilde{I}_{\mu}^{(\mathrm{e})}(t)$ denotes the interference caused by other overlapped blocks.

The Fourier transform of (4.12) yields

$$\tilde{Z}_{\mu}^{(\mathrm{e})}(f) = \int_{0}^{T_{\mathrm{bl}}} \tilde{z}_{\mu}^{(\mathrm{e})}(t) e^{-j2\pi f t} dt \tag{4.13}$$

$$= \tilde{Y}_{\mu}^{(\mathrm{e})}(f) + \tilde{\mathcal{I}}_{\mu}^{(\mathrm{e})}(f) + \tilde{N}_{\mu}^{(\mathrm{e})}(f), \tag{4.14}$$

where

$$\tilde{\mathcal{I}}_{\mu}^{(\mathrm{e})}(f) := \sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} \int_{0}^{T_{\mathrm{bl}}} \tilde{y}_{\nu}^{(\mathrm{e})} \left(t - \xi_{\nu}^{(\mathrm{e})} + \xi_{\mu}^{(\mathrm{e})}\right) e^{-j2\pi f t} dt, \tag{4.15}$$

and $\tilde{N}_{\mu}^{(\mathrm{e})}(f)$ is similarly defined.

Assume that the ambient noise in the frequency domain is white and follows a zero-mean complex Gaussian distribution, i.e., $\tilde{N}_{\mu}^{(\mathrm{e})}(f) \sim \mathcal{CN}(0, N_0/2)$. The SINR at the eavesdropper can be formulated as

$$\lambda_{\mu}^{(\mathrm{e})} := \frac{\frac{1}{T_{\mathrm{bl}}} \int_{\mathcal{B}} \left| \tilde{Y}_{\mu}^{(\mathrm{e})}(f) \right|^2 df}{\frac{1}{T_{\mathrm{bl}}} \int_{\mathcal{B}} \left| \tilde{\mathcal{I}}_{\mu}^{(\mathrm{e})}(f) \right|^2 df + N_0 B/2}. \tag{4.16}$$

Based on the Parseval's Theorem [132], we have

$$\int_{\mathcal{B}} \left| \tilde{\mathcal{I}}_{\mu}^{(\mathrm{e})}(f) \right|^2 df = \frac{1}{2} \int_0^{T_{\mathrm{bl}}} \left| \sum_{\nu=1, \nu \neq \mu}^{N_{\mathrm{DAE}}} \tilde{y}_{\nu}^{(\mathrm{e})} \left(t - \xi_{\nu}^{(\mathrm{e})} + \xi_{\mu}^{(\mathrm{e})} \right) \right|^2 dt. \tag{4.17}$$

Moreover, to obtain a closed-form representation of (4.17), we approximate the transmitted waveform as white Gaussian noise and assume that all the blocks are independently distributed. Hence,

$$\begin{aligned} \int_{\mathcal{B}} \left| \tilde{\mathcal{I}}_{\mu}^{(\mathrm{e})}(f) \right|^2 df &\approx \frac{1}{2} \sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} \int_0^{T_{\mathrm{bl}}} \left| \tilde{y}_{\nu}^{(\mathrm{e})} \left(t - \xi_{\nu}^{(\mathrm{e})} + \xi_{\mu}^{(\mathrm{e})} \right) \right|^2 dt \\ &= \frac{1}{2} \sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} P_{\mathrm{rx},\nu}^{(\mathrm{e})} \left[T_{\mathrm{bl}} - |\xi_{\nu}^{(\mathrm{e})} - \xi_{\mu}^{(\mathrm{e})}| \right]^+, \end{aligned} \tag{4.18}$$

where $[\cdot]^+ := \max\{\cdot, 0\}$.

The SINR in (4.16) can be rewritten as

$$\lambda_{\mu}^{(\mathrm{e})} = \frac{P_{\mathrm{rx},\mu}^{(\mathrm{e})}}{N_0 B + \frac{1}{T_{\mathrm{bl}}} \sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} P_{\mathrm{rx},\nu}^{(\mathrm{e})} \left[T_{\mathrm{bl}} - |\xi_{\nu}^{(\mathrm{e})} - \xi_{\mu}^{(\mathrm{e})}| \right]^{+}}. \tag{4.19}$$

Following the same derivation, the SNR of each received block at the legitimate user can be obtained as

$$\lambda_{\mu} := \frac{P_{\mathrm{rx},\mu}}{N_0 B}. \tag{4.20}$$

4.4 Signal Alignment for Transmission Secrecy

To minimize the amount of information leaked to the eavesdropper, the active DAE set $\mathcal{T}_{\mathrm{DAE}}$, and the transmission starting time and transmission power of each active DAE, can be jointly optimized under a constraint that the received SNR at the legitimate user is sufficiently large for successful decoding. In this work, we take the maximal SINR of the blocks received by the eavesdropper as an indicator of the eavesdropper's interception performance.

4.4.1 Signal Alignment with Eavesdropper's Location Information

Sort the time-of-arrivals of signals from all the active DAEs at the legitimate user in an increasing order, i.e., $\{\xi_{\ell_1}, \xi_{\ell_2}, \cdots, \xi_{\ell_M}\}$, where the time-of-arrival ξ_{ℓ_μ} is related to the transmission starting time as $\xi_{\ell_\mu} = t_{\ell_\mu} + d_{\ell_\mu}/c$. The optimization problem can be cast as

$$\min_{\mathcal{T}_{\rm DAE}, \{t_\mu, P_{{\rm tx},\mu}:\ \mu \in \mathcal{T}_{\rm DAE}\}} \ \max_{\mu \in \mathcal{T}_{\rm DAE}} \lambda_\mu^{(\rm e)} \tag{4.21a}$$

$$\text{s.t.} \qquad \lambda_\mu \geq \Gamma_{\rm th}, \tag{4.21b}$$

$$\xi_{\ell_{\mu+1}} - \xi_{\ell_\mu} \geq T_{\rm bl}, \tag{4.21c}$$

$$0 \leq P_{{\rm tx},\mu} \leq P_{\rm th},\ \forall \mu \in \mathcal{T}_{\rm DAE} \tag{4.21d}$$

where $\Gamma_{\rm th}$ is the lower bound of the required decoding SNR at the legitimate user, (4.21c) ensures collision-free at the legitimate user, and $P_{\rm th}$ is the maximal transmission power at each DAE.

To make the optimization problem in (4.21) tractable, we introduce an auxiliary variable

$$\varrho := \max_{\mu \in \mathcal{T}_{\rm DAE}} \lambda_\mu^{(\rm e)}, \tag{4.22}$$

with a new constraint $\lambda_\mu^{(\mathrm{e})} \le \varrho$, $\forall \mu \in \mathcal{T}_{\mathrm{DAE}}$. The received SNR constraint in (4.21b) can be translated into a lower bound of the transmission power of each active DAE based on (4.7) and (4.20), which combined with (4.21d) yields $P_{\mathrm{tx},\mu,\mathrm{L}} \le P_{\mathrm{tx},\mu} \le P_{\mathrm{th}}$.

Define

$$\theta_{\nu,\mu} := \left[1 - \frac{|\xi_\nu^{(\mathrm{e})} - \xi_\mu^{(\mathrm{e})}|}{T_{\mathrm{bl}}}\right]^+, \tag{4.23}$$

which satisfies $\theta_{\mu,\nu} = \theta_{\nu,\mu}$. Define a column vector $\mathbf{q}$ of length M, with its μth element

$$[\mathbf{q}]_\mu = \frac{1}{N_0 B} P_{\mathrm{tx},\mu} \cdot P_{\mathrm{loss}}(f_{\mathrm{c}}, d_\mu^{(\mathrm{e})}), \tag{4.24}$$

for $\mu = 1, 2, \cdots, M$. Denote $\boldsymbol{\xi} := [\xi_{\ell_1}, \xi_{\ell_2}, \cdots, \xi_{\ell_M}]^{\mathrm{T}}$. As detailed in Appendix A.1, the optimization problem in (4.21) can be reformulated as

$$\min_{\varrho, \mathcal{T}_{\mathrm{DAE}}, \{t_\mu, P_{\mathrm{tx},\mu}:\ \mu \in \mathcal{T}_{\mathrm{DAE}}\}} \quad \varrho \tag{4.25a}$$

$$\text{s.t.} \qquad \tilde{\mathbf{G}}_1(\boldsymbol{\xi}, \varrho)\mathbf{q} \le \varrho \mathbf{1}, \tag{4.25b}$$

$$\mathbf{G}_2 \boldsymbol{\xi} \ge T_{\mathrm{bl}} \mathbf{1}, \tag{4.25c}$$

$$\mathbf{q}_{\mathrm{L}} \le \mathbf{q} \le \mathbf{q}_{\mathrm{U}}, \tag{4.25d}$$

where $\mathbf{1}$ is a column vector of length M with unity elements, $\mathbf{q}_{\mathrm{L}}$ and $\mathbf{q}_{\mathrm{U}}$ are similarly defined as $\mathbf{q}$ in (4.24) through replacing $P_{\mathrm{tx},\mu}$ by $P_{\mathrm{tx},\mu,\mathrm{L}}$ and P_{th}, respectively, and

$\tilde{\mathbf{G}}_1(\boldsymbol{\xi}, \varrho)$ and $\mathbf{G}_2$ are two matrices of size $M \times M$ and $(M-1) \times M$, respectively,

$$\tilde{\mathbf{G}}_1(\boldsymbol{\xi}, \varrho) := \begin{bmatrix} 1 & -\varrho\theta_{1,2} & \cdots & -\varrho\theta_{1,M} \\ -\varrho\theta_{2,1} & 1 & \cdots & -\varrho\theta_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ -\varrho\theta_{M,1} & -\varrho\theta_{M,2} & \cdots & 1 \end{bmatrix} \tag{4.26}$$

and

$$\mathbf{G}_2 := \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}. \tag{4.27}$$

As detailed in Appendix A.2, through an appropriate modification of (4.21) based on the receiver processing algorithms at the legitimate user and at the eavesdropper, the proposed signal alignment method can be extended to the scenario with multiple users and multiple eavesdroppers and the scenario when an eavesdropper has multiple geographically distributed antenna elements.

4.4.2 Optimization Problem Solver

Although the optimization problem in (4.25) is derived for quasi-synchronous signal alignment at the eavesdropper (see Fig. 4.3 (a)), there are scenarios that synchronous signal alignment can be achieved when the blocks can arrive simultaneously at the eavesdropper while stay collision-free at the legitimate user; see an example in Fig. 4.3 (b). Under those scenarios, the transmission starting time of each active DAE can be immediately determined based on the sound propagation delay to the eavesdropper. For ease of exposition, in this section we will first present a solution to the optimization problem in (4.25) for synchronous signal alignment, and then proceed to an optimization problem solver for quasi-synchronous signal alignment.

4.4.2.1 Synchronous Signal Alignment

When synchronous signal alignment can be achieved, all the blocks have an identical time-of-arrival at the eavesdropper, i.e., $\xi_\mu^{(\mathrm{e})} = \xi_\nu^{(\mathrm{e})},\ \forall \mu, \nu$, which can be assumed zero without loss of generality. The transmission starting time of each active DAE, e.g., the μth DAE, can be determined as $t_\mu = -d_\mu^{(\mathrm{e})}/c$. Furthermore, $\theta_{\nu,\mu} = 1, \forall \mu, \nu$, and

the matrix $\tilde{\mathbf{G}}_1(\boldsymbol{\xi}, \varrho)$ can be simplified as

$$\mathbf{G}_1(\varrho) := \begin{bmatrix} 1 & -\varrho & \cdots & -\varrho \\ -\varrho & 1 & \cdots & -\varrho \\ \vdots & \vdots & \ddots & \vdots \\ -\varrho & -\varrho & \cdots & 1 \end{bmatrix}. \tag{4.28}$$

The optimization problem for synchronous signal alignment can be cast as

$$\min_{\varrho, \mathcal{T}_{\mathrm{DAE}}, \{P_{\mathrm{tx},\mu}:\ \mu \in \mathcal{T}_{\mathrm{DAE}}\}} \quad \varrho \tag{4.29a}$$

$$\text{s.t.} \qquad \mathbf{G}_1(\varrho)\mathbf{q} \leq \varrho \mathbf{1}, \tag{4.29b}$$

$$\mathbf{q}_{\mathrm{L}} \leq \mathbf{q} \leq \mathbf{q}_{\mathrm{U}}. \tag{4.29c}$$

To solve the optimization problem in (4.29), we first enumerate all the possible combinations of DAEs. Taking each combination as the active DAE set $\mathcal{T}_{\mathrm{DAE}}$, the optimal transmission power of each active DAE can be computed via the bisectional search method and the Simplex method [133] as stated in Algorithm IV-1. The combination which yields the minimal value of ϱ^{opt} provides the final optimal solution. We compared the above optimization method with a grid search method. The two methods yield identical solutions.

