WARWICK
THE UNIVERSITY OF WARWICK

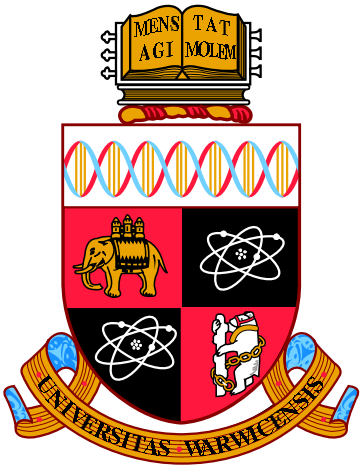
A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/184060

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications



Optimising Signal Detection Techniques in Wireless Ultraviolet Communication Systems

by

Wenxiu Hu

Thesis

submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Engineering

School of Engineering

February 2023

THE UNIVERSITY OF WARWICK

Contents

List of Figures v

Acknowledgments x

Declarations xii

List of Publications xiii

Abstract xvi

Abbreviations xviii

Chapter 1 Introduction 1

1.1 Background 1

1.2 Motivation 4

1.3 Outline of Existing Approaches 5

1.3.1 Channel Estimation Methods 5

1.3.2 Coherent Signal Detection Schemes 6

1.3.3 Non-coherent Signal Detection Methods 7

1.3.4 Other Techniques to Improve Detection Accuracy . . . 8

1.4 Contributions of this Thesis 9

1.4.1 1D Non-coherent Detection 9

1.4.2 HD non-coherent detection 11

1.4.3 NN-based Nonlinear Feature for Non-coherent Detection 12
1.5 Organization of the Rest of this Thesis 12

Chapter 2 Literature Review on Existing Signal Detection Methods for OWC 15

2.1 Channel Estimation Technique for Coherent Signal Detection . 16
2.1.1 Least Square Channel Estimator 16
2.1.2 Simulations of LS Channel Estimator 18
2.2 Linear Equalizer based on Minimum Mean Squared Error Criterion 20
2.2.1 MMSE equalizer 21
2.2.2 Simulation Results of MMSE Equalizer 23
2.3 Maximum Likelihood Sequence Detection 26
2.3.1 Construction of CSI and TPM 26
2.3.2 Likelihood PDF of each Received Signal 28
2.3.3 Theory of MLSD 28
2.3.4 Simulation Result of MLSD 30
2.4 Sub-optimal Bayesian Sequential Detection 31
2.4.1 Theory of Bayesian Sequential Detection 31
2.4.2 Particle Filter Realisation of Bayesian Framework . . . 32
2.4.3 Simulation Results of Bayesian Sequential Detection . . 35
2.5 Conclusions & Discussions 38

Chapter 3 Ultraviolet Signal Model & System 40

3.1 System Model 40
3.1.1 UV Transmitting Module 40
3.1.2 UV Channel 41
3.1.3 UV Receiver Module 42
3.2 Simulations of UV Channel 43

3.3 Conclusions & Discussions 48

Chapter 4 One-Dimensional Non-Coherent Signal Detection 50

4.1 System Models and Problem Formulation 51

4.2 Feature Extraction 52

4.2.1 Feature Designs 52

4.2.2 Advantages of feature-based metrics 57

4.3 Equal-Weighted Non-Coherent Detection Method 58

4.3.1 EWNCD Design 58

4.3.2 Influence of Weights 60

4.4 Optimally-Weighted Non-Coherent Detection Method 61

4.4.1 OWNCD Design 62

4.4.2 Distribution of weighted combining metric 62

4.4.3 Theoretical BER computation 63

4.4.4 Optimal weight computation 64

4.4.5 Heuristic Algorithm 66

4.5 Performance Analysis and Numerical Simulations 68

4.5.1 Complexity Analysis 68

4.5.2 Numerical Simulations 71

4.6 Conclusions & Discussions 80

Chapter 5 High-Dimensional Non-Coherent Detection for Multi-Intensity Modulated Communications 83

5.1 System Models and Problem Formulation 84

5.2 HD Feature Construction 85

5.3 Theoretical Symbol Rate Error 89

5.4 Proof of Better than Lower-Dimensional Methods 94

5.4.1 Better SER than 1D feature 94

5.4.2 Better SER than Lower Dimensional feature 95

5.5 HD Non-Coherent Signal Detection Scheme 97
5.5.1 Constrained GMM Clustering 99
5.5.2 Overall Non-Coherent Detection Algorithm 103
5.5.3 Convergence Analysis 105
5.6 Numerical Simulations 106
5.6.1 GMM Clustering accuracy 107
5.6.2 Communication Performance 111
5.6.3 Complexity Comparison 118
5.7 Conclusions & Discussions 119

Chapter 6 Machine Learning based Nonlinear Feature for Non-coherent UV Signal Detection 121

6.1 Supervised NN based UV Detector 122
6.1.1 Implementation Details 122
6.1.2 Simulation Results 124
6.2 Half-Supervised NN UV Detector with Explainable AI 126
6.2.1 Parzen Window PDF Approximation 127
6.2.2 Parzen Window-based Neural Network Detector 128
6.2.3 Simulation Results 131
6.2.4 Conclusion & Discussion 134

Chapter 7 Summary & Future Works 135

7.1 Summary of this Thesis 135
7.2 Future Works 139
7.2.1 Reinforcement Learning the Optimal Feature Construction 139
7.2.2 Expand to Explainable AI 143

Bibliography 143

List of Figures

2.1 Performance of LS channel estimator: RMSE versus SNR and length of the pilot sequence L'. 19

2.2 Performance of signal detection via MMSE equalizer, under different SNR, with symbol transmission rate as $R_s = 10^6\ bit/s$. . 24

2.3 Performance of signal detection via MMSE equalizer, under different ISI intensity (in terms of symbol rate R_s), with SNR as 20dB. 25

2.4 Performance of signal detection via MLSD scheme, under different SNR, with transmission rate as $R_s = 10^6\ bit/s$. 30

2.5 Performance of Bayesian sequential detection scheme: BER versus SNR. 36

2.6 Complexity of proposed scheme as opposed to imperfect MLSD. 37

3.1 Illustration of the basic UVC system with a LED transmitter, dynamic UV channel, and a UV photon-counting receiver. . . 41

3.2 Illustrations of binary (OOK) symbols that will be transmitted, the CIR of UVC system, and the received signals with SNR=8dB. 45

3.3 Illustrations of 8-PAM symbols that will be transmitted, the CIR of the UVC system, and the received signals with SNR=15dB. 46

3.4 Illustrations of same OOK sequence undergoing different transmitted interval (ISI level) with SNR=15dB. 47

4.1 Illustration of UVC system with a LED transmitter, dynamic UV channel, and a UV photon-counting receiver. The 1D non-coherent detection scheme is deployed at the end of the receiver, whereby metrics are constructed at first, followed by the Heuristic detection. 51
4.2 Illustrations of the feature-based metrics: the intra-symbol rising-edge. 53
4.3 Illustrations of the feature-based metric: the intra-symbol energy. 54
4.4 Illustrations of the feature-based metric: the inter-symbol minimum inflection. 56
4.5 Illustrations of the feature-based metric: the inter-symbol energy difference. 56
4.6 The BER performance with different settings of weights, normalised by $w_1 + w_2 + w_3 = 1$. The minimum BER corresponds to the optimal weights computed in Eq. (4.25), i.e., $w_1 = 0.6$, $w_2 = 0.32$, and $w_3 = 0.08$. 61
4.7 Illustration of received UVC signals and the corresponding EWNCD and OWNCD metrics. 73
4.8 Illustration of statistic histograms of unprocessed signals, proposed EWNCD and OWNCD metrics. 75
4.9 BER versus different intensities of ISI (whose increase is reflected by the increase of symbol rate R_b). The slower growth of BER with the increase of the ISI indicates the abilities of the proposed OWNCD to combat ISI. 76
4.10 BER comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1] in a static channel. 77

4.11 BER comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1], in a time-varying channel. 79
4.12 complexity comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1] in a static channel, based on 10^4 bits. 81

5.1 Schematic flow illustrates the UVC system and the High-Dimensional non-coherent detection scheme. 85
5.2 Illustrations of the feature-based metrics: the intra-symbol rising-edge, in 4-PAM mode. 86
5.3 Illustrations of the feature-based metric: the intra-symbol energy, in 4-PAM mode. 86
5.4 Illustrations of the feature-based metric: the inter-symbol minimum inflection, in 4-PAM mode. 87
5.5 Illustrations of the feature-based metric: the inter-symbol energy difference, in 4-PAM mode. 87
5.6 Illustration of clustering errors using classic GMM in Eq. (5.28): four cases of GMM trapped in local optimal solutions. 101
5.7 Illustration of non-coherent schemes that transform the ISI-caused sequential detection space into an ISI-released multi-hypothesis detection framework. Under the same received UV signals, (a) shows the poor detection accuracy from 1D non-coherent schemes, (b) gives the higher detection accuracy of the proposed HD non-coherent scheme. 102
5.8 Illustration of revised GMM on OOK signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=8dB, (b) SNR=15dB. 107

5.9 Illustration of revised GMM on 4-PAM signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=15dB, (b) SNR=25dB. 109
5.10 Illustration of revised GMM on 8-PAM signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=20dB, (b) SNR=30dB. 110
5.11 BER comparison versus SNR under OOK modulation. the proposed HD non-coherent detection outperforms the MMSE equalizer and the previous 1D non-coherent scheme, and approaches that of the MLSD method. 112
5.12 BER comparison versus ISI (rendered by different symbol rates R_s), under OOK modulation, with SNR=15dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method. 113
5.13 BER versus SNR under J = 4-PAM symbol. the proposed scheme shows comparable accuracy with the MLSD and MMSE equalizer, suggesting its suitability for multi-symbol UVC scenarios. 114
5.14 BER comparison versus ISI (rendered by different symbol rates R_s), under 4-PAM modulation, with SNR=20dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method. 115
5.15 BER versus SNR under J = 8-PAM symbol. the proposed scheme shows comparable accuracy with the MLSD and MMSE equalizer, suggesting its suitability for multi-symbol UVC scenarios. 116

5.16 BER comparison versus ISI (rendered by different symbol rates R_s), under 8-PAM modulation, with SNR=25dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method. 117
5.17 Computational complexity compared to the MLSD method, emphasizing the computational efficiency of the proposed HD non-coherent scheme. 119

6.1 Illustration fully connected NN based detector, where the input layer is the received UV signals, and the output layer is the symbol classes (here the symbols have two classes, representing the OOK modulations for transmitter). 123
6.2 Performance of supervised NN based signal detector: BER versus SNR in dB. 125
6.3 Illustration of Parzen-PNN that consists of three layers, i.e., the input layer fed by the designed metric $\mathbf{z}_k$, the pattern layer trained by the labels $s_{k-L}, s_{k-L+1}, \cdots, s_{k-1}$, and the class layer for classification. 129
6.4 Performance of PNN based signal detector. 132
6.5 Selection of virtual variance η^2, in which the suggested virtual variance η^2 gives the lowest BER. 133

7.1 Illustration fully connected NN-based detector, where the input layer is the received UV signals, and the output layer is the symbol classes (here the symbols have two classes, representing the OOK modulations for transmitter). 140

Acknowledgments

First, I would like to appreciate my two supervisors Dr Tianhua Xu and Dr Mark Leeson when I am completing my PhD study at the School of Engineering, University of Warwick.

Dr. Tianhua is my first supervisor, who offered me the opportunity to study at this fantastic group and helped me apply for scholarships to cover my tuition fees. During my PhD life at Warwick, Dr Tianhua guided me to do a literature review, helped me find an interesting topic, and taught me the writing skills for academic papers. His professionalism and kindness helped me and keep providing a great deal and are very worth learning.

Prof. Mark is my second supervisor, who has been offering me a great deal of help in revising papers and how to deal with Reviewers' comments. Also, his expertise in optical wireless communication and the specialty in numerical optimisation and digital signal processing provided me a great help for the system modeling and algorithm designs in my research.

Apart from these, I appreciate my family members' supports which bring me great confidence. Besides, I would like to appreciate Dr. Zhuangkun Wei for his expertise on linear algebra and nonlinear optimisation theories and approaches, Prof. Min Zhang for his professional guidance on the Monte Carlo simulation testbed, Ms. Liping Wan, Ms. Cenqin Jin, Ms. Xueyun Gu, Mr. Zichao Zhang, Ms. Bailu Jin and Juju for their kindness, company and help. Next, I would like to appreciate my panels: Prof. Jihong and Dr.

Subhash, who make a good guide on my school life. Finally, I owe a great deal to the School of Engineering, University of Warwick who offer me the chance to pursue my PhD study here, and who cover my tuition fees.

Declarations

I, Wenxiu Hu confirm that, this thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Engineering. It has been composed by myself and has not been submitted for any degree at other universities.

List of Publications

Inclusion of Published Works During PhD

1. **W. Hu**, Z. Wei, S. Popov, M. Leeson, M. Zhang and T. Xu, "Non-coherent detection for ultraviolet communications with inter-symbol interference." Journal of Lightwave Technology 38.17 (2020): 4699-4707.

2. **W. Hu**, M. Zhang, Z. Li, S. Popov, M. Leeson, and T. Xu, "High-dimensional feature based non-coherent detection for multi-intensity modulated ultraviolet communications." Journal of Lightwave Technology 40, no. 7 (2021): 1879-1887.

3. **W. Hu**, Z. Wei, S. Popov, M. Leeson, and T. Xu, "Tapping Eavesdropper Designs against Physical Layer Secret Key in Point-to-Point Fiber Communications." Journal of Lightwave Technology (2022 Nov 17).

4. **W. Hu**, Z. Wei, Z. Li, S. Popov, M. Leeson, T. Xu, "Eavesdropping Against Bidirectional Physical Layer Secret Key Generation in Fibre Communications," In 2022 IEEE Photonics Conference (IPC), pp. 1-2. IEEE, 2022.

Other Publications During PhD

1. J. Ding, T. Liu, T. Xu, **W. Hu**, S. Popov, M. Leeson, J. Zhao, and T. Xu, "Intra-Channel Nonlinearity Mitigation in Optical Fiber Transmission Systems Using Perturbation-based Neural Network." Journal of Lightwave Technology, vol. 40, no. 21, pp. 7106-7116, 2022.

2. Z. Wei, B. Li, W. Guo, **W. Hu** and C. Zhao, "On the Accuracy and Efficiency of Sensing and Localization for Robotics," IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 2480-2492, 2022.

3. M. Li, T. Xu, S. Wang, **W. Hu**, J. Jiang, and T. Liu, "Probe pulse design in Brillouin optical time-domain reflectometry." IET Optoelectronics 16, no. 6 (2022): 238-252.

4. T. Xu, M. Li, N. Shevchenko, M. Tan, **W. Hu**, Z. Li, and T. Liu, "Achievable Information Rates in C-band Nonlinear Coherent Optical Communication Systems," 2022 IEEE Photonics Society Summer Topicals Meeting Series (SUM), Cabo San Lucas, Mexico, pp. 1-2, 2022.

5. M. Li, T. Xu, J. Ding, Z. Liu, **Wenxiu Hu**, J. Jiang, and T. Liu. "Evaluation of transmission reach and information rates in nonlinear optical fiber communication systems." In Semiconductor Lasers and Applications XI, vol. 11891, pp. 51-58. SPIE, 2021.

6. Z. Wei, B. Li, **W. Hu**, W. Guo, C. Zhao, "Hamming–Luby Rateless Codes for Molecular Erasure Channels", Nano Communication Networks, vol. 23, art. 100280, 8 pages, 2020.

7. T. Xu, J. Ding, Y. Liu, **W. Hu**, B. Karanov, N. Shevchenko, Z. Li, and T. Liu, "Influence of spectral broadening on nonlinearity compensation in ultra-wideband optical fiber communications." In 2020 22nd International Conference on Transparent Optical Networks (ICTON), pp. 1-4. IEEE, 2020.

8. Z. Wei, B. Li, W. Guo, **W. Hu** and C. Zhao, "Sequential Bayesian Detection of Spike Activities From Fluorescence Observations," IEEE Transactions on Molecular, Biological and Multi-Scale Communications, vol. 5, no. 1, pp. 3-18, 2019.

Abstract

Wireless ultraviolet (UV) communication is regarded as a promising supplement to conventional wireless communications. The challenges lie in the inter-symbol-interference (ISI) and the time-varying channel impulse response (CIR), which deteriorate the detection/estimation of transmitted symbols from the received signals. The existing coherent detection schemes that leverage CIR estimation for ISI compensation, fall into two camps. They present either an overhead burden of pilot sequences and computational complexity or poor CIR acquisition that further hinders the detection performance.

The aim of this thesis is to design non-coherent detection schemes that can transform the ISI contaminated sequential detection process into discrete binary or multiple detection framework. This is achieved by extracting the UV communication signal-related geometrical features that are inherently resistant to ISI. Then, one-dimensional and high-dimensional non-coherent detection schemes are proposed, by designing optimal linear and high-dimensional combinations of these features that minimize the theoretical bit error rate (BER). Both theoretical and simulation verification are performed to validate the proposed scheme, showing the comparable detection accuracy of the state-of-the-art coherent schemes but at the expense of lower computational complexity.

To further expand the scope of the ISI-resistant features, machine learning is employed to discover nonlinear features that can express hidden relations from received signals to transmitted symbols. This is done by (i) a supervised neural network (NN) based detector, and (ii) a more explainable Parzen window technique based NN to approximate the detection likelihoods. For future work, deep reinforcement learning will be utilized to explore better ISI-resistant features for detection purposes. As such, by casting the complex sequential detection into the concise discrete detection framework, and

combining the manual and machine learning-based ISI-resistant feature construction, this work provides a novel idea, not only for UV communications but can also be applied to other communication and signal detection scenarios suffering from ISI.

Abbreviations

IoT	Internet of Things
M2M	Machine-to-Machine
CPSs	Cyber-Physical Systems
RF	Radio Frequency
OWC	Optical Wirless Communication
IR	Infrared
VLC	Visible Light Communication
UVC	Ultraviolet Communication
NLOS	Non-Line-of-Sight
ISI	Inter-Symbol Interference
CIR	Channel Impulse Response
CSI	Channel Statement Information
MLSD	Maximum Likelihood Sequence Detection
ML	Maximum Likelihood
BER	Bit Error Rate
MCvD	Molecular Communication via Diffusion
1D	One-Dimensional
HD	High-Dimensional
EWNCD	Equally-Weighted Non-Coherent Detection
OWNCD	Optimally-Weighted Non-Coherent Detection
SNR	Signal-to-Noise-Ratio

FEC	Forward-Error-Correction
GMM	Gaussian Mixture Model
MMSE	Minimum Mean Squared Error Criterion
PAM	Pulse Amplitude Modulation
NNs	Neural Networks
LIT	Linear Time Invariant
LS	Least Square
RMSE	Root Mean Square Error
FIR	Finite Impulse Response
ZF	Zero-Forcing
OOK	On-Off Keying
TPM	Transitional Probability Matrix
PDF	Probability Distribution Function
EM	Expectation Maximization
PNN	Parzen Window based Neural Network
RNN	Recurrent Neural Network
RBFN	Radial basis Function Network
DDPG	Deep Deterministic Policy Gradien
DQN	Deep Q-learning
X-AI	Explainable Artificial Intelligence

Chapter 1

Introduction

1.1 Background

With the rapid development of various electronic technologies, lifestyles become deeply reliant on digital media, during the last twenty years. In this age of data, communication networks not only supply social and entertainment media, but also provides services to a massive number of new devices and applications related to the Internet of Things (IoT) [2], Machine-to-Machine (M2M) [3], and Cyber-Physical Systems (CPSs) [4].

Fibre optic builds the backbone of the communication networks, feeding sufficient bandwidth to carry the exponential growth of the data demand. It is noted that a fibre optic network can only deliver the message to the devices that are tethered to it, while there are as many transmission demands for mobile communication as that for wired communication. Most wireless communications are conducted by radio frequency (RF) communication systems, consequently, the shortage of spectrum resources is becoming a major obstacle to the development of wireless communications [5]. According to Cooper's Law, the data-carrying capacity of the RF spectrum double every 2.5 years on average [6]. An increase in research concentrating on exploiting new commu-

nication technologies to promote transmission capacity is thus required.

Optical Wireless Communication (OWC) has been widely studied and researched in recent years, as an alternative or complementary to traditional RF communication systems [7, 8]. OWC refers to the use of light-emitting devices(e.g., LEDs and lasers) to transmit information through the free air space, offering a number of benefits including higher bandwidth, security, license-free, and lower interference [9]. The main advantage of OWC is the higher bandwidth compared to the limited and congested RF spectrum. This allows for faster data transfer and increased capacity for a large number of users.

There are different forms of OWC such as infrared (IR) [10], visible light communication (VLC) [11], and ultraviolet communication (UVC) [12]. IR and VLC are commonly used OWC technologies and have been applied in short-range communication scenarios, such as television remote controls and personal computer-to-peripheral communication [13–15]. UVC is a relatively new technology, that offers improved bandwidth and security, and has recently been proposed for use in communication networks and IoT devices.

Compared with visible light, ultraviolet light has a shorter wavelength and its spectrum length ranges from 100nm to 400nm. According to the wavelength ranges, ultraviolet light is divided into ultraviolet A rays (UV-A 315nm to 400nm), ultraviolet B rays (UV-B 280nm-315nm) ultraviolet C rays (UV-C 190nm-280nm) [16].

As the atmosphere has a strong absorption effect on ultraviolet light, the UV-C rays generated by solar radiation are almost absorbed in the near-Earth atmosphere, consequently, the UV-C is also known as the solar blind zone [17]. Besides, different from UV-A and UV-B, the UV-C band is harmless to humans. As such, the UV-C band is selected for supporting wireless ultraviolet communication. Because of no interference from natural sunlight radiation, UVC has the advantage of low interference from background noise [18]. In

addition, the ultraviolet photons exhibit strong particle properties, due to the short wavelength of the ultraviolet beam. This then leads to the strong scattering characteristics of the ultraviolet communications. [19].

Due to the unique solar blind and scattering characteristics of UVC, ultraviolet communication has acquired unique channel characteristics, which will remedy the defects of IR and VLC. The main advantages of UVC are summarized as follows.

i) Low background interference promises work day and night

The UV-C band light from solar radiation is easily absorbed by the atmospheric layer, as a consequence it can hardly reach the near-earth atmosphere. Thus, ultraviolet communications working in the UV-C band are free from the background noise caused by solar radiation, both day and night. It means this absorption characteristic promises UVC qualified for day and night work scenarios [20].

ii) Non-line-of-sight communication adapts to complex communication environment

When ultraviolet light in the UVC band propagates in the atmosphere, it shows strong scattering characteristics, which enable it to communicate in the non-line-of-sight (NLOS) scene. It means the deployment of the transmitter and the receiver could be in a relaxed alignment, whereas the IR and VLC require strict alignment[10]. Due to this property, UVC presents a promising candidate for complex and RF denial communication scenarios, e.g., in the above sea marine communication scene, which is difficult to guarantee a strict alignment, when suffering wave turbulence.

iii) Good Confidentiality

The atmosphere produces strong absorption and scattering effect on ultraviolet photons, which makes the fading of the ultraviolet communication channel increase exponentially with distance. Therefore, UV wireless optical communication is mainly used for short-distance communication. Attributed to the fast fading characteristic, signals cannot be detected from long-distance, avoiding the transmitted information leaking in a wide rage. Thus, as a short-distance communication system, it obtains good concealment [11].

Overall, UVC is a promising technology that has the potential to provide NLOS, secure wireless communication solutions, utilized in a variety of environments and applications. However, further research is needed to address the technical and practical challenges associated with its implementation and deployment.

1.2 Motivation

As aforementioned, UVC is regarded as a promising supplement for conventional wireless communication systems, due to its NLOS transmission characteristic [21, 22] and solar-blind property [12, 21–24]. This has raised a large number of researches, ranging from the developments of UVC devices [25–28], the model formulations [22,29–33], designs of channel coding [34–37], to detection scheme designs [1,38–41], all of which aim to enhance the communication accuracy for UVC scenarios.

From the perspective of signal detection, there are three challenges that deteriorate communication performance and that are thereby required to be addressed. (i) Due to the channel dispersion leading to the long impulse response tail, the photon-detector suffers from the intensive inter-symbol interference (ISI) at the receiving point, especially when the communication rate is

high [1,21,40]. (ii) The counting noise as well as the thermal noise induced by the transceivers may deteriorate the communication performance [21,36,40]. (iii) Most of the research overlooked the time-varying channel effects, which although can be addressed by repeated channel estimation via the usages of the pilot sequences [1], may result in large overhead and extensive computational and storage complexity.

In short, improving the detection accuracy and reducing the computational complexity would be meaningful for the application of UVC in real communication systems. Thus, the goal of this thesis is to design appropriate detection methods, which have high detection accuracy in terms of BER but with low computing consumption.

1.3 Outline of Existing Approaches

To overcome the aforementioned challenges in UVC signal detection, existing methods are mainly classified into two genres: coherent detection and non-coherent detection, according to whether or not the method relies on the channel impulse response (CIR) in the detection process. A brief introduction of what they are given, along with their drawbacks in this part. Their detailed theories and implementations are provided in Chapter 2.

1.3.1 Channel Estimation Methods

For the coherent detection method, the CIR plays a significant role, as it is inevitably necessary to acquire the underlying channel statement information (CSI) for ISI compensation before the process of signal detection. Typical channel estimation approaches include the least square (LS) channel estimator, and the minimum mean square error (MMSE) estimator with the knowledge of the CIR distribution. These techniques rely on a pre-designed UV channel-

related pilot sequence to estimate the whole CIRs.

1.3.2 Coherent Signal Detection Schemes

Leveraging the estimated CIR, the coherent detection schemes in [1, 39–42] aim to design either equalizer or construct the posterior or likelihood probability density functions (PDFs) to determine the transmitted symbols. To be specific, a linear equalizer is proposed in [42], which tries to recover the transmitted symbols from a linear filter of a window of received signals and their corresponding CIRs, by minimizing the statistical mean squared error. Then, to further improve the detection accuracy, the state-of-the-art maximum likelihood sequence detection (MLSD) is proposed in [1] for the UVC system, which firstly computes the likelihood PDFs conditioned on transmitted symbols using the estimated CIRs, and then select the optimal paths of the sequentially transmitted symbols to maximize the likelihood trajectory. To implement this idea, a Viterbi-like algorithm is proposed in [39], and demonstrates the optimal performance over all model-based detection schemes.

Although MLSD is a practical method to combat the ISI in a time-invariant system, however it has several drawbacks. The MLSD method gives rise to either the long overhead of pilot sequences and huge computational complexity for accurate channel estimations in time-varying scenarios or poor CIR acquisition that further deteriorates the signal detection performance. To realize a more computationally and resourcefully efficient MLSD scheme, the study in [40] devised a UV channel-related Bayesian scheme. The idea is to use a particle filter to adapt to the time-variant UV system, which can simultaneously detect transmitted symbols and estimate the channel impulses. However, the efficiency problem, though being sorted out to some extent, still remains, given the trade-off between the number of particles and the estimation/detection accuracy.

1.3.3 Non-coherent Signal Detection Methods

In contrast to coherent detection, the non-coherent detection schemes do not rely on the exact CIR to compensate ISI. Instead, they focus on first simplifying the complex sequential detection variables, e.g., the posterior or likelihood PDFs, and then design signal detection methods specific to these simplified variables. The popular non-coherent schemes for OWC are listed in [38,43–45], in which the advantages and drawbacks of these papers are recapped in the following. The optimal non-coherent maximum likelihood (ML) detectors (for multiple-symbol MS, and symbol-by-symbol SS) have been proposed and studied in [38, 43], which provided optimal (or suboptimal) detection algorithms for OWC applications. The results showed a promisingly lower bit error rate (BER) in the case of the negligible ISI.

However, the obvious drawback is their vulnerabilities to combat the strong ISI. This is because their simplification processes omit the long tail interference of the previously transmitted symbol. As such, new non-coherent detection techniques that can resist ISI are needed, and this is what will be done in this thesis.

It is noteworthy that non-coherent detection methods have been studied in molecular communications, whereby the properties from chemical reactions, and molecule structures and movements (e.g., diffusion, and flow) are leveraged to reduce the ISI from previously transmitted molecules. One popular example comes from the utilisation of enzyme [46, 47], which, properly designed by the receptor, can mitigate the long tail nature of the diffusive particles in the free-space and vessel channels. A similar idea has been exploited by the work in [48], which relies on the chemical reactions of the calcium ions to reduce their lifetime for ISI cancelation. However, such non-coherent idea is hard to be implemented on the OWC, given the lack of chemical cancelation properties for ultraviolet photons.

1.3.4 Other Techniques to Improve Detection Accuracy

Apart from the designing of signal detection schemes, other techniques to protect the transmitted symbols from being contaminated by noise and ISI include channel coding, and noise reduction approaches, which serve as the pre/post-processing before/after the signal detection process. Here, a brief introduction to these schemes will be provided, but the implementation details will not be discussed in this thesis, as these do not involve in the designs of the detection schemes.

Channel coding refers to schemes that introduce redundancy of the information bits (e.g., generate the check bits of the information bits), in order to protect the correctness of the information bits. The popular and widely used schemes include but are not limited to the polar code [49–52], the low-density parity check (LDPC) code [53], and the Luby transform (LT) code [54, 55]. These channel coding techniques provide a forward error correction (FEC) limit, which means if the signal detection schemes can reach an order of 10^{-2} BER, then by deploying these codes for bit check and recovery, the BER can be further reduced and suitable for practical usages.

Noisy reduction aims to filter the received signals, by the difference in frequencies of signals and background noise. One example is the work in [56], which uses the wavelet technique to remove the noise components by filtering out high-frequency components from the frequency domain. By doing so as a pre-processing step, the detection accuracy of signal detection schemes will be promoted. Yet, one should notice the statistical modification of the probability distribution of the received signals, e.g., the likelihood PDFs of MLSD may not fit well as the Gaussian process and may lead to detection error given the inaccurate computation of likelihood PDFs.