Algorithm IV-1 Synchronous Signal Alignment

Input: Active DAE set $\mathcal{T}_{\rm DAE}$
Output: Optimal ϱ and the transmission power of each active DAE

Generate initial $\varrho_{\min}$, $\varrho_{\max}$, ϵ, $N_{\max}$, and $i \leftarrow 1$
while $\frac{\varrho_{\max}-\varrho_{\min}}{(\varrho_{\max}+\varrho_{\min})/2} \geq \epsilon$ **and** $i \leq N_{\max}$ **do**
    $\varrho \leftarrow (\varrho_{\min} + \varrho_{\max})/2$
    With a known ϱ, compute $\mathbf{G}_1(\varrho)$, $\mathbf{q}_{\rm L}$, and $\mathbf{q}_{\rm U}$ in (4.29), and use the Simplex method to solve for the transmission power vector $\tilde{\mathbf{p}}_{\rm tx} := [\tilde{P}_{{\rm tx},\ell_1}, \tilde{P}_{{\rm tx},\ell_2}, \cdots, \tilde{P}_{{\rm tx},\ell_{\rm M}}]^{\rm T}$ whose summation is minimal.
    if a feasible solution of $\tilde{\mathbf{p}}_{\rm tx}$ can be found **then**
        $\varrho_{\max} \leftarrow \varrho$, $\mathbf{p}_{\rm tx}^{\rm opt} \leftarrow \tilde{\mathbf{p}}_{\rm tx}$
    else
        $\varrho_{\min} \leftarrow \varrho$
    $i \leftarrow i + 1$
$\varrho^{\rm opt} \leftarrow \varrho_{\max}$
return $\varrho^{\rm opt}$ and $\mathbf{p}_{\rm tx}^{\rm opt}$

4.4.2.2 Quasi-synchronous Signal Alignment

When the synchronous signal alignment at the eavesdropper and the non-overlapping constraint at the legitimate user cannot be simultaneously satisfied, quasi-synchronous alignment at the eavesdropper can be performed. Similar to Algorithm IV-1, to solve the optimization problem in (4.25), we first enumerate all the possible combinations of DAEs. Taking each combination as the active DAE set $\mathcal{T}_{\rm DAE}$, the optimal transmission starting time and transmission power of each active DAE that yield the minimal value of ϱ can be computed via the simulated annealing method [134] and a random search method [135]. The DAE combination which leads to the minimal value of $\varrho^{\rm opt}$ provides the final optimal solution. The overall algorithm is described in Algorithm IV-2. Specifically, in the ith iteration of Algorithm 2,

a feasible transmission schedule $\tilde{t}$ is obtained based on the transmission starting time $\mathbf{t}_{i-1}$ in the $(i-1)$th iteration. Based on $\tilde{\mathbf{t}}$, the optimal transmission power $\tilde{\mathbf{p}}_{\text{tx}}$ and the minimal SINR $\tilde{\varrho}$ at the eavesdropper are calculated. The transmission starting time $\tilde{\mathbf{t}}$ and transmission power $\tilde{\mathbf{p}}_{\text{tx}}$ are recorded if the minimal SINR $\tilde{\varrho}$ is smaller than the minimal SINRs obtained in previous iterations. Based on the accepting criteria of the simulated annealing method, the transmission starting time $\tilde{\mathbf{t}}$ will be accepted, i.e., $\mathbf{t}_i = \tilde{\mathbf{t}}$ and otherwise $\mathbf{t}_i = \mathbf{t}_{i-1}$.

To initialize Algorithm 2, $\mathbf{t}_0$ is taken as the one that allows consecutive arrivals of blocks at the legitimate user with zero intervals, and the arrival sequence of blocks at the legitimate user is identical to the arrival sequence when the blocks arrive simultaneously at the eavesdropper.

Set $\beta = 0.98$ and $N_{\max} = 200$. We compared Algorithm 2 with a grid search method in simulations, which showed that the normalized optimal SINR difference is 1.9×10^{-4} in decimal, averaged over 1000 random system layouts.

Remark 3: In the proposed method, all the transmissions from DAEs are constrained to be useful signals. In fact, some DAEs near the eavesdropper can serve as friendly jammers. The jamming transmission, however, has to be carefully designed to minimize its impact at the legitimate user. A detailed design, including (1) DAE selection for useful signal transmission and for jamming transmission, (2) transmission

Algorithm IV-2 Quasi-synchronous Signal Alignment

Input: Active DAE set $\mathcal{T}_{\mathrm{DAE}}$

Output: Optimal ϱ, and the transmission starting time and power of each active DAE

1: Generate initial $\mathbf{t}_0$, $T_{\mathrm{emp}}^{(0)}$, $N_{\max}$, ϱ^{opt}, and $i \leftarrow 1$
2: **while** $i \leq N_{\max}$ **do**
3: $T_{\mathrm{emp}}^{(i)} \leftarrow \beta T_{\mathrm{emp}}^{(i-1)}$ $(0 < \beta < 1)$, $\tilde{\mathbf{t}} \leftarrow$ **SubFunction**$(\mathbf{t}_{i-1})$ as described from Line 13
4: Use Algorithm IV-1 to obtain the optimal $\tilde{\varrho}$ and $\tilde{\mathbf{p}}_{\mathrm{tx}}$ based on $\tilde{\mathbf{t}}$
5: **if** $\tilde{\varrho} < \varrho^{\mathrm{opt}}$ **then**
6: $\varrho^{\mathrm{opt}} \leftarrow \tilde{\varrho}$, $\mathbf{t}^{\mathrm{opt}} \leftarrow \tilde{\mathbf{t}}$, and $\mathbf{p}_{\mathrm{tx}}^{\mathrm{opt}} \leftarrow \tilde{\mathbf{p}}_{\mathrm{tx}}$
7: **if** $\tilde{\varrho} - \varrho_{i-1} < 0$ **or** $\exp\left(-\frac{\tilde{\varrho}-\varrho_{i-1}}{T_{\mathrm{emp}}^{(i)}}\right) > u$ where $u \sim \mathcal{U}[0,1]$ **then**
8: $\mathbf{t}_i \leftarrow \tilde{\mathbf{t}}$, $\varrho_i \leftarrow \tilde{\varrho}$
9: **else**
10: $\mathbf{t}_i \leftarrow \mathbf{t}_{i-1}$, $\varrho_i \leftarrow \varrho_{i-1}$
11: $i \leftarrow i + 1$
12: **return** ϱ^{opt}, $\mathbf{t}_{\mathrm{tx}}^{\mathrm{opt}}$ and $\mathbf{p}_{\mathrm{tx}}^{\mathrm{opt}}$

SubFunction: Generate $\tilde{\mathbf{t}}$ based on $\mathbf{t}_{i-1}$

13: Based on $\mathbf{t}_{i-1}$, calculate the time-of-arrivals of blocks at the eavesdropper and sort them in an increasing order $\{\xi_{\kappa_1}^{(\mathrm{e})}, \xi_{\kappa_2}^{(\mathrm{e})}, \cdots, \xi_{\kappa_M}^{(\mathrm{e})}\}$
14: Randomly pick up blocks κ_m and κ_n, and assume $m < n$ and $\kappa_m = \ell_\mu$, $\kappa_n = \ell_\nu$
15: $\Delta t_{\kappa_m} \leftarrow \min\left([\xi_{\ell_{\mu+1}} - \xi_{\ell_\mu} - T_{\mathrm{bl}}]^+, \left|\xi_{\kappa_m}^{(\mathrm{e})} - \xi_{\kappa_n}^{(\mathrm{e})}\right|\right)$, $\tilde{\mathbf{t}}_1 \leftarrow \mathbf{t}_{i-1}$
16: $\Delta t_{\kappa_n} \leftarrow \min\left([\xi_{\ell_\nu} - \xi_{\ell_{\nu-1}} - T_{\mathrm{bl}}]^+, \left|\xi_{\kappa_m}^{(\mathrm{e})} - \xi_{\kappa_n}^{(\mathrm{e})}\right|\right)$, $\tilde{\mathbf{t}}_2 \leftarrow \mathbf{t}_{i-1}$
17: **if** $\Delta t_{\kappa_m} = 0$ **then**
18: $[\tilde{\mathbf{t}}_2]_{\ell_\nu} \leftarrow [\mathbf{t}_{i-1}]_{\ell_\nu} - u_1 \Delta t_{\kappa_n} (u_1 \sim \mathcal{U}[0,1])$
19: **else**
20: $[\tilde{\mathbf{t}}_1]_{\ell_\mu} \leftarrow [\mathbf{t}_{i-1}]_{\ell_\mu} + u_1 \Delta t_{\kappa_m}$, $[\tilde{\mathbf{t}}_2]_{\ell_\nu} \leftarrow [\mathbf{t}_{i-1}]_{\ell_\nu} - u_2 \Delta t_{\kappa_n}$ $(u_1, u_2 \sim \mathcal{U}[0,1])$
21: **return** $\tilde{\mathbf{t}}_1$ or $\tilde{\mathbf{t}}_2$ which results in a lower maximal SINR at the eavesdropper

scheduling and power control, and (3) jamming signal lengths, warrants another piece of work.

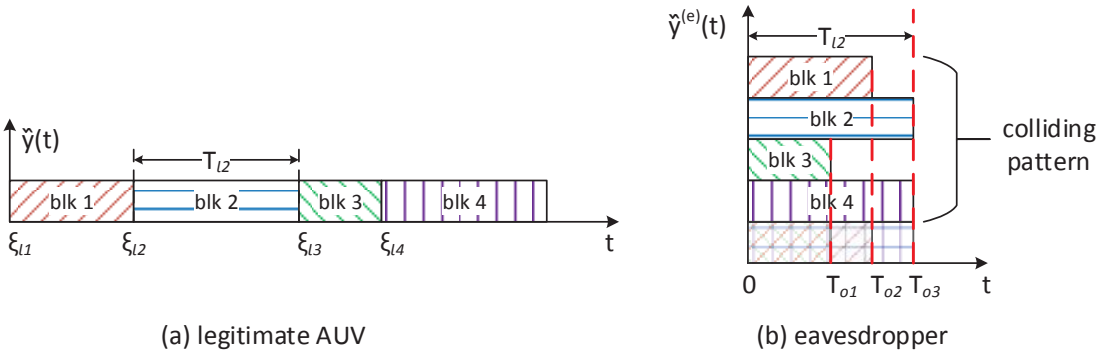


Figure 4.4: Illustration of time-of-arrivals of DAE transmitted signals

4.4.3 Signal Alignment without Eavesdropper's Location

When the eavesdropper's location information is not available, we propose to select a set of active DAEs that satisfy both the lower bound constraint of the received SNR at the legitimate user and the maximal transmission power constraint. The transmission starting time of each active DAE is determined by allowing blocks arriving consecutively with zero interval at the legitimate user in an arbitrary order. Exploiting the spatial distribution of DAEs and the low sound speed in water, signals from multiple DAEs have a large probability of collision at a randomly located eavesdropper. We will examine the security performance of this scheme in simulations.