1.4 Contributions of this Thesis

This thesis focuses on non-coherent detection, which aims to extract the features from the received signals that are inherently resistant to the ISI, and then transform the sequential detection problem into a discrete binary or multiple detection framework. To implement this idea, four ISI-resistant features are first designed by exploiting the geometry shapes of the received UV signals. Then, two signal detection schemes are designed, respectively leveraging the optimal feature linear combination (termed the *optimally-weighted non-coherent detection scheme* in Chapter 4), and the high-dimensional feature construction (called the *high-dimensional non-coherent detection scheme* in Chapter 5).

To further expand the scope of the ISI-resistant features, a machine learning based technique is employed to discover nonlinear features that can express the hidden mapping from the received signals to the transmitted symbols. This is done by (i) a supervised neural network (NN) based detector, and (ii) a more explainable Parzen window technique based NN to approximate the detection likelihood PDFs. The contributions of these detection methods are elaborated in detail as follows.

1.4.1 1D Non-coherent Detection

In Chapter 4, the 1D non-coherent method tries to find the optimal linear combination of the designed ISI-resistant features, by minimizing the theoretical BER. To sum up, the main contributions of 1D method are listed as follows.

(1) Four linear features of UV signals are designed and extracted based on the geometry shapes of the received signals. These include the intra-symbol rising edge, the intra-symbol energy, the inter-symbol minimum inflection, and the inter-symbol energy difference of the signal. Such features are inherently

insensitive to ISI contamination.

(2) Leveraging the extracted four features, two novel non-coherent schemes are proposed called *equally-weighted non-coherent detection (EWNCD)* and *optimally-weighted non-coherent detection (OWNCD)*, respectively by equally and optimally weighting and combining such features. The proposed 1D non-coherent schemes are able to convert the detection scenarios with ISI into a binary detection framework, which avoids the complex channel estimation process. Also, the proposed OWNCD method is able to adjust the optimal weights with respect to the different levels of the ISI intensities and therefore delivers a better detection performance.

(3) For the time-varying channel response, a heuristic algorithm is designed to adaptively achieve the optimal weights. Such a heuristic algorithm relies on the previous data that can reflect the continuous changes in the channel environment, thereby capable of addressing the time-varying UV channel challenges, but with lower computational complexity.

(4) The detection performance of the proposed EWNCD and OWNCD schemes are numerically evaluated. Compared to the state-of-the-art coherent MLSD in the cases of static and time-varying CSI, the proposed OWNCD can gain $\sim$ 1 dB and $\sim$ 8 dB in signal-to-noise-ratio (SNR) at %7 overhead FEC limit (BER of 4.5×10^{-3} according to [57]), respectively. This means that given appropriate channel coding schemes, our proposed OWNCD can conduct practical communication (BER less than 10^{-6}) with lower SNR, compared to the coherent MLSD with imperfect CSI estimation. Also, the computational complexity required by the proposed OWNCD is four order of magnitude reduction when compared to the coherent MLSD scheme.

1.4.2 HD non-coherent detection

In Chapter 5, a high dimensional combination of the designed geometrical features is constructed for further non-coherent detection. The detection process is then pursued by the revised Gaussian mixture model (GMM) based clustering method. The main contributions of the HD method are listed in the following.

(1) Leveraged on the extracted four features, a novel high-dimensional non-coherent scheme is proposed, by constructing a high-dimensional feature space. It transforms the ISI-caused sequential detection into the ISI-released multi-hypothesis detection framework, therefore avoiding the high computational CIR estimation and sequential likelihood computations required by the coherent schemes (a reduction of 4 magnitude orders of complexity is achieved).

(2) The BER performance of the HD method is analyzed, by deriving the theoretical BER of the HD feature-based non-coherent detection. It mathematically shows the reason it is better than the previously proposed one-dimensional feature-based non-coherent schemes.

(3) A clustering algorithm is designed to conduct the classification of the HD features for signal determination, based on the Gaussian mixture model (GMM) clustering framework.

(4) Numerical simulation is conducted to evaluate the HD method. This illustrates that the proposed HD method is able to provide higher signal noise ratio (SNR) for signal detection than current non-coherent schemes, which leads to a very-close detection accuracy with the state-of-the-art coherent MLSD ($< 1dB$ in SNR at the FEC limit BER=4.5×10^{-3}). These thereby demonstrate the promising pathway of the proposed non-coherent scheme towards a low-computational but high-accurate detection paradigm for optical wireless communications with intensive ISI.

1.4.3 NN-based Nonlinear Feature for Non-coherent Detection

To deeply explore the hidden nonlinear characteristic of UV signals, machine learning techniques are utilised to train the detector or feature construction process via the combinations of linear weights and bias of neurons and the nonlinear activation functions. To do so, the following two works will be elaborated on in this thesis.

(1) A supervised fully connected NN is used to map the received UV signals to their transmitted symbols. The neurons of the output layer of the trained NN present the nonlinear features for non-coherent detection, which outperforms the state-of-the-art MLSD in the low SNR region and gives a comparable BER (the gap is less than 1dB) in high SNR region.

(2) The supervised NN-based detector requires data overhead for training and re-training in the time-varying channel scenarios and is difficult to be explainable due to its black-box nature. This further makes the NN-based features in-explainable. To solve this problem, a physical-aware Parzen window-based NN detector is designed, which aims to approximate the likelihood PDF of the received signals, and thereby is easily explainable via mathematical deduction.

1.5 Organization of the Rest of this Thesis

In Chapter 2, there is a literature review of existing signal detection methods that combat noisy and ISI contaminated received signals. The channel estimation technique for coherent signal detection is introduced in detail in Section 2.1. Leveraging this, the minimum mean squared error criterion (MMSE) equalizer for symbol detection is introduced in Section 2.2. Then, the state-of-the-art MLSD method is provided in Section 2.3, which bases on my previous

paper [39]. The sub-optimal Bayesian sequential detection approach based on the particle filter is provided in Section 2.4, which relies on my previous paper [40].

Chapter 3 conducts an introduction to the communication system, based on which the non-coherent detection schemes are proposed in the later chapters. In section 3.1, the generalized system model for ultraviolet communication is illustrated in detail. Then, section 3.2 evaluates the effect of ISI on UV channel, in terms of varying the symbol rate and the different modulation methods (i.e., J-PAM).

In Chapter 4, 1D non-coherent detection schemes are proposed, and the relevant paper has been published, seen in [58]. In Section 4.1, the framework of the 1D non-coherent detection method is given, following which is a problem formulation. Section 4.2 focuses on the feature extraction of optical signals. There are four features obtained (i.e., the intra-symbol rising edge, the intra-symbol energy, the inter-symbol minimum inflection, and the inter-symbol energy difference of the signal), which are insensitive to the ISI contamination. Besides, the advantages of such feature-based metrics are evaluated. Section 4.3 introduces the EWNCD scheme which combines the metrics with equal weight and demonstrates the influence of the weights by simulation. Then, Section 4.4 elaborates on the OWNCD scheme, containing: i) the construction of the OWNCD design, ii) the theoretical analysis of the BER performance, iii) the optimal weight computation method, iv) the design of a heuristic algorithm. Afte that, Section 4.5 evaluates the EWNCD and OWNCD schemes, in terms of BER accuracy and computational complexity, by numerical simulation. Finally, a conclusion of 1D non-coherent detection methods will be given in Section 4.6.

In Chapter 5, a novel HD feature-based non-coherent detection is proposed, and the relevant paper has been published, seen in [59]. In section 5.1,

there is an illustration of the system model for the HD scheme along with the problem formulation. Then, the construction of HD UV features will be illustrated in section 5.2. Next, Section 5.4 will compute and compare the theoretical signal error ratio (SER) from this proposed HD non-coherent scheme with the 1D methods. In Section 5.5, the proposed HD non-coherent detection scheme and the algorithm design are elaborated. The simulation results are provided in Section 5.6. The chapter finally concludes in Section 5.7.

Chapter 6 explores the nonlinear characteristic of UV signals, resorting to machine learning techniques to train the detector or feature construction process. The process and results of the supervised neural network (NN) framework are provided in Section 6.1. Then, a half-supervised and explainable NN is provided in 6.2.

Chapter 7 conducts a summary of this thesis in Section 7.1, followed by the future work in 7.2. For the future work, I will try to utilize deep reinforcement learning method for extracting the nonlinear features from the received signals, based on which builds a reliable and explainable artificial intelligence (X-AI) framework for signal detection combating ISI and adversarial environment.

Chapter 2

Literature Review on Existing Signal Detection Methods for OWC

In this chapter, the literature on existing signal detection methods that combat noisy and ISI contaminated received signals will be introduced. Typically, a linear time-invariant (LTI) system-based ISI-contaminated signal can be expressed in terms of the CIR and the transmitted symbols. By denoting the transmitted symbols from discrete time steps $\tau = 1, 2, \cdots, k$ as $s_1, s_2, \cdots, s_k$, the received signal for current k time-step is:

$$\bar{r}_k = \bar{q}_0 \cdot s_k + \sum_{\tau=1}^{L} \bar{q}_\tau \cdot s_{k-\tau} + \epsilon_k. \tag{2.1}$$

In Eq. (2.1), $\bar{q}_\tau$ are the channel impulses of time step $k - \tau$ that influence the current k time step, which causes ISI effect on current signal $\bar{r}_k$. $L \in \mathbb{N}^+$ is the integer number of ISI components. ϵ_k are the noise effect at current k time step.

With this signal model, the aim of signal detection is to detect/estimate

the transmitted symbol s_k given the received signal r_k. By categorizing whether or not the channel impulses $\bar{q}_0, \bar{q}_1, \cdots, \bar{q}_{k-1}$ will be used, a signal detection method can be categorised as coherent and non-coherent families. From the conceptual point of view, coherent detection refers to the scheme to estimate the transmitted symbols in the awareness of the channel state information (or channel impulse response), while non-coherent detection is to estimate the transmitted symbols without CSI and CIR. These schemes will be considered respectively in the following.

2.1 Channel Estimation Technique for Coherent Signal Detection

The coherent signal detection method at first requires channel estimation techniques to estimate the CIR parameters, i.e., $\bar{q}_0, \bar{q}_1, \cdots, \bar{q}_L$. This can be done by sending $L' \in \mathbb{N}^+$ predefined pilot symbols $s_1, s_2, \cdots, s_{L'}$ known by the receiver, and computing the CIR parameters using the least square (LS) criterion [60–62].

2.1.1 Least Square Channel Estimator

An LS-based channel estimator resorts to linear algebra and the quadratic optimisation framework. To be specific, Eq. (2.1) may be rewritten in the

matrix formulation as follows:

$$\begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_{L+1} \\ \vdots \\ \bar{r}_{L'} \end{bmatrix} = \begin{bmatrix} s_1 & 0 & 0 & \cdots & 0 \\ s_2 & s_1 & 0 & \cdots & 0 \\ & & \vdots & & \\ s_{L+1} & s_L & s_{L-1} & \cdots & s_1 \\ & & \vdots & & \\ s_{L'} & s_{L'-1} & s_{L'-2} & \cdots & s_{L'-L} \end{bmatrix} \cdot \begin{bmatrix} \bar{q}_0 \\ \bar{q}_1 \\ \vdots \\ \bar{q}_L \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{L+1} \\ \vdots \\ \epsilon_{L'} \end{bmatrix} \tag{2.2}$$

$$\bar{\mathbf{r}}_{L'} = \mathbf{S} \cdot \mathbf{q} + \boldsymbol{\epsilon}_{L'}$$

where $\bar{\mathbf{r}}_{L'}$, $\mathbf{S}$, $\mathbf{q}$ and $\boldsymbol{\epsilon}_{L'}$ are to denote the corresponding parts in the matrix equation. Then, the channel estimation of $\mathbf{q}$ is pursued by the following unconstrained quadratic optimisation problem:

$$\min_{\mathbf{q}\in\mathbb{R}^{L+1}} f \triangleq \|\bar{\mathbf{r}}_{L'} - \mathbf{S}\cdot\mathbf{q}\|_2^2, \tag{2.3}$$

where $\|\cdot\|_2$ represents 2-norm.

Next, to solve Eq. (2.3), the first order derivative is computed and made zero, i.e.,

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{q}} =& \frac{\partial \left[\left(\bar{\mathbf{r}}_{L'}^T - \mathbf{q}^T\cdot\mathbf{S}^T\right)\cdot\left(\bar{\mathbf{r}}_{L'} - \mathbf{S}\cdot\mathbf{q}\right)\right]}{\partial \mathbf{q}} \\ =& \frac{\partial\left(\mathbf{q}^T\cdot\mathbf{S}^T\mathbf{S}\cdot\mathbf{q} - 2\mathbf{q}^T\cdot\mathbf{S}^T\bar{\mathbf{r}}_{L'}\right)}{\partial\mathbf{q}} \\ =& 2\cdot\left(\mathbf{S}^T\mathbf{S}\cdot\mathbf{q} - \mathbf{S}^T\cdot\bar{\mathbf{r}}_{L'}\right) \\ =& 0. \end{aligned} \tag{2.4}$$

By solving Eq. (2.4), the channel estimation result, denoted as $\hat{\mathbf{q}}$, can be computed as:

$$\hat{\mathbf{q}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T\cdot\bar{\mathbf{r}}_{L'}. \tag{2.5}$$

From Eqs. (2.4)-(2.5), the selection of L' should make the existence of the inverse matrix, i.e., $(\mathbf{S}^T\mathbf{S})^{-1}$ exists. This requires $L' \geq L+1$.

2.1.2 Simulations of LS Channel Estimator

After the description of the channel estimation technique, the performance of estimation accuracy versus both the SNR and the number of pilot symbols are provided in the following, respectively. In this simulation, On-Off Keying (OOK) modulation is considered, i.e., $s_k \in \{0, 1\}$, and the length of ISI is $L = 12$. The estimation accuracy is defined in terms of the normalised root mean square error (RMSE), i.e.,

$$RMSE_{LS} \triangleq \frac{\|\hat{\mathbf{q}} - \mathbf{q}\|_2}{\|\mathbf{q}\|_2}. \tag{2.6}$$

In Fig. 2.1, the x-coordinate represents the different values of SNR, and the y-coordinate is the normalised RMSE. It is first seen that with the increase of SNR, the channel estimation RMSE decreases, because of the reduction of the noise effect.

Then, it is observed that the length of pilot sequences L' affects drastically the channel estimation accuracy. For example, when SNR is 11dB, the RMSE of $L' = 2L$ is 0.39, which is larger than that of $L' = 10L$, i.e., 0.05, not to mention the RMSE= 0.01 at $L' = 100L$. This is because with the increase of the length of pilot sequences, i.e., L', the estimation error becomes less, as more measurements, $\bar{\mathbf{r}}_{L'} = [\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_{L'}]^T$, can be used to estimate the CIR $\bar{q}_0, \bar{q}_1, \cdots, \bar{q}_L$ (with fixed length $L+1$). This can be also explained in a mathematical manner. The numerator of the normalised RMSE of Eq. (2.6) can be further deduced and expended by taking the expression of $\bar{\mathbf{r}_{L'}}$ in Eq. (2.2). By defining the function $\lambda_{\min}(\cdot)$ as the minimum eigenvalues of a

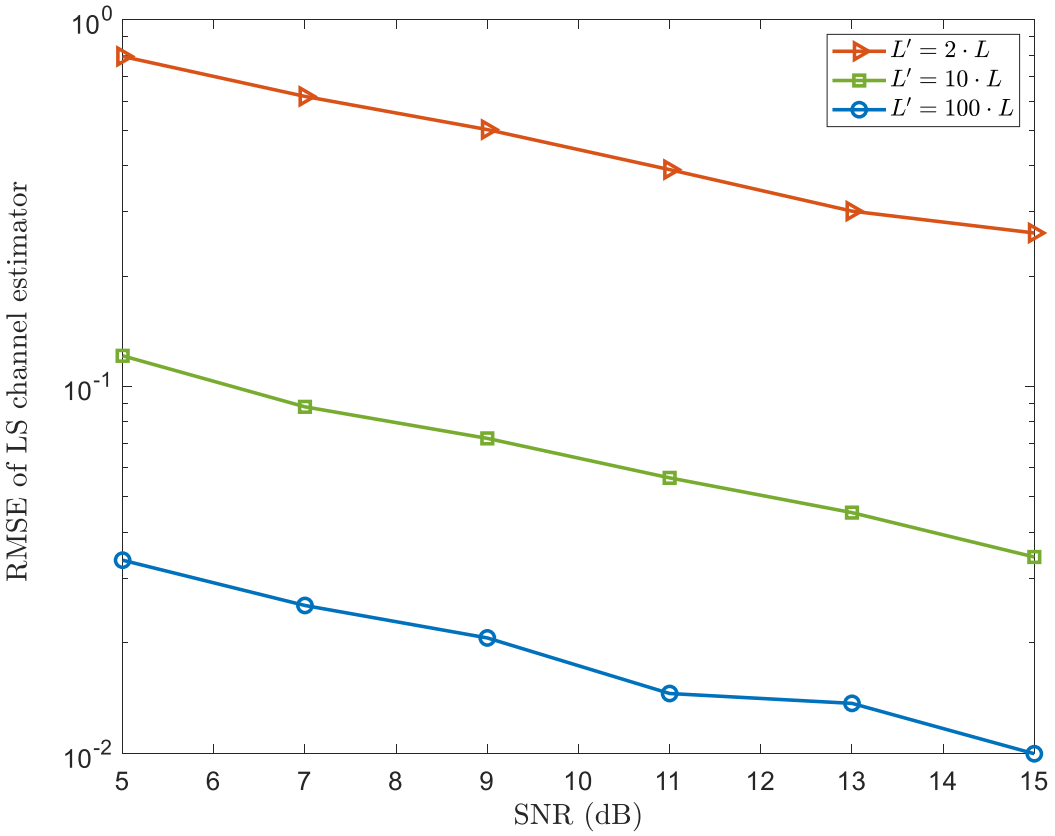


Figure 2.1: Performance of LS channel estimator: RMSE versus SNR and length of the pilot sequence L'.

matrix, the detailed deduction of $\|\hat{\mathbf{q}} - \mathbf{q}\|_2^2$ is provided in the following, i.e.,

$$\begin{aligned}
\|\hat{\mathbf{q}} - \mathbf{q}\|_2^2 &= \left\| \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T \cdot \bar{\mathbf{r}}_{L'} - \mathbf{q} \right\|_2^2 \\
&= \left\| \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T \cdot (\mathbf{S}\mathbf{q} + \boldsymbol{\epsilon}_{L'}) - \mathbf{q} \right\|_2^2 \\
&= \left\| \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T \cdot \boldsymbol{\epsilon}_{L'} \right\|_2^2 \\
&\leq \left\| \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T \right\|_2^2 \cdot \|\boldsymbol{\epsilon}_{L'}\|_2^2 \\
&\propto \frac{1}{\lambda_{\min}(\mathbf{S}^T\mathbf{S})}.
\end{aligned} \tag{2.7}$$

From Eq. (2.7), the LS estimation RMSE is directly dependent on the minimum eigenvalues of matrix $\mathbf{S}^T\mathbf{S}$. The increase of the length of the pilot

sequence L' makes $\mathbf{S}^T\mathbf{S}$ full rank with a small conditional number, which then decreases the upper bound of the estimation RMSE.

After the description of the channel estimation technique, the coherent signal detection schemes then can be pursued relying on the channel estimation results. In the following, the essential procedures of three popular coherent detection schemes will be introduced, which are the minimum mean square error-based linear equalizer, the state-of-the-art maximum likelihood sequence detection method, and the sequential Bayesian detection approach.

2.2 Linear Equalizer based on Minimum Mean Squared Error Criterion

A linear equalizer is an algorithm used in wireless communication systems to equalize the impulse response of the channel and mitigate the effect of ISI [42, 63]. To be specific, the received signals at first go through a finite impulse response (FIR) linear filter for CIR and ISI compensation, and then the signal detection is pursued relying on the compensated signals, which can lead to a more accurate estimation of the transmitted symbols. The goal of the linear equalizer is to recover the transmitted symbols by correcting the distortion caused by the channel, improving the BER performance of the communication system. Linear equalization can be achieved through various criteria such as zero-forcing (ZF), minimum mean square error, etc [64–66]. These methods differ in terms of the criteria they use to design the equalization filter and the computational complexity they require. In the following, the MMSE equalizer will be introduced in details.

2.2.1 MMSE equalizer

The MMSE equalizer is a linear equalization technique used in communication systems to mitigate ISI caused by channel distortion. The MMSE equalizer tries to recover the transmitted symbol of $(k-d)$th time-step, denoted as s_{k-d}, via the usages of the received signals corresponding to $(k-L)$th to kth transmitted symbols, i.e., $\bar{r}_{k-L}, \bar{r}_{k-L+1}, \cdots, \bar{r}_k$, i.e., [63]

$$\hat{s}_{k-d} = \sum_{n=0}^{L} \alpha_n \cdot \bar{r}_{k-n}, \tag{2.8}$$

where $\hat{s}_{k-d}$ is the recovered symbol of s_{k-d}. The setting of d depends on the length of ISI L, which is typically assigned as $d = round((L+1)/2)$ [42, 67].

The implementation of such MMSE equalizer in Eq. (2.8) is converted to compute the taps $\alpha_0, \alpha_1, \cdots, \alpha_L$, which relies on the minimisation of the mean square error between the recovered and the transmitted symbols, whereby the error is expressed as follows:

$$\varepsilon_{k-d} = s_{k-d} - \hat{s}_k. \tag{2.9}$$

With the help of the error expression in Eq. (2.9), the computation of the taps $\alpha_0, \alpha_1, \cdots, \alpha_L$ is to make $\mathbb{E}(\varepsilon_k \cdot \bar{r}_{k-m}) = 0$ for all $m = \{0, 1, \cdots, L\}$, where $\mathbb{E}(\cdot)$ denotes the expectation operator of a random variable, i.e.,

$$\begin{aligned} &\mathbb{E}\left(\varepsilon_{k-d} \cdot \bar{r}_{k-m}\right) = \mathbb{E}\left[\left(s_{k-d} - \hat{s}_{k-d}\right) \cdot \bar{r}_{k-m}\right] \\ =&\mathbb{E}\left[\left(s_{k-d} - \sum_{n=0}^{L} \alpha_n \cdot \bar{r}_{k-n}\right) \cdot \bar{r}_{k-m}\right] \\ =&\mathbb{E}\left(s_{k-d} \cdot \bar{r}_{k-m}\right) - \sum_{n=0}^{L} \alpha_n \mathbb{E}\left(\bar{r}_{k-n} \cdot \bar{r}_{k-m}\right) = 0. \end{aligned} \tag{2.10}$$

Then, $V_{m,n} \triangleq \mathbb{E}(\bar{r}_{k-n} \cdot \bar{r}_{k-m})$, and $g_m \triangleq \mathbb{E}(s_{k-d} \cdot \bar{r}_{k-m})$ are defined for a compact

notation. $V_{m,n}$ is computed as:

$$\begin{aligned} & V_{m,n} = \mathbb{E}(\bar{r}_{k-n} \cdot \bar{r}_{k-m}) \\ =& \mathbb{E}\left[\left(\sum_{\tau=0}^{L} \bar{q}_\tau \cdot s_{k-n-\tau} + \epsilon_{k-n}\right) \cdot \left(\sum_{\tau'=0}^{L} \bar{q}_{\tau'} \cdot s_{k-m-\tau'} + \epsilon_{k-m}\right)\right] \\ =& \sum_{\tau=0}^{L}\sum_{\tau'=0}^{L} \bar{q}_\tau \cdot \bar{q}_{\tau'} \cdot \mathbb{E}(s_{k-n-\tau} \cdot s_{k-m-\tau'}) + \sigma_\epsilon^2 \delta_{m,n} \\ =& \sum_{\tau=n}^{L+n}\sum_{\tau'=m}^{L+m} \bar{q}_{\tau-n} \cdot \bar{q}_{\tau'-m} \cdot \mathbb{E}(s_{k-\tau} \cdot s_{k-\tau'}) + \sigma_\epsilon^2 \delta_{m,n} \end{aligned} \tag{2.11}$$

where $\delta_{m,n}$ is the Kronecker delta function, i.e., $\delta_{m,n} = 1$ when $m = n$ and otherwise $\delta_{m,n} = 0$. σ_ϵ^2 is the variance of noise component. g_m is computed as:

$$\begin{aligned} & g_m = \mathbb{E}(s_{k-d} \cdot \bar{r}_{k-m}) \\ =& \mathbb{E}\left[s_{k-d} \cdot \left(\sum_{\tau=0}^{L} \bar{q}_\tau \cdot s_{k-m-\tau} + \epsilon_{k-m}\right)\right] \\ =& \sum_{\tau=0}^{L} \bar{q}_\tau \cdot \mathbb{E}(s_{k-d} \cdot s_{k-m-\tau}) \end{aligned} \tag{2.12}$$

For a specific J-PAM ($J \in \mathbb{N}^+$) transmitted symbols (e.g., OOK, 4-PAM) where $s_1, s_2, \cdots, s_k \in \{0, 1, \cdots, J-1\}$, $\mathbb{E}(s_m \cdot s_n)$ in Eqs. (2.11)-(2.12) can be computed as:

$$\mathbb{E}(s_m \cdot s_n) = \begin{cases} \dfrac{(J-1)\cdot(2J-1)}{6} & m = n, \\ \dfrac{(J-1)^2}{4} & m \neq n. \end{cases} \tag{2.13}$$

By taking Eq. (2.13) into Eqs. (2.11)-(2.12), the matrix $\mathbf{V} \triangleq [V_{m,n}]$ of size $(L+1) \times (L+1)$ and the vector $\mathbf{g} = [g_0, g_2, \cdots, g_L]^T$ of size $L+1$ are derived. As such, the MMSE equalizer is performed by re-written $\mathbb{E}(\varepsilon_k \cdot \bar{r}_{k-m}) = 0$ for

all $m = \{0, 1, \cdots, L\}$ in Eq. (2.10), i.e.,

$$\mathbf{g} = \mathbf{V} \cdot \boldsymbol{\alpha}, \tag{2.14}$$

where $\boldsymbol{\alpha} \triangleq [\alpha_0, \alpha_1, \cdots, \alpha_L]^T$ are the tap vector of the equalizer and can be computed as:

$$\boldsymbol{\alpha} = \mathbf{V}^{-1} \cdot \mathbf{g}. \tag{2.15}$$

Then, the recovered transmitted symbols can be recovered by taking the computed taps in vector $\boldsymbol{\alpha}$ back into Eq. (2.8).

2.2.2 Simulation Results of MMSE Equalizer

After the description of the MMSE equalizer implementation, its symbol recovery performance in terms of the BER will be evaluated, via different SNR and ISI intensities respectively. The existing MMSE signal detection method is evaluated in different modulation types (e.g., OOK, 4-PAM), with ideal CIRs. The reason that higher-order modulations are needed is because of their abilities to increase the bit rates. So, here, making different modulations at the same bit rates may confuse the readers, but making them with the same symbol rate is more reasonable (as is done in Fig. 2.2).

Firstly, the detection accuracy is illustrated under different SNR, in case of a fixed symbol transmission rate $R_s = 10^6\ bit/s$.

In Fig. 2.2, the horizontal axis respects SNR, and the vertical axis is BER. From left to right, the three lines indicate the BER performance via MMSE equalizer, in the cases of OOK, 4-PAM and 8-PAM modulation, respectively. It can be observed that a higher SNR can lead to a lower BER when the modulation mode is fixed. This phenomenon is reasonable, as a higher SNR means lesser noise contamination. Besides, if the SNR is a certain value, the OOK has the best BER performance, followed by that of the 4-PAM,

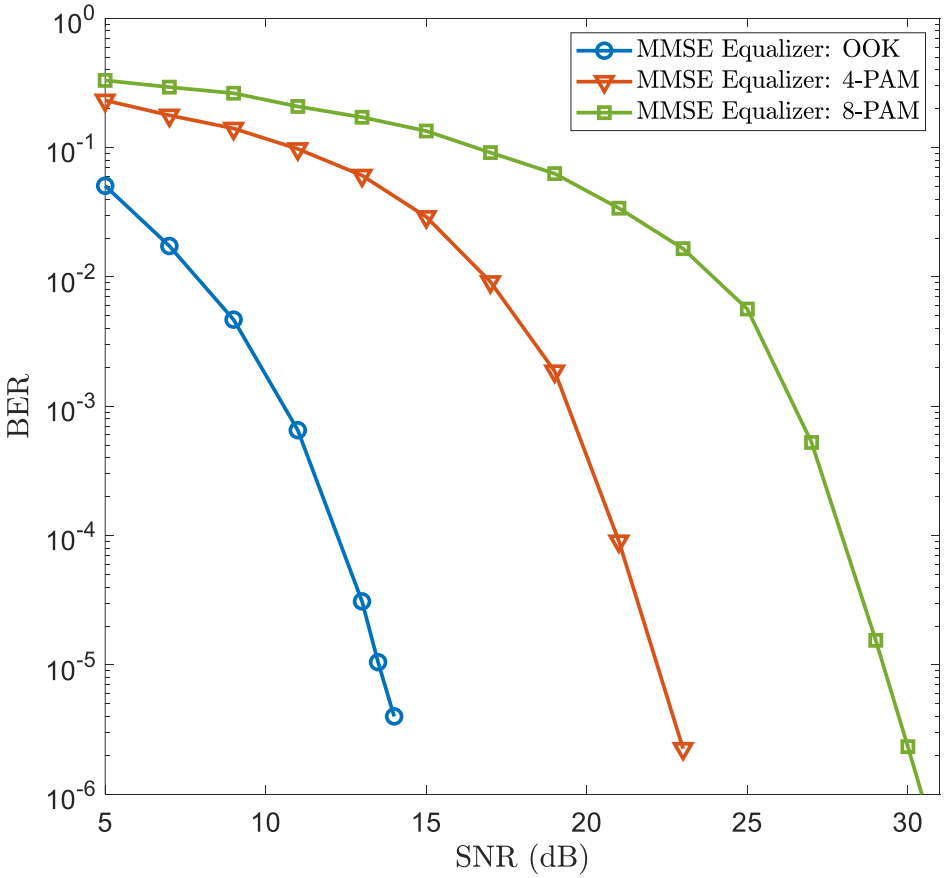


Figure 2.2: Performance of signal detection via MMSE equalizer, under different SNR, with symbol transmission rate as $R_s = 10^6\ bit/s$.

while 8-PAM conducts the worst BER performance. For example, when SNR is $20dB$, the BER of OOK is below 10^{-6}, while that of 4-PAM and 8-PAM are approximately $10^{-3.5}$ and 10^{-1}. It is because of multi-intensity modulation brings more ISI than OOK modulation, and the higher the multi-intensity, the stronger the ISI influence, consequently leading to worse BER performance.