4.5 Secrecy Capacity in AWGN Channels

In this section, we investigate the secrecy capacity of the signal alignment method in AWGN channels. The secrecy capacity per CoMP transmission is defined as

$$C_{\mathrm{s}} = \max_{\{\tilde{s}_\mu(t):\mu\in\mathcal{T}_{\mathrm{DAE}}\}} \left[I\left(\{\tilde{s}_\mu(t):\mu\in\mathcal{T}_{\mathrm{DAE}}\};\tilde{y}(t)\right) - I\left(\{\tilde{s}_\mu(t):\mu\in\mathcal{T}_{\mathrm{DAE}}\};\tilde{y}^{(\mathrm{e})}(t)\right)\right]. \tag{4.30}$$

It has been shown that with a Gaussian input, the secrecy capacity equals to the difference between the legitimate link capacity C_{U} and the wiretap link capacity C_{E} [92], i.e.,

$$C_{\mathrm{s}} = C_{\mathrm{U}} - C_{\mathrm{E}}. \tag{4.31}$$

Different from the signal alignment method for the block transmission with an identical length, we will let each active DAE transmit for a maximal time duration as long as there is no collision at the legitimate user, as illustrated in Fig. 4.4. The transmission time duration of the ℓ_μth DAE is

$$T_{\ell_\mu} := \xi_{\ell_{\mu+1}} - \xi_{\ell_\mu}, \qquad 1 \le \mu \le M-1. \tag{4.32}$$

Define

$$T_{\max} := \max\{T_{\ell_1}, \cdots, T_{\ell_{M-1}}\}. \tag{4.33}$$

Without loss of generality, we take the time duration of the signal arriving last at the legitimate user as

$$T_{\ell_M} = T_{\max}, \tag{4.34}$$

which leads to a total reception time duration of the legitimate user

$$T_{\text{total}} := \sum_{\mu=1}^{M} T_{\ell_\mu} = \xi_{\ell_M} - \xi_{\ell_1} + T_{\ell_M}. \tag{4.35}$$

Denote σ_{w}^2 as the ambient noise variance. The legitimate link capacity per CoMP transmission is

$$C_{\text{U}} = \sum_{\mu \in \mathcal{T}_{\text{DAE}}} \frac{T_\mu}{2} \log_2 \left(1 + \frac{P_{\text{rx},\mu}}{\sigma_{\text{w}}^2} \right). \tag{4.36}$$

Consider that the wiretap link can be regarded as a multi-access (MAC) channel. Its capacity is lower bounded by

$$C_{\text{E}} \geq \sum_{\mu \in \mathcal{T}_{\text{DAE}}} T_\mu \log_2 \left(1 + \frac{P_{\text{rx},\mu}^{(\text{e})}}{\sigma_{\text{w}}^2 + P_{\text{I},\mu}^{(\text{e})}} \right), \tag{4.37}$$

where the power of the inter-block-interference at the eavesdropper can be approximated as

$$P_{\text{I},\mu}^{(\text{e})} \approx \frac{1}{T_\mu} \sum_{\nu \in \mathcal{T}_{\text{DAE}}, \nu \neq \mu} P_{\text{rx},\nu}^{(\text{e})} \min(T_\mu, T_\nu). \tag{4.38}$$

To derive the upper bound of the wiretap link capacity, we sort the transmission time duration of active DAEs in an increasing order $\{T_{o_1}, T_{o_2}, \ldots, T_{o_M}\}$, and for expression

convenience, we denote $T_{o_0} = 0$. As illustrated in Fig. 4.4 (b), by truncating the received signal at the eavesdropper into multiple segments based on the number of collided blocks, the upper bound of the wiretap link capacity can be obtained as the sum rate of each truncation,

$$C_{\mathrm{E}} \leq \sum_{\mu=1}^{M-1} \frac{(T_{o_\mu} - T_{o_{\mu-1}})}{2} \log_2 \left(1 + \frac{\sum_{\nu \geq \mu} P_{\mathrm{rx},o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2} \right). \tag{4.39}$$

The lower bound $C_{\mathrm{s,Lo}}$ and the upper bound $C_{\mathrm{s,Up}}$ of the secrecy capacity per CoMP transmission can be obtained as

$$C_{\mathrm{s,Up}} = \left[\sum_{\mu \in \mathcal{T}_{\mathrm{DAE}}} \frac{1}{2} \left\{ T_\mu \log_2 \left(1 + \frac{P_{\mathrm{rx},\mu}}{\sigma_{\mathrm{w}}^2} \right) - T_\mu \log_2 \left(1 + \frac{P_{\mathrm{rx},\mu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{I},\mu}^{(\mathrm{e})}} \right) \right\} \right]^+, \tag{4.40}$$

$$\begin{aligned} C_{\mathrm{s,Lo}} = \Bigg[& \left\{ \sum_{\mu \in \mathcal{T}_{\mathrm{DAE}}} \frac{T_\mu}{2} \log_2 \left(1 + \frac{P_{\mathrm{rx},\mu}}{\sigma_{\mathrm{w}}^2} \right) \right\} \\ & - \left\{ \sum_{\mu=1}^{M-1} \frac{(T_{o_\mu} - T_{o_{\mu-1}})}{2} \log_2 \left(1 + \frac{\sum_{\nu \geq \mu} P_{\mathrm{rx},o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2} \right) \right\} \Bigg]^+. \end{aligned} \tag{4.41}$$

Due to the lack of closed-form expression of the secrecy capacity, we will bound the maximal secrecy capacity $C_{\mathrm{s}}^{\mathrm{opt}}$ by maximizing its lower bound in bits per second, i.e.,

$$\max_{\mathcal{T}_{\mathrm{DAE}},\{P_{\mathrm{tx},\mu}\}_{\mu\in\mathcal{T}_{\mathrm{DAE}}}} \frac{C_{\mathrm{s,Lo}}}{T_{\mathrm{total}}} \tag{4.42a}$$

$$\text{s.t.} \qquad 0 \leq P_{\mathrm{tx},\mu} \leq P_{\mathrm{th}},\ \forall \mu. \tag{4.42b}$$

To solve the optimization problem, we first enumerate all the possible combinations of DAEs. Taking each combination as an active DAE set, the total transmission duration T_{total} can be calculated *a priori*. With the known T_{total}, the objective function of (4.42) can be rewritten as a difference of two convex functions (DC). With the constraints being affine, the optimization problem (40) is a standard DC programming problem [136]. Hence, the global optimal transmission power of each active DAE that maximizes the lower bound $C_{\mathrm{s,Lo}}$ can be obtain by the DC programming algorithm proposed in [137]. Furthermore, it can be shown that in both the high transmission SNR regime and the low transmission SNR regime, $C_{s,\mathrm{Lo}}$ can be approximated as *weak* concave, hence a low-complex solver, such as the interior point method can be used. The combination that has the maximal lower bound of the secrecy capacity leads to the optimal solution.

It is worth noting that, in the very low transmission SNR regime where noise dominates the receiver decoding performance, instead of aligning multiple useful signals at the eavesdropper to create self interference, it can be easily show that the maximal secrecy capacity is achieved by taking the DAE having the largest difference

between its distance to the legitimate user and to the eavesdropper for transmission at the maximal power level. In this scenario, the secrecy capacity of Gaussian wiretap channels applies [92].

To quantify the cost for transmission secrecy, the secure d.o.f., which is defined as the ratio of the secrecy capacity to the capacity without secrecy constraint as the transmission power goes to infinite, is often used. Here, we define the secure d.o.f. as the ratio of the secrecy capacity to the capacity of the channel between the legitimate user and its nearest DAE μ' with a transmission power P_{th},

$$\eta := \lim_{P_{\text{th}} \to \infty} \frac{C_{\text{s}}^{\text{opt}} / T_{\text{total}}^{\text{opt}}}{\frac{1}{2} \log_2 \left(1 + \frac{P_{\text{rx},\mu'}}{\sigma_{\text{w}}^2}\right)}. \tag{4.43}$$

The upper and lower bounds of the secure d.o.f. are derived in Appendix A.3, by substituting the upper and lower bounds of the secrecy capacity in (4.40) and (4.41) into (4.43). The derivation yields an upper bound of 1, and a lower bound of $1 - \frac{T_{\max}}{T_{\text{total}}}$, where $\frac{T_{\max}}{T_{\text{total}}}$ is the ratio of received signal lengths at the eavesdropper and the legitimate user. Since $T_{\max} \leq \frac{T_{\text{total}}}{2}$ (see (4.33), (4.34) and (4.35)), the lower bound indicates that a secure d.o.f. greater than $\frac{1}{2}$ can be achieved.

4.6 Simulation Results

We consider a system with $N_{\mathrm{DAE}} = 4$ DAEs, and assume that all the DAEs, the legitimate user and the eavesdropper are uniformly distributed within a disk of radius 4 km. In particular, the distance between any two DAEs is constrained to vary from 100 m to 8 km, and each DAE can cover the entire disk area. The simulation results are averaged over 1000 random system layouts. The transmission loss of the acoustic signal is modeled as

$$P_{\mathrm{loss}}(f, d) \propto e^{-\alpha(f)d} d^{-\beta}, \tag{4.44}$$

where $\alpha(f)$ is the frequency-dependent absorption coefficient (see [138] for a simplified formula of $\alpha(f)$), and β is the path-loss exponent taking a practical value of 1.5. Despite the simple form of the transmission loss in (4.44), it suffices to validate the performance of the proposed signal alignment methods. In practical systems, the transmission loss can be estimated based on the received signal strength in field [139].

We assume that each signal block is modulated by the zero-padded (ZP) OFDM technique. Denote K as the total number of subcarriers, and denote T as the time duration of each OFDM symbol. The kth subcarrier frequency is $f_k := f_{\mathrm{c}} + k/T$, for $k = -K/2, \cdots, (K/2) - 1$. The signal bandwidth is therefore $B = K/T$. To avoid the inter-symbol-interference caused by multiple channel paths, a guard interval of

length $T_{\rm g}$ is padded at the end of each OFDM symbol. The total time duration of each OFDM block is thus $T_{\rm bl} = T + T_{\rm g}$. Denote $s_\mu[k]$ as the data (or pilot) symbol at the kth subcarrier of the μth DAE. The transmitted signal at the μth DAE can be formulated as

$$\tilde{s}_\mu(t) = 2\Re\left\{\sum_{k=-K/2}^{K/2-1} s_\mu[k] e^{j2\pi f_k t} g(t)\right\}, \qquad t \in [0,\ T_{\rm bl}] \tag{4.45}$$

where $g(t)$ is a rectangular window being one for $t \in [0,\ T]$ and zero elsewhere.

The ZP-OFDM parameters are listed in Table 4.1. Out of 1024 subcarriers, 96 are null subcarriers with 24 on each edge for band protection and 48 distributed evenly in the middle, 256 are pilot subcarriers for channel estimation, and the remaining 672 are data subcarriers for information delivery. The data symbols are encoded with a rate-1/2 nonbinary low-density parity-check (LDPC) code [140] and modulated with a QPSK constellation, which leads to a data rate

$$R = \frac{1}{2}\frac{|\mathcal{S}_{\rm D}|}{T + T_{\rm g}}\log_2 4 = 5.2 \text{ kb/s}, \tag{4.46}$$

where $|\mathcal{S}_{\rm D}|$ denotes the number of data subcarriers.

To simulate underwater acoustic channels, we assume that the channel between each transmit and receive pair consists of 10 discrete paths, where the inter-arrival time of paths follows an exponential distribution with a mean of 1 ms. The path amplitudes

Table 4.1
OFDM Parameters in Simulations.

center frequency: $f_{\rm c}$	13 kHz
bandwidth: B	9.77 kHz
# of subcarriers: K	1024
time duration: T	104.86 ms
frequency spacing: $\Delta f := 1/T$	9.54 Hz
guard interval: $T_{\rm g}$	24.6 ms

are Rayleigh distributed with an average power decreasing exponentially with the delay, where the difference between the beginning and the end of the guard time is 20 dB. We assume that all the paths are time-invariant with zero Doppler rates.

The sparse channel estimation method [141] and the linear minimum mean square error (LMMSE) estimator are adopted for channel estimation and symbol detection, respectively. The block-error-rate (BLER) performance is used as the decoding performance metric. The average transmission SNR to be used in the sequel is defined as

$$\text{TSNR} := \frac{1}{|\mathcal{T}_{\rm DAE}|} \sum_{\mu \in \mathcal{T}_{\rm DAE}} \frac{P_{{\rm tx},\mu}}{N_0 B}. \tag{4.47}$$

4.6.1 BLER Performance

We compare the decoding performance of the eavesdropper and the legitimate user in three configurations.

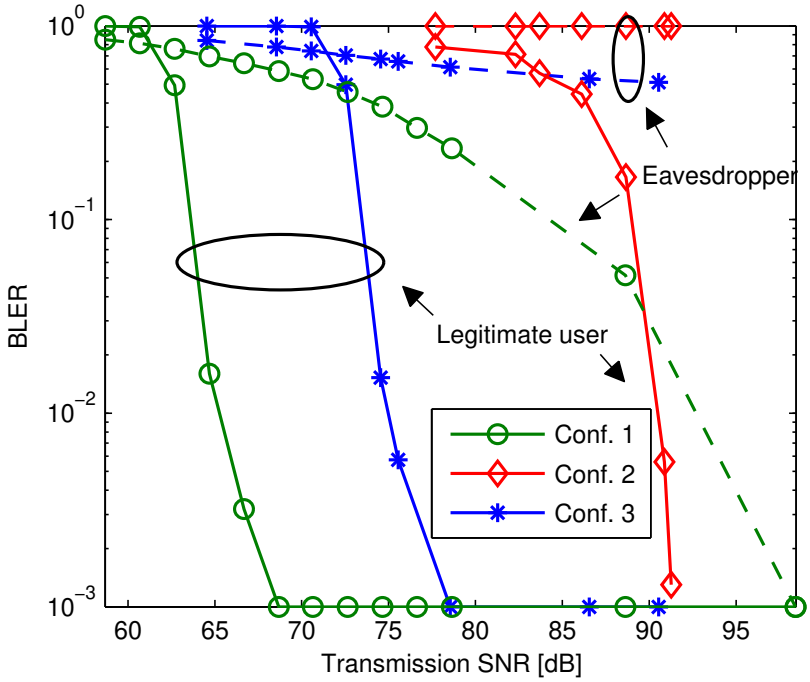


Figure 4.5: Decoding BLERs of the legitimate user and the eavesdropper. Both have single receive antenna.