Then, the ISI's influence on communication performance will be evaluated directly, by changing the symbol rate. To be specific, the BER performance of the MMSE equalizer is evaluated via different symbol rate R_s, under OOK, 4-PAM and 8-PAM modulation, as shown in Fig. 2.3.

The horizontal axis and the vertical axis indicate R_s and BER, in Fig.

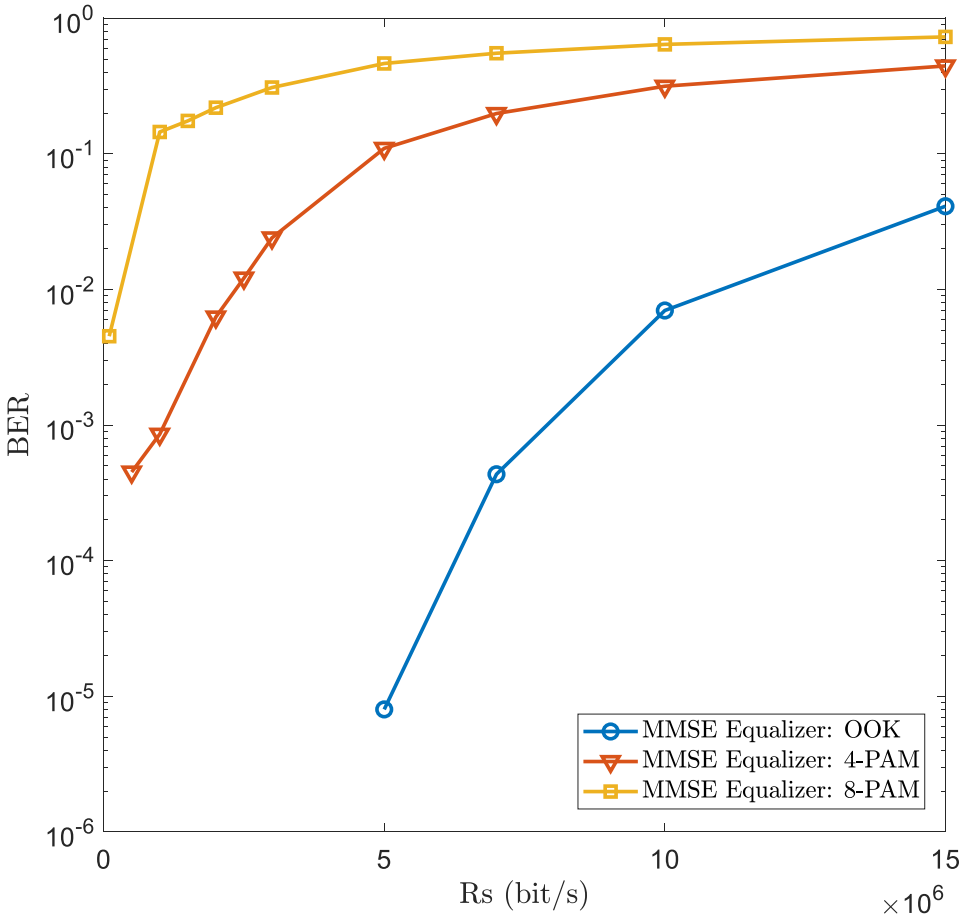


Figure 2.3: Performance of signal detection via MMSE equalizer, under different ISI intensity (in terms of symbol rate R_s), with SNR as 20dB.

2.3. The blue, red and yellow lines illustrate the BER performance of the MMSE equalizer, in cases of OOK, 4-PAM, and 8-PAM modulation, respectively. It can be observed that a higher R_s can lead to a higher BER, with a fixed modulation mode. The reason for it is the faster the symbol rate the stronger the ISI, resulting in more contamination on the received signals, and consequently worse BER performance. Besides, in terms of the same symbol rate, the BER performance of OOK is better than that of the 4-PAM, while 8-PAM conducts the worst BER performance. For instance, when R_s is $5 \times 10^6 bit/s$, the BER of OOK and 4-PAM are 10^{-6} and 10^{-1}, while that

of the 8-PAM are higher than them. It is because a higher multi-intensity modulation (i.e., using multiple levels of photon intensities to modulate bit streams, e.g., 4-PAM, and 8-PAM), gives rise to a stronger ISI, resulting in a worse BER performance.

Here, MMSE is not focused on in this thesis; the aim of these figures is to provide a comparison baseline for the non-coherent detection schemes elaborated in this thesis. In the following comparison between the proposed non-coherent detection and MMSE, the setting of the same modulation types is considered for the fair evaluation.

2.3 Maximum Likelihood Sequence Detection

As a coherent detection scheme, the prerequisite of MLSD is the awareness of CIR (or estimated CIR), i.e., $\bar{q}_0, \bar{q}_1, \cdots, \bar{q}_L$ in Eq. (2.1), which is required to construct the CSI and its transition probability matrix (TPM). Then, MLSD tries to traverses all paths of likelihood, and find the optimal symbol paths with maximum likelihood PDFs [1, 39].

2.3.1 Construction of CSI and TPM

The channel state for kth transmitted symbol s_k is defined as:

$$\mathbf{s}_k \triangleq [s_k, s_{k-1}, \cdots, s_{k-L}]^T, \tag{2.16}$$

which characterises the all the transmitted symbols that interferes with the current received signals, i.e., in the region of L ISI length.

The transitional TPM $\mathbf{P}$ captures the dynamic relationship between

$\mathbf{s}_{k-1}$ and $\mathbf{s}_k$. Define the state space of the CSI for J-PAM symbols as :

$$\begin{aligned} \mathcal{S} = \{\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \cdots, \tilde{\mathbf{s}}_S\} \triangleq \{0, 1, \cdots, J-1\}^{L+1}, \\ S \triangleq |\mathcal{S}| = J^{L+1} \end{aligned} \tag{2.17}$$

where $|\mathcal{S}|$ is the cardinality of the set $\mathcal{S}$, and therefore $\mathbf{s}_k \in \mathcal{S}$. Then TPM can be defined as the transitional mapping of probability values, indicating the likelihood of transitioning from one state in $\mathbb{S}$ to another state in $\mathbb{S}$. The TPM is crucial in the Bayesian scheme, as it helps update the understanding of CSI in each time step based on the previous observations and to make optimal decisions regarding sequential signal detection. TPM is defined as follows:

$$\mathbf{P} = \begin{bmatrix} P_{1\to1} & P_{1\to2} & \cdots & P_{1\to S} \\ P_{2\to1} & P_{2\to2} & \cdots & P_{2\to S} \\ \vdots & \vdots & \ddots & \vdots \\ P_{S\to1} & P_{S\to2} & \cdots & P_{S\to S} \end{bmatrix}, \tag{2.18}$$

where TPM element $P_{l\to j}$ represents the transitional probability from lth state to jth state of set $\mathcal{S}$, i.e., from $\tilde{\mathbf{s}}_l$ to $\tilde{\mathbf{s}}_j$. Based on Eq. (2.18), the transitional PDF of the CSI $\mathbf{s}_k$ from $\mathbf{s}_{k-1}$, denoted as $p(\mathbf{s}_k|\mathbf{s}_{k-1})$, can be derived. For example, if $\mathbf{s}_{k-1} = \tilde{\mathbf{s}}_i$ and $\mathbf{s}_k = \tilde{\mathbf{s}}_j$, then $p(\mathbf{s}_k|\mathbf{s}_{k-1}) = P_{i\to j}$. This transitional PDF of CSI provides a representation of the probability distribution of the CSI at time slot k given the UV CSI at time slot $k-1$. By utilizing this PDF in the Bayesian estimation process, the system can perform sequential detection of the transmitted symbols.

2.3.2 Likelihood PDF of each Received Signal

In the receiver, CSI at kth time-step $\mathbf{s}_k$ is measured as $\bar{r}_k$ with noise ϵ_k, as is shown in Eq. (2.1). Given the knowledge of the PDF of the noise ϵ_k (e.g., zero mean Gaussian distributed with variance σ_ϵ^2), the likelihood PDF of the received signal $\bar{r}_k$ conditioned on the CSI $\mathbf{s}_k$ can be expressed as:

$$p(\bar{r}_k|\mathbf{s}_k) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\bar{r}_k - \mathbf{q}^T \cdot \mathbf{s}_k\right)^2}{2\sigma_\epsilon^2}\right). \tag{2.19}$$

Here, the construction of the likelihood PDF uses the CIRs in vector $\mathbf{q} = [\bar{q}_0, \bar{q}_1, \cdots, \bar{q}_L]^T$, which is estimated by the channel estimation approach (e.g., LS in Section 2.1.1). The estimation accuracy therefore serves as the prerequisite for the further detection accuracy of transmitted symbols.

2.3.3 Theory of MLSD

The coherent MLSD is an algorithm maximum *a posteriori*. It is a widely used algorithm that maximizes the likelihood function by taking into account the prior knowledge of the CSI. It is applied to the sequential detection problem with ISI, where the goal is to determine the most possible symbol sequence that maximizes the likelihoods, given the received signals.

To be specific, a sequence of transmitted symbols of length $K \in \mathbb{N}^+$ is denoted as $s_1, s_2, \cdots, s_K$, with the corresponding received signal sequence as $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_K$. Given Eq. (2.16), the corresponding CSIs are created as $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K$. In this view, the overall posterior PDF can be defined as $p(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K|\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_K)$, and the desired symbol sequence in terms of the state sequence will be derived by maximizing the *posteriori*, i.e., [39]

$$[\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, ..., \hat{\mathbf{s}}_K] = \underset{\substack{\mathbf{s}_k \in \mathcal{S} \\ k \in \{1,2,\cdots,K\}}}{\operatorname{argmax}} \left\{p(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K|\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_K)\right\}. \tag{2.20}$$

where $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, ..., \hat{\mathbf{s}}_K$ are the desired CSI sequences, and whose first elements are the desired symbol sequence.

The implementation of Eq. (2.20) is via the Viterbi-like searching algorithm, which is to present the right-hand side of Eq. 2.20 in a recursive mode, i.e.,

$$\begin{aligned}
&\max_{\substack{\mathbf{s}_k \in \mathcal{S} \\ k \in \{1,2,\cdots,K\}}} p(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K | \bar{r}_1, \bar{r}_2, \cdots, \bar{r}_K) \\
=&\max_{\mathbf{s}_1 \in \mathcal{S}} p(\mathbf{s}_1) p(\bar{r}_1|\mathbf{s}_1) \cdot \max_{\mathbf{s}_2 \in \mathcal{S}} p(\mathbf{s}_2|\mathbf{s}_1) p(\bar{r}_2|\mathbf{s}_2) \cdot \cdots \cdot \max_{\mathbf{s}_K \in \mathcal{S}} p(\mathbf{s}_K|\mathbf{s}_{K-1}) p(\bar{r}_K|\mathbf{s}_K),
\end{aligned} \tag{2.21}$$

From Eq. (2.21), the MLSD can be pursued in an iterative manner, by iteratively maximizing $p(\mathbf{s}_{k+1}, \mathbf{s}_{k+2}, \cdots, \mathbf{s}_K, \bar{r}_k, \bar{r}_{k+1}, \cdots, \bar{r}_K | \mathbf{s}_k)$, i.e.,

$$\begin{aligned}
\Theta_k(\mathbf{s}_k) &\triangleq \max_{\mathbf{s}_{k+1}, \cdots, \mathbf{s}_K \in \mathcal{S}} \ln\left(p(\mathbf{s}_{k+1}, \mathbf{s}_{k+2}, \cdots, \mathbf{s}_K, \bar{r}_k, \bar{r}_{k+1}, \cdots, \bar{r}_K | \mathbf{s}_k)\right) \\
&= \ln\left(p(\bar{r}_k|\mathbf{s}_k)\right) + \max_{\mathbf{s}_{k+1}, \cdots, \mathbf{s}_K \in \mathcal{S}} \left\{\ln\left(p(\mathbf{s}_{k+1}|\mathbf{s}_k)\right) + \Theta_{k+1}(\mathbf{s}_{k+1})\right\},
\end{aligned} \tag{2.22}$$

with initial $\Theta_K(\mathbf{s}_K) = \ln\left(p(\bar{r}_K|\mathbf{s}_K)\right)$.

For each value of $\mathbf{s}_k$, the best transition from $\mathbf{s}_k$ to $\mathbf{s}_{k+1}$ that maximizes $\ln(p(\mathbf{s}_{k+1}|\mathbf{s}_k)) + \Theta_{k+1}(\mathbf{s}_{k+1})$ is stored in memory. Once the iteration ends at $k = 1$, $\hat{\mathbf{s}}_1$ can be determined by finding the maximum value of $\Theta_1(\mathbf{s}_1)$ among all possible values in set $\mathcal{S}$. Then, $\mathbf{s}_2$ can be identified given the stored best trajectory and so does $\hat{\mathbf{s}}_k, 2 \leq k \leq K$ in a recursive mode. The estimated CSI sequence, i.e., $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, ..., \hat{\mathbf{s}}_K$, can then be used to obtain the corresponding transmitted symbol sequence, via their first value, i.e., $\hat{s}_1 = \hat{\mathbf{s}}_1(1), \hat{s}_2 = \hat{\mathbf{s}}_2(1), \cdots, \hat{s}_K = \hat{\mathbf{s}}_K(1)$.

It is noteworthy that, accurate estimation of the CIR is vital for the detection accuracy of the coherent MLSD, as it serves as the basis to construct the likelihood PDF in Eq. (2.19). This process will give rise to large consumption of computation and storage resources, especially in the case of

time-varying environment [68].

2.3.4 Simulation Result of MLSD

After the description of the theory and the implementation of MLSD in ISI sequential symbol detection, its detection accuracy in terms of BER is illustrated here, versus the values of SNR, under different accuracy of channel estimation results.

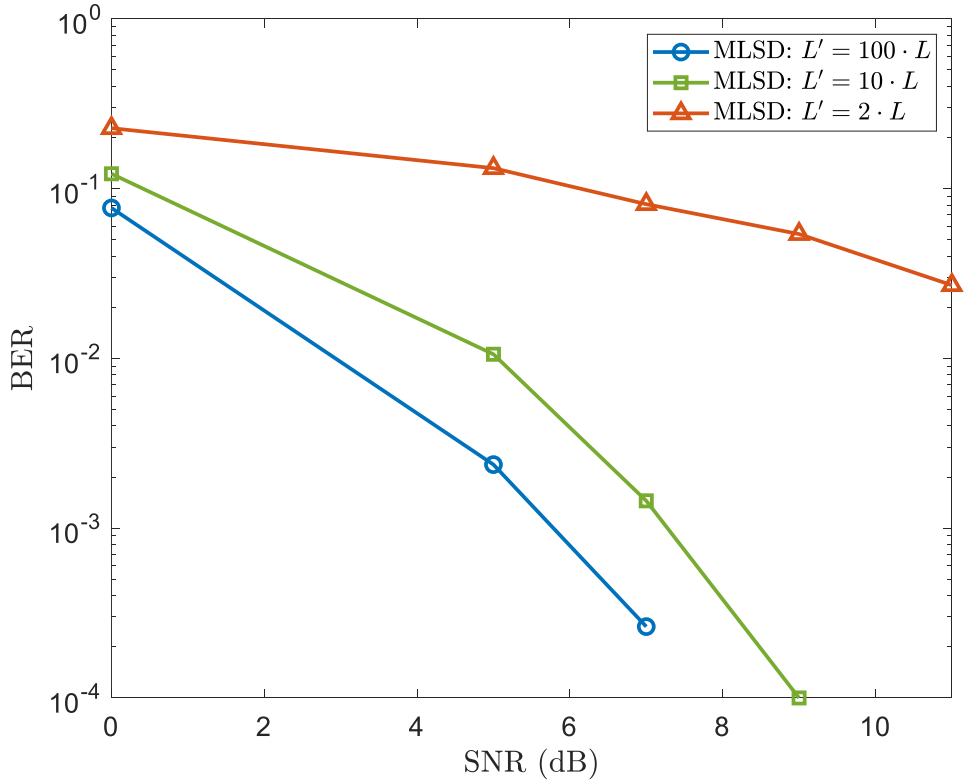


Figure 2.4: Performance of signal detection via MLSD scheme, under different SNR, with transmission rate as $R_s = 10^6\ bit/s$.

In Fig. 2.4, x-coordinate represents the values of SNR, and y-coordinate is the BER. Here, different L' represents the length of pilot sequence used for LS channel estimation, which is expressed in terms of the length of ISI L. As is shown in Fig. 2.4, the BER of MLSD goes down with the increase of

the SNR, showing that the noise reduction betters the detection performance. Then, it is observed that the detection accuracy of MLSD depends heavily on the accuracy of the channel estimation results. For example, with the same SNR as 5dB, the BERs of MLSD with $L' = 2L$ and with $L' = 10L$ lengths for channel estimation are greatly different (a BER gap from 0.2 to 10^{-2} is shown). This is due to the fact that the estimated CIR directly contributes to the likelihood computations in MLSD scheme, shown in Eq. (2.19), which, if biases, will inevitably deteriorate the further symbol detection results. This process will give rise to large consumption of computation and storage resources, especially in the case of time-varying environment [68], which impairs its applicable usages in real-time UV communications.

2.4 Sub-optimal Bayesian Sequential Detection

In practice, MLSD has some implementation issues. The major one is when dealing with $(J > 2)$-PAM symbols, e.g., 4-PAM in UV communications, the cardinality of the state space can be as huge as $|\mathcal{S}| = 4^{12} = 1.6 \times 10^7$, which is tremendous for Vieterbi trajectory searching in Eq. (2.21), i.e., finding maximal value of likelihood PDF for one iteration requires the searching in 4^{12} space. To cope with this huge computational and storaging challenge, an sub-optimal Bayesian sequential detection method is proposed in [40].

2.4.1 Theory of Bayesian Sequential Detection

In Bayesian detection framework in [40,69], the aim is to sequentially estimate the transmitted symbol detection using the current and previous received signals. This is different from the MLSD which requires the whole search of all possible sequential trajectories.

For kth time step, given the transitional PDF of the CSI as $p(\mathbf{s}_k|\mathbf{s}_{k-1})$,

its likelihood density as $p(\bar{r}_k|\mathbf{s}_k)$, and the received signals as $\bar{\mathbf{r}}_k \triangleq [\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k]^T$, a Bayesian framework is provided to estimate the CSI, i.e.,

$$p(\mathbf{s}_k|\bar{\mathbf{r}}_{k-1}) = \int p(\mathbf{s}_k|\mathbf{s}_{k-1}) \cdot p(\mathbf{s}_{k-1}|\bar{\mathbf{r}}_{k-1}) \cdot \mathrm{d}(\mathbf{s}_{k-1}), \tag{2.23}$$

$$p(\mathbf{s}_k|\bar{\mathbf{r}}_k) = \frac{p(\bar{r}_k|\mathbf{s}_k) \cdot p(\mathbf{s}_k|\bar{\mathbf{r}}_{k-1})}{\int p(\bar{r}_k|\mathbf{s}_k) \cdot p(\mathbf{s}_k|\bar{\mathbf{r}}_{k-1}) \cdot \mathrm{d}(\mathbf{s}_k)}, \tag{2.24}$$

where Eq. (2.23) refers to as the predict stage, and Eq. (2.24) accounts for the update stage. In predict stage, the predicted density $p(\mathbf{s}_k|\mathbf{z}_{1:k-1})$ is constructed by evolving the previously computed posterior density $p(\mathbf{s}_{k-1}|\bar{\mathbf{r}}_{k-1})$ via the transitional PDF $p(\mathbf{s}_k|\mathbf{s}_{k-1})$. Then, in the update stage, the posterior PDF, $p(\mathbf{s}_k|\bar{\mathbf{r}}_k)$ is derived by updating the predicted PDF through the likelihood PDF. In Eqs. (2.23)-(2.24), the transitional PDF $p(\mathbf{s}_k|\mathbf{s}_{k-1})$, and the likelihood PDF $p(\bar{r}_k|\mathbf{s}_k)$ are as same as those of MLSD.

After the computation of the posterior PDF $p(\mathbf{s}_k|\bar{\mathbf{r}}_k)$, the estimated CSI, denoted as $\hat{\mathbf{s}}_k$ can be obtained by maximizing the *posteriori*, i.e.,

$$\hat{\mathbf{s}}_k = \underset{\mathbf{s}_k}{\operatorname{argmax}} \{p(\mathbf{s}_k|\bar{\mathbf{r}}_k)\} \tag{2.25}$$

Then, the current kth state, i.e., $\hat{\mathbf{s}}_k$ is derived.

2.4.2 Particle Filter Realisation of Bayesian Framework

It is noteworthy that the computations of the predicted PDF $p(\mathbf{s}_k|\bar{\mathbf{r}}_{1:k-1})$ in Eq. (2.23) requires the complicated integral operation, which is infeasible in real implementation. To cope with that, a particle filter based algorithm is deployed in [40] to realise the two stages of the Bayesian approach. In essence, particle filter relies on a bundle of random and discrete particles, denoted as $\mathbf{s}^{(i)}, i = 1, 2, ..., I$, with evolving probability weights, denoted as $\bar{w}^{(i)}$, to

approximate the evolution process of complex PDFs, i.e., [70, 71]

$$p(\mathbf{s}) \backsimeq \sum_{i=1}^{I} \delta\left(\mathbf{s} - \mathbf{s}^{(i)}\right) \cdot \bar{w}^{(i)} \tag{2.26}$$

where $\delta(\cdot)$ is the delta function with $\delta(0) = 1$ and otherwise 0.

For sequential detection/estimation framework with ISI, the particles for kth time step are assigned according to the constructed CSI in Eq. (2.16), i.e.,

$$\mathbf{s}_k^{(i)} = \left[s_k^{(i)}, s_{k-1}^{(i)}, \cdots, s_{k-L}^{(i)}\right]^T, \ i \in \{1, 2, \cdots, I\}, \tag{2.27}$$

which will be then propagated to approximate the predict and update stages in Eqs. (2.23)-(2.24).

Particle Filter for Predict Stage

At the beginning of the kth time step, the initial particles are drawn from the posterior PDF at $(k-1)$th time step, i.e.,

$$\mathbf{s}_{k-1|k-1}^{(i)} \sim p\left(\mathbf{s}_{k-1}^{(i)} | \bar{\mathbf{r}}_{k-1}\right), \tag{2.28}$$

whose corresponding weight $\bar{w}_{k-1|k-1}^{(i)}$ satisfies

$$p(\mathbf{s}_{k-1}|\bar{\mathbf{r}}_{k-1}) \backsimeq \sum_{i=1}^{I} \bar{w}_{k-1|k-1}^{(i)} \cdot \delta\left(\mathbf{s}_{k-1} - \mathbf{s}_{k-1|k-1}^{(i)}\right). \tag{2.29}$$

With the initial particles and their weights, the predict stage can be pursued by simulating the transitional evolution of particles at $(k-1)$th time step, i.e.,

$$\begin{aligned} \mathbf{s}_{k|k-1}^{(i)} = &\left[s_k^{(i)} = j, s_{k-1}^{(i)}, \cdots, s_{k-L}^{(i)}\right]^T, \\ &j \in \{0, 1, \cdots, J-1\} \text{ with identical probability} \frac{1}{J} \end{aligned} \tag{2.30}$$

with their weights unchanged, i.e.,

$$\bar{w}_{k|k-1}^{(i)} = \bar{w}_{k-1|k-1}^{(i)}. \tag{2.31}$$

Particle Filter for Update Stage

After the predicted particles and their weights are acquired, i.e., $\left\{\mathbf{s}_{k|k-1}^{(i)}, \bar{w}_{k|k-1}^{(i)}\right\}_{i=1}^{I}$, the update stage will be performed by a) remaining the particles, i.e.,

$$\mathbf{s}_{k|k}^{(i)} = \mathbf{s}_{k|k-1}^{(i)}, \tag{2.32}$$

and b) updating the weights, i.e.,

$$\bar{w}_{k|k}^{(i)} = \frac{p(\bar{r}_k|\mathbf{s}_{k|k}^{(i)}) \cdot \bar{w}_{k|k-1}^{(i)}}{\sum_{i=1}^{I} p(\bar{r}_k|\mathbf{s}_{k|k}^{(i)}) \cdot \bar{w}_{k|k-1}^{(i)}}, \tag{2.33}$$

where the likelihood PDF $p(\bar{r}_k|\mathbf{s}_{k|k}^{(i)})$ is computed by replacing $\mathbf{s}_k$ with $\mathbf{s}_{k|k}^{(i)}$ in Eq. (2.19).

Following the acquisition of the posterior PDF and the estimation results, a re-sample process should be pursued in order to eliminate particles with negligible weights and reproduce those whose weights are of importance [70, 72, 73]. Finally, as the complete of the re-sample process, all particles remaining are of identical importance and should be assigned with identical weights, which will be used for the next round, i.e.$(k+1)$th time-slot, particle filter algorithm.

Algorithm flow of the UV channel-related Bayesian scheme

To make the implementation process clear and easily to follows, a schematic flow of the particle filter based Bayesian jointly channel estimation and symbol detection algorithm is provided in Algorithm 1.

Algorithm 1 Particle Filter scheme at kth time step

Input: Received signal of kth transmitted symbol, i.e., $\bar{r}_k$, particles and weights of previous $(k-1)th$ time step, i.e., $\{\mathbf{s}_{k-1|k-1}^{(i)}, \bar{w}_{k-1|k-1}^{(i)}\}_{i=1}^{I}$.

1: **for** $i \in \{1, \cdots, I\}$ **do**
2: Draw predicted particles $\mathbf{s}_{k|k-1}^{(i)}$ via Eq. (2.30).
3: Remain weight $\bar{w}_{k|k-1}^{(i)} = \bar{w}_{k-1|k-1}^{(i)}$.
4: Remain updated particles $\mathbf{s}_{k|k}^{(i)} = \mathbf{s}_{k|k-1}^{(i)}$.
5: Update weight $\bar{w}_{k|k}^{(i)}$ via Eq. (2.33).
6: **end for**
7: Pursue re-sampling procedure via [70, 72, 73].
8: Normalise the weights as $\bar{w}_{k|k}^{(i)} = 1/I$.

Output: Return the particles and the weights $\{\mathbf{s}_{k|k}^{(i)}, \bar{w}_{k|k}^{(i)}\}_{i=1}^{I}$ for next $(k+1)$th round.

At the beginning of the kth time step, the inputs are the received signal of kth transmitted symbol, and the particles and their corresponding weights of the $(k-1)$th time step. For iterating all the particles in line 1-6, line 2-3 are to obtain the predicted particles and their weights. Line 4-5 are to derive the updated particles and the corresponding weights. Line 7 provides the re-sample process to reproduce the particles with relatively large weights, and eliminate those whose weights are negligible. Line 8 finally normalises the weights as $\bar{w}_{k|k}^{(i)} = 1/I$. The outputs are the particles and their weights which will be used for the next-round Bayesian scheme.

2.4.3 Simulation Results of Bayesian Sequential Detection

In this section, the performance of the Bayesian sequential detection scheme leveraging the particle filter is provided.

Detection Accuracy

In Fig. 2.5, the detection BER versus different values of SNR is illustrated, where x-coordinate represents the SNR, and the y-coordinate is BER. From Fig. 2.5, it is at first seen that with the increase of the values of SNR, the detection BERs become lower, showing better detection performance in the sense of a smaller noise.

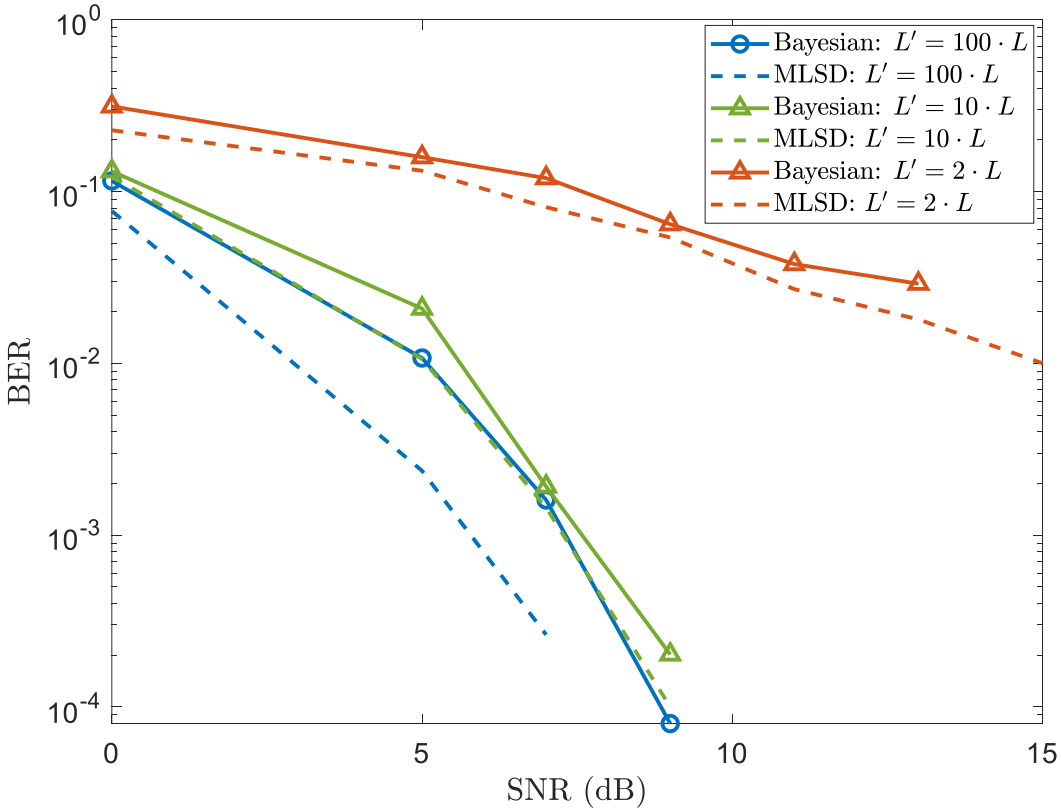


Figure 2.5: Performance of Bayesian sequential detection scheme: BER versus SNR.