- *Conf. 1*: A benchmark method that takes the DAE nearest to the legitimate user to transmit consecutively all the OFDM blocks;
- *Conf. 2*: The signal alignment method with the eavesdropper's location information;
- *Conf. 3*: The signal alignment method without the eavesdropper's location information.

Assume that both the legitimate user and the eavesdropper have one receive hydrophone. The BLER performance of the eavesdropper and the legitimate user in the three configurations are shown in Fig. 4.5. The simulation results reveal that the

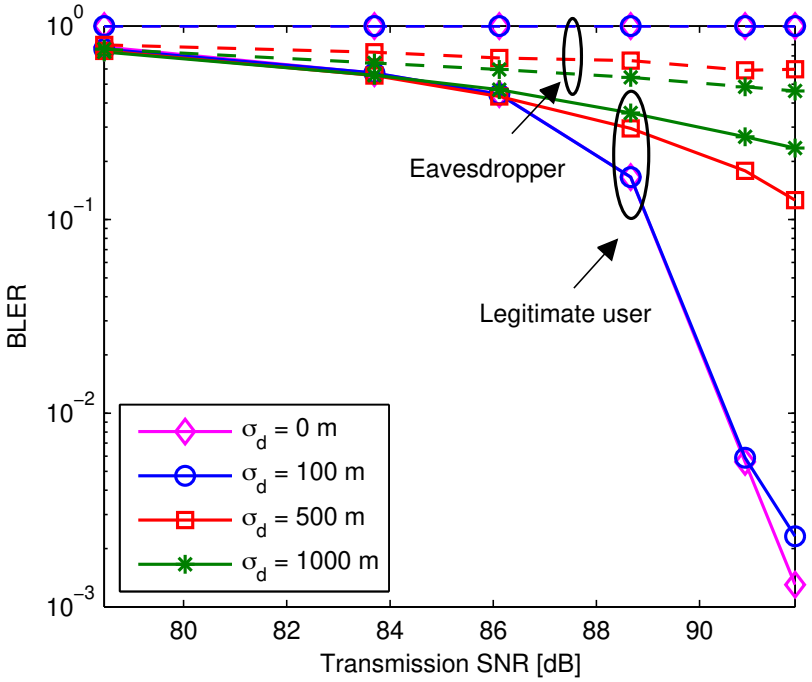


Figure 4.6: Decoding BLERs of the legitimate user and the eavesdropper with location inaccuracy.

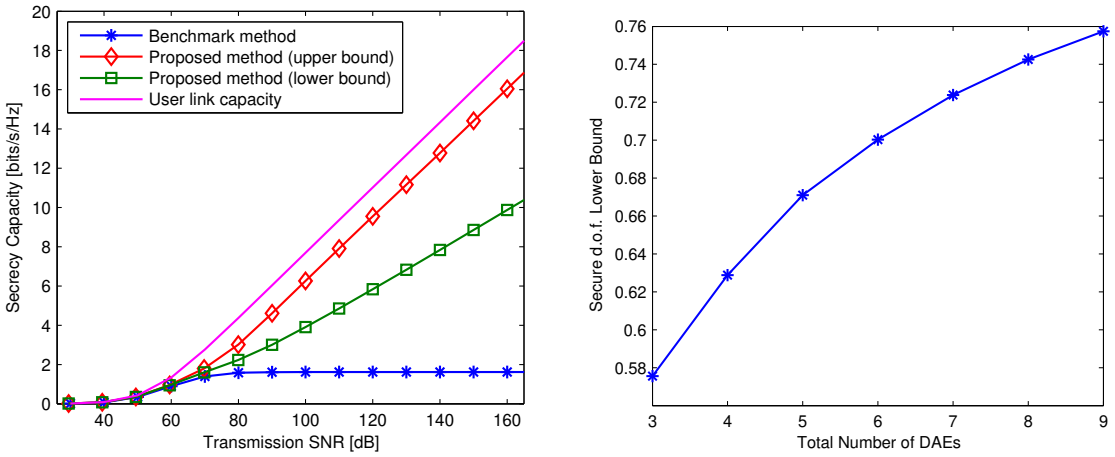
signal alignment method with the eavesdropper's location information achieves the highest level of transmission secrecy — the eavesdropper cannot decode any block, whereas it consumes vast transmission power in order to minimize the eavesdropper's interception capability. Compared to the benchmark method, better secrecy performance can be achieved by the signal alignment method without the eavesdropper's location information, benefited from the spatial diversity of DAEs and the legitimate user.

4.6.2 Sensitivity Analysis

To evaluate the performance sensitivity of the signal alignment method with the eavesdropper's location information to the user's and the eavesdropper's location inaccuracy, we assume that the user's and the eavesdropper's location errors both are independently and identically distributed in longitude and latitude, and in each direction the error follows a Gaussian distribution $\mathcal{N}(0, \sigma_d^2)$. The BLER performance of the eavesdropper and the legitimate user with different values of σ_d are shown in Fig. 4.6, which reveals that as σ_d increases, the BLER performance gap between the legitimate user and the eavesdropper degrades. However, with a fairly large location error level $\sigma_d = 100$ m, only slight performance degradation is observed, which indicates the robustness of the proposed method to the location inaccuracy.

4.6.3 Secrecy Capacity and Secure DOF

For a narrowband system at frequency 13 kHz, Fig. 4.7 (a) shows the secrecy capacity of the benchmark method and the signal alignment method with the eavesdropper's location information, as well as the legitimate user link capacity in the benchmark method. In the signal alignment method, the active DAE set and the transmission power of each active DAE are obtained in (4.42a) through maximizing the lower



(a) Secrecy capacity in AWGN channels

(b) Lower bound of the secure d.o.f.

Figure 4.7: Secrecy capacity and secure degrees-of-freedom (d.o.f.) with different total number of DAEs.

bound of the secrecy capacity. As the transmission SNR increases, one can observe that (1) the secrecy capacity of the benchmark method converges, and is further less than the secrecy capacity lower bound of the signal alignment method; and (2) the secrecy capacity upper bound of the signal alignment method increases almost at the same rate as that of the user link capacity in the benchmark method.

Fig. 4.7 (b) shows the lower bound of the secure d.o.f. of the signal alignment method with different total numbers of randomly distributed DAEs. The result reveals that the average lower bound of the secure d.o.f. increases monotonically with the total number of DAEs, which agrees with intuition, as more DAEs promise more freedom to enlarge the received signal quality difference between the legitimate user and the eavesdropper.

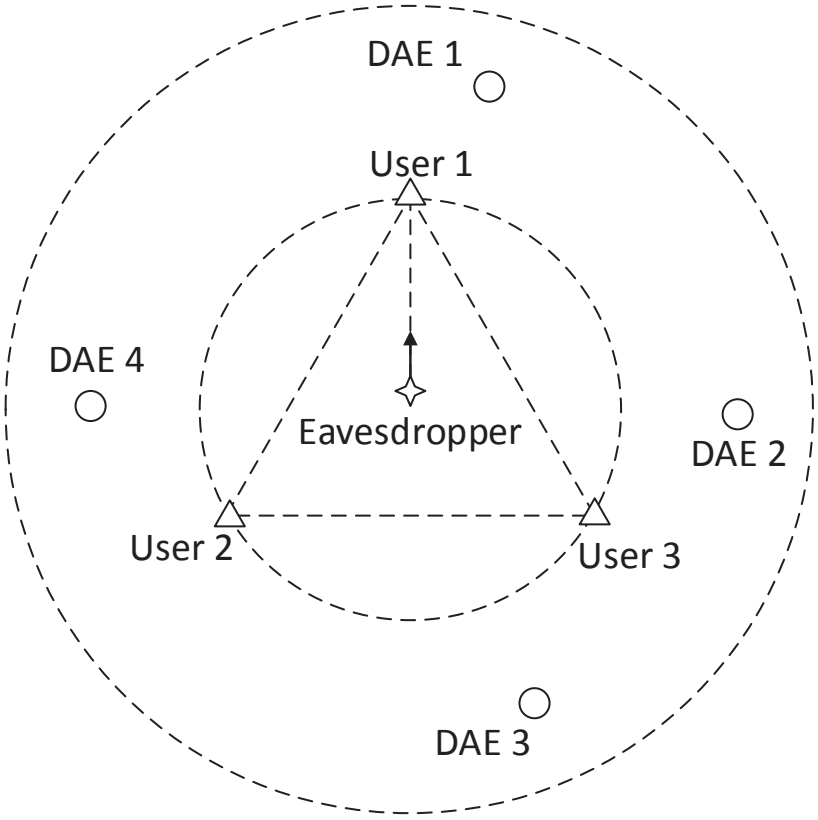
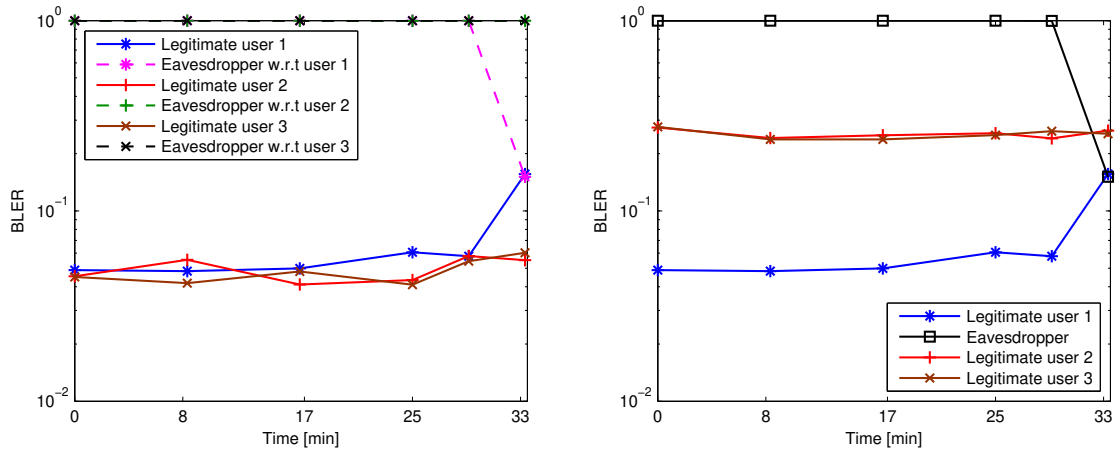


Figure 4.8: A case study: An example of the system layout in one Monte Carlo run.

4.6.4 A Case Study

Some insights about the proposed signal alignment method can be revealed in a case study. Specifically, we consider 4 DAEs uniformly distributed in a ring area defined by a circle of radius 4 km and the circumcircle (with a radius of 2 km) of a triangle formed by three legitimate users, as illustrated in Fig. 4.8. An eavesdropper moves from the the center of the triangle to user 1 at a speed of 1 m/s.

Two types of transmissions are considered: (i) the unicast transmission, where the



(a) Performance for unicast transmissions

(b) Performance for broadcast transmissions

Figure 4.9: Decoding BLERs of the legitimate user and the eavesdropper in the case study.

DAEs transmit to each legitimate user individually based on the proposed signal alignment method; and (ii) the broadcast transmission, where the proposed signal alignment method is applied only with respect to user 1 and the eavesdropper. The minimal received SNR at each individual user in the unicast transmission and at user 1 in the broadcast transmission is constrained to be $\Gamma = 7$ dB. 400 system layouts are tested. The performance of different transmission strategies is presented in Fig. 4.9.

The following observations can be obtained based on Fig. 4.9 (a) about the unicast transmission. Firstly, the proposed signal alignment method provides decent transmission secrecy for users 2 and 3, and none of the blocks can be decoded at the eavesdropper. Secondly, when the distance between the eavesdropper and user 1 is large, the secrecy performance of user 1 is similar to that of users 2 and 3. The performance gap between the eavesdropper and user 1 decreases as the eavesdropper

Table 4.2
OFDM Parameters in the MACE10 Experiment.

center frequency:	f_c	13 kHz
bandwidth:	B	4.883 kHz
# of subcarriers:	K	1024
time duration:	T	209.7 ms
frequency spacing:	$\Delta f := 1/T$	4.77 Hz
guard interval:	T_g	40.3 ms

moves closer to user 1. When the eavesdropper and user 1 are at the same location, the decoding performances of the two are identical.

About the broadcast transmission, Fig. 4.9 (b) shows that when the distance between the eavesdropper and user 1 is large, user 1 achieves better decoding performance than users 2 and 3, since signals from active DAEs could collide at the latter two users. In addition, users 2 and 3 exhibit similar decoding performance due to their symmetric location with respect to the eavesdropper.