Then, it is observed that the performance of Bayesian sequential detection also heavily depends on the accuracy of channel estimation results. For example, at a given SNR as 9 dB, the BER with the length of pilot sequence as $L' = 100 \cdot L$ reaches 10^{-4}, which is much lower than that with $L' = 2 \cdot L$, i.e., BER= 10^{-1}. This is because that the estimated CIR is directly used in

the construction of the likelihood PDF, whose accuracy inevitability affects the detection accuracy of the transmitted symbols.

Next, it is noticed that the Bayesian scheme can be treated as a sub-optimal solution to solve the maximization problem of the sequential likelihood PDF. As is shown in Fig. 2.5, the BERs of the Bayesian detection scheme with specific pilot sequences are closer to their counterpart MLSD scheme. For example, when the pilot sequence with $L' = 10L$ is considered, the BER gap between MLSD and Bayesian schemes is less than 2dB, which suggests the potential of Bayesian sequential detection scheme as a sub-optimal solution of the complex MLSD estimation process.

Complexity analysis

The computational complexity of the Bayesian sequential detection is further evaluated in comparison to that of MLSD.

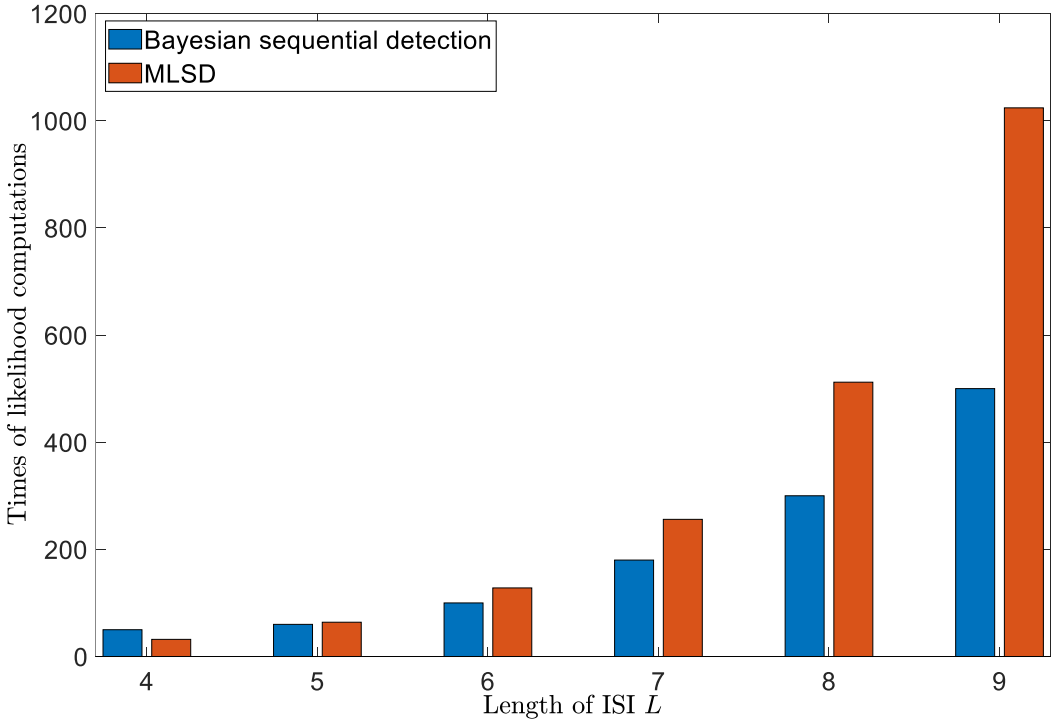


Figure 2.6: Complexity of proposed scheme as opposed to imperfect MLSD.

In Fig. 2.6, x-coordinate represents the length of ISI L, and the y-coordinate represents the number of computations of likelihood PDF. In general, the complexity of the posteriori based schemes (e.g., MLSD, Bayesian) can be roughly measured by the total times of computations of the likelihood PDFs, each of which has been given an unit complexity as ϑ [74]. For the Bayesian sequential detection scheme in this section, the times of likelihood computations are proportional to the number of particles used in the particle filter algorithm, i.e., I. In this view, the total computational complexity in terms of the likelihood computations can be expressed as $I \cdot \vartheta$ multiplying the number of transmitted symbols K, which is $I \cdot K \cdot \vartheta$ By contrast, the total times of likelihood computations in the MLSD scheme are actually the cardinality of the state-space $S = J^L$, multiplying the number of transmitted symbols K, which is $J^L \cdot K \cdot \vartheta$.

In this simulation, the number of particles for Bayesian sequential detection is $I = 100$, and the length of ISI $L = 12$ with OOK modulations, i.e., $J = 2$. In this case, a huge gap of computational complexity can be computed as $I = 100$ versus $J^L = 4096$ between Bayesian sequential detection scheme and the MLSD scheme. This gap is increased drastically when using $J > 2$-PAM modulations (e.g., 4-PAM and 8-PAM in further chapters). To make a practical comparison with MLSD in the following chapter, the sub-optimal Bayesian sequential detection is used to avoid the searching of $4^{12} = 1.6 \times 10^7$ or $8^{12} = 6.8 \times 10^{10}$ state-space for each time-step k in total K.

2.5 Conclusions & Discussions

In this chapter, the existing literature on channel estimation and signal detection approaches for ISI contaminated signals has been introduced with detailed theories, implementation procedures, and simulation results. The state-of-the-

art MLSD with perfect channel estimation results gives the best symbol detection accuracy, yet at the expense of tremendous computational and storage complexity, rendering its infeasibility for $J > 2$-PAM transmitted symbols. Its sub-optimal Bayesian joint channel estimation and signal detection method, sacrifices the a little amount of accuracy but makes it practical for detecting general PAM symbols in the sequential ISI detection framework. However, the challenges still lie in time, complexity, and resources consumption in sequential detection procedures, not to mention the overhead cost and the inaccuracy channel estimation caused by the channel estimation schemes, which further impairs their usages in real-time and low-cost signal detection with ISI. This therefore motivates the design of non-coherent detection schemes, which do not require the channel estimation and which is able to convert the sequential ISI signal detection problem into a discrete binary/multiple detection framework.

Chapter 3

Ultraviolet Signal Model & System

This chapter conducts an introduction to the ultraviolet communication system, based on which the non-coherent detection schemes are proposed in the later chapters. In section 3.1 the generalised system model for ultraviolet communication is illustrated in detail. Then, section 3.2 evaluates the effect of ISI on the UV channel, in terms of varying the symbol rate and the different modulation methods (i.e., J-PAM).

3.1 System Model

The generalised system model for ultraviolet communications is shown in Fig. 3.1, which consists of a UV transmitting module, a UV channel, and a UV detector. These are elaborated on in the following.

3.1.1 UV Transmitting Module

The UV transmitting module aims at modulating and transmitting the input information into the UV channel. In this work, the input information is the

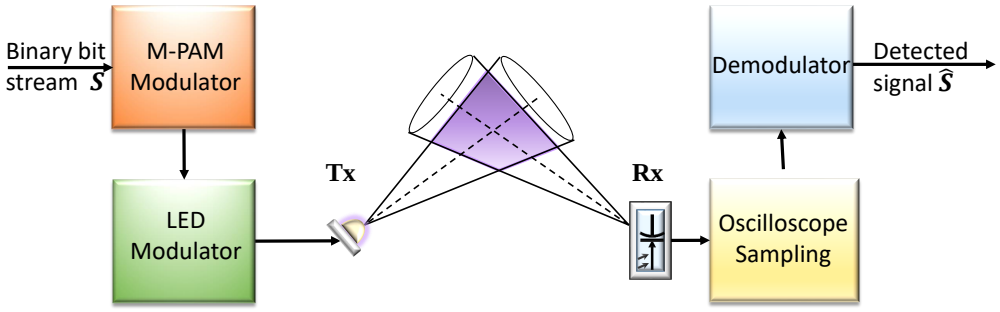


Figure 3.1: Illustration of the basic UVC system with a LED transmitter, dynamic UV channel, and a UV photon-counting receiver.

binary signal string and will be modulated by the amplitude modulator (i.e., J-PAM). By denoting the binary information as $s_k = j \in \{0, J-1\}$ ($k = 1, \cdots, +\infty$), the modulated signal $s(t)$ is expressed as:

$$s(t) = \sum_{k=1}^{\infty} s_k \cdot \delta\left(t - kT_b\right), \tag{3.1}$$

where T_b is the symbol interval, and $\delta(\cdot)$ represents the Dirac function. Then, the modulated signal in Eq. (3.1) will be transmitted by the DC-supplied ultraviolet LED transmitter, which transforms the signal into the intensities of the ultraviolet photons and emits such photons into the UV channel. After that, the photons are detected by the photo-multiplier tubes, which is photovoltaic conversion. Such detected electric signals are sampled by an oscilloscope. Finally, digital signal processing is conducted, to recover the transmitted signals. Such schematic flow of the UV transmitting module is illustrated in Fig. 3.1.

3.1.2 UV Channel

The UV CIR, denoted by $h(t)$, is governed by the transmission distance (i.e., r), the transmitter's beam divergence θ_T and apex angle β_T, the receiver's half-field of view θ_R and apex angle β_R, and the atmospheric based channel parameters [1,40]. It is noted that the exact CIR formulas are **not** required for

the proposed non-coherent schemes, but only the transient shape of the CIR. For the simulation part, a Monte Carlo process in [29, 30] to obtain different CIRs $h(t)$ with different channel parameters.

An illustration of two CIRs is provided in Fig. 3.2. Note that although the exact shapes of the CIRs may change drastically with their channel parameters, the transient features of such CIRs are in common. For example, there is an obvious rising edge for the appearance of a 1-bit, the inflections, and the signal difference between two adjacent bits. These give inspiration for designing the non-coherent detection scheme in the receiver, by exploiting such inherent features of the UV signal.

3.1.3 UV Receiver Module

For this work, the UV receiver contains four components: (i) the Rx that receives the emitted photons, (ii) the photodetector (e.g., photo-multiplier tubes, PMT [21]) that aims to detect the arrived photons based on photoelectric-digital conversion, (iii) the oscilloscope sampling module for signal sampling, and (iv) the non-coherent signal detection unit. It is worth noting that the photo-multiplier tubes play a crucial role in ensuring high sensitivity and signal-to-noise ratio in the UV receiver. The oscilloscope sampling module, on the other hand, enables the accurate measurement of signal characteristics. The non-coherent signal detection unit, in combination with the other components, allows for effective detection and estimation of the UV signals. The non-coherent detection unit is the critical part that will be focused on. All of these components work together to achieve the desired performance in the UV receiver.

The received signal is expressed in the form of the number of received photons. To be specific, it is assumed the emitted symbol interval as T_b, and the sampling rate for the oscilloscope sampling unit as $T_s = T_b/M$. Then, for

any kth symbol s_k, the UV receiver will evenly collect M samples, of which any ith sample, denoted as r_i, $i \in \{(k-1)M+1, \cdots, kM\}$ is the number of received photons within time $((i-1)T_s, iT_s]$. Given the studies in [1, 29, 40], r_i follows a Poisson distribution and can be expressed as:

$$r_i \sim \mathcal{P}\left(\sum_{l=k-L}^{k} N_T \cdot q_{i-(l-1)M} \cdot s_l + \bar{\epsilon}\right). \tag{3.2}$$

In Eq. (3.2), $\mathcal{P}(\cdot)$ denotes the Poisson distribution. N_T is the number of emitted photons for 1-bit. L is the length of ISI, indicating the influence of the previous L symbols on the currently received signal. $\bar{\epsilon}$ is the expectation of the number of received background noise photons. $q_{i-(l-1)M}$ is the probability of receiving one photon emitted at lth symbol interval, and can be expressed as [40]:

$$q_{i-(l-1)M} = \int_{(i-(l-1)M-1)T_s}^{(i-(l-1)M)T_s} h(t)dt. \tag{3.3}$$

Further, it is assumed that the background noise dominates (i.e., $\bar{\epsilon} \gg \sum_{l=k-L}^{k} N_T \cdot q_{i-(l-1)M} \cdot s_l$), such that the variance of r_i is approximately $\bar{\epsilon}$. In this view, by re-writing Eq. (3.2) as

$$r_i = \underbrace{N_T \cdot q_{i-(k-1)M} \cdot s_k}_{\text{current signal}} + \underbrace{\sum_{l=k-L}^{k-1} N_T \cdot q_{i-(l-1)M} \cdot s_l}_{\text{ISI}} + \underbrace{\epsilon_i}_{\text{noise}}, \tag{3.4}$$

an independently and identically distribution (i.i.d.) of r_i with a constant noise variance is offered.

3.2 Simulations of UV Channel

Here, the Monte Carlo method is adopted to simulate the UV channel [29, 30], of which the parameter assignment is listed in Table 3.1. As is shown in Table

Table 3.1: Simulation Parameters

Parameters	**Configurations**
Wavelength range	$\lambda \in [190nm, 280nm]$
Number of Emitted Photons for Each '1'-bit	$N_T = 10^{10}$
Transmission Distance between Tx and Rx	$r = 250m$
Transmitter's Beam Divergence	$\theta_T = \pi/12$
Transmitter's Apex Angle	$\beta_T = \pi/4$
Receiver's Half-field of view	$\theta_R = \pi/12$
Receiver's Apex Angle	$\beta_R = \pi/4$
Absorption Coefficient	$k_a = 5 \times 10^{-4} m^{-1}$
Scattering Coefficient	$k_s = 4.9 \times 10^{-4} m^{-1}$
Number of Sampling	$M = 20$
Range of transmitted symbol interval	$T_b \in [10^6, 5 \times 10^6]$bit/s
Range of SNR	5dB to 20dB

3.1, the number of emitted photons for a single symbol-unit (i.e., $s_k = 1$) is $N_T = 10^{10}$, over a transmission distance between Tx and Rx of $r = 250$m. The Tx beam divergence and apex angle are set to be $\theta_T = \pi/12$ and $\beta_T = \pi/4$, and the Rx half-field of view and apex angle are $\theta_R = \pi/12$ and $\beta_R = \pi/4$. The atmospheric coefficients are $\kappa_a = 5 \times 10^{-4} m^{-1}$ and $\kappa_s = 4.9 \times 10^{-4} m^{-1}$. For simulating different ISI intensities, the range of the symbol interval is $T_b \in [10^6,\ 5 \times 10^6]$ (with units of bit/s). The SNR is assigned to range from 5dB to 25dB to illustrate the BER versus different levels of ambient noise.

In this section, the ISI in the transmission will be demonstrated, in terms of varying the symbol rate and the different modulation methods (i.e., J-PAM), both of which have an interaction effect on the received signals.

Firstly, the ISI effect on the transmission signals will be illustrated on the OOK modulation, with a given transmitted interval at $10^{-6}s$ and a fixed SNR at 8dB, as seen in Fig. 3.2.

It can be observed the sending signals, the CIR of the UVC channel and the received signals are respectively shown in Fig. 3.2 (a)-(c). For Fig. 3.2 (a), the horizon axis is time with a unit 's', and the vertical axis is defined

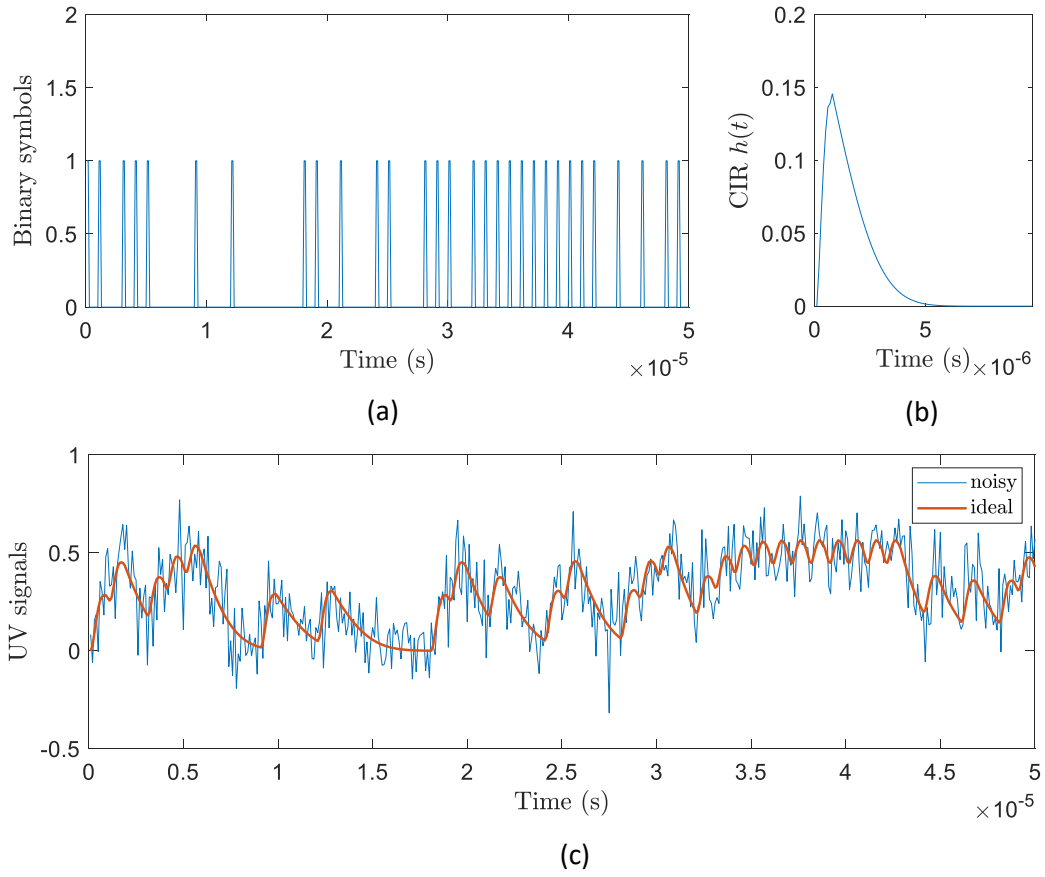


Figure 3.2: Illustrations of binary (OOK) symbols that will be transmitted, the CIR of UVC system, and the received signals with SNR=8dB.

as the amplitude of the OOK symbols (i.e., 0 or 1). It can be seen, '1'-bit signal last about $10^{-6}s$ (therefore will affect the following signal processing if the transmitted data interval is less than $10^{-6}s$). However, based on the CIR in Fig. 3.2 (b), the impulse response of one bit is lasting for at least $5 \times 10^{-6}s$, which will result in a symbol ISI. Then, in Fig. 3.2 (c) the ideal signal is calculated by multiplying the transmitted symbols and the CIR, and the noise-contaminated signal with the addition of the Gaussian noise is shown. The existence of ISI can be observed, by the ideal signal. For instance, the third bit has a high level of voltage, which should be a '0-bit' following the first two '1-bits'. Meanwhile, compared with the ideal signal, based on the noise-

contaminated signal, it is more difficult to distinguish '0-bit' from '1-bit'. It is reasonable the added noise leads to extra damage to the transmitted signal, consequently influencing signal detection.

Then, the symbol ISI will be demonstrated on the 8-PAM modulation, with a given transmitted interval $T_s = 10^{-6}s$ and a fixed signal noise rate $SNR = 15$dB, seen in Fig. 3.3.

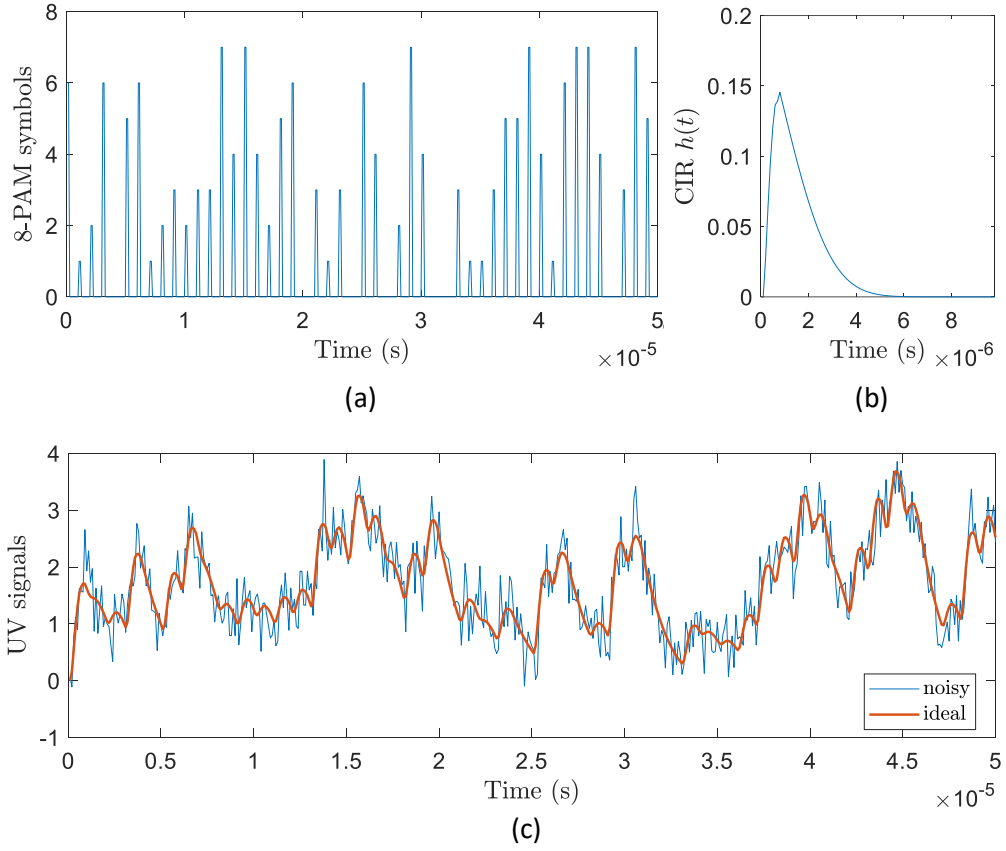


Figure 3.3: Illustrations of 8-PAM symbols that will be transmitted, the CIR of the UVC system, and the received signals with SNR=15dB.

In Fig. 3.3 (a)-(c), the transmitted 8-PAM symbol stream, the CIR of the UVC channel and the received signals are respectively illustrated. Different from the previous OOK modulation, in Fig. 3.3 (a), the vertical axis represents the amplitude of the 8-PAM symbols, with 8 different levels of voltage. In the simulation, a single bit symbol has a $10^{-6}s$ transmitted interval, while the

imposed response of one bit is lasting for about $5 \times 10^{-6}s$ according to the CIR in Fig. 3.3 (b). Then, in Fig. 3.3 (c) the ideal signal and the noise-contaminated signal are shown. The ideal signal suffers a serious ISI, as it is difficult to figure out 8 types of amplitude.

Finally, the effect of transmission speed on ISI is illustrated by comparing the signals under different symbol rates, in the case of OOK modulation, and a fixed SNR is employed.

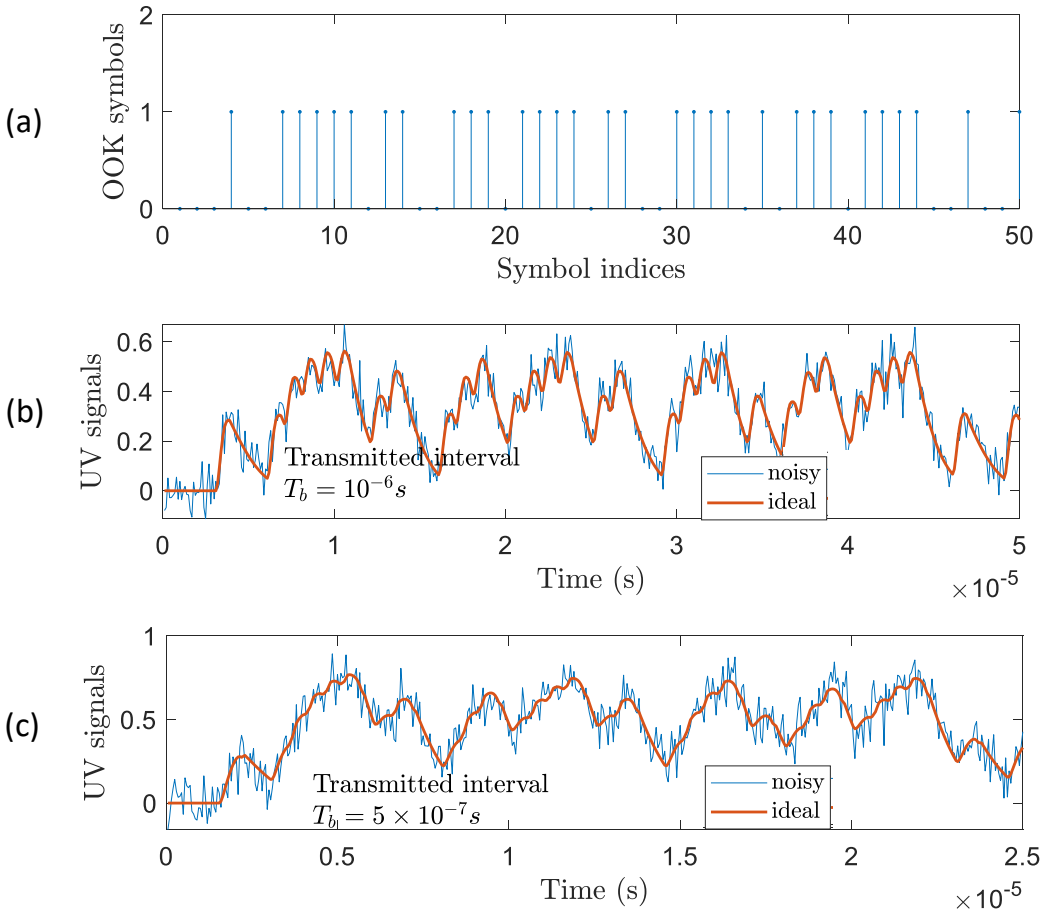


Figure 3.4: Illustrations of same OOK sequence undergoing different transmitted interval (ISI level) with SNR=15dB.

As shown in Fig. 3.4, it demonstrates the influence of symbol rate on the ISI, via transmitting the same OOK sequence with two different transmitted

intervals (ISI levels). In Fig. 3.4(a), the horizon axis is the symbol indices, and the vertical axis is defined as the amplitude of the OOK symbols (i.e., 0 or 1). It can be seen that 50 symbols are listed, which construct the given symbol sequence for transmission.

Then, the symbol sequence is transmitted by two different transmission speeds whose transmitted intervals are $T_b = 10^{-6}$ and $T_b = 5 \times 10^{-7}$, under the same $SNR = 15dB$. The ideal and real signals are displayed in the following Fig. 3.4 (b) and Fig. 3.4 (c), separately indicated by the red and blue lines. In both of these two figures, the horizontal axis indicates the time with a unit 's', and the vertical axis is the value of the UV signals. The ideal signals are derived from the CIR, and real signals include additive noise. It shows that the signals in fig. 3.4 (b) is easier to distinguish '0-bit' from '1-bit' than that in fig. 3.4 (c). For instance, in the sequence's $30th$ to $35th$ symbols '111101', in Fig. 3.4 (b) five rising edges can be figured out, while there are only four small rising edges that can be observed in Fig. 3.4 (c). Besides, the '0-bit' in the $34th$ slot has a high energy level in Fig. 3.4 (c) at 0.5 which is similar to that of the $35th$ slot with '1-bit', while in Fig. 3.4 (b) the energy of $34th$ slot is 0.2 clearly distinct with the following '1-bit'. The reason is that Fig. 3.4 (c) has a smaller transmitted interval than Fig. 3.4 (b), consequently leading to more ISI to deteriorate the transmitted signals. It means the transmission speed gives a rise to the ISI intensity. However, a higher transmission speed is required in real communication systems.

3.3 Conclusions & Discussions

The basic ultraviolet communication system has been illustrated in this chapter. Then, the effect of ISI on the UV channel has been evaluated by numerical simulation, in terms of varying the symbol rate and the different modulation

methods (i.e., J-PAM). Based on this model, the next 3 chapters will show the designed non-coherent detection schemes, which are able to convert the sequential detection challenges induced by ISI into ISI-resistant discrete detection framework.

Chapter 4

One-Dimensional Non-Coherent Signal Detection

In this chapter, 1D non-coherent detection schemes are proposed, which leveraged on the linearly weighted combinations of the transient features of the UV signal. It includes the EWNCD and the OWNCD schemes, in terms of the addition methods that are respectively combined by equal weight and optimal weights.

The main content of this chapter are listed as follows. In Section 4.1, the framework of the 1D non-coherent detection method is given, following which is a problem formulation. Section 4.2 focuses on the feature extraction of optical signals. There are four features obtained (i.e., the intra-symbol rising edge, the intra-symbol energy, the inter-symbol minimum inflection, and the inter-symbol energy difference of the signal), which are insensitive to the ISI contamination. Besides, the advantages of such feature-based metrics are evaluated. Section 4.3 introduces the EWNCD scheme which combines metrics with equal weight, and demonstrates the influence of the weights by simulation. Then, Section 4.4 elaborates on the OWNCD scheme, containing: i) the construction of the OWNCD design, ii) the theoretical analysis of the

BER performance, iii) the optimal weight computation method, iv) the design of a heuristic algorithm. After that, Section 4.5 evaluates the EWNCD and OWNCD scheme, in terms of BER accuracy and computational complexity, by numerical simulation. Finally, a conclusion of 1D non-coherent detection methods will be given in Section 4.6.

4.1 System Models and Problem Formulation

In this part, there is an introduction to the framework of 1D non-coherent detection, as seen in Fig. 4.1.

The 1D non-coherent detection communication system is based on the basic UVC system, introduced by Chapter 3. As it is only suitable for OOK, the signal is modulated by OOK modulation at the Tx point. Then, the modulated UV signal passes through the UV channel. At the Rx point, the non-coherent signal detection unit contains feature-based metric construction and heuristic detection. The non-coherent signal detection unit, in combination with the other components, allows for effective detection and estimation of the UV signals.

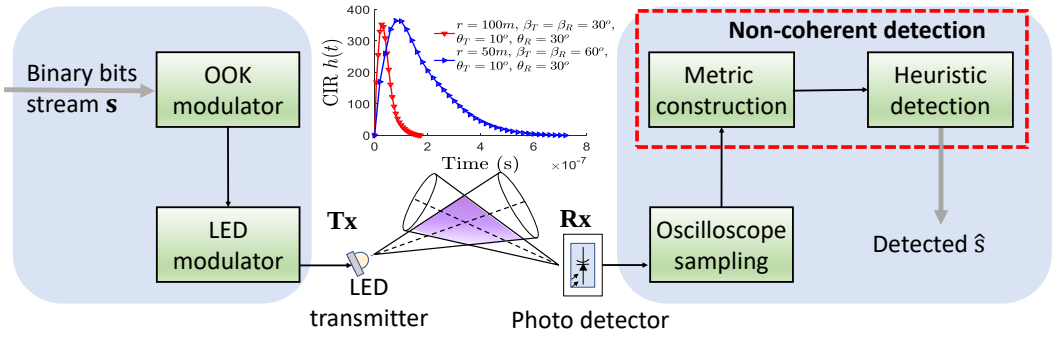


Figure 4.1: Illustration of UVC system with a LED transmitter, dynamic UV channel, and a UV photon-counting receiver. The 1D non-coherent detection scheme is deployed at the end of the receiver, whereby metrics are constructed at first, followed by the Heuristic detection.

Given the expressions of the received signals in Eq. (3.4), the purpose of this work is to detect the current information bit s_k from the received signal $\mathbf{r}_{1:kM} = [r_1, ..., r_{kM}]^T$. Here, it is noteworthy that the expression of the CIR $h(t)$, and the variance of the noise ε_i are of unavailability, varying with the changes of the channel parameters and the uncertainties of the Rx, and the PMT detector.

As aforementioned, coherent detection heavily relays on the CSI which will lead to (i) the heavy overheads of the pilot sequences in order to track the changes of the CIR, and (ii) the computational and storage complexity for the ISI compensation and the computation of posterior (or likelihood) probability. This gives the motivation of this work, whereby non-coherent schemes will be elaborated in the following.

4.2 Feature Extraction

Before elaborating on the proposed 1D non-coherent detection schemes, feature extraction of UV signals is first illustrated in Section 4.2.1. Then, Section 4.2.2 will analyse the advantages of these features.

4.2.1 Feature Designs

Four features are extracted via the exploration of 1) the intra-symbol rising edge, 2) the intra-symbol energy, 3) the inter-symbol minimum inflection, and 4) the inter-symbol energy difference. By exploiting the transient features of the optical signals, the decision metric is directly derived from the observations r_i for signal detection, which completely excludes the complex CSI estimations.

a) Intra-symbol rising-edge

Taking the kth interval with $M = T_b/T_s$ samples as an example, in the case of $s_k = 1$, the output r_i will produce a distinct rising edge. Otherwise, when $s_k = 0$, r_i shows a slow decay.

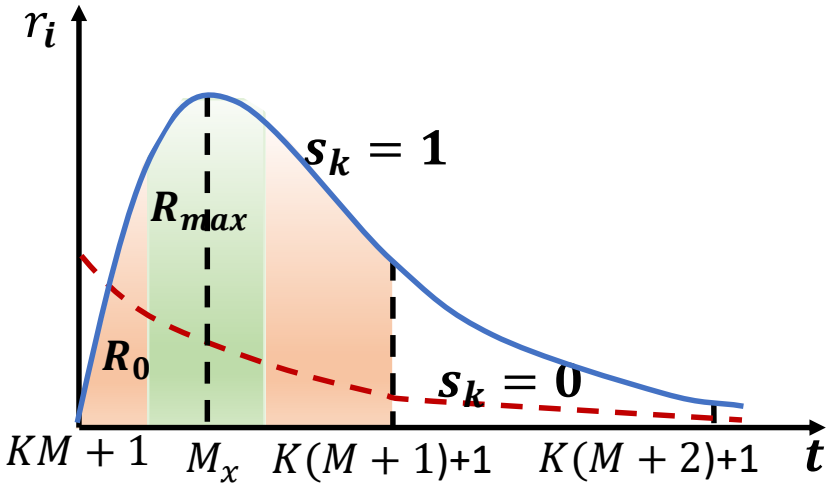


Figure 4.2: Illustrations of the feature-based metrics: the intra-symbol rising-edge.

As shown in Fig. 4.2, this rising edge can be expressed by the difference of its maximum (computed by averaging its neighbourhood $\mathcal{R}_{\max}$) from the initially $\mathcal{R}_0$, i.e.,

$$z_{1,k} \triangleq \sum_{i \in \mathcal{R}_{\max}} r_i - \sum_{i \in \mathcal{R}_0} r_i. \tag{4.1}$$

Here, the widths of $\mathcal{R}_{\max}$ and $\mathcal{R}_0$ are assigned to be $|\mathcal{R}_{\max}| = |\mathcal{R}_0| = M/4$, where $|\cdot|$ denotes the width (number of elements) of a set.

The distribution of $z_{1,k}$ is analyzed as follows. Given Eq. (4.1), $z_{1,k}$ is the linear summation of the independent observations r_i, whose distribution is Gaussian according to Eq. (3.4). As such, the probabilistic distribution of $z_{1,k}$ can be regarded as a Gaussian distribution, given the central limit theorem

(CLT), i.e.,

$$z_{1,k}|s_k \sim \begin{cases} \mathcal{N}(\mu_{1,0},\ \sigma_1^2) & s_k = 0, \\ \mathcal{N}(\mu_{1,1},\ \sigma_1^2) & s_k = 1, \end{cases} \tag{4.2}$$

where $\mu_{1,0}$ and $\mu_{1,1}$ are the expectations for cases $s_k = 0$ and $s_k = 1$ respectively, and σ_1^2 is the variance of metric $z_{1,k}$.

b) Intra-symbol energy

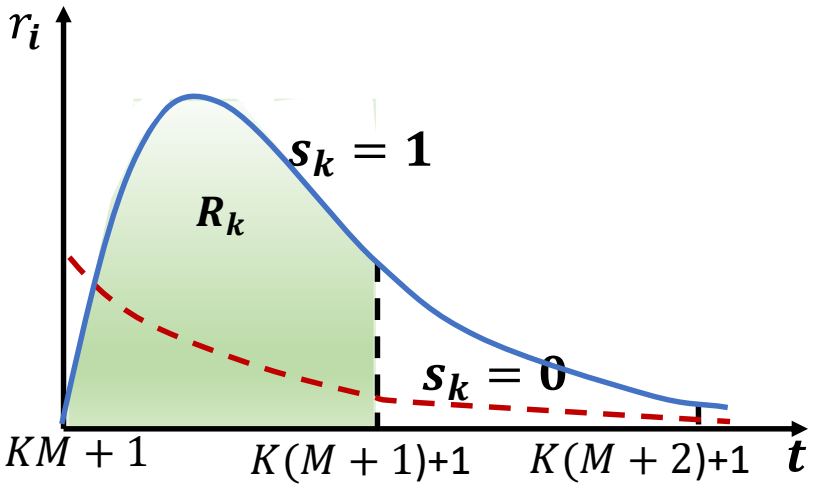


Figure 4.3: Illustrations of the feature-based metric: the intra-symbol energy.

Intra-symbol energy refers to the summation of samples in the current kth slot signal. This feature, overlooked by the previous study of a non-coherent scheme for molecular communications in [75], cannot be ignored in the context of UVCs. This is because the UV beam exhibits less of a particulate nature than molecules, which suggests a shorter averaging suspended period for each particle (i.e., UV photon), and thereby a reduced intensity of ISI. As such, given the smaller ISI effect in UVCs, the intra-symbol energy is to some extent able to distinguish the difference between $s_k = 1$ and $s_k = 0$. As is

shown in Fig. 4.3, the Intra-symbol energy is defined as as:

$$z_{2,k} = \sum_{i \in \mathcal{R}_k} r_i, \tag{4.3}$$

with $\mathcal{R}_k = \{kM + 1, \cdots, k(M + 1)\}$.

Similar to the distribution of $z_{1,k}$, the distribution of $z_{2,k}$ is also Gaussian, and can be expressed as:

$$z_{2,k}|s_k \sim \begin{cases} \mathcal{N}\left(\mu_{2,0},\ \sigma_2^2\right) & s_k = 0, \\ \mathcal{N}\left(\mu_{2,1},\ \sigma_2^2\right) & s_k = 1, \end{cases} \tag{4.4}$$

where $\mu_{2,0}$ and $\mu_{2,1}$ are the expectations for cases $s_k = 0$ and $s_k = 1$ respectively, and σ_2^2 is the variance of metric $z_{2,k}$.

Then, the transient features between two adjacent symbols are analyzed. For two successive symbols in the slots $k - 1$ and k, it may be noticed from Fig. 4.4 and 4.5.

c) Inter-Symbol Minimum inflection

The transient shape at the beginning of the kth symbol is distinguishable between $s_k = 1$ and $s_k = 0$.

$$z_{3,k} \triangleq -\sum_{i \in \mathcal{R}_{\text{inflx}}} r_i + \frac{1}{2}\left(\sum_{i \in \mathcal{R}_{\text{left}}} r_i + \sum_{i \in \mathcal{R}_{\text{right}}} r_i\right), \tag{4.5}$$

where the widths are assigned as $|\mathcal{R}_{\text{inflx}}| = |\mathcal{R}_{\text{left}}| = |\mathcal{R}_{\text{right}}| = M/4$. Then, the distributions of $z_{3,k}$, given the CLT, may be given i.e.,

$$z_{3,k}|s_k \sim \begin{cases} \mathcal{N}\left(\mu_{3,0},\ \sigma_3^2\right) & s_k = 0, \\ \mathcal{N}\left(\mu_{3,1},\ \sigma_3^2\right) & s_k = 1, \end{cases} \tag{4.6}$$

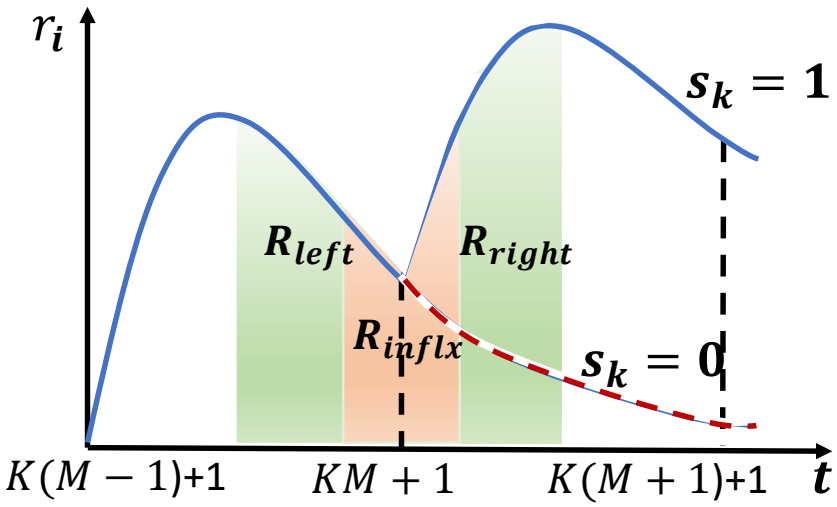


Figure 4.4: Illustrations of the feature-based metric: the inter-symbol minimum inflection.

In Eq. (4.6), $\mu_{3,0}$ and $\mu_{3,1}$ are the expectations for cases $s_k = 0$ and $s_k = 1$ respectively, and σ_3^2 is the variance of metric $z_{3,k}$.

d) Inter-Symbol Energy Difference

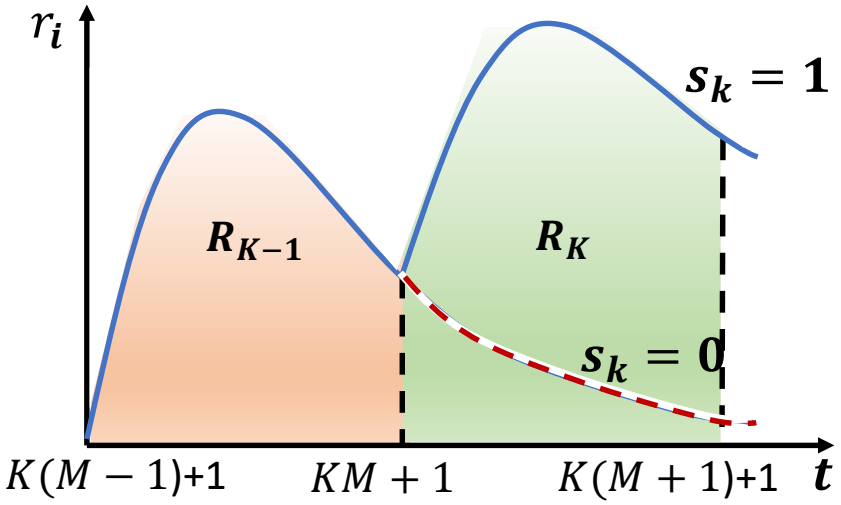


Figure 4.5: Illustrations of the feature-based metric: the inter-symbol energy difference.

There exists an obvious successive energy difference. As such, two sub-metrics are designed (i.e., the inter-symbol minimum inflection, and the inter-symbol energy difference), which are given as follows:

$$z_{4,k} \triangleq \sum_{i \in \mathcal{R}_k} r_i - \sum_{i \in \mathcal{R}_{k-1}} r_i, \tag{4.7}$$

Then, the distributions of $z_{3,k}$ and $z_{4,k}$ given the CLT are,

$$z_{4,k}|s_k \sim \begin{cases} \mathcal{N}\left(\mu_{4,0},\ \sigma_4^2\right) & s_k = 0, \\ \mathcal{N}\left(\mu_{4,1},\ \sigma_4^2\right) & s_k = 1. \end{cases} \tag{4.8}$$

In Eq. (4.8), $\mu_{4,0}$ and $\mu_{4,1}$ are the expectations for cases $s_k = 0$ and $s_k = 1$, and σ_4^2 are the variance of $z_{4,k}$.

4.2.2 Advantages of feature-based metrics

The derived features in Eqs. (4.1)-(4.7) have three advantages.

First, each one transforms the observation r_i that is heavily contaminated by ISI into a new domain that is insensitive to ISI. For instance, given a fixed intensity of the ISI and any of s_k, the expectations of $z_{1,k}$ $z_{2,k}$, $z_{3,k}$ and $z_{4,k}$ remain quasi-constant, as they reflect only the common features of the signal shapes. This suggests the inherent ability of the four features to counteract the ISI, without any process of complex channel estimations for ISI mitigation (as the coherent schemes do). As such, the extracted features convert the detection with ISI into a binary detection framework, enabling the potential designs of the non-coherent schemes.

Second, the utilisation of the neighbouring sets i.e., $\mathcal{R}_{\max}$, $\mathcal{R}_0$, $\mathcal{R}_{\text{inflx}}$, $\mathcal{R}_{\text{left}}$, $\mathcal{R}_{\text{right}}$ and $\mathcal{R}_k$ avoids the sensitivity to channel parameter perturbation induced time-varying channel response. Such neighbourhoods in which the

expectations of the elements change smoothly help neutralize the noise effects by the mechanism of averaging filters.

Third, the four features $z_{1,k}$ $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$ share the same trends of their conditional expectations. That is, $\mu_{1,1} - \mu_{1,0}$, $\mu_{2,1} - \mu_{2,0}$, $\mu_{3,1} - \mu_{3,0}$, and $\mu_{4,1} - \mu_{4,0}$ are with the same sign (positive or negative). This sharing trend gives rise to the idea of combining such four features in order to strengthen the detection accuracy of the input s_k. There now follows an elaboration of the design of the optimally-weighted combination strategy.

4.3 Equal-Weighted Non-Coherent Detection Method

Based on the UV feature-based metrics, the OWNCD method is designed, detailed in Section 4.3.1. Then, the evaluation of the influence of combination weights is conducted in Section 4.3.2.

4.3.1 EWNCD Design

Given the aforementioned mutual positiveness/negativeness property of the four metrics, one intuitive combination strategy is the linear summation with equal weights. As such, the EWNCD scheme is designed by deriving the equally combined metric, and its probabilistic distribution, i.e.,:

$$z_k^{\text{EWNCD}} = z_{1,k} + z_{2,k} + z_{3,k} + z_{4,k}, \tag{4.9}$$

$$z_k^{\text{EWNCD}} | s_k \sim \begin{cases} \mathcal{N}\left(\sum_{d=1}^{D} \mu_{d,0},\ \sum_{d=1}^{D} \sigma_d^2\right) & s_k = 0, \\ \mathcal{N}\left(\sum_{d=1}^{D} \mu_{d,1},\ \sum_{d=1}^{D} \sigma_d^2\right) & s_k = 1. \end{cases} . \tag{4.10}$$

Then, the detection process for EWNCD scheme can be regarded as a binary detection framework, i.e.,

$$\hat{s}_k = \begin{cases} 0, & z_k^{\text{EWNCD}} \leq \gamma, \\ 1, & z_k^{\text{EWNCD}} > \gamma, \end{cases} . \tag{4.11}$$

where $\hat{s}_k$ is the decision of the kth symbol, and γ is the detection threshold and can be computed as $\gamma = (\sum_{d=1}^{4} \mu_{d,0} + \sum_{d=1}^{4} \mu_{d,1})/2$.

It is noteworthy that for the case where the CIR $h(t)$ is unknown, the direct computation $\gamma = (\sum_{d=1}^{4} \mu_{d,0} + \sum_{d=1}^{4} \mu_{d,1})/2$ is infeasible. To deal with it, for any kth symbol, γ is approximated by

$$\gamma \approx \frac{1}{k} \sum_{k'=1}^{k} z_{k'}^{\text{EWNCD}}, \tag{4.12}$$

where a straightforward proof can be performed as:

$$\begin{aligned} &\lim_{k=+\infty} \frac{1}{k} \sum_{k'=1}^{k} z_{k'}^{\text{EWNCD}} \\ =&\lim_{k=+\infty} \frac{1}{k} \left(\sum_{s_{k'}=1, k' \leq k} z_{k'}^{\text{EWNCD}} + \sum_{s_{k'}=1, k' \leq k} z_{k'}^{\text{EWNCD}} \right) \\ =&\frac{1}{2} \left(\sum_{d=1}^{4} \mu_{d,0} + \sum_{d=1}^{4} \mu_{d,1} \right) \\ =&\gamma. \end{aligned} \tag{4.13}$$

However, one obvious drawback of the EWNCD scheme cannot be overlooked. The contribution of $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$ may not be the same for different cases of the ISI intensity. For example, in the face of less intensive ISI scenarios, the inter-symbol energy difference performs badly, as $z_{3,k}$ of the successive two 1 bits, i.e., $s_{k-1} = s_k = 1$ is close to that of successive two 0

bits i.e., $s_{k-1} = s_k = 0$. As such, this raises the necessity of assigning different weights for the appropriate contributions of $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$, and there now follows computation of the optimal weights.

4.3.2 Influence of Weights

First, the influences of the weights on the BER performance is examined in Fig. 4.6. To display the result in terms of 3D which demonstrates the result in a more intuitive perspective, three features are selected to conduct linear combinations.

Here, the linear combination of these metrics with different weights w_i is evaluated. The weights are normalised by setting $w_1 + w_2 + w_3 = 1$, aiming to illustrate the change of BER via only two of the three weights. To be specific, w_1 and w_2 are recognised as two variates, based on which w_3 can be easily figured out, given that the values of w_1, w_2, w_3 meet the condition $w_1 + w_2 + w_3 = 1$. Thus, in this simulation, only w_1 and w_2 will be shown, as seen in Fig. 4.6.

In Fig. 4.6, the x-axis and y-axis represent w_1 and w_2, respectively, $w_1, w_2 \in [0, 1]$. The z-axis indicts the BER, the value of which is also shown by the gradient ramp with the color bar on the right. In can be observed that there is a BER surface according to the continuous variable of the weights setting. That means the BER performance is affected by the different values of the weights. Besides, a minimum BER point exists in the BER surface illustrating that minimum BER can be obtained by optimal assignment of the weights.

To obtain a higher detection accuracy, an optimally-weighted non-coherent detection scheme is designed, by calculating the optimal weight setting. The following part provides more detail of the proposed OWNCD scheme.

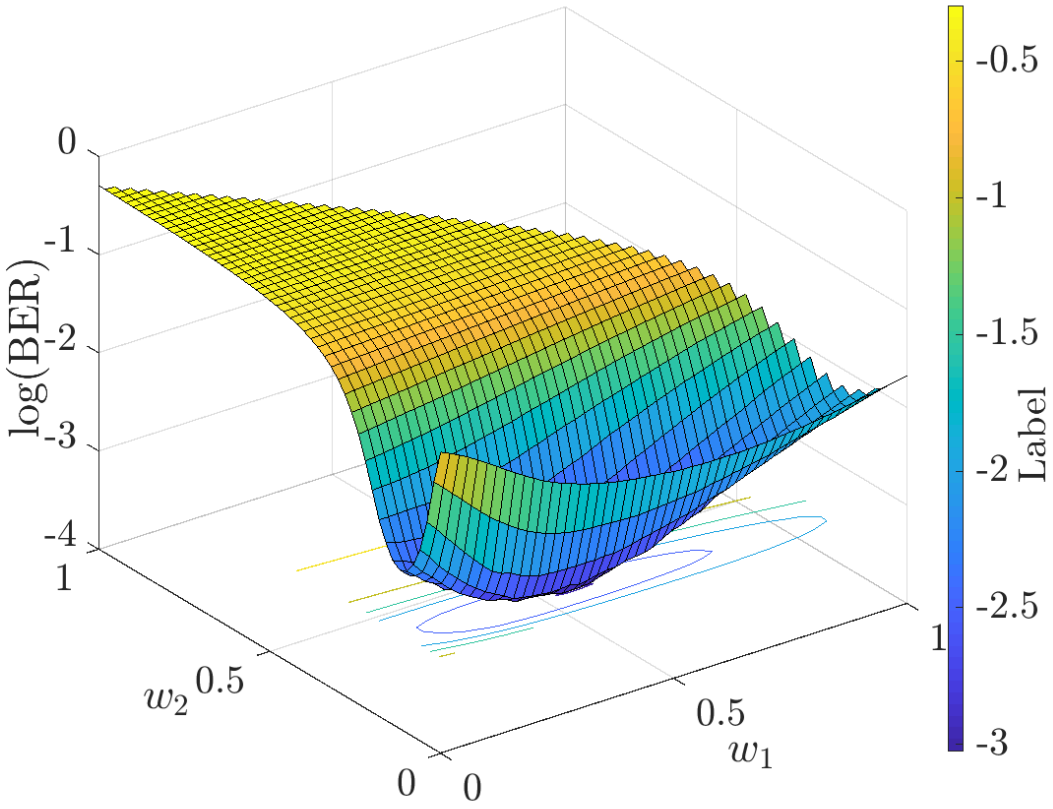


Figure 4.6: The BER performance with different settings of weights, normalised by $w_1 + w_2 + w_3 = 1$. The minimum BER corresponds to the optimal weights computed in Eq. (4.25), i.e., $w_1 = 0.6$, $w_2 = 0.32$, and $w_3 = 0.08$.

4.4 Optimally-Weighted Non-Coherent Detection Method

This part explains the details of the OWNCD method, which combines the metrics with optimal weights. Then, the optimal weight calculation method and application algorithm design are detailed in Section 4.4.4 and Section 4.4.5, respectively. Finally, the performance of the OWNCD is provided, by deriving the theoretical BER, seen in Section 4.4.3.

4.4.1 OWNCD Design

The key point of OWNCD is to derive optimal weights for the four extracted features, in order to minimize the BER. To do so, the weighted combination of the features is expressed as (and the decision process is leveraged on such a metric):

$$z_k^{\text{OWNCD}} = w_1 \cdot z_{1,k} + w_2 \cdot z_{2,k} + w_3 \cdot z_{3,k} + w_4 \cdot z_{4,k} \underset{\hat{s}_k=0}{\overset{\hat{s}_k=1}{\gtrless}} \zeta, \tag{4.14}$$

where w_1, w_2, w_3, and w_4 are the corresponding weights, and ζ is the detection threshold. Next, the optimal weights are derived and also the detection threshold by 1) analyzing the probabilistic distribution of the metric z_k^{OWNCD}, 2) computing the theoretical BER with respect to the optimal detection threshold and different weights, and 3) computing the optimal weights w_1, w_2, w_3, and w_4 by minimizing the theoretical BER.

4.4.2 Distribution of weighted combining metric

With the help of Eq. (4.14) and the distributions of $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$, the distribution of z_k^{OWNCD} conditioned on s_k is straightforward, i.e.,

$$z_k^{\text{OWNCD}} | s_k \sim \begin{cases} \mathcal{N}(\mu_0,\ \sigma^2) & s_k = 0, \\ \mathcal{N}(\mu_1,\ \sigma^2) & s_k = 0, \end{cases} \tag{4.15}$$

where μ_0, μ_1 is the the expectations under conditions of $s_k = 0$ and $s_k = 1$ respectively, and σ^2 is the variance, i.e., $\mu_0 = \sum_{d=1}^{4} w_d \cdot \mu_{d,0}$, $\mu_1 = \sum_{d=1}^{4} w_d \cdot \mu_{d,1}$ and $\sigma^2 = \sum_{d=1}^{4} w_d^2 \cdot \sigma_d^2$.

4.4.3 Theoretical BER computation

Given the Gaussian distribution of z_k^{OWNCD} in Eq. (4.15), the theoretical BER of Eq. (4.14) can be computed as:

$$\begin{aligned}\text{BER} =& \Pr\{s_k = 0\} \cdot \Pr\{\hat{s}_k = 1 | s_k = 0\} \\ &+ \Pr\{s_k = 1\} \cdot \Pr\{\hat{s}_k = 0 | s_k = 1\} \\ =& \frac{1}{2\sqrt{2\pi}\sigma} \left(\int_{\zeta}^{+\infty} e^{-\frac{(z-\mu_0)^2}{2\sigma^2}} dz + \int_{-\infty}^{\zeta} e^{-\frac{(z-\mu_1)^2}{2\sigma^2}} dz \right)\end{aligned} \tag{4.16}$$

where $\Pr\{\cdot\}$ represents the probability of the event. Here, without loss of generality, $\Pr\{s_k = 0\} = \Pr\{s_k = 1\} = 0.5$ is assumed, corresponding to equal probabilities of the emitted symbols $s_k = 0$ and $s_k = 1$.

Accordingly, the optimal detection threshold that minimizes the BER in Eq. (4.16) is straight forward, being

$$\zeta = \frac{\mu_1 + \mu_0}{2}. \tag{4.17}$$

By taking Eq. (4.17) back into Eq. (4.16), the BER can be further deduced as a function of the weights w_1, w_2, w_3, and w_4, i.e.,

$$\begin{aligned}&\text{BER}(w_1, w_2, w_3, w_4) \\ &= Q\left(\sqrt{\frac{(\mu_1 - \mu_0)^2}{4\sigma^2}} \right) \\ &Q\left(\sqrt{\frac{\left(\sum_{d=1}^{4} w_d(\mu_{d,1} - \mu_{d,0})\right)^2}{4\sum_{d=1}^{4} w_d^2 \sigma_d^2}} \right),\end{aligned} \tag{4.18}$$

where $Q(x) = \int_x^{+\infty} 1/\sqrt{2\pi} e^{-z^2/2} dz$. As such, given the relationship between BER and the weights in Eq. (4.18), the optimal weights are computed by minimizing Eq. (4.18).

4.4.4 Optimal weight computation

First, the existence of the minimum BER is explained. This is due to the fact that values of SNR for the four metrics $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$ are finite, and therefore the SNR of the linear combination of such metrics is inevitable bounded and there exists an upper-bound of SNR that is associated to the lowest BER.