4.7 Emulated Experiment Results

We use the data set collected from a field experiment to emulate OFDM transmissions from coordinated DAEs. The *mobile acoustic communication experiment* (MACE10) was carried out off the coast of Martha's Vineyard, Massachusetts, in June 2010. The water depth was about 95 to 100 meters. The receiver array was stationary, while the source was towed slowly away from the receiver from 500 m to 4.5 km and then

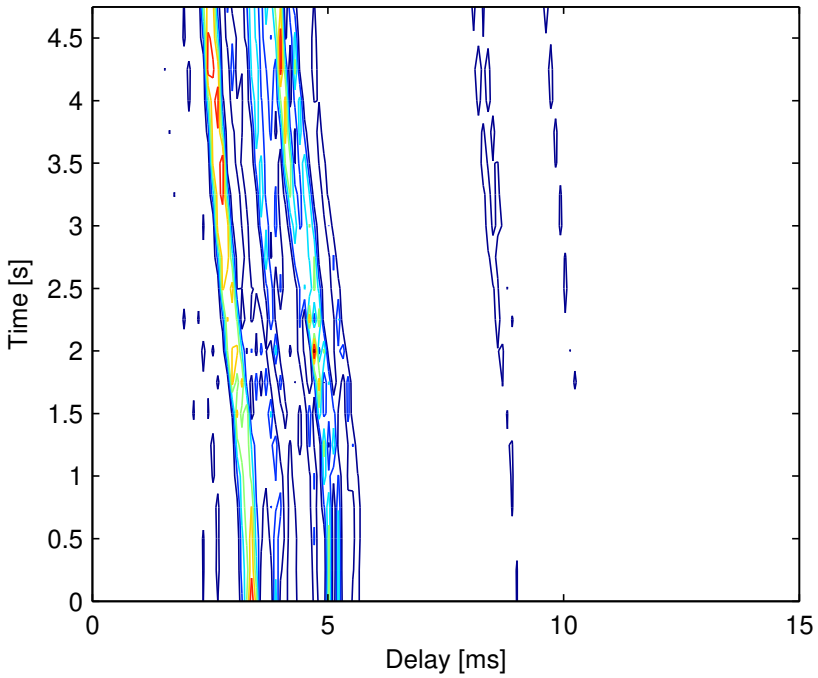


Figure 4.10: Evolution of the channel impulse response in the MACE10 experiment.

towed back, at a speed around 1 m/s. Out of two tows in the experiment, we only consider the data set collected in the first tow, where 30 recorded transmissions are used for emulation and each transmission has 20 OFDM blocks. The average received SNR of the recorded transmissions is around 20 dB.

The ZP-OFDM parameters are listed in Table 4.2. The subcarrier distribution is identical to that in Section 4.6. The data symbols are encoded with a rate-1/2 nonbinary LDPC code and modulated with a QPSK constellation. The data rate is

$$R = \frac{1}{2}\frac{|\mathcal{S}_{\mathrm{D}}|}{T + T_{\mathrm{g}}}\log_2 4 = 2.7 \text{ kb/s}. \tag{4.48}$$

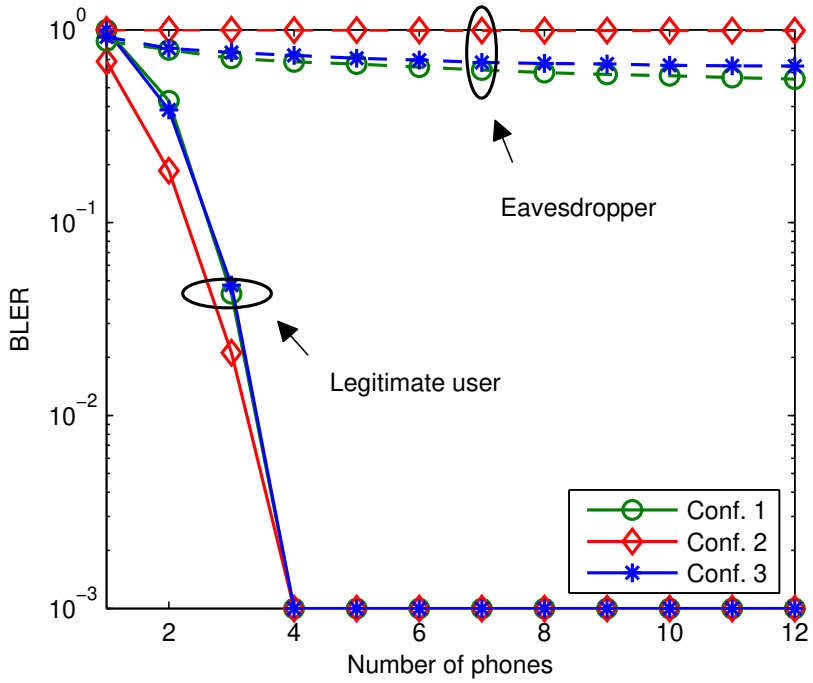


Figure 4.11: Decoding BLERs of the legitimate user and the eavesdropper with MACE10 data.

We recycle the simulated system layouts in Section 4.6, and assume that the maximal transmission power level of each DAE is 180 dB and the noise level is 60 dB. The received signals at the legitimate user and the eavesdropper are emulated based on the recorded waveforms in MACE10, where prior to the emulation, a resampling operation to remove the Doppler effect caused by the source mobility. Fig. 4.10 illustrates an example of the evolution of the underwater acoustic channel in MACE10. For each simulation layout, white Gaussian noise of appropriate variance levels are introduced to control the received SNR at the legitimate user and the eavesdropper.

With different numbers of receive hydrophones, Fig. 4.11 shows the BLER performance of the eavesdropper and the legitimate user in the three configurations defined

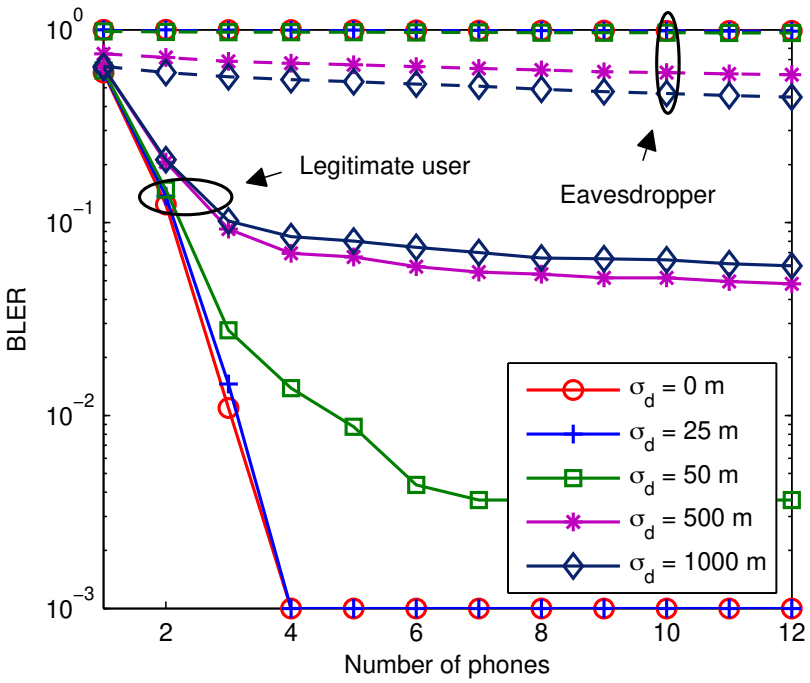


Figure 4.12: Decoding BLERs of the legitimate user and the eavesdropper with location inaccuracy.

in Section 4.6.1. Similar to our observations in simulation, the signal alignment method with the eavesdropper's location information achieves the highest level of transmission secrecy. The signal alignment method without the eavesdropper's location information exhibits better secrecy performance than the benchmark method.

For the signal alignment method with the eavesdropper's location information, the impact of location inaccuracy of both the legitimate user and the eavesdropper on the transmission secrecy is depicted in Fig. 4.12, where the location error of the legitimate user and the eavesdropper follows an independent zero-mean Gaussian distribution with a standard deviation of σ_d. Similar to the observation in Fig. 4.6, the BLER performance gap between the legitimate user and the eavesdropper degrades as the

location error increases. Nevertheless, the proposed method achieves decent secrecy performance with a location error less than 50 m, and localization techniques with an error less than 10 m have been demonstrated in real underwater acoustic systems [121].

4.8 Summary

In this chapter, we investigated signal alignment for secure underwater CoMP transmissions. Exploiting the low sound speed in water and the spatial diversity of DAEs, transmission secrecy was achieved by overlapping signals at the eavesdropper while keeping them free of collision at the legitimate user. Practical designs of the above signal alignment concept were pursued. The eavesdropper's interception capability was minimized through jointly optimizing relevant transmission parameters, including the transmit DAE set, and the transmission schedule and power level of each DAE, under a lower bound constraint of the received SNR at the legitimate user. Taking OFDM as the underlying modulation, both simulation and emulated experimental results showed that the proposed method has much higher data confidentiality than a benchmark method. From an information-theoretic perspective, we further derived the secrecy capacity and the secure d.o.f. of the signal alignment method, which revealed that a secure d.o.f. greater than $\frac{1}{2}$ can be achieved, and the lower bound of the secure d.o.f. increases as the number of DAEs increases.

Chapter 5

Conclusions

We have witnessed emerging underwater applications from ocean monitoring to recent underwater data centers. UWA communication networks are promising techniques for medium and long ranges wireless information transfer in underwater applications. However, with challenges induced by complex underwater environments and high dynamic UWA channels, we are still at the early stage to make UWA communication networks more reliable and efficient in demanding underwater applications.

This dissertation provides three research works towards intelligent and secure UWA communication networks exploiting machine learning and signal processing. The three research problems studied in the dissertation and their solutions are summarized as follows.

- RL-based adaptive transmission: A RL-based adaptive transmission framework was proposed for point-to-point UWA communication networks. The transmission strategies were determined adaptively to balance energy efficiency and information delivery latency according to UWA channel dynamics.

- RL-based adaptive trajectory planning: A RL-based adaptive trajectory planning of multiple AUVs for water parameter field reconstruction was developed in the under-ice environment. The optimal trajectories were learned online with a goal to minimize a long-term system cost for field uncertainty reduction, where the field knowledge was learned on the fly.

- Secure underwater CoMP: A signal alignment method was designed for underwater CoMP to secure a legitimate user against eavesdropping. The transmission schedule, the active DAE, and the transmission power were determined to degrade the decoding performance of the eavesdropper.

The proposed algorithms were validated based on simulation and/or experimental data sets. Compared to the benchmarks which had perfect system knowledge, the proposed algorithms achieved decent system performance.

The machine learning techniques offer a new perspective to rethink about research problems in UWA communication networks. In this dissertation, the RL framework

was exploited to study the sequential decision-making problem for underwater applications which operate in the long term. It showed that applying machine learning techniques to UWA communication networks significantly improved their system performance. About UWA network security, due to large sound prorogation delay in water, conventional security mechanisms designed for terrestrial radio networks cannot be applied directly to UWA networks, and many algorithms were proposed to overcome that delay challenge. However, in this dissertation, we took the advantage of the large sound prorogation delay to secure UWA communication networks. The solutions in this dissertation provided innovative perspectives for research challenges and opportunities in UWA communication networks.

Appendix A

Detailed Derivation and Extension for Chapter 2

A.1 Reformulation of Optimization Problem (4.21)

The optimization problem in (4.21) can be recast as

$$\min_{\varrho, \mathcal{T}_{\mathrm{DAE}}, \{P_{\mathrm{tx},\mu}\}_{\mu \in \mathcal{T}_{\mathrm{DAE}}}} \quad \varrho \tag{A.1a}$$

$$\text{s.t.} \qquad \lambda_{\mu}^{(\mathrm{e})} \leq \varrho, \tag{A.1b}$$

$$\xi_{\ell_{\mu+1}} - \xi_{\ell_{\mu}} \geq T_{\mathrm{bl}}, \tag{A.1c}$$

$$P_{\mathrm{tx},\mu,\mathrm{L}} \leq P_{\mathrm{tx},\mu} \leq P_{\mathrm{th}}, \ \forall \mu \in \mathcal{T}_{\mathrm{DAE}}. \tag{A.1d}$$

Based on the definition of $\theta_{\nu,\mu}$ in (4.23), the constraint in (A.1b) becomes

$$\lambda_{\kappa_\mu}^{(\mathrm{e})} = \frac{P_{\mathrm{rx},\kappa_\mu}^{(\mathrm{e})}}{N_0 B + \sum_{\nu \neq \mu} P_{\mathrm{rx},\kappa_\nu}^{(\mathrm{e})} \theta_{\nu,\mu}} \leq \varrho. \tag{A.2}$$

Substituting (4.10) into (A.2) yields

$$\frac{P_{\mathrm{tx},\kappa_\mu} \cdot P_{\mathrm{loss}}(f_{\mathrm{c}}, d_{\kappa_\mu}^{(\mathrm{e})})}{N_0 B + \sum_{\nu \in \mathcal{T}_{\mathrm{DAE}}, \nu \neq \mu} P_{\mathrm{tx},\kappa_\nu} \cdot P_{\mathrm{loss}}(f_{\mathrm{c}}, d_{\kappa_\nu}^{(\mathrm{e})}) \theta_{\nu,\mu}} \leq \varrho. \tag{A.3}$$

With the definition of vector $\mathbf{q}$ in (4.24), the constraint in (A.1b) can be rewritten as

$$\underbrace{\begin{bmatrix} 1 & -\varrho\theta_{1,2} & \cdots & -\varrho\theta_{1,M} \\ -\varrho\theta_{2,1} & 1 & \cdots & -\varrho\theta_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ -\varrho\theta_{M,1} & -\varrho\theta_{M,2} & \cdots & 1 \end{bmatrix}}_{=\tilde{\mathbf{G}}_1(\boldsymbol{\xi},\varrho)} \mathbf{q} \leq \varrho \mathbf{1}, \tag{A.4}$$

where $\tilde{\mathbf{G}}_1(\boldsymbol{\xi}, \varrho)$ is a generic matrix of size $M \times M$. The constraint (A.1c) can be

rewritten as

$$\underbrace{\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}}_{=\mathbf{G}_2} \boldsymbol{\xi} \geq T_{\mathrm{bl}}\mathbf{1}, \tag{A.5}$$

where the size of $\mathbf{G}_2$ is $(M-1)\times M$. Define $\mathbf{q}_{\mathrm{L}}$ and $\mathbf{q}_{\mathrm{U}}$ through replacing $P_{\mathrm{tx},\mu}$ by $P_{\mathrm{tx},\mu,\mathrm{L}}$ and $P_{\mathrm{tx},\mu,\mathrm{th}}$ in (4.24) , respectively. The optimization problem in (A.1) can be reformulated into the matrix form in (4.25).