Given the theoretical BER in Eq. (4.18), the computation of its minimum is to find the solution of w_1, w_2, w_3, and w_4 such that:

$$\frac{\partial \mathrm{BER}(w_1, w_2, w_3, w_4)}{\partial w_n} = 0, \quad n = 1, 2, 3, 4. \tag{4.19}$$

Eq. (4.19) is further simplified by analyzing the SNR in the function of $Q(\sqrt{\mathrm{SNR}})$ in Eq. (4.18), where $Q(x)$ is monotonically decreasing with the increment of x. As such, by expressing the SNR as:

$$\mathrm{SNR}(w_1, w_2, w_3, w_4) = \frac{\left(\sum_{d=1}^{4} w_d(\mu_{d,1} - \mu_{d,0})\right)^2}{4\sum_{d=1}^{4} w_d^2 \sigma_d^2}, \tag{4.20}$$

solving Eq. (4.19) is converted to solving

$$\begin{cases} \frac{\partial \mathrm{SNR}}{\partial w_1} = 0, \\ \frac{\partial \mathrm{SNR}}{\partial w_2} = 0, \\ \frac{\partial \mathrm{SNR}}{\partial w_3} = 0, \\ \frac{\partial \mathrm{SNR}}{\partial w_4} = 0. \end{cases} \tag{4.21}$$

By denoting $\varrho_d = \mu_{d,1} - \mu_{d,0}$ by $d = 1, 2, 3, 4$, the left hand side of this equation

is further computed as:

$$\begin{aligned}
&\frac{\partial \mathrm{SNR}}{\partial w_n} \\
&= \frac{\varrho_n \sum_{d=1}^{4} w_d \varrho_d \sum_{d=1}^{4} w_d^2 \sigma_d^2 - w_n \sigma_n^2 \left(\sum_{d=1}^{4} w_d \varrho_d\right)^2}{2\left(\sum_{d=1}^{4} w_d^2 \sigma_d^2\right)^2} \\
&= \frac{\sum_{d=1}^{4} w_d \varrho_d}{2\left(\sum_{d=1}^{4} w_d^2 \sigma_d^2\right)^2} \left(\varrho_n \sum_{d=1}^{4} w_d^2 \sigma_d^2 - w_n \sigma_n^2 \sum_{d=1}^{4} w_d \varrho_d\right).
\end{aligned} \tag{4.22}$$

Thus, it can be converted to:

$$\varrho_n \sum_{d=1}^{4} w_d^2 \sigma_d^2 - w_n \sigma_n^2 \sum_{d=1}^{4} w_d \varrho_d = 0. \tag{4.23}$$

Noticing from Eq. (4.23), that:

$$\frac{w_1 \sigma_1^2}{\varrho_1} = \frac{w_2 \sigma_2^2}{\varrho_2} = \frac{w_3 \sigma_3^2}{\varrho_3} = \frac{w_4 \sigma_4^2}{\varrho_4} = \frac{\sum_{d=1}^{4} w_d^2 \sigma_d^2}{\sum_{d=1}^{4} w_d \varrho_d} = \xi, \tag{4.24}$$

which indicates the proportional relationship among w_1, w_2, w_3, and w_4. To be specific, by assigning ξ as any positive (real) value, a group of weights can be derived to be:

$$w_d = \xi \cdot \frac{\mu_{d,1} - \mu_{d,0}}{\sigma_d^2},\ d = 1, 2, 3, 4 \tag{4.25}$$

which ensures $\frac{\partial \mathrm{SNR}}{\partial w_1} = \frac{\partial \mathrm{SNR}}{\partial w_2} = \frac{\partial \mathrm{SNR}}{\partial w_3} = \frac{\partial \mathrm{SNR}}{\partial w_4} = 0$. Therefore, such derived w_1, w_2, w_3, and w_4 can guarantee a minimum BER. It is noteworthy that, the solutions are a series of values with any positive ξ, all of which give the same BER that is not related to ξ in Eq. (4.26) by taking Eq. (4.25) into Eq. (4.18), i.e.,

$$\mathrm{BER}_{\min} = Q\left(\sqrt{\frac{\sum_{d=1}^{4} \frac{\left(\mu_{d,1} - \mu_{d,0}\right)^2}{\sigma_d^2}}{4\sum_{d=1}^{4} \left(\mu_{d,1} - \mu_{d,0}\right)^2}}\right). \tag{4.26}$$

Thus, the best BER performance will be obtained when the weights of the

metrics are derived from Eq. (4.25).

Compared with previous non-coherent detection methods, OWNCD is able to achieve better communication accuracy, attributed to the derivation of the optimal weights. However, such computation of the optimal weights requires partial channel information (i.e., the expectations and the variances of $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$). Different from the coherent method which relies on complex channel estimation to obtain the exact CIR, the OWNCD method does not require the whole information of CIR, but only extracts the partial channel information (i.e., the ISI insensitive geometrical channel features) via simple mathematical operation. The traditional channel estimation technologies (e.g.the ML approach) require a large pilot sequence overhead (repeatedly for time-varying UVC channels), and extensive computational resources due to the computation of the exponential/logarithm -form of likelihood PDF, which makes them unattractive. In real application, it is difficult to guarantee the channel invariant through a communication period. In order to ensure detection accuracy, the repeat extraction of the signal features is necessary, which is similar to the repeat channel estimation in a coherent detection method. To do so, a heuristic algorithm is designed to recursively compute the designed signal features from time-varying channel, which can ensure the detection performance with the changes of the channels.

4.4.5 Heuristic Algorithm

Details of the heuristic algorithm now follow. For the first k_0 symbols, the assignment $w_1 = w_2 = w_3 = w_4 = 1$ is made, and EWNCD adopted for metric combination $z_k^{\text{EWNCD}}, 1 \leq k \leq k_0$ and signal detection i.e., $\hat{s}_k$. Then, for $(k > k_0)$th symbol, the expectations and variances are computed (i.e., $\mu_{d,1}$,

Algorithm 2 Heuristic Detection Algorithm

Input: Samples of kth symbol $[r_{(k-1)M+1}, \cdots, r_{kM}]$, previous metrics $z_{1,k'}$, $z_{2,k'}$, $z_{3,k'}$, $z_{4,k'}$ and decisions $\hat{s}_{k'}$ $(k' < k)$

if $k = 1$ **then**
  Compute $z_{1,k}$ via Eq. (4.1).
  Initialize $z_k = z_{1,k}$, and threshold $\gamma = 0$.
  Determine $\hat{s}_k$ via Eq. (4.11).
else
  Compute $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, and $z_{4,k}$ via Eqs. (4.1)-(4.5),
  if $1 < k \leq k_0$ **then**
    Compute $z_k^{\text{EWNCD}} = z_{1,k} + z_{2,k} + z_{3,k} + z_{4,k}$.
    Compute decision threshold γ via Eq. (4.12).
    Determine $\hat{s}_k$ via Eq. (4.11).
  else
    Compute w_1, w_2 and w_3 via Eq. (4.25).
    Compute $z_k^{\text{OWNCD}} = w_1 z_{1,k} + w_2 z_{2,k} + w_3 z_{3,k} + w_4 z_{4,k}$.
    Compute decision threshold ζ via Eq. (4.17).
    Determine $\hat{s}_k$ via Eq. (4.14).
  end if
end if

Output: Return metrics $z_{1,k}$, $z_{2,k}$, $z_{3,k}$, $z_{4,k}$ and decision $\hat{s}_k$

$\mu_{d,0}$, σ_d^2 with $d = 1, 2, 3$) via the previous k_0 data, i.e.,

$$\mu_{d,0} = \frac{1}{|\mathcal{K}_0|} \sum_{k' \in \mathcal{K}_0} z_{d,k'}, \quad d = 1, 2, 3 \tag{4.27}$$

$$\mu_{d,1} = \frac{1}{|\mathcal{K}_1|} \sum_{k' \in \mathcal{K}_1} z_{d,k'}, \quad d = 1, 2, 3 \tag{4.28}$$

$$\sigma_d^2 = \frac{1}{k_0} \left(\sum_{k' \in \mathcal{K}_0} \left(z_{d,k'} - \mu_{d,0} \right)^2 + \sum_{k' \in \mathcal{K}_1} \left(z_{d,k'} - \mu_{d,1} \right)^2 \right), \tag{4.29}$$

where

$$\mathcal{K}_0 = \{k' | \hat{s}_{k'} = 0, k - k_0 \leq k' < k\} \tag{4.30}$$

$$\mathcal{K}_1 = \{k' | \hat{s}_{k'} = 11, k - k_0 \leq k' < k\}, \tag{4.31}$$

with $|\mathcal{K}_0| + |\mathcal{K}_1| = k_0$. Leveraged on such computed partial channel information, the OWNCD metric c_k^{OWNCD} can be derived and the decision can be pursued.

The detailed algorithm is provided in Algo. 2. The inputs are the samples of kth symbol (i.e., $[r_{(k-1)M+1}, \cdots, r_{kM}]$), and the previously computed metrics $z_{1,k'}$, $z_{2,k'}$, $z_{3,k'}$, $z_{4,k'}$ with the decisions $\hat{s}_{k'}$ $(k' < k)$. Steps 1-4 are to detect the first symbol via only the intra-symbol maximal inflection $z_{1,1}$, as there is no previous data for the construction of inter-symbol based metrics $z_{2,1}$, $z_{3,1}$ and $z_{4,1}$. Steps 5-10 are to process the first k_0th symbols by EWNCD, since there is not enough information for the computation of partial channel information that will be used for OWNCD. Steps 11-17 are to detect any $(k > k_0)$th symbol using OWNCD. The outputs are the decision for current kth symbol, and the metrics $z_{1,k}$, $z_{2,k}$, $z_{3,k}$ and $z_{4,k}$ for the next $k+1$ round of signal detection.

As such, the heuristic algorithm avoids the usages of the pre-defined pilot sequence and the complex channel estimation process, rather it relies on only the basic adder and multiplier, which makes it easy for implementation and energy-saving. Next, there is and analysis of the complexity of the proposed non-coherent heuristic scheme.

4.5 Performance Analysis and Numerical Simulations

4.5.1 Complexity Analysis

The complexity of the proposed non-coherent scheme contains the computational cost and the storage consumption. Here, the computational cost is evaluated by counting the number of multiplication operations, and the stor-

Table 4.1: Comparison of complexity among three detection schemes.

Methods	Total complexity	Storage
EWNCD (proposed)	$O(1)$	$O(k)$
OWNCD (proposed)	$O(k_0)$	$O(k_0)$
Coherent MLSD [1]	$O(k \cdot 2^L + NL^2)$	$O(k \cdot 2^L)$

age cost is measured by the number of previous data saved. The complexity comparison among different schemes is provided in Table. 4.1.

A) The complexity of EWNCD

For the EWNCD scheme, for any kth symbol, the computational complexity comes from (i) the computations of the metrics in Eqs. (4.1)-(4.5), and (ii) the computation of the decision threshold γ in Eq. (4.12). As such, the number of multiplications is 7, computed as $O(1)$. Given that the computation of γ requires the summations of $c_1^{\text{EWNCD}}, \cdots, c_k^{\text{EWNCD}}$, the storage needed for the EWNCD scheme is measured as $O(k)$.

B) The complexity of OWNCD

The algorithm complexity of OWNCD is caused by four aspects: (i) the computations of the three metrics in Eqs. (4.1)-(4.5), (ii) the calculation of the weights in Eq. (4.25) and Eqs. (4.27)-(4.29), (iii) the combination of these metrics in Eq. (4.14), and (iv) the computation of the threshold ζ in Eq. (4.17). As such, the total number of multiplications for OWNCD is $5 + (2 + 1 + k_0) \cdot 3 + 3 + 3 = O(k_0)$. According to the inputs of Algo. 2, the storage required for OWNCD is measured as $4 \cdot k_0 = O(k_0)$ for the storing of previous k_0 metrics and decisions.

C) The complexity of coherent MLSD

Now, the computational complexity and the storage consumption of the coherent MLSD approach are analysed, which is provided in [1, 39]. The complexity and storage costs contains two aspects. The first is the costs of the Viterbi implementation. According to [39], finding the maximum posterior density for current k time-step requires the recursive computations from $\max_{u_{k-1}} p(u_k|u_{k-1})p(\tilde{y}_k|u_k)$ to $\max_{u_1} p(u_2|u_1)p(\tilde{y}_2|u_2)$, with $u_k = \sum_{l=k-L+1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot s_l$ for $k \geq L$, and $u_k = \sum_{l=1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot s_l$ for $1 \leq k < L$. Hence, for each $L \leq k' \leq k$, 2^L likelihood densities will be computed for enumerating all potential $\mathbf{s}_{k'-L+1:k'} \in \{0,1\}^L$. For $1 \leq k' < L$, $2^{k'}$ likelihoods are computed. Given from [1], each logarithm-form likelihood PDF includes one logarithm computation. As such, the total number of multiplications for the MLSD detection can be measured as:

$$(k - L + 2) \times 2^L - 2 = O(k2^L). \tag{4.32}$$

The second is the consumption for channel estimation. Take the ML channel estimation approach as an example. In [1], a predefined pilot sequence of length N is used to estimate the CSI. To do so, three matrices (of size $L \times (N - L + 1)$) multiplications, and a multiplication of such matrices with a $(N - L + 1) \times 1$ vector are used, which requires $(N - L + 1) \times (2L^2 + L)$ times of multiplications.

Combined with the detection process, the total number of multiplications for coherent MLSD with ML channel estimation is

$$(k - L + 2) \times 2^L - 2 + (N - L + 1) \times (2L^2 + L) = O(k \times 2^L + N \times L^2). \tag{4.33}$$

For the storage needed, the Viterbi tracing requires k observations, L CSI, as well as $2^L(2^{k'})$ likelihood densities for each $L \leq k' \leq k (1 \leq k' < L)$. As such,

the storage required for the coherent MLSD is $(k - L + 2) \times 2^L - 2 + L + k = O(k \times 2^L)$. Such complexity is compared with the proposed OWNCD in Table 1, and in Fig. 4.12 .

4.5.2 Numerical Simulations

In this part, the communication performance of both the EWNCD and the OWNCD schemes are evaluated, by the numerical simulation based on the MATLAB platform. Here, the detection accuracy is measured via the BER, and the computational complexity is approximated by the number of multiplications needed.

The environment and parameters involved are configured in Table 3.1, which is typical for a UVC scattering system using a UV-LED transmitter. Briefly, the number of emitted photons for each '1'-bit is taken as $N_T = 10^{10}$. The transmission distance is set at 250m. The transmitter's beam divergence and apex angle are $\pi/12$ and $\pi/4$, respectively. The receiver's half-field of view and apex angle are $\pi/12$ and $\pi/4$, respectively. The atmospheric coefficients are $\kappa_a = 5 \times 10^{-4} m^{-1}$ and $\kappa_s = 4.9 \times 10^{-4} m^{-1}$. For simulating different levels of ISI, the symbol rate is set in the range $R_b = 1/T_b \in [10^6 \text{bit/s},\ 5 \times 10^6 \text{bit/s}]$. Similarly, SNR values from 5 dB and 25 dB are considered in order to test the BER versus different intensities of ambient noise. Given the aforementioned channel parameters, different CIRs of the UV channel are generated via the Monte Carlo simulator [29, 30]. For both the EWNCD and the OWNCD methods, a constant number of samples $M = 20$ are assigned, evenly sampled for each symbol.

For comparison, EWNCD and the state-of-the-art coherent MLSD for generic FSOs in [1] was chosen, for the following two reasons. First, the aim was to show the necessity of the design of OWNCD, by testing the performance of equally weighted combination. Second, in most cases (including UVC),

MLSD has been proven to have the lowest BER when CSI is accurately derived, but this consumes substantial computational and storage resources for CSI estimation and signal detection. In this view, the communication performance of the proposed OWNCD for UVCs could be demonstrated by comparing both BER and resource consumption.

A) Analysis of Received UVC signals and the corresponding Metric

Here, the evaluation of the proposed EWNCD and OWNCD schemes will be conducted, by analysing the received signals and the corresponding metrics generated by them. In the following, the received signals and the derived metrics are demonstrated in Fig. 4.7, meanwhile, the statistic histograms of them will be calculated in Fig. 4.8.

In Fig. 4.7(a), the vertical axis is the amplitude of the received signals, and the horizontal axis indicates time with a unit s. The ideal signals and the received are represented by the red and blue lines. Following that, there are three figures to illustrate the amplitude of unprocessed signal, EWNCD metric, and OWNCD metric, all of which have the same horizon and vertical axis.

It can be seen that the amplitudes of unprocessed signal are in the range of -0.2 to 0.6, which is smaller than that of EWNCD at a range of $[-0.3, 0.9]$ approximately. Besides, the values of OWNCD have the largest range, almost from -2.2 to 5. It is noted that the signals randomly indicate '0-bit' and '1-bit' with an equal ratio. The unprocessed signal in Fig. 4.8, is hard to distinguish a '0-bit' from a '1-bit'. On the contrary, the metric values of EWNCD and OWNCD are easier to be divided into two kinds of signals, especially for the metric values based on OWNCD. For instance, the last symbol should be a '0-bit', inferred by the ideal signal, however, it has high energy in Fig. 4.7 (b) easily to be miss-determined as 1-bit". However, the metric values accounted

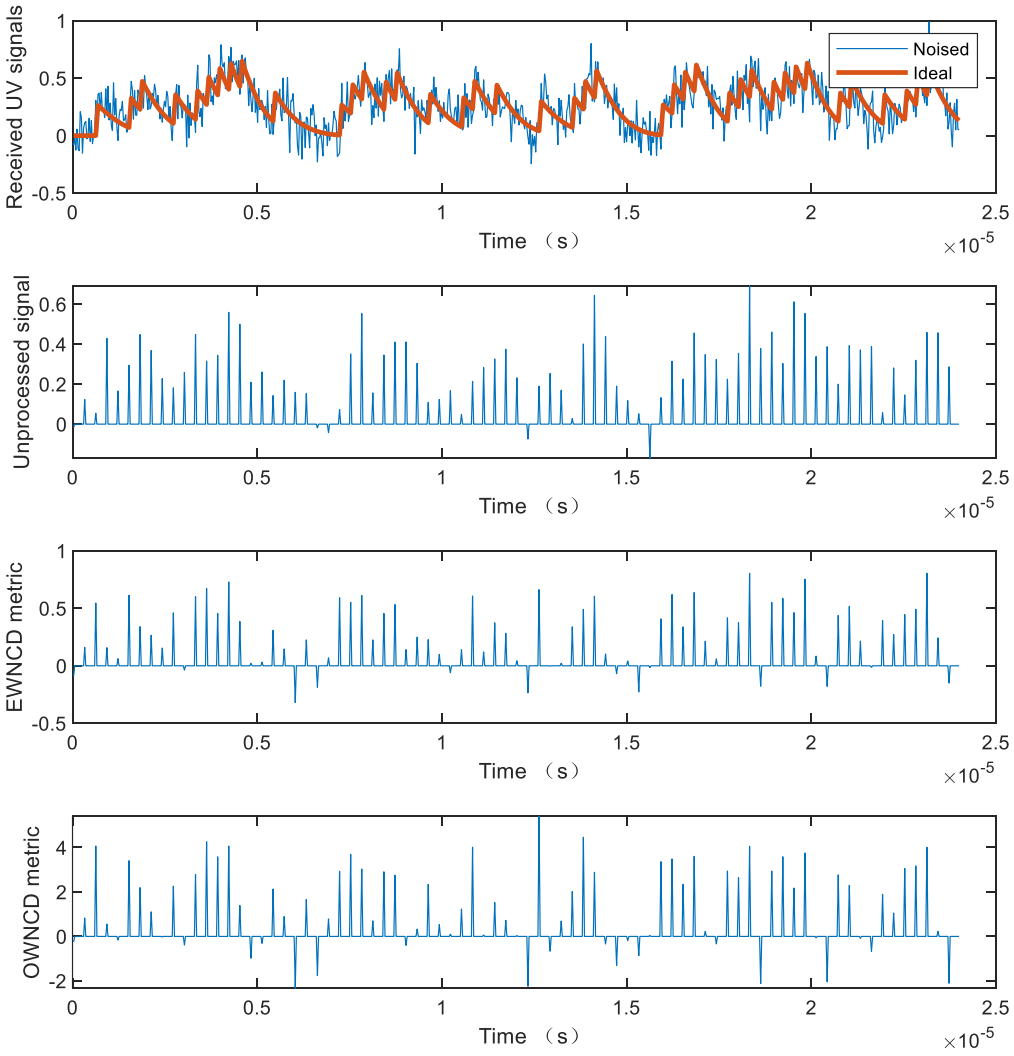


Figure 4.7: Illustration of received UVC signals and the corresponding EWNCD and OWNCD metrics.

for by both the EWNCD and the OWNCD schemes may easily be recognised as '0-bit', as shown in the Fig. 4.7 (c) and the Fig. 4.7 (d). For the marked symbol (near the 1) in the figure, it should be a '1-bit', as there is an obvious rising edge and a peak, in the ideal signal. But, the unprocessed signal of this point is at a low amplitude, even apparently smaller than the amplitude of the last symbol ('0-bit'). Based on the unprocessed amplitude, it is no doubt

to be misjudged as '0-bit', while EWNCD scheme can achieve a little gain and OWNCD can recover it to a high amplitude. In short, both the EWNCD and OWNCD schemes are capable of improving the signal amplitudes for a more accurate signal decision, and the OWNCD performance is better than EWNCD.

In signal detection, a decision surface could be determined, if the probability distribution function of the amplitude or metric of the received signals is given, based on which the theoretical BER will be derived. Thus, in the following, received signals' amplitudes, the EWNCD and the OWNCD metrics will be evaluated, in terms of the probability distribution.

Here, the statistic histograms of the unprocessed signals, the EWNCD and the OWNCD metrics are illustrated, respectively by Fig. 4.8 (a-c). In these figures, the horizontal axis represents the amplitudes or the metrics of the received signals, and the vertical axis indicates the possibility distribution function. It can be observed that the metrics generated by the EWNCD and the OWNCD obey Gaussian distributions, as well as the unprocessed signals. It matches the previous analysis that the linear sum of several Gaussian distributions is still a Gaussian distribution.

As is well-known, before conducting the signal determination, a decision threshold should be placed in the middle of these two distributions, as shown in the figures. Then, the amplitude over this threshold is '1-bit', and the amplitude below it is '0-bit'. It would be reasonable that a smaller overlap can derive a better determination result, as the overlap part will result in a misjudgment. In Fig. 4.8 (a), the PDF of unprocessed signals is demonstrated, suffering a large overlap of the '0-bit' and the '1-bit', which consequently results in a bad determination performance. It can be seen, the overlap is dramatically reduced, by the EWNCD metric. Moreover, the OWNCD metric has the least overlap, compared with the unprocessed signal and the EWNCD metric. It

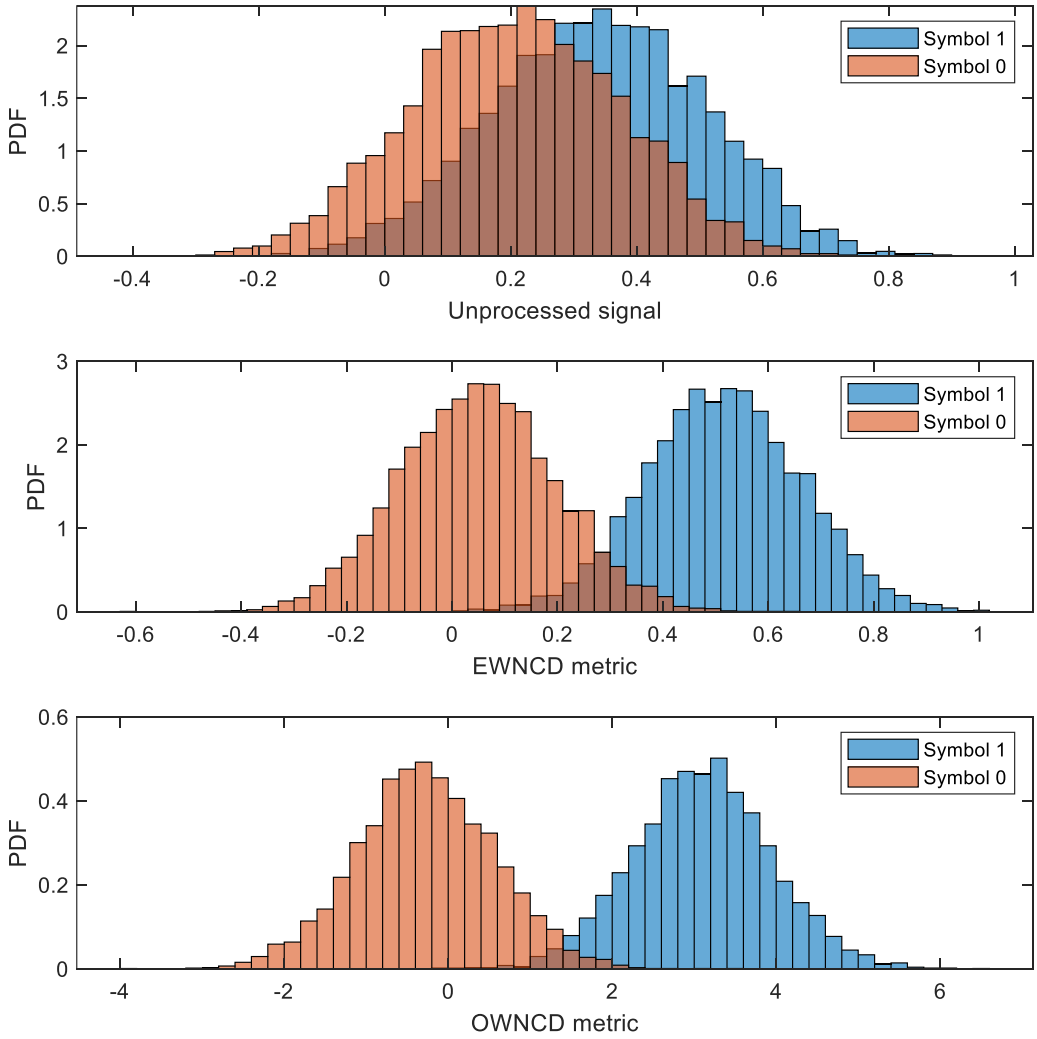


Figure 4.8: Illustration of statistic histograms of unprocessed signals, proposed EWNCD and OWNCD metrics.

means the proposed EWNCD and OWNCD schemes are capable of promoting detection accuracy, and OWNCD can lead to an extra promotion.

B) BER performance evaluation in static channel

In this part, BER performance and computational complexity are conducted to evaluate the communication performance of EWNCD and OWNCD schemes,in the case with a static CIR, while the BER performance comparison in the face

of changing CIR will be pursued in next part.

First, the detection accuracy of the proposed scheme with respect to different levels of ISI intensity is evaluated. For this investigation, the SNR is kept constant at 18 dB, and the symbol rate changed R_b to reflect different intensities of ISI.

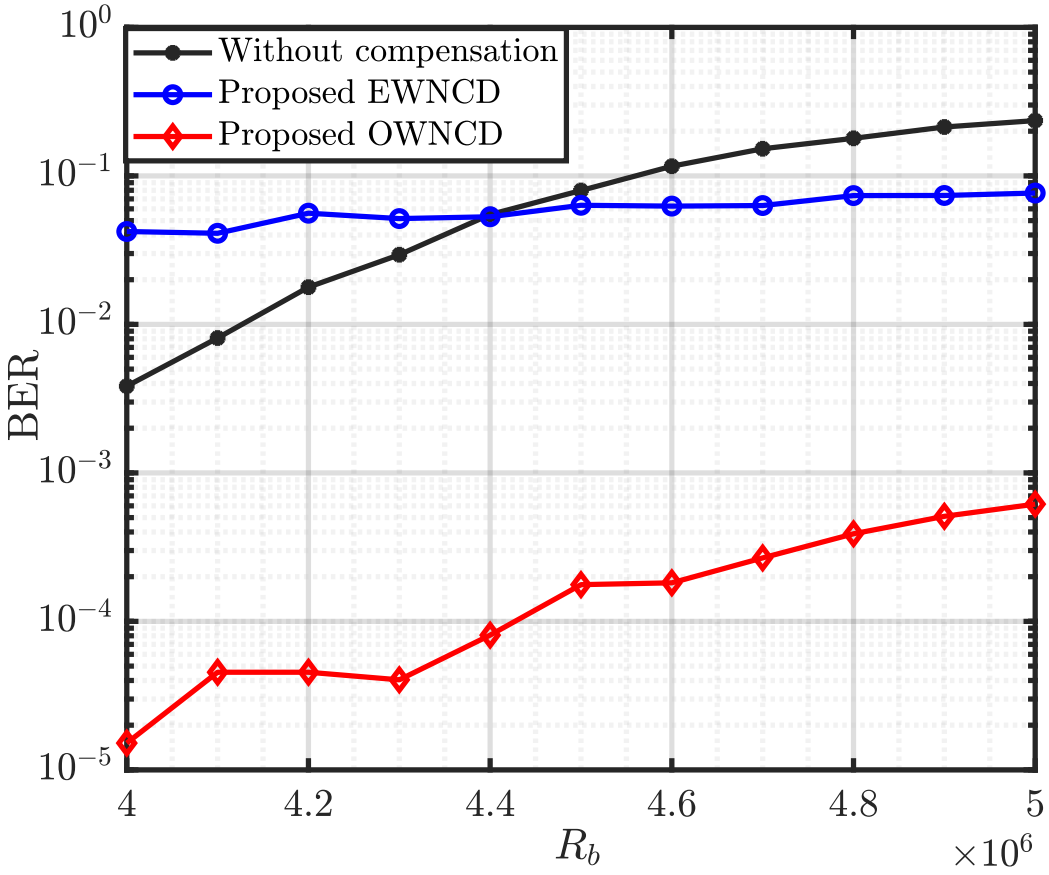


Figure 4.9: BER versus different intensities of ISI (whose increase is reflected by the increase of symbol rate R_b). The slower growth of BER with the increase of the ISI indicates the abilities of the proposed OWNCD to combat ISI.

In Fig. 4.9, it is observed that the proposed OWNCD improves the BER, by at least two orders of magnitude, compared to the direct detection and the proposed EWNCD method. Such BER advantage over direct detection is attributed to the design of ISI-insensitive features, the combination of which enables a robust performance to ISI contamination. Then, when it comes to

the comparison with EWNCD method, the BER advantage comes from the proposed OWNCD can derive the optimal weights, which indicates its ability to appropriately adjust the four metrics in accordance with the time-varying channel environment.

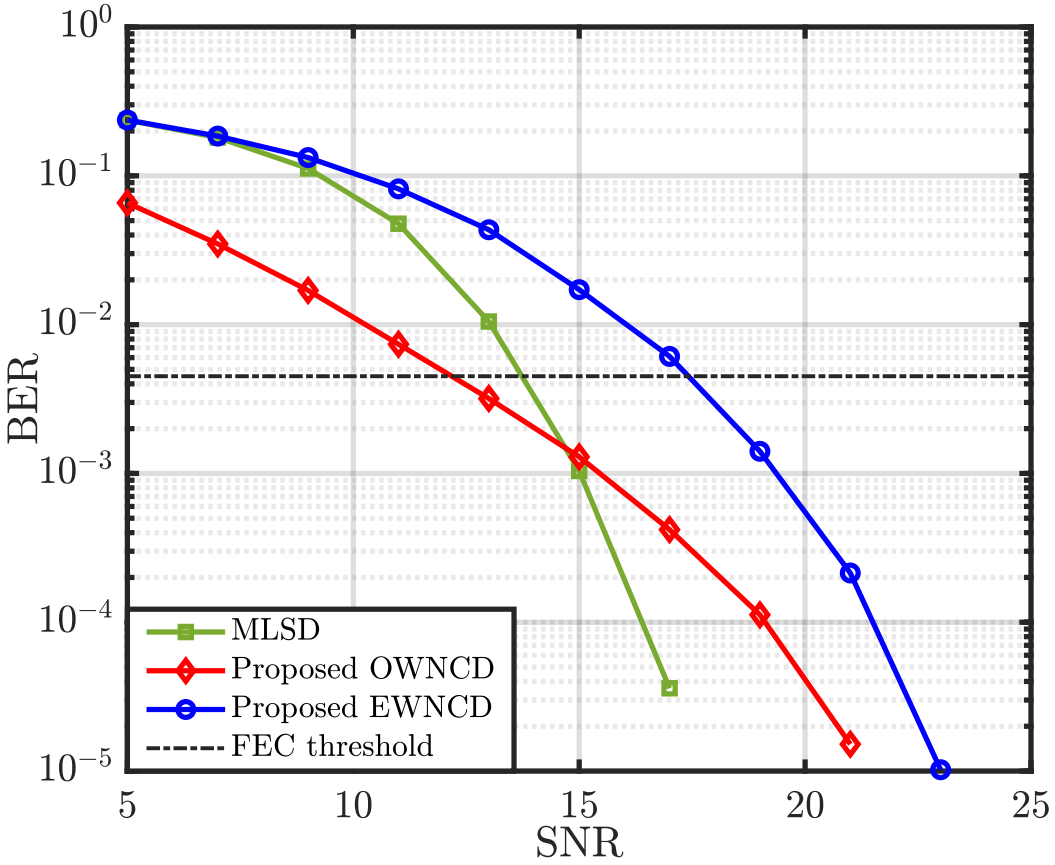


Figure 4.10: BER comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1] in a static channel.

It is seen from Fig. 4.10 that, the proposed OWNCD gains $\sim$ 4 dB in SNR compared to the EWNCD method. This is for the same reasons that the OWNCD scheme considers more crucially assigns optimal weights to each feature with respect to the specific ISI intensity rather than equal weights as in EWNCD method.

Then, compared to the state-of-the-art MLSD, the proposed OWNCD

provides a lower BER in low SNR region (i.e., SNR $<$ 15 dB), but higher BER when SNR is large. For the high SNR regions, the simulation result matches the common perception that the MLSD method approximates the BER benchmark in signal detection, by maximizing the likelihood. It is because in the high SNR regions, channel estimation can achieve more accurate CSI than the low SNR regions. Based on the accurate CSI, MLSD can obtain its best detection performance. However, in the low SNR regions, the OWNCD outperforms the MLSD. This phenomenon is caused by a combination of three factors. Firstly, the MLSD method heavily relies on channel estimation accuracy, while the proposed OWNCD scheme does not need the CSI. In the low SNR regions, the transmitted signals suffer from strong ambient noises. As channel estimation is vulnerable to ambient noise, it leads to inaccuracy channel estimation results and consequently influences the detection accuracy. Second, as a sequence detection method, MLSD method has to face error propagation. Especially in this situation with inaccuracy CSI, error propagation gives rise to further misjudgement. On the contrary, OWNCD method transforms the ISI sequential detection into discrete and binary detection framework, and therefore does not rely on channel estimation, and is free from error propagation as a discrete and binary detection method. Furthermore, the extracted features of the proposed OWNCD can combat ISI inherently, which guarantees it is less sensitive to ISI compared with MLSD.

However, it is noted that although the coherent MLSD shows a better performance in the high SNR region, the proposed OWNCD can reach the 7% overhead hard-decision FEC limit (BER = 4.5×10^{-3}) [57] faster than MLSD (by more than 1 dB). This indicates that with a proper FEC code, the OWNCD can achieve an error-free transmission faster compared to the coherent MLSD.

C) BER performance evaluation in time-varying channel

A BER comparison with changing CIR is illustrated in Fig. 4.11. Here, the time-varying CIR is performed by adding process randomness to the channel parameters, e.g., $r(k) = r(k-1) + \Delta_r$, and $\theta_R(k) = \theta_R(k-1) + \Delta_{\theta_R}$, indicating the slight disturbance given the flexible deployment of UVC.

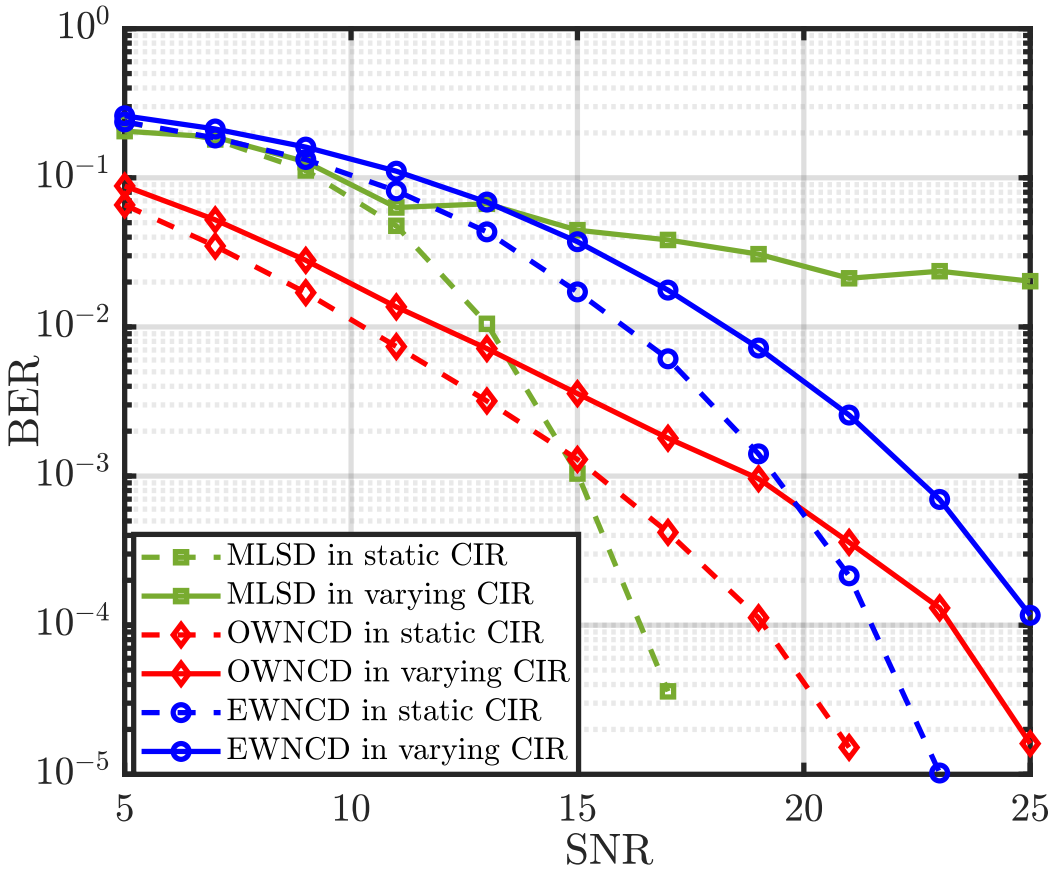


Figure 4.11: BER comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1], in a time-varying channel.

In Fig. 4.11, it may be observed that the proposed OWNCD is robust in the case of the time-varying CIR, since its BER does not deteriorate drastically when compared with static CIR. This is mainly attributed to the heuristic algorithm, which updates the optimal weights and the threshold by the previous data that can reflect the continuous changes in the CIR. Meanwhile,

it can be observed that the coherent MLSD performs unreliably in dealing with the continuously varying channel (>8 dB loss in SNR is shown compared to the proposed OWNCD). This is because MLSD relies heavily on the estimated CIR from the pilot sequence, which may be unreliable in the face of the time-varying channel environment, thereby resulting in unattractive detection accuracy.

As such, the proposed OWNCD is more suitable for counteracting the time-varying channel environment, which is closer to real UVC scenarios. Also, compared with the coherent schemes, the proposed OWNCD does not rely on complex channel estimation of the CIR, making them capable of saving computational and storage resources.

D) Complexity comparison

Moreover, when combined with the cost in Fig. 4.12, it can be seen that the proposed OWNCD has a great advantage in terms of both the computational complexity and storage resources. The proposed OWNCD has a complexity of the order of 10^6, which is much less than that of the MLSD (10^{10}). Meanwhile, the storage consumption required for the proposed scheme is smaller compared with the coherent MLSD (around 10^4 versus 10^6).

Thus, the proposed OWNCD provides a low-complexity and low-storage-based detection paradigm for the UVC scenarios, which can approach the detection accuracy of the coherent MLSD but with the benefit of lower computational resources.

4.6 Conclusions & Discussions

For most sequential optical wireless detection scenarios, signal detection without explicit channel information is challenging. Current bottlenecks lie in how

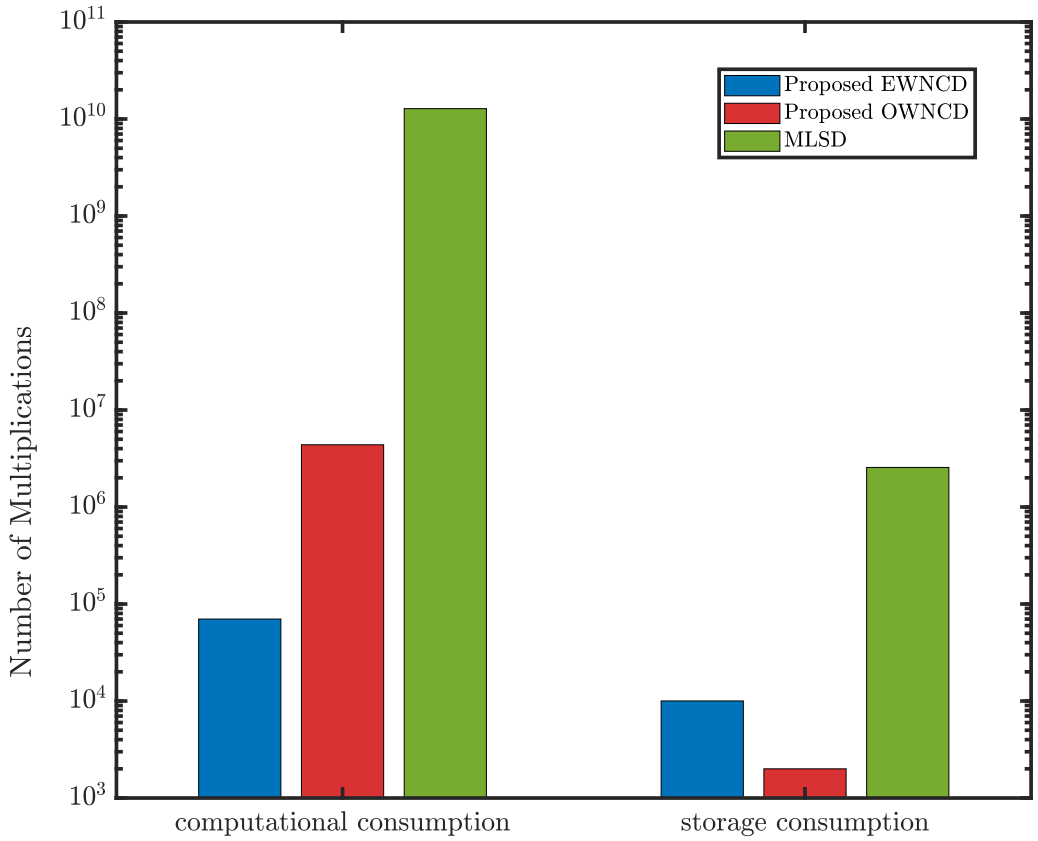


Figure 4.12: complexity comparison between the proposed OWNCD, EWNCD, and the state-of-the-art coherent MLSD [1] in a static channel, based on 10^4 bits.

to address the ISI from the long-tail and time-varying channel response. In this work, a novel non-coherent paradigm has been proposed by exploiting the inherent features, which are innately insensitive to ISI, of UV signals. Leveraged on the features, the OWNCD is developed, which enables the conversion of signal detection with ISI into a binary detection framework. It is found that, compared to the state-of-the-art coherent MLSD, the proposed OWNCD scheme improves the detection accuracy by $\sim$1 dB for the static CSI and $\sim$8 dB for the time-varying CSI, in terms of SNR at the 7% overhead FEC limit i.e., BER=4.5×10^{-3}. In addition, it reduces the computational and storage

resources by four orders of magnitude. As such, the proposed feature-based non-coherent technique provides a promising pathway and novel insight for the efficient detection in complex and dynamic optical channels with intensive ISI.

Chapter 5

High-Dimensional Non-Coherent Detection for Multi-Intensity Modulated Communications

In the last Chapter, four ISI insensitive features of the UV signals' geometric shapes (e.g., rising edge and inflection) were extracted and exploited, showing a reliable performance to counteract the intensive ISI. However, the linear combination of the features does not fully employ the signal's properties and results in a loss of SNR and BER, e.g., a $>$ 5dB loss in SNR when comparing with the MLSD having exact CIR estimation. Besides, the linearly combined 1D non-coherent method is only suitable for OOK modulation systems.

In this chapter, a novel high-dimensional (HD) feature-based non-coherent detection is proposed, which leads to a more accurate detection by constructing a high-dimensional feature space. Moreover, the HD method can be deployed in J-PAM communication systems. In section 5.1, there is an illustration of the system model for the HD scheme along with the problem formulation.

Then, the construction of HD UV features will be illustrated in section 5.2. Next, Section 5.4 will compute and compare the theoretical SER from this proposed HD non-coherent scheme with the 1D methods proposed in Chapter 4. In Section 5.5, the proposed HD non-coherent detection scheme is elaborated with the algorithm design. The simulation results are provided in Section 5.6. This chapter finally concludes in Section 5.7.

5.1 System Models and Problem Formulation

In this part, there is an introduction about the framework of the HD non-coherent detection, as seen in Fig. 5.1.

The HD non-coherent detection system is based on the basic UVC system, introduced by Chapter 3. Different from the 1D non-coherent methods, the HD scheme is not only suitable for OOK but also the $(J > 2)$-PAM modulation. In Tx, the binary input is modulated by the $J \in \mathbb{N}^+$ pulse amplitude modulation (J-PAM), which then will be converted to different intensities of UV photons for transmission. At the Rx point, the HD non-coherent signal detection unit contains feature-based metric construction and clustering algorithm.

Given the expressions of the received signals in Eq. (3.4), the aim of this work is to detect the current J-PAM s_k from the received signals $r_1, ..., r_{kM}$. The main challenge lies in how to combat the ISI with the unknown CIR coefficients $q_{i-(l-1)M}$, which, if multiplied by the multi-PAM symbols, i.e., $q_{i-(l-1)M} \cdot s_l$, makes the ISI more severe. To cope with this, coherent detection schemes (e.g., MLSD [1,39] and particle-filter methods [40]) demand heavy pilot overheads for CIR estimation, and high computational complexity to search the J^L symbol space for *a posteriori* computation and symbol detection (given J-PAM with L length ISI). Non-coherent schemes are all designed for binary

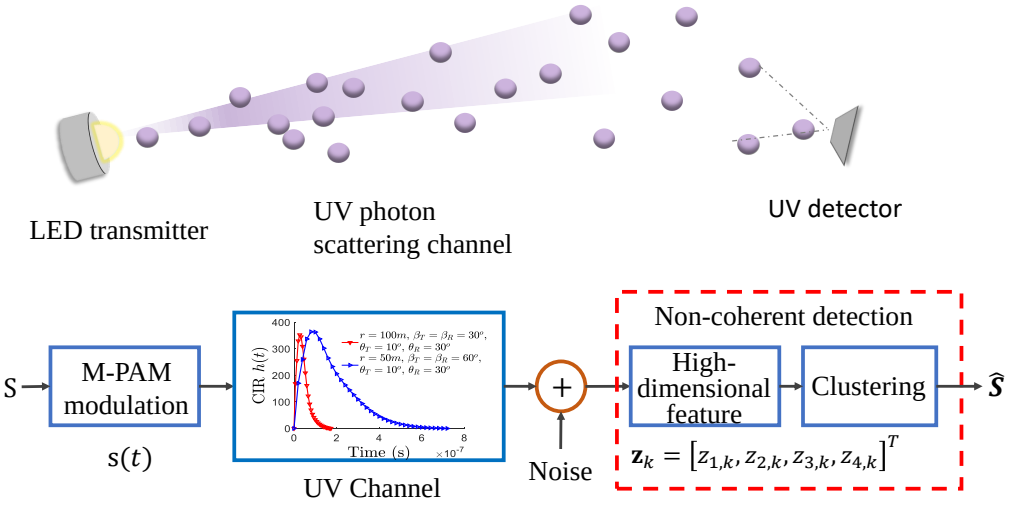


Figure 5.1: Schematic flow illustrates the UVC system and the High-Dimensional non-coherent detection scheme.

inputs [76], and offer poor detection accuracy when dealing with multi-symbol detection. This thereby motivates us to design a pilot-less, computationally efficient and more reliable non-coherent detection scheme for J-PAM UVC systems.

5.2 HD Feature Construction

A D-dimensional feature space $\mathbb{R}^D$ is constructed via the previously designed UV sub-features, i.e.,

$$\mathbf{z}_k = [z_{1,k}, z_{2,k} \cdots, z_{D,k}]^T \in \mathbb{R}^D. \tag{5.1}$$

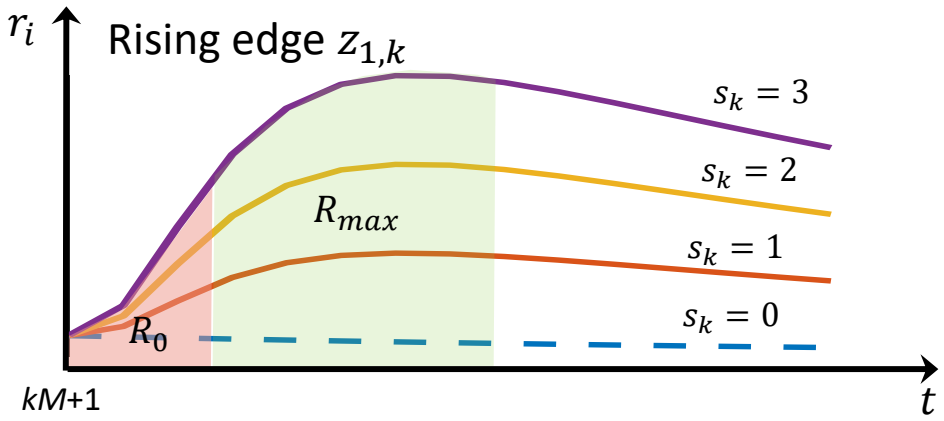


Figure 5.2: Illustrations of the feature-based metrics: the intra-symbol rising-edge, in 4-PAM mode.

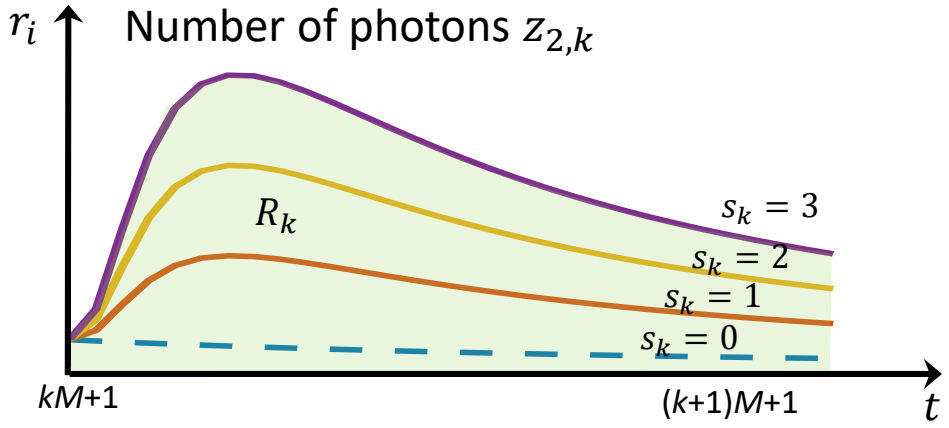


Figure 5.3: Illustrations of the feature-based metric: the intra-symbol energy, in 4-PAM mode.

Here, $D = 4$ of the previously designed UV sub-features are used, seen in Figs. 5.2-5.5:

$$z_{1,k} \triangleq \frac{1}{|\mathcal{R}_{\max}|} \sum_{i \in \mathcal{R}_{\max}} r_i - \frac{1}{|\mathcal{R}_0|} \sum_{i \in \mathcal{R}_0} r_i, \tag{5.2a}$$

$$z_{2,k} \triangleq \frac{1}{|\mathcal{R}_k|} \sum_{i \in \mathcal{R}_k} r_i, \tag{5.2b}$$

$$z_{3,k} \triangleq -\frac{1}{|\mathcal{R}_{\text{inflx}}|} \sum_{i \in \mathcal{R}_{\text{inflx}}} r_i + \frac{1}{2} \left(\frac{\sum_{i \in \mathcal{R}_{\text{left}}} r_i}{|\mathcal{R}_{\text{left}}|} + \frac{\sum_{i \in \mathcal{R}_{\text{right}}} r_i}{|\mathcal{R}_{\text{right}}|} \right), \tag{5.2c}$$

$$z_{4,k} \triangleq \frac{1}{|\mathcal{R}_k|} \sum_{i \in \mathcal{R}_k} r_i - \frac{1}{|\mathcal{R}_{k-1}|} \sum_{i \in \mathcal{R}_{k-1}} r_i, \tag{5.2d}$$

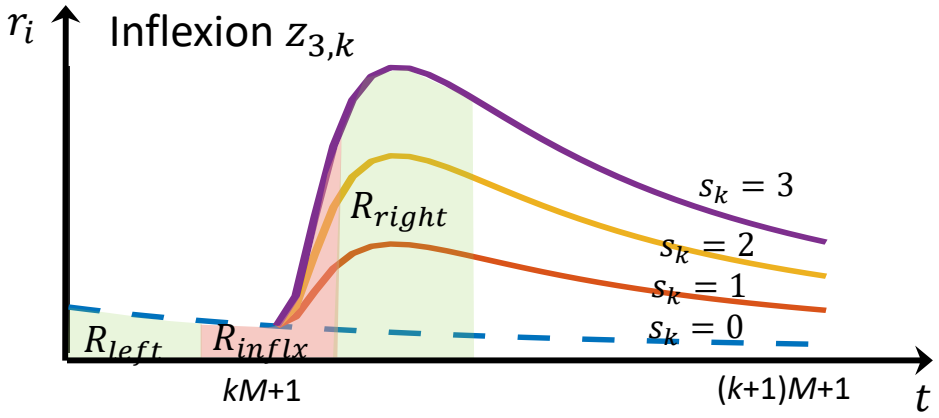


Figure 5.4: Illustrations of the feature-based metric: the inter-symbol minimum inflection, in 4-PAM mode.

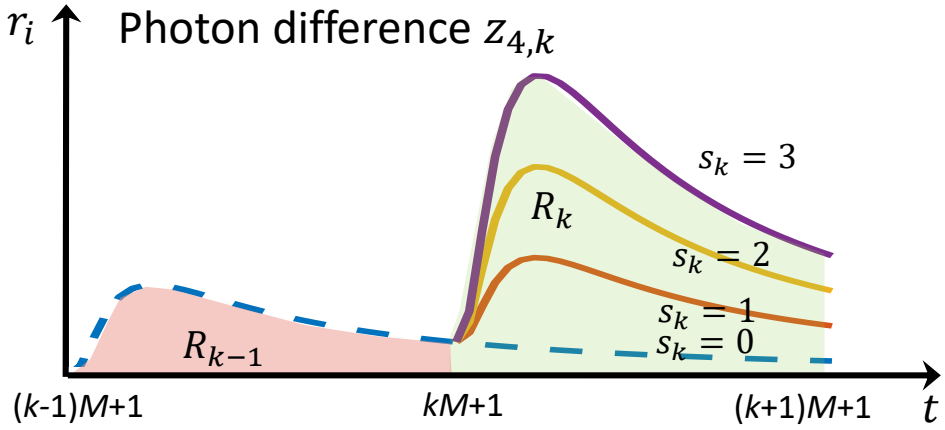


Figure 5.5: Illustrations of the feature-based metric: the inter-symbol energy difference, in 4-PAM mode.

In Eq. (5.2a), the sub-feature $z_{1,k}$ is referred to the rising edge of kth symbol-interval, whose value is ISI-insensitive but only depends on the different PAM input symbols s_k. $\mathcal{R}_{\max} \triangleq (k-1)M + \{i_{\max} - \lfloor M/4 \rfloor, \cdots, i_{\max} + \lfloor M/4 \rfloor\}$ and $\mathcal{R}_0 \triangleq (k-1)M + \{1, \cdots, \lfloor M/4 \rfloor\}$ are set to determine the related index areas for rising-edge computation. $i_{\max}$ is the index corresponding to the maximum value of discrete CIRs. Approximation of $i_{\max}$ is by using the index of the maximum value of the non-zero symbols related to received signals in training data. Here, $i_{\max}$ does not need to be very accurate, since $\mathcal{R}_{\max}$ is

defined to include the adjacent area of $i_{\max}$ for rising edge computation. In Eq. (5.2b), $z_{2,k}$ is referred to as the number of received photons within kth symbol-interval, where $\mathcal{R}_k \triangleq \{(k-1)M+1, \cdots, kM\}$. In Eq. (5.2c), $z_{3,k}$ accounts for the inflection feature, whose value is also ISI-insensitive and increases with a larger symbol pulse s_k. Here, the inflection feature is computed by the left and right neighbours' values subtracting the central area's values. The quantities below are hereby defined $\mathcal{R}_{\text{inflx}} \triangleq (k-1)M+\{-\lfloor M/4 \rfloor, \cdots, \lfloor M/4 \rfloor\}$, $\mathcal{R}_{\text{left}} \triangleq (k-1)M+\{-\lfloor M/2 \rfloor, \cdots, -\lfloor M/4 \rfloor - 1\}$, and $\mathcal{R}_{\text{right}} \triangleq (k-1)M+\{\lfloor M/4 \rfloor + 1, \cdots, \lfloor M/2 \rfloor\}$. $z_{4,k}$ gives the received photon number difference between successive symbol intervals k and $k-1$.

From Eq. (5.2) and the corresponding sub-feature shapes in Figs. 5.2-5.5, each of the sub-features is able to transform the received ISI contaminated signals r_i into a new detection space that is insensitive to ISI. For instance, given a fixed ISI intensity and any of s_k, the expectations of $z_{k,1}$ $z_{k,2}$, $z_{k,3}$ and $z_{k,4}$ remain quasi-constant, as they reflect only the common geometric shapes of the UV signals (e.g., the rising edge and the inflection). In this view, by stacking them into an HD signal feature, i.e., $\mathbf{z}_k = [z_{1,k}, \cdots, z_{4,k}]^T$, it is possible to counteract the ISI effect, and convert the sequential detection with ISI into an ISI-insensitive detection framework (see Fig. 5.7(b)). Also, differing from the previous 1D feature, i.e., $z_k = \sum_{d=1}^{D} w_d z_{d,k}$, the HD feature $\mathbf{z}_k$ can provide an HD decision surface and result in a lower SER for signal detection.

With the help of the HD feature $\mathbf{z}_k$ in Eqs. (5.1)-(5.2), the likelihood PDF of $\mathbf{z}_k$ conditioned on different J-PAM symbol $s_k = j \in \{0, 1, \cdots, J-1\}$

is provided, i.e.,

$$\begin{aligned} & p(\mathbf{z}_k \big| s_k = j) \\ =& \mathcal{N}(\mathbf{z}_k; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}) \\ \triangleq& \frac{1}{\sqrt{(2\pi)^d \cdot det(\boldsymbol{\Sigma})}} \exp\left(-0.5\left(\mathbf{z}_k - \boldsymbol{\mu}_j\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{z}_k - \boldsymbol{\mu}_j\right)\right), \end{aligned} \tag{5.3}$$

where $\boldsymbol{\mu}_j = [\mu_{1,j} \cdots, \mu_{D,j}]^T$ represents the expectation of $\mathbf{z}_k = [z_{1,k}, \cdots z_{D,k}]^T$ conditioned on $s_k = j$, $\boldsymbol{\Sigma}$ gives the covariance matrix (of size $D \times D$) among the total $D = 4$ features in Eq. (5.2), and $det(\cdot)$ is the determinant operator of a matrix. As such, the signal detection process is converted to determine $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$, and select $s_k \in \{0, 1, \cdots, J-1\}$ that corresponds to the maximal likelihood PDF $p(\mathbf{z}_k \big| s_k = j)$, which can be viewed as a clustering problem, see Fig. 5.7(b) for illustration.