A.2 Extension to General Scenarios

In the scenario with multiple users, the constraints (4.21b) and (4.21c) can be modified to ensure that the blocks arrived at each user are well separated and with sufficiently large received SNRs. Denote $\mathcal{T}_{\mathrm{user}}$ as the user set with N_{user} users. The optimization problem can be formulated as

$$\min_{\mathcal{T}_{\mathrm{DAE}},\{t_\mu,P_{\mathrm{tx},\mu}:\ \mu\in\mathcal{T}_{\mathrm{DAE}}\}} \ \max_{\mu\in\mathcal{T}_{\mathrm{DAE}}} \ \lambda_\mu^{(\mathrm{e})} \tag{A.6a}$$

$$\text{s.t.} \qquad \lambda_{\mu\nu} \geq \Gamma_{\mathrm{th}}, \tag{A.6b}$$

$$\xi_{\ell_{\mu+1},\nu} - \xi_{\ell_\mu,\nu} \geq T_{\mathrm{bl}},\ \forall \nu \in \mathcal{T}_{\mathrm{user}} \tag{A.6c}$$

$$0 \leq P_{\mathrm{tx},\mu} \leq P_{\mathrm{th}},\ \forall \mu \in \mathcal{T}_{\mathrm{DAE}} \tag{A.6d}$$

where $\lambda_{\mu\nu}$ and $\xi_{\mu,\nu}$ is the received SNR and the time-of-arrival of the signal from the μth DAE at the νth legitimate user, respectively. The optimization problem (A.6) can be solved by the proposed Algorithm 1 and Algorithm 2 for synchronous and asynchronous signal alignment, respectively.

Furthermore, from the legitimate user's perspective, the scenario with multiple eavesdroppers is equivalent to the scenario where an eavesdropper has multiple geographically distributed antennas. Modification of the objective function (4.21a) to the scenario where an eavesdropper has multiple distributed antennas depends on the diversity combining technique used by the eavesdropper. For instance, when the maximal ratio combining (MRC) is used by an eavesdropper equipped with N_{ant} distributed antennas, the optimization problem can be recast as

$$\min_{\mathcal{T}_{\text{DAE}},\{t_\mu,P_{\text{tx},\mu}:\ \mu\in\mathcal{T}_{\text{DAE}}\}} \ \max_{\mu\in\mathcal{T}_{\text{DAE}}} \sum_{\nu=1}^{N_{\text{ant}}} \lambda_{\mu\nu}^{(\text{e})} \tag{A.7a}$$

$$\text{s.t.} \qquad \lambda_\mu \geq \Gamma_{\text{th}}, \tag{A.7b}$$

$$\xi_{\ell_{\mu+1}} - \xi_{\ell_\mu} \geq T_{\text{bl}}, \tag{A.7c}$$

$$0 \leq P_{\text{tx},\mu} \leq P_{\text{th}},\ \forall \mu \in \mathcal{T}_{\text{DAE}} \tag{A.7d}$$

where $\lambda_{\mu\nu}^{(\text{e})}$ is the received SINR of the signal from the μth DAE at the νth antenna

of the eavesdropper. By introducing an auxiliary variable

$$\varrho := \max_{\mu \in \mathcal{T}_{\mathrm{DAE}}} \sum_{\nu=1}^{N_{\mathrm{ant}}} \lambda_{\mu\nu}^{(\mathrm{e})} \tag{A.8}$$

and a new constraint $\sum_{\nu=1}^{N_{\mathrm{ant}}} \lambda_{\mu\nu}^{(\mathrm{e})} \leq \varrho$, the optimization problem (A.7) can be solved based on Algorithm 1 and Algorithm 2 for synchronous and asynchronous signal alignment, respectively.

It can be easily shown that the scenario with multiple legitimate users and the scenario where an eavesdropper has multiple distributed antenna elements can be jointly considered through combining the optimization problems (A.6) and (A.7).

A.3 Derivation of Secure Degrees of Freedom

Assume the transmission powers of active DAEs going to infinity in the same order. Accordingly, the received signal power at the legitimate user and the eavesdropper can be rewritten as $P_{\mathrm{rx},\mu} = \alpha_\mu P$ and $P_{\mathrm{rx},\mu}^{(\mathrm{e})} = \alpha_\mu^{(\mathrm{e})} P$, respectively, where α_μ and $\alpha_\mu^{(\mathrm{e})}$ are constant. For upper bound of secrecy d.o.f., substituting (4.40) into (4.43) yields

$$\eta_{\mathrm{U}} = \lim_{P \to +\infty} \left[\sum_{\mu \in \mathcal{T}_{\mathrm{DAE}}} \left\{ \frac{T_\mu}{T_{\mathrm{total}}} \frac{\log_2\left(1 + \frac{\alpha_\mu P}{\sigma_{\mathrm{w}}^2}\right)}{\log_2\left(1 + \frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} - \frac{T_\mu}{T_{\mathrm{total}}} \frac{\log_2\left(1 + \frac{\alpha_\mu^{(\mathrm{e})} P}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{I},\mu}^{(\mathrm{e})}}\right)}{\log_2\left(1 + \frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} \right\} \right]^+ . \tag{A.9}$$

Due to the inter-symbol-interference the denominator of the second term is bounded as $P \to +\infty$, hence leading the second term to be zero. Applying the L'Hospital's Rule to the first term yields

$$\begin{aligned}\eta_{\mathrm{U}} &= \lim_{P\to+\infty} \sum_{\mu\in\mathcal{T}_{\mathrm{DAE}}} \left\{ \frac{T_\mu}{T_{\mathrm{total}}} \frac{\frac{1}{\ln 2}\frac{1}{1+\alpha_\mu P/\sigma_{\mathrm{w}}^2}\frac{\alpha_\mu}{\sigma_{\mathrm{w}}^2}}{\frac{1}{\ln 2}\frac{1}{1+\alpha_{\mu'} P/\sigma_{\mathrm{w}}^2}\frac{\alpha_{\mu'}}{\sigma_{\mathrm{w}}^2}} \right\} \\ &= \sum_{\mu\in\mathcal{T}_{\mathrm{DAE}}} \left\{ \frac{T_\mu}{T_{\mathrm{total}}} \right\} = 1. \end{aligned} \tag{A.10}$$

For lower bound of secrecy d.o.f., substituting (4.41) into (4.43) yields

$$\begin{aligned}\eta_{\mathrm{L}} = \lim_{P\to+\infty} \Bigg[&\sum_{\mu\in\mathcal{T}_{\mathrm{DAE}}} \frac{T_\mu}{T_{\mathrm{total}}} \frac{\log_2\left(1+\frac{\alpha_\mu P}{\sigma_{\mathrm{w}}^2}\right)}{\log_2\left(1+\frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} \\ &- \left\{ \sum_{\mu=1}^{M-1} \frac{(T_{o_\mu} - T_{o_{\mu-1}})}{T_{\mathrm{total}}} \frac{\log_2\left(1+\frac{P\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2}\right)}{\log_2\left(1+\frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} \right\} \Bigg]^+. \end{aligned} \tag{A.11}$$

For the first term, from (A.10), it goes to 1 as $P \to +\infty$. For the second term, it comes to

$$\sum_{\mu=1}^{M-1} \left\{ \frac{(T_{o_\mu} - T_{o_{\mu-1}})}{T_{\mathrm{total}}} \lim_{P\to+\infty} \frac{\log_2\left(1+\frac{P\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2}\right)}{\log_2\left(1+\frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} \right\}. \tag{A.12}$$

By using L'Hospital's Rule, it yields

$$\lim_{P\to+\infty} \frac{\log_2\left(1+\frac{P\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2}\right)}{\log_2\left(1+\frac{\alpha_{\mu'} P}{\sigma_{\mathrm{w}}^2}\right)} = \lim_{P\to+\infty} \frac{\frac{1}{\ln 2}\frac{1}{1+P\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\mathrm{e})}/\sigma_{\mathrm{w}}^2}\frac{\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\mathrm{e})}}{\sigma_{\mathrm{w}}^2}}{\frac{1}{\ln 2}\frac{1}{1+\alpha_{\mu'} P/\sigma_{\mathrm{w}}^2}\frac{\alpha_{\mu'}}{\sigma_{\mathrm{w}}^2}} = 1. \tag{A.13}$$

Thus, the second term is

$$\sum_{\mu=1}^{M-1}\left\{\frac{(T_{o_\mu}-T_{o_{\mu-1}})}{T_{\text{total}}}\lim_{P\to+\infty}\frac{\log_2\left(1+\frac{P\sum_{\nu\geq\mu}\alpha_{o_\nu}^{(\text{e})}}{\sigma_{\text{w}}^2}\right)}{\log_2\left(1+\frac{\alpha_{\mu'}P}{\sigma_{\text{w}}^2}\right)}\right\}=\sum_{\mu=1}^{M-1}\frac{(T_{o_\mu}-T_{o_{\mu-1}})}{T_{\text{total}}}$$
$$=\frac{T_{o_{\text{M}-1}}}{T_{\text{total}}}. \tag{A.14}$$

Consider that $T_{o_{M-1}}=T_{\max}$. The secrecy d.o.f. is therefore lower bounded by

$$\eta_{\text{L}}=1-\frac{T_{\max}}{T_{\text{total}}}. \tag{A.15}$$

Appendix B

Permission Letters

B.1 Permission Letter for Chapter 2

Title: Reinforcement Learning-Based Adaptive Transmission in Time-Varying Underwater Acoustic Channels
Author: Chaofeng Wang
Publication: IEEE Access
Publisher: IEEE
Date: 2018

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

B.2 Permission Letter for Chapter 4

Title: Signal Alignment for Secure Underwater Coordinated Multipoint Transmissions
Author: Chaofeng Wang
Publication: Signal Processing, IEEE Transactions on
Publisher: IEEE
Date: 1 Dec.1, 2016

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

References

[1] C. Wang, Z.-H. 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, 2017.

[2] C. Wang, L. Wei, Z. Wang, M. Song, and N. Mahmoudian, "Reinforcement learning-based adaptive trajectory planning for AUVs in under-ice environments," in *Proc. of MTS/IEEE OCEANS Conf.*, Charleston SC, Oct. 2018.

[3] C. Wang and Z.-H. Wang, "Signal alignment for secure underwater coordinated multipoint transmissions," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6360–6374, Dec. 2016.

[4] W. Duan and Y. R. Zheng, "Soft direct-adaptive turbo equalization for MIMO underwater acoustic communications," in *Proc. of MTS/IEEE OCEANS Conf.*, Washington DC, Oct. 2015.

[5] M. Chitre, M. Motani, and S. Shahabudeen, "Throughput of networks with

large propagation delays," *IEEE J. Ocean. Eng.*, vol. 37, no. 4, pp. 645 –658, 2012.

[6] Z.-H. Wang, X. Xu, and C. Wang, "Multi-flow scheduling in multi-hop underwater acoustic networks," in *Proc. of MTS/IEEE OCEANS Conf.*, Shanghai China, Apr. 2016.

[7] W. Duan and Y. R. Zheng, "Experimental evaluation of turbo receivers in single-input single-output underwater acoustic channels," in *Proc. of MTS/IEEE OCEANS Conf.*, Shanghai China, Apr. 2016.

[8] H. Zeng, Y. T. Hou, Y. Shi, W. Lou, S. Kompella, and S. F. Midkiff, "Shark-IA: An interference alignment algorithm for multi-hop underwater acoustic networks with large propagation delays," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Rome Italy, 2014.

[9] Z.-H. Wang, C. Wang, and W. Sun, "Adaptive transmission scheduling in time-varying underwater acoustic channels," in *Proc. of MTS/IEEE OCEANS Conf.*, Washington DC, Apr. 2015.

[10] W. Duan, J. Tao, and Y. R. Zheng, "Efficient adaptive turbo equalization for multiple-input-multiple-output underwater acoustic communications," *IEEE J. Ocean. Eng.*, 2017.

[11] C. Wang, Z.-H. Wang, and S. Nooshabadi, "Signal alignment for secure underwater coordinated multipoint transmissions," in *Proc. of IEEE Conference*

on Communications and Network Security (CNS): Wokshop on Physical-layer Methods for Wireless Security (PhySec'14), San Francisco CA, Oct. 2014.

[12] "Ocean Networks Canada," http://www.oceannetworks.ca/.

[13] "The Ocean Observatory Initiative (OOI)," http://oceanobservatories.org/.

[14] D. V. Djonin and V. Krishnamurthy, "MIMO transmission control in fading channels - A constrained Markov decision process formulation with monotone randomized policies," *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 5069–5083, Oct. 2007.

[15] M. H. Ngo and V. Krishnamurthy, "Optimality of threshold policies for transmission scheduling in correlated fading channels," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2474–2483, Aug. 2009.

[16] R. Srivastava and C. E. Koksal, "Energy optimal transmission scheduling in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1550–1560, May 2010.

[17] N. Ding, P. Sadeghi, and R. A. Kennedy, "On monotonicity of the optimal transmission policy in cross-layer adaptive m-QAM modulation," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3771–3785, Sep. 2016.

[18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Cambridge:Cambridge Univ. Press, 2011.

[19] C. Pandana and K. Liu, "Near-optimal reinforcement learning framework for energy-aware sensor communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 788–797, Apr. 2005.

[20] D. V. Djonin and V. Krishnamurthy, "Q-learning algorithms for constrained Markov decision processes with randomized monotone policies: Application to MIMO transmission control," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 2170–2181, May 2007.

[21] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE J. Select. Areas Commun.*, vol. 26, no. 4, pp. 732–742, May 2008.

[22] M. H. Ngo and V. Krishnamurthy, "Monotonicity of constrained optimal transmission policies in correlated fading channels with ARQ," *IEEE Trans. Signal Processing*, vol. 58, no. 1, pp. 438–451, Jan. 2010.

[23] N. Mastronarde and M. van der Schaar, "Joint physical-layer and system-level power management for delay-sensitive wireless communications," *IEEE Trans. Mobile Computing*, vol. 12, no. 4, pp. 694–709, Apr. 2013.

[24] L. Lei, Y. Kuang, X. S. Shen, K. Yang, J. Qiao, and Z. Zhong, "Optimal reliability in energy harvesting industrial wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5399–5413, Aug. 2016.

[25] I. Ahmed, K. T. Phan, and T. Le-Ngoc, "Optimal stochastic power control for energy harvesting systems with delay constraints," *IEEE J. Select. Areas Commun.*, vol. 34, no. 12, pp. 3512–3527, Dec. 2016.

[26] D. T. Hoang, D. Niyato, P. Wang, D. I. Kim, and L. B. Le, "Optimal data scheduling and admission control for backscatter sensor networks," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2062–2077, May 2017.

[27] M. Hirzallah, W. Afifi, and M. Krunz, "Full-duplex-based rate/mode adaptation strategies for Wi-Fi/LTE-U coexistence: A POMDP approach," *IEEE J. Select. Areas Commun.*, vol. 35, no. 1, pp. 20–29, Jan. 2017.

[28] 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 Journal of Oceanic Engineering*, vol. 39, no. 2, pp. 357–370, Apr. 2014.

[29] 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.

[30] 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 J. Ocean. Eng.*, vol. 40, no. 2, pp. 327–336, Apr. 2015.

[31] E. Demirors, J. Shi, R. Guida, and T. Melodia, "SEANet G2: Toward a high-data-rate software-defined underwater acoustic networking platform," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Oct. 2016.

[32] B. Tomasi and J. C. Preisig, "Energy-efficient transmission strategies for delay constrained traffic with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 14, pp. 1369–1379, Mar. 2015.

[33] ——, "Efficient heuristic scheduling with partial queue and channel state information," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Rome, Italy, Sep. 2014.

[34] L. Jin and D. D. Huang, "A slotted CSMA based reinforcement learning approach for extending the lifetime of underwater acoustic wireless sensor networks," *Computer Communications*, vol. 36, no. 9, pp. 1094–1099, 2013.

[35] V. D. Valerio, C. Petrioli, L. Pescosolido, and M. V. D. Shaar, "A reinforcement learning-based data-link protocol for underwater acoustic communications," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Washington DC, Oct. 2015.

[36] 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 Computing*, vol. 9, no. 6, pp. 796–809, Jun. 2010.

[37] N. M. Carbone and W. S. Hodgkiss, "Effects of tidally driven temperature fluctuations on shallow-water acoustic communications at 18 kHz," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 84–94, Jan. 2000.

[38] M. Badiey, Y. Mu, J. A. Simmen, and S. E. Forsythe, "Signal variability in shallow-water sound channels," *IEEE J. Ocean. Eng.*, vol. 25, no. 4, pp. 492–500, Oct. 2000.

[39] A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, and M. B. Porter, "Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx)," *J. Acoust. Soc. Am.*, vol. 123, no. 2, pp. 856–865, Feb. 2008.

[40] P. van Walree, "Propagation effects in underwater acoustic communication channels," in *Proc. of the Workshop on Underwater Communications: Channel Modelling & Validation*, Italy, Sep. 2012.

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

[42] J. Llor and M. P. Malumbres, "Statistical modeling of large-scale signal path loss in underwater acoustic networks," *Sensors*, vol. 13, no. 2, pp. 2279–2294, 2013.

[43] B. Tomasi, P. Casari, L. Badia, and M. Zorzi, "A study of incremental redundancy hybrid ARQ over Markov channel models derived from experimental data," in *Proc. of the ACM Intl. Workshop on Underwater Networks (WUWNet)*, Woods Hole, MA, Sep. 2010.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* New Jersey: Prentice Hall, 1993, vol. 2.

[45] S. Atapattu, C. Tellambura, and H. Jiang, "A mixture Gamma distribution to model the SNR of wireless channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4193–4203, Dec. 2011.

[46] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 348–361, Feb. 2004.

[47] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[48] M. Ghavamzadeh, S. Mannor, J. Pineau, and A. Tamar, "Bayesian reinforcement learning: A survey," *Found. Trends Mach. Learn.*, vol. 8, no. 5-6, pp. 359–483, Nov. 2015.

[49] M. Kearns, Y. Mansour, and A. Ng, "A sparse sampling algorithm for near-optimal planning in large Markov decision processes," *Machine Learning*, vol. 49, no. 2, pp. 193–208, Nov. 2002.

[50] A. Guez, D. Silver, and P. Dayan, "Scalable and efficient Bayes-adaptive reinforcement learning based on Monte-Carlo tree search," *Journal of Artificial Intelligence Research*, vol. 48, pp. 841–883, Nov. 2013.

[51] C. M. Bishop, *Pattern Recognition and Machine Learning*, 6th ed. Springer-Verlag New York, 2006.

[52] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[53] U. Mitra, S. Choudhary, F. Hover, R. Hummel, N. Kumar, S. Naryanan, M. Stojanovic, and G. Sukhatme, "Structured sparse methods for active ocean observation systems with communication constraints," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 88–96, Nov. 2015.

[54] J. Sun, A. Song, J. Yu, A. Zhang, and F. Zhang, "Underwater acoustic field reconstruction using a compressive sensing approach," in *Proc. of MTS/IEEE OCEANS Conf.*, Anchorage AK, Sep. 2017.

[55] W. Sun, C. Wang, Z.-H. Wang, and M. Song, "Estimation of the under-ice acoustic field in AUV communication networks," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Halifax Canada, 2017.

[56] A. Kukulya, A. Plueddemann, T. Austin, R. Stokey, M. Purcell, B. Allen, R. Littlefield, L. Freitag, P. Koski, E. Gallimore, J. Kemp, K. Newhall, and

J. Pietro, "Under-ice operations with a REMUS-100 AUV in the Arctic," in *IEEE/OES Autonomous Underwater Vehicles*, Monterey CA, 2010.

[57] P. Norgre and R. Skjetne, "Using autonomous underwater vehicles as sensor platforms for ice-monitoring," *Modeling, Identification and Control*, vol. 35, no. 4, pp. 269–277, Nov. 2014.

[58] W. Sun, C. Wang, Z.-H. Wang, and M. Song, "Experimental comparison between under-ice and open-water acoustic channels," in *Proc. of the ACM Intl. Conf. on Underwater Networks (WUWNet)*, Washington DC, Oct. 2015.

[59] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research chanllenges," *Ad Hoc Networks*, vol. 3, pp. 257–279, 2005.

[60] R. Marchant and F. Ramos, "Bayesian optimisation for informative continous path planning," in *Proc. Proc. of Int'l. Conf. on Robotics and Automation*, Hong Kong, China, Jun. 2014.

[61] Y. Xu, J. Choi, and S. Oh, "Mobile sensor network navigation using gaussian processes with truncated observations," *IEEE Trans. on Robotics*, vol. 27, no. 6, pp. 1118–1131, Dec. 2011.

[62] L. Nguyen, S. Kodagoda, R. Ranasinghe, and G. Dissanayake, "Information-driven adaptive sampling strategy for mobile robotic wireless sensor network," *IEEE Trans. on Control Systems Technology*, vol. 24, no. 1, pp. 372–379, Jan. 2016.

[63] R. Martinez-Cantin, N. Freitas, E. Brochu, J. Castellanos, and A. Doucet, "A Bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot," *Autonomous Robots*, vol. 27, no. 2, pp. 93–103, 2009.

[64] A. Singh, A. Krause, and W. Kaiser, "Nonmyopic adaptive informative path planning for multiple robots," in *Proc. of IJCAI*, Pasadena, CA, Jul. 2009.

[65] R. Marchant, F. Ramos, and S. Sanner, "Sequential Bayesian optimization for spatial-temporal monitoring," in *Proc. of UAI*, Singapore, Jul. 2014.

[66] P. Morere, R. Marchant, and F. Ramos, "Sequential Bayesian optimization as a POMDP for environment monitoring with UAVs," in *Proc. Proc. of Int'l. Conf. on Robotics and Automation*, Singapore, Jun. 2017.

[67] G. Hollinger, B. Englot, F. Hover, U. Mitra, and G. Sukhatme, "Undertainty-driven view planning for underwater inspection," in *Proc. Proc. of Int'l. Conf. on Robotics and Automation*, Saint Pau,l MN, May 2012.

[68] J. Binney, A. Krause, and G. S. Sukhatme, "Informative path planning for an autonomous underwater vehicle," in *Proc. Proc. of Int'l. Conf. on Robotics and Automation*, Anchorage AK, May 2010.

[69] N. Yilmaz, C. Evangelinos, P. Lermusiaux, and N. Patrikalakis, "Path planning of autonomous underwater vehicles for adaptive sampling using mixed integer

linear programming," *IEEE J. Ocean. Eng.*, vol. 33, no. 4, pp. 522 – 537, Oct. 2008.

[70] A. Marino, G. Antonelli, A. Aguiar, A. Pascoal, and S. Chiaverini, "A decentralized strategy for multirobot sampling/patrolling: Theory and experiments," *IEEE Trans. on Control Systems Technology*, vol. 23, no. 1, pp. 313–322, Jan. 2015.

[71] C. E. Rasmussen, "Gaussian processes in machine learning," *Advanced lectures on machine learning*, pp. 63–71, 2004.

[72] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in neural information processing systems*, 1996, pp. 514–520.

[73] R. Bellman, "A Markovian decision process," *Journal of Mathematics and Mechanics*, pp. 679–684, 1957.

[74] S. P. Mertikas, "Error distributions and accuracy measures in navigation: An overview," *Tech. Rep.*, 1985.

[75] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[76] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and

D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. of the International Conference on Learning Representations (ICLR)*, 2016.

[77] D. P. Kroese and Z. I. Botev, *Spatial Process Simulation.* Cham: Springer International Publishing, 2015, pp. 369–404. [Online]. Available: https://doi.org/10.1007/978-3-319-10064-7_12

[78] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 108–113, 2003.

[79] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 161–167, 2013.

[80] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-Advanced," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 26–34, June 2010.

[81] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 44–50, Nov. 2012.

[82] C. L. I, J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li, "Recent progress on C-RAN centralization and cloudification," *IEEE Acess*, vol. 2, pp. 1030–1039, Sept. 2014.

[83] A. Saleh, A. Rustako, and R. Roman, "Distributed antennas for indoor radio communications," *IEEE Trans. Commun.*, vol. 35, no. 12, pp. 1245–1251, Dec. 1987.

[84] R. Heath, T. Wu, Y. H. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4885–4899, Oct. 2011.

[85] S.-R. Lee, S.-H. Moon, J.-S. Kim, and I. Lee, "Capacity analysis of distributed antenna systems in a composite fading channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1076–1086, 2012.

[86] R. Hayford, D. Nagle, and J. Catipovic, "AUTEC undersea cellular network," in *Proc. of the ACM Intl. Workshop on Underwater Networks (WUWNet)*, Berkeley, California, 2009.

[87] L. Wei, P. Zheng, H. Zhou, J.-H. Cui, S. Zhou, Z. Shi, and J. O'Donnell, "Long Island Sound testbed and experiments," in *Proc. of MTS/IEEE OCEANS Conf.*, San Diego, CA, 2013.

[88] W. Stallings, *Cryptography and Network Security: Principles and Practices*, 6th ed. Prentice Hall, 2006.

[89] M. Chitre, S. Shahabudeen, and M. Stojanovic, "Underwater acoustic communications and networking: Recent advances and future challenges," *Marine Technology Society Journal*, vol. 42, no. 1, pp. 103–116, 2008.

[90] S. Al-Dharrab, M. Uysal, and T. M. Duman, "Cooperative underwater acoustic communications," *IEEE Communications Magazinel*, vol. 51, no. 7, pp. 146–153, July 2013.

[91] S. Zhou and Z.-H. Wang, *OFDM for Underwater Acoustic Communications*. Wiley, 2014.

[92] A. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[93] R. Bassily, E. Ekrem, X. He, E. Tekin, J. Xie, M. Bloch, S. Ulukus, and A. Yener, "Cooperative security at the physical layer: A summary of recent advances," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 16–28, 2013.

[94] Y. Hong, P.-C. Lan, and C. Kuo, *Signal Processing Approaches to Secure Physical Layer Communications in Multi-Antenna Wireless Systems*. Springer, 2013.

[95] G. Geraci, H. S. Dhillon, J. G. Andrews, J. Yuan, and I. B. Collings, "Physical layer security in downlink multi-antenna cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2006–2021, 2014.

[96] M. Zhang, R. Xue, H. Yu, H. Luo, and W. Chen, "Secrecy capacity optimization in coordinated multi-point processing," in *Proc. of Intl. Conf. on Commun.*, Budapest, 2013.

[97] H. Jin and J. Kin, "Combined antenna selection and beamforming for secure transmission in distributed antenna systems," in *Proc. of Intl. Conf. on ICT conergence*, Jeju Island, South Korea, 2013.

[98] G. Zheng, I. Kriskidis, J. Li, A. P. Petropulu, and B. Ottersten, "Improving physical layer secrecy using full-duplex jamming receivers," *IEEE Trans. Signal Processing*, vol. 61, no. 20, pp. 4962–4974, 2013.

[99] F. Zhu, F. Gao, M. Yao, and H. Zou, "Joint information- and jamming-beamforming for physical layer security with full duplex base station," *IEEE Trans. Signal Processing*, vol. 62, no. 24, pp. 6391–6401, 2014.

[100] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channels," *IEEE Trans. Antennas and Propagation*, vol. 53, no. 11, pp. 3776–3784, Nov. 2005.

[101] L. Lai, Y. Liang, , and H. V. Poor, "A unified framework for key agreement over wireless fading channels," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 2, pp. 480–490, Apr. 2012.

[102] J. Xie and S. Ulukus, "Secure degrees of freedom of the Gaussian wiretap channel with helpers," in *50th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, 2012.

[103] ——, "Secure degrees of freedom of the Gaussian wiretap channel with helpers and no eavesdropper CSI: Blind cooperative jamming," in *Proc. of Conf. Information Sciences Systems*, 2013.

[104] O. Koyluoglu, H. El Gamal, L. Lai, and H. V. Poor, "Interference alignment for secrecy," *IEEE Trans. Inform. Theory*, vol. 57, no. 6, pp. 3323 –3332, Jun. 2011.

[105] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys and Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

[106] G. Zheng, I. Krikidis, C. Masouros, S. Timotheou, D. Toumpakaris, and Z. Ding, "Rethinking the role of interference in wireless networks," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 152–158, 2014.

[107] M. Domingo, "Securing underwater wireless communication networks," *IEEE Wireless Communications*, vol. 18, no. 1, pp. 22 –28, Feb. 2011.

[108] G. Han, J. Jiang, N. Sun, and L. Shu, "Secure comminication for underwater acoustic sensor networks," *IEEE Communications Magazine*, vol. 61, no. 16, pp. 231–252, Aug. 2015.

[109] M. Zuba, Z. Shi, Z. Peng, J. Cui, and S. Zhou, "Vulnerabilities of underwater acoustic networks to denial-of-service jamming attacks," *Wiley Security and Communication Networks*, pp. 1–10, 2012.

[110] M. Zuba, M. Fagan, J. Cui, and Z. Shi, "A vulnerability study of geographic routing in underwater acoustic networks," in *Proc. of IEEE Conference on Communications and Network Security*, National Harbor, MD, 2013.

[111] X. Lu, M. Zuba, J. Cui, and Z. Shi, "Uncooperative localization improves attach performance in underwater acoustic networks," in *Proc. of IEEE Conference on Communications and Network Security*, San Francisco, CA, 2014.

[112] H. Kulhandjian, T. Melodia, and D. Koutisonikolas, "Securing underwater acoustic communications through analog network coding," in *Proc. of IEEE Intl. Conf. on Sensing, Communication, and Netorking*, Singapore, 2014.

[113] L. Xiao, Q. Li, T. Chen, E. Cheng, and H. Dai, "Jamming games in underwater sensor networks with reinforcement learning," in *Proc. of the IEEE Global Communications Conference (GLOBECOM)*, San Diego CA, 2015.

[114] Y. Luo, L. Pu, Z. Peng, and Z. Shi, "Rss-based secret key generation in underwater acoustic networks: Advantages, challenges, and performance improvements," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 32–38, 2016.

[115] Y. Huang, S. Zhou, Z. Shi, and L. Lai, "Experimental study of secret key generation in underwater acoustic channels," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2014.

[116] J. Chen, R. Zhang, L. Song, Z. Han, and B. Jiao, "Joint relay and jammer selection for secure two-way relay networks," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 1, pp. 310–320, 2012.

[117] M. Dehghan, D. L. Goechel, M. Ghaderi, and Z. Ding, "Energy efficient of cooperative jamming strategies in secure wireless networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3025–3029, Sept. 2012.

[118] W. Saad, X. Zhou, B. Maham, T. Basar, and H. V. Poor, "Tree formation with physical layer security considerations in wireless multi-hop netoworks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3980–3991, Nov. 2012.

[119] H. Wang, X. Zhou, and M. C. Reed, "On the physical layer security in large scale cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2776–2787, June 2013.

[120] C. Liu, N. Yang, J. Yuan, and R. Malaney, "Location-based secure transmission for wiretap channels," *IEEE J. Select. Areas Commun.*, vol. 33, no. 7, pp. 1458–1470, July 2015.

[121] P. Carroll, K. Mahmood, S. Zhou, H. Zhou, X. Xu, and J.-H. Cui, "On-demand asynchronous localization for underwater sensor networks," *IEEE Trans. Signal Processing*, vol. 62, no. 13, pp. 3337–3348, July 2014.

[122] W. R. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," *J. Acoust. Soc. Am.*, vol. 58, no. 1, pp. 201–207, 1975.

[123] G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 463–470, June 1981.

[124] B. H. Maranda, *Passive Sonar*. New York, NY: Springer New York, 2008, pp. 1757–1781. [Online]. Available: https://doi.org/10.1007/978-0-387-30441-0_97

[125] Y. T. Chan and F. L. Jardine, "Target localization and tracking from doppler-shift measurements," *IEEE J. Ocean. Eng.*, vol. 15, no. 3, pp. 251 –257, 1990.

[126] V. R. Cadambe and S. A. Jafar, "Degrees of freedom of wireless networks - What a difference delay makes," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove CA, 2007.

[127] P. Qarabaqi and M. Stojanovic, "Modeling the large scale transmission loss in underwater acoustic channels," in *Proc. of the 49th Annual Allerton Conference on Communication, Control, and Computing*, 2011.

[128] T. A. Thomas and F. W. Vook, "Asynchronous interference suppression in broadband cyclic-prefix communications," *IEEE Wireless Communication and Networking*, vol. 1, pp. 568–572, 2003.

[129] H. Jung and M. D. Zoltowski, "Semiblind multichannel identification in asynchronous multiuser OFDM systems," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove CA, Oct. 2004.

[130] ——, "On the equalization of asynchronous multiuser OFDM signals in fading channels," in *Proc. of Intl. Conf. on ASSP*, vol. 4, May 2004, pp. 765–768.

[131] Q. Li, J. Zhu, X. Guo, and C. N. Georghiades, "Asynchronous co-channel interference suppression in MIMO OFDM systems," in *Proc. of Intl. Conf. on Commun.*, Jun. 2007, pp. 5744 –5750.

[132] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*, 4th ed. Englewood Cliffs: Prentice Hall, 2006.

[133] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific, 1997.

[134] D. Bertsimas and J. Tsitsiklis, "Simulated annealing," *Statistical science*, vol. 8, no. 1, pp. 10–15, 1993.

[135] F. J. Solis and R. J.-B. Wets, "Minimization by random search techniques," *Mathematics of operations research*, vol. 6, no. 1, pp. 19–30, Feb. 1981.

[136] R. Horst and N. V. Thoai, "DC programming: Overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, Oct. 1999.

[137] R. Horst, T. Q. Phong, N. V. Thoai, and J. de Vries, "On solving a DC programming problem by a sequence of linear programs," *Journal of Global Optimization*, vol. 1, no. 2, pp. 183–203, Oct. 1991.

[138] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *Journal of Global Optimization*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[139] W. Sun, Z.-H.Wang, M. Jamalabdollahi, and S. A. Zekavat, "Experimental study on the difference between acoustic communication channels in freshwater rivers/lakes and in oceans," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2014.

[140] J. Huang, S. Zhou, and P. Willett, "Nonbinary LDPC coding for multicarrier underwater acoustic communication," *IEEE J. Select. Areas Commun.*, vol. 26, no. 9, pp. 1684–1696, Dec. 2008.

[141] C. R. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[142] W. Sun and Z.-H. Wang, "Modeling and prediction of large-scale temporal variation in underwater acoustic channels," in *Proc. of MTS/IEEE OCEANS Conf.*, Shanghai China, Apr. 2016.

[143] C. Manss, D. Shutin, T. Wiedemann, A. Alberto, and J. Mueller, "Decentralized multi-agent entropy-driven exploration under sparsity constraints," in *Proc. of*

Intl' Workshop on Compressed Sensing Theory and its Application to Radar, Sonar and Remote Sensing, Aachen, Germany, Sep. 2016.

[144] R. Marchant, F. Ramos, and S. Sanner, "Exploration under spasity constraints," in *Proc. of European Conf. on Mobile Robots*, Lincoln, UK, Sep. 2015.

[145] B. Chen, P. Pandey, and D. Pompili, "A distributed adaptive sampling soluting using autonomous underwater vehicles," in *Proc. of the ACM Intl. Workshop on Underwater Networks (WUWNet)*, Los Angeles, CA, Nov. 2012.

[146] R. Cui, Y. Li, and W. Yan, "Mutual information-based multi-AUV path planning for scalar field sampling using mutidimentional RRT*," *IEEE Trans. System, Man, and Cybernetics: Systems*, vol. 46, no. 7, pp. 993 – 1004, Jul. 2016.

[147] F. Fazel, M. Fazel, and M. Stojanovic, "Randnom access comppresed sensing over fading and noisy communication channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2114 – 2125, May 2013.

[148] F. Hover, R. Hummel, U. Mitra, and G. Sukhatme, "One-step-Ahead kinematic compressive sensing," in *Proc. of IEEE GLOBECOM Wi-UAV Wksp.*, Houston, TX, Dec. 2011.

[149] R. Hummel, S. Poduri, F. Hover, U. Mitra, and G. Sukhatme, "Mission design for compressive sensing with mobile robots," in *Proc. Proc. of Int'l. Conf. on Robotics and Automation*, Shanghai, China, May 2011.