5.3 Theoretical Symbol Rate Error

The theoretical SER of the HD feature $\mathbf{z}_k$ will be computed and studied in this part. This is done by firstly computing the theoretical BER in the OOK symbol case, which then will be extended for J-PAM symbols.

The first case to consider is the binary one with OOK modulation. According to the Bayesian rule, the detection of transmitted symbols is equivalent to computing and comparing the posterior PDFs of $\mathbf{z}_k$ corresponding to different transmitted symbols of $s_k = 0$ and $s_k = 1$, i.e.,

$$\frac{p(s_k = 1|\mathbf{z}_k)}{p(s_k = 0|\mathbf{z}_k)} = \frac{P(s_k = 1) \cdot p(\mathbf{z}_k|s_k = 1)}{P(s_k = 0) \cdot p(\mathbf{z}_k|s_k = 0)} \overset{\text{(a)}}{=} \frac{p(\mathbf{z}_k|s_k = 1)}{p(\mathbf{z}_k|s_k = 0)} \underset{\hat{s}_k = 0}{\overset{\hat{s}_k = 1}{\gtrless}} 1. \tag{5.4}$$

In Eq. (5.4), $p(s_k|\mathbf{z}_k)$ represents the posterior PDF. (a) accounts for the fact the probabilities of sending 1-bit and 0-bit are identical, i.e., $P(s_k = 0) =$

$P(s_k = 1) = 0.5$.

The computation of the theoretical BER requires the decision surface from Eq. (5.4), which can be transformed as the logarithm difference i.e.,

$$\ln p(\mathbf{z}_k|s_k = 1) - \ln p(\mathbf{z}_k|s_k = 0) = 0, \tag{5.5}$$

Taken Eq. (5.3) into the left-hand side of Eq. (5.5), the decision surface is defined and then computed as:

$$\mathcal{H}(\mathbf{z}_k) \triangleq (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_0^T)\boldsymbol{\Sigma}^{-1}\mathbf{z}_k - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0) = 0. \tag{5.6}$$

Given the expression of the decision surface in Eq. (5.6), the theoretical BER for OOK symbols can be computed via two parts of error probabilities, i.e.,

$$\begin{aligned}
BER_{OOK}(\mathbf{z}_k) &= \overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{z})\geqslant 0} p(\mathbf{z}|s_k = 0)d\mathbf{z} + \overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{z})<0} p(\mathbf{z}|s_k = 1)d\mathbf{z} \\
&= 2\overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{z}+\boldsymbol{\mu}_0)\geqslant 0} \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}\right) dz_1 z_2 \cdots z_D \\
&\overset{(a)}{=} 2\overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{Q}\cdot\mathbf{x}+\boldsymbol{\mu}_0)\geqslant 0} \frac{|\mathbf{J}|}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Lambda}^{-1}\mathbf{x}\right) dx_1 x_2 \cdots x_D \\
&\overset{(b)}{=} 2\overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{Q}\cdot\boldsymbol{\Lambda}^{0.5}\cdot\mathbf{u}+\boldsymbol{\mu}_0)\geqslant 0} \frac{1}{\sqrt{(2\pi)^D}} \exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right) du_1 u_2 \cdots u_D \\
&\overset{(c)}{=} 2\cdot\Phi\left(-\frac{1}{2}\sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}\right).
\end{aligned} \tag{5.7}$$

The computation of Eq. (5.7) is explained in the following.

(a) It is noticed that $\mathbf{z}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{z}$ is the quadratic form, with a positive-

definite symmetric covariance matrix $\mathbf{\Sigma}$ of size $D \times D$. As such, eigen-decomposition can be performed as $\mathbf{\Sigma} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{-1} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{T}$, where $\mathbf{\Lambda} = diag\{[\lambda_1, \cdots, \lambda_D]\}$ is the composition of D different eigenvalues of $\mathbf{\Sigma}$, and the columns of $\mathbf{Q}$ are composed of the corresponding D normalised eigenvectors. Then, $\mathbf{z}^T \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{z} = \mathbf{x}^T \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{x}$ by denoting $\mathbf{z} = \mathbf{Q} \cdot \mathbf{x}$. Next, to transform the integral space from $\mathbf{z}$ to $\mathbf{x}$, the determinant of the Jacobian matrix is needed, i.e.,

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_d} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_d}{\partial x_1} & \frac{\partial z_d}{\partial x_2} & \cdots & \frac{\partial z_d}{\partial x_d} \end{vmatrix} = |\mathbf{Q}| = 1. \tag{5.8}$$

(b) Further expressing $\mathbf{u} = \mathbf{\Lambda}^{-0.5}\mathbf{x}$ to replace $\mathbf{x}$, the equation of (ii) can be obtained by offsetting the determinant of the Jacobian matrix and $|\mathbf{\Sigma}|$, i.e., $|\mathbf{\Sigma}| = |\mathbf{\Lambda}|$.

(c) Before explaining the reason that holds for(c), the following equation is proved with specific proof, i.e.,

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot \Phi\left(az + b\right) dz \equiv \Phi\left(\frac{b}{\sqrt{1 + a^2}}\right), \tag{5.9}$$

where

$$\Phi(x) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \tag{5.10}$$

is the cumulative probability function (CDF) of the normal distribution $\mathcal{N}(0, 1)$,

at the given point x. This is because

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot \Phi\left(az+b\right) dz \\
=&\int_{-\infty}^{b} \frac{\partial\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot \Phi\left(az+x\right) dz\right)}{\partial x} dx \\
=&\int_{-\infty}^{b} \frac{1}{\sqrt{2\pi(1+a^2)}} \exp\left(-\frac{x^2}{2(1+a^2)}\right) dx \\
=&\Phi\left(\frac{b}{\sqrt{1+a^2}}\right).
\end{aligned} \tag{5.11}$$

Then, by taking Eq. (5.9) back to (b) of Eq. (5.6), proof of (c) can be

completed by the following deduction, i.e.,

$$
\begin{aligned}
&2\cdot\overbrace{\int\cdots\int}^{D}_{\mathcal{H}(\mathbf{Q}\cdot\mathbf{\Lambda}^{0.5}\cdot\mathbf{u}+\boldsymbol{\mu}_0)\geqslant 0}\frac{1}{\sqrt{(2\pi)^D}}\exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right)du_1\cdots u_D\\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{D-1}^2}{2}\right)\\
&\cdot\int_{-\frac{\sum_{i=1}^{D-1}\xi_i\cdot u_i-B}{\xi_D}}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_D^2}{2}\right)du_1\cdots du_D\\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{D-1}^2}{2}\right)\\
&\cdot\Phi\left(\frac{\sum_{i=1}^{D-1}\xi_i\cdot u_i-B}{\xi_D}\right)du_1\cdots du_{D-1}\\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{D-1}^2}{2}\right)\\
&\cdot\Phi\left(\frac{\xi_{D-1}}{\xi_D}u_{D-1}+\sum_{i=1}^{D-2}\frac{\xi_i}{\xi_D}u_i-\frac{B}{\xi_D}\right)du_1\cdots du_{D-1}\\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{D-2}^2}{2}\right)\\
&\cdot\Phi\left(\frac{\sum_{i=1}^{D-2}\xi_i u_i-B}{\sqrt{\xi_D^2+\xi_{D-1}^2}}\right)du_1\cdots du_{D-2}\\
=&\cdots=2\cdot\Phi\left(-\frac{B}{\sum_{i=1}^{D}\xi_i^2}\right)=2\cdot\Phi\left(-\frac{0.5\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}}{\boldsymbol{\xi}^T\boldsymbol{\xi}}\right)=2\cdot\Phi\left(-\frac{1}{2}\sqrt{\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}}\right),
\end{aligned}
\tag{5.12}
$$

where $\boldsymbol{\xi}=[\xi_1,\ldots,\xi_D]^T=\mathbf{\Lambda}^{-0.5}\mathbf{Q}^{-1}\cdot(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)$, and $B=0.5\cdot\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}$.

After the proof, by moving back to Eq. (5.7), the general expression of BER for any designed D-dimensional feature $\mathbf{z}_k$ is provided, which shows the insight to extend it to the $J>2$-PAM symbol cases. Similar to the binary

case (e.g., OOK), the theoretical SER for the J-PAM symbol is expressed by the summation of the detection error probability for each symbol, i.e., $p(\hat{s}_k \neq m | s_k = m)$, which then can be expressed in term of the SNR given the Gaussian typed likelihood PDF in Eq. (5.1). As such, the theoretical SER is provided as follows:

$$\begin{aligned} & SER(\mathbf{z}_k) \\ =& \sum_{j=0}^{J-1} P\left(\hat{s}_k \neq j | s_k = j\right) \\ =& \sum_{j=0}^{J-1} \Phi\left(-0.5\sqrt{\mathrm{SNR}}\right) \\ =& \sum_{j=0}^{J-1} \Phi\left(-0.5\sqrt{\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}\right)}\right) \end{aligned} \tag{5.13}$$

5.4 Proof of Better than Lower-Dimensional Methods

In this part, the aim is to prove that the detection accuracy (in terms of SER) based the higher dimensional features outperforms those with lower dimensional features. Specially, the HD non-coherent detection scheme proposed in this chapter is better than the 1D OWNCD method.

5.4.1 Better SER than 1D feature

This is to prove that the proposed HD non-coherent detection can provide a lower SER compared with the 1D non-coherent scheme, i.e.,

$$SER(\mathbf{z}_k) < SER(\mathbf{w}^T \cdot \mathbf{z}_k), \tag{5.14}$$

where the latter is the SER of the 1D non-coherent scheme with $\mathbf{w} = [w_1, w_2, \cdots, w_D]^T$ the optimal weights recalled in Chapter 4. By taking Eq. (5.13) into Eq. (5.14) and using the monotonically increasing property of $\Phi(x)$, a stricter inequality can be proved as follows:

$$\boldsymbol{\varrho}^T \cdot \left(\boldsymbol{\Sigma}^{-1} - \mathbf{w} \left(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \right)^{-1} \mathbf{w}^T \right) \cdot \boldsymbol{\varrho} \geq 0, \tag{5.15}$$

where $\boldsymbol{\varrho} = \boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1},\ \forall j \in \{0, \cdots, J-1\}$. Given the symmetric positive definite property of the covariance matrix $\boldsymbol{\Sigma}$, the eigenvalue decomposition is $\boldsymbol{\Sigma} = \mathbf{Q} \cdot diag([\lambda_1, \cdots, \lambda_D]^T) \cdot \mathbf{Q}^T$. Then, by assigning $\mathbf{v} = [v_1, \cdots, v_D]^T = \mathbf{Q}^T \mathbf{w}$ and $\mathbf{x} = [x_1, \cdots, x_D]^T = \mathbf{Q}^T \boldsymbol{\varrho}$, by rewriting the left-hand side of Eq. (5.15), it remains to prove that it is no less than 0, i.e.,

$$\begin{aligned} & \mathbf{x}^T \cdot \left(diag\left(\left[\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_D} \right]^T \right) - \frac{\mathbf{v}\mathbf{v}^T}{\sum_{d \in \mathcal{D}} \lambda_d \cdot v_d^2} \right) \cdot \mathbf{x} \\ = & \sum_{d \in \mathcal{D}} \sum_{d' \in \mathcal{D} \setminus \{d\}} \lambda_{d'} v_{d'}^2 \cdot \frac{x_d^2}{\lambda_d} - 2 \sum_{d, d' \in \mathcal{D}, d \neq d'} v_d v_{d'} x_d x_{d'} \\ = & \sum_{d, d' \in \mathcal{D}, d \neq d'} \left(\frac{\lambda_{d'} v_{d'}}{\lambda_d} x_d - \frac{\lambda_d v_d}{\lambda_{d'}} x_{d'} \right)^2 \geq 0. \end{aligned} \tag{5.16}$$

As such, it has been proved that the proposed HD feature-based non-coherent detection scheme has a lower SER as opposed to that of the 1D non-coherent scheme. This will be further demonstrated in Fig. 5.7, and via the Simulation part.

5.4.2 Better SER than Lower Dimensional feature

The proof is next generalised to show that the SER of high-dimensional features is lower than that of lower-dimensional features. SER comparison is performed between the non-coherent schemes with the D-dimensional feature $\mathbf{z}_k$, and a lower $(D_1 \leq D)$-dimensional feature, i.e., $\mathbf{H} \cdot \mathbf{z}_k$, where $\mathbf{H}$ is the

construction matrix of size $D_1 \times D$ with $rank(\mathbf{H}) = D_1 < D$. The comparison result is:

$$SER(\mathbf{z}_k) \leq SER(\mathbf{H} \cdot \mathbf{z}_k), \tag{5.17}$$

where the equality holds when $D_1 = D$.

The proof of Eq. (5.17) provided in the following. Given the SER expression in Eq. (5.13), and the monotonically increasing property of Gaussian CDF $\Phi(x)$ with respect to x, the comparison between $SER(\mathbf{z}_k)$ and $SER(\mathbf{H} \cdot \mathbf{z}_k)$ can be converted as follows:

$$\begin{aligned} & SER(\mathbf{z}_k) - SER(\mathbf{H} \cdot \mathbf{z}_k) \leq 0 \\ \Leftarrow \quad & \left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}\right)^T \left(\boldsymbol{\Sigma}^{-1} - \mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T)^{-1}\mathbf{H}\right)\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}\right) > 0 \\ & \qquad\qquad \forall j \in \{0, 1, \cdots, J-1\}, \\ \Leftrightarrow \quad & \boldsymbol{\Sigma}^{-1} - \mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T)^{-1}\mathbf{H} \text{ is positive semidefinite} \end{aligned} \tag{5.18}$$

where $\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T$ gives the covariance matrix of $\mathbf{H}\mathbf{z}_k$.

First, when $rank(\mathbf{H}) = D_1 = D$, then Eq. (5.18) can be computed as the zero matrix, i.e.,

$$\boldsymbol{\Sigma}^{-1} - \mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T)^{-1}\mathbf{H} = \boldsymbol{\Sigma}^{-1} - \mathbf{H}^T \cdot \left(\mathbf{H}^T\right)^{-1} \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{H}^{-1} \cdot \mathbf{H} = \mathbf{O}. \tag{5.19}$$

For $D_1 < D$, only $D_1 = D - 1$ is studied, as other cases $D_1 < D - 1$ can be recursively proved by replacing D as $D - 1$. First, $\mathbf{H} = \mathbf{H}_1 \cdot \mathbf{Q}^T$ is divided (recalling $\mathbf{Q}$ is the matrix composed of D eigenvectors of $\boldsymbol{\Sigma}$, and therefore $rank(\mathbf{Q}) = D$), with $\mathbf{H}_1 = \mathbf{H} \cdot \mathbf{Q}$. Then, the lower dimensional feature can be re-written as $\mathbf{H} \cdot \mathbf{z}_k = \mathbf{H}_1 \cdot \mathbf{Q}^T\mathbf{z}_k$ whereby the covariance matrix of $\mathbf{Q}^T \cdot \mathbf{z}_k$ is the diagonal eigenvalue matrix of the original $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Lambda}$. Also, note that $\mathbf{H}_1 = \boldsymbol{\Pi} \cdot \mathbf{H}_2$, where $\boldsymbol{\Pi}$ is the multiplications of elementary row transformation operators ($rank(\boldsymbol{\Pi}) = D_1 = D - 1$), and $\mathbf{H}_2 = [\mathbf{I}_{D_1 \times D_1}\ \mathbf{a}]$

where $\mathbf{a} = [a_1, \cdots a_{D-1}]^T$. As such, by using $SER(\mathbf{Q}^{-1}\mathbf{z}_k) = SER(\mathbf{z}_k)$, and $SER(\mathbf{H}\cdot\mathbf{z}_k) = SER(\mathbf{\Pi}\cdot\mathbf{H}_2\cdot\mathbf{Q}^{-1}\mathbf{z}_k) = SER(\mathbf{H}_2\cdot\mathbf{Q}^{-1}\mathbf{z}_k)$, the proof is converted to proving $\mathbf{\Lambda}^{-1} - \mathbf{H}_2^T(\mathbf{H}_2\mathbf{\Lambda}\mathbf{H}_2^T)^{-1}\mathbf{H}_2$ is positive semidefinite. This will be done by showing that it has only non-negative eigenvalues:

$$\begin{aligned}
&\mathbf{\Lambda}^{-1} - \mathbf{H}_2^T\left(\mathbf{H}_2\mathbf{\Lambda}\mathbf{H}_2^T\right)^{-1}\mathbf{H}_2 \\
=&\mathbf{\Lambda}^{-1} - \mathbf{H}_2^T\left(diag([\lambda_1, \ldots, \lambda_{D-1}])^{-1} - \zeta\cdot\boldsymbol{\rho}\boldsymbol{\rho}^T\right)\mathbf{H}_2 \\
=&\zeta\cdot\begin{bmatrix}\boldsymbol{\rho}\boldsymbol{\rho}^T & -\boldsymbol{\rho}\\ -\boldsymbol{\rho}^T & 1\end{bmatrix} \\
=&\zeta\cdot\mathbf{\Upsilon}\cdot diag\left(\left[tr\left(\begin{bmatrix}\boldsymbol{\rho}\boldsymbol{\rho}^T & -\boldsymbol{\rho}\\ -\boldsymbol{\rho}^T & 1\end{bmatrix}\right), 0, \cdots 0\right]\right)\cdot\mathbf{\Upsilon}^{-1} \\
=&\mathbf{\Upsilon}\cdot diag\left(\left[\zeta\cdot tr\left(1+\boldsymbol{\rho}^T\boldsymbol{\rho}\right), 0, \cdots 0\right]\right)\cdot\mathbf{\Upsilon}^{-1},
\end{aligned} \tag{5.20}$$

where

$$\boldsymbol{\rho} = \left[a_1\frac{\sqrt{\lambda_d}}{\lambda_1}, a_2\frac{\sqrt{\lambda_d}}{\lambda_2}, \ldots, a_{d-1}\frac{\sqrt{\lambda_d}}{\lambda_{D-1}}\right]^T \tag{5.21}$$

$$\zeta = \frac{1}{1+\lambda_D\sum_{i=1}^{D-1}\frac{a_i^2}{\lambda_i}}, \tag{5.22}$$

with $\mathbf{\Upsilon}$, the eigenvector matrix. Therefore, Eq. (5.17) has been proved.

5.5 HD Non-Coherent Signal Detection Scheme

The initialization of $\boldsymbol{\mu}_j$ is leveraging the increasingly equal difference characteristics, i.e., $\boldsymbol{\mu}_{j-1} - 2\boldsymbol{\mu}_j + \boldsymbol{\mu}_{j+1} = \mathbf{0}$. This is because the increasingly equal difference among the J-PAM symbols, i.e., $j \in \{0, 1, \cdots, J-1\}$, remains after passing through the LTI signal model in Eq. (3.4), and the linear sub-feature constructions in Eq. (5.2). Here, this increasingly equal difference is charac-

terised by the following equations:

$$[\boldsymbol{\mu}_0, \cdots, \boldsymbol{\mu}_{J-1}] \cdot \boldsymbol{\Psi} = \mathbf{O}, \quad \boldsymbol{\mu}_j \prec \boldsymbol{\mu}_{j+1} \tag{5.23}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ & & & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}^T . \tag{5.24}$$

where $\prec$ is the elemental-wise less than operator, and matrix $\boldsymbol{\Psi}$ of size $J \times (J-2)$ is to ensure the equal difference.

The initialization of $\boldsymbol{\mu}_j$, denoted as $\boldsymbol{\mu}_j^{(0)} = [\mu_{1,j}^{(0)}, \mu_{2,j}^{(0)}, \cdots, \mu_{D,j}^{(0)}]^T$ is to implement Eqs. (5.23)-(5.24) via the first L' (unlabelled) features, i.e.,

$$z_{d,k_1} < \cdots < z_{d,k_{L'}}, \text{ with } \{k_1, \cdots, k_{L'}\} = \{1, \cdots, L'\} \tag{5.25a}$$

$$\mathcal{L}_{d,j} = \left\{k_{(j-1)L'/J+1}, \cdots, k_{jL'/J}\right\}, \tag{5.25b}$$

$$c_{d,j} = \frac{1}{|\mathcal{L}_{d,j}|} \sum_{k \in \mathcal{I}_{d,j}} z_{d,k}, \quad \mathbf{c}_j = [c_{1,j}, \cdots, c_{D,j}]^T. \tag{5.25c}$$

$$[\boldsymbol{\mu}_0^{(0)}, \cdots, \boldsymbol{\mu}_{J-1}^{(0)}] = [\mathbf{c}_0, \cdots, \mathbf{c}_{J-1}] \left([\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2]^\dagger\right)^T [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2]^T. \tag{5.25d}$$

Here, Eq. (5.25a) is to re-order the first L' sub-features $z_{d,1}, \cdots, z_{d,L'}$ into a ascending order. Eqs. (5.25b)-(5.25c) are to assign J sets (clusters) given the ascending order of z_{d,k_l}, and compute each cluster's centre (mean). Eq. (5.25d) is to ensure the increasingly equal difference of $\boldsymbol{\mu}_0, \cdots, \boldsymbol{\mu}_{J-1}$, or saying to ensure $[\boldsymbol{\mu}_0^{(0)}, \cdots, \boldsymbol{\mu}_{J-1}^{(0)}]^T$ in the null-space of $\boldsymbol{\Psi}$ according to Eq. (5.23). This is done by firstly computing the null-space of $\boldsymbol{\Psi}^T$ (of size $(J-2) \times J$), i.e., its two eigenvectors $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ corresponding to 0 eigenvalue. Then, the null-space basis is used to approximate the centre of each set by least square error, and assign the approximation as the initialization. As such, the initial $[\boldsymbol{\mu}_0^{(0)}, \cdots, \boldsymbol{\mu}_{J-1}^{(0)}]$ can be computed as Eq. (5.25d) following the increasingly

equal difference and also located around the centres of the feature set.

After the initialization, the following part is the clustering algorithm, where a heuristic clustering algorithm is designed, based on the GMM clustering framework.

5.5.1 Constrained GMM Clustering

The GMM model is selected as the basic clustering model, according to the characteristic of the extracted features. GMM is suitable for situations where the data distribution is a mixture of multiple Gaussian distributions. In Eq. (5.3), the likelihood PDF of $\mathbf{z}_k$ conditioned on J-PAM symbol is given, which is a multiple Gaussian distributions. Moreover, to overcome the issue of classic GMM potentially getting stuck in local optimal solutions, additional constraints based on the features are incorporated into the algorithm to ensure convergence to a more reasonable region.

After the initialization of $\boldsymbol{\mu}_j$, the GMM clustering [77, 78] framework is adopted to identify the expectation $\boldsymbol{\mu}_j$ and the covariance $\boldsymbol{\Sigma}$ of the UV feature $\mathbf{z}_k|(s_k = j)$ in Eq. (5.3). In essence, GMM clustering is to maximize the log-likelihood, i.e.,

$$\max_{\boldsymbol{\mu}_{0:J-1},\boldsymbol{\Sigma}} \mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1},\boldsymbol{\Sigma}), \tag{5.26}$$

where $\mathbf{z}_{1:L'} \triangleq [\mathbf{z}_1, \cdots, \mathbf{z}_{L'}]$, $\boldsymbol{\mu}_{0:J-1} \triangleq [\boldsymbol{\mu}_0, \cdots, \boldsymbol{\mu}_{J-1}]$. For this work, $\mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1},\boldsymbol{\Sigma})$

is computed as:

$$\mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1},\boldsymbol{\Sigma}) = \sum_{k=1}^{I}\log\sum_{j=0}^{J-1}\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j,\boldsymbol{\Sigma})$$

$$= \sum_{k=1}^{L'}\log\sum_{j=0}^{J-1} w_{j,k}\frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j,\boldsymbol{\Sigma})}{w_{j,k}} \tag{5.27a}$$

$$\geq \sum_{k=1}^{L'}\sum_{j=0}^{J-1} w_{j,k}\log\frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j,\boldsymbol{\Sigma})}{w_{j,k}} \tag{5.27b}$$

$$= \sum_{k=1}^{L'}\sum_{j=0}^{J-1} w_{j,k}\left(-\log(w_{j,k}) - \frac{1}{2}\log(\boldsymbol{\Sigma})\right)$$

$$- \sum_{k=1}^{L'}\sum_{m=0}^{J-1} w_{j,k}\left(\frac{1}{2}(\mathbf{z}_k-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}_k-\boldsymbol{\mu}_j)\right). \tag{5.27c}$$

In Eq. (5.27), $w_{j,k}$ is referred to as a hidden variable. Eq. (5.27b) is due to the Jensen inequality and concavity of logarithms, where equality holds if $\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j,\boldsymbol{\Sigma})/w_{j,k}$ is constant, i.e., $w_{j,k} = \mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_j,\boldsymbol{\Sigma})/\sum_{j=0}^{J-1}\mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_j,\boldsymbol{\Sigma})$. Then, the maximization problem in Eqs. (5.26)-(5.27) can be pursued via the expectation maximization (EM) algorithm, i.e., [79, 80]

$$w_{j,k}^{(i)} = \frac{\mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_j^{(i)},\boldsymbol{\Sigma}^{(i)})}{\sum_{j=0}^{J-1}\mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_m^{(j)},\boldsymbol{\Sigma}^{(j)})}, \tag{5.28a}$$

$$\boldsymbol{\mu}_j^{(i+1)} = \frac{\sum_{k=1}^{L'} w_{j,k}^{(i)}\cdot\mathbf{z}_k}{\sum_{k=1}^{L'} w_{j,k}^{(i)}}, \tag{5.28b}$$

$$\boldsymbol{\Sigma}^{(i+1)} = \frac{\sum_{j=0}^{J-1}\sum_{k=1}^{L'} w_{j,k}^{(i)}(\mathbf{z}_k-\boldsymbol{\mu}_j^{(i+1)})^T\cdot(\mathbf{z}_k-\boldsymbol{\mu}_j^{(i+1)})}{\sum_{j=0}^{J-1}\sum_{k=1}^{L'} w_{j,k}^{(i)}}. \tag{5.28c}$$

where i is the iteration step. Eq. (5.28a) refers to the E-step, which rises the previous jth iteration maximized lower-bound to the original log-likelihood, by equalizing the Jensen Inequality. Eqs. (5.28b)-(5.28c) are to maximize the lower-bound in Eq. (5.27c) given the fixed hidden variable $w_{j,k}^{(i)}$. As such, an

optimal (local or global) solution of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$ can be obtained by iteration using Eqs. (5.28a)-(5.28c).

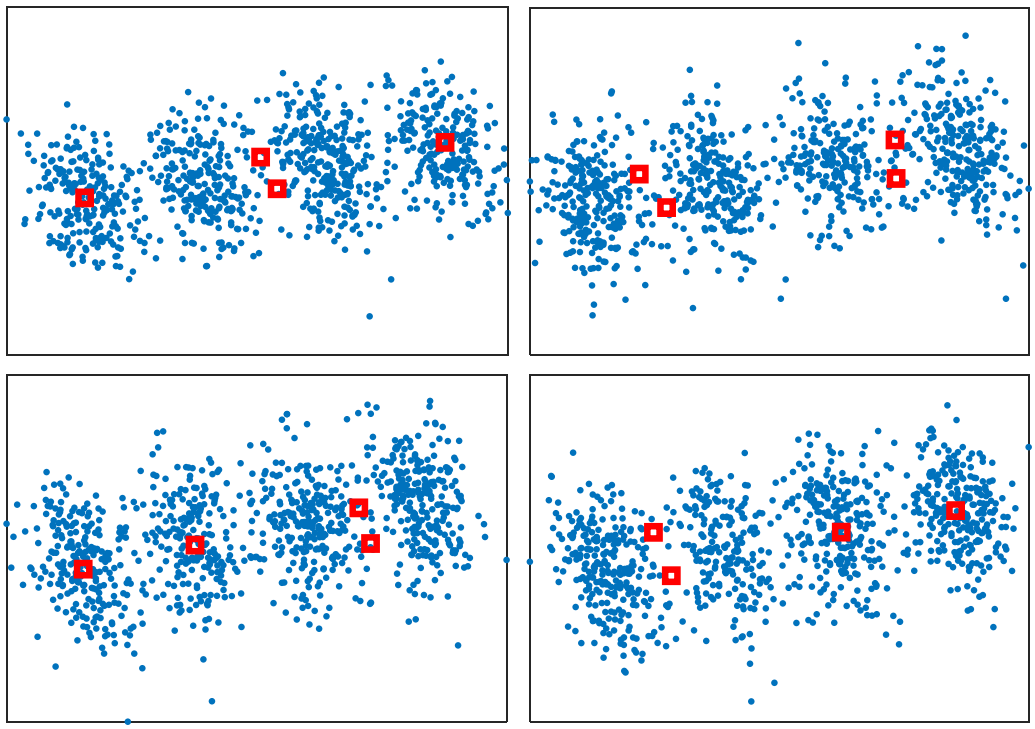


Figure 5.6: Illustration of clustering errors using classic GMM in Eq. (5.28): four cases of GMM trapped in local optimal solutions.

However, it is noteworthy that directly using Eq. (5.28) may fall into a local optimal solution, due to the likelihood approximation by Jensen's inequality and the following successive optimization steps [79, 80]. This is illustrated in Fig. 5.6.

To avoid such local optimal solutions, the properties of the extracted UV features are exploited and Eq. (5.28b) is converted into a convex problem. This is done by re-writing the objective function corresponding to $\boldsymbol{\mu}_j$ (the second term in Eq. (5.27c)), and then adding convex constraints in accordance with the characteristics of the UV features. The second term in Eq. (5.27c)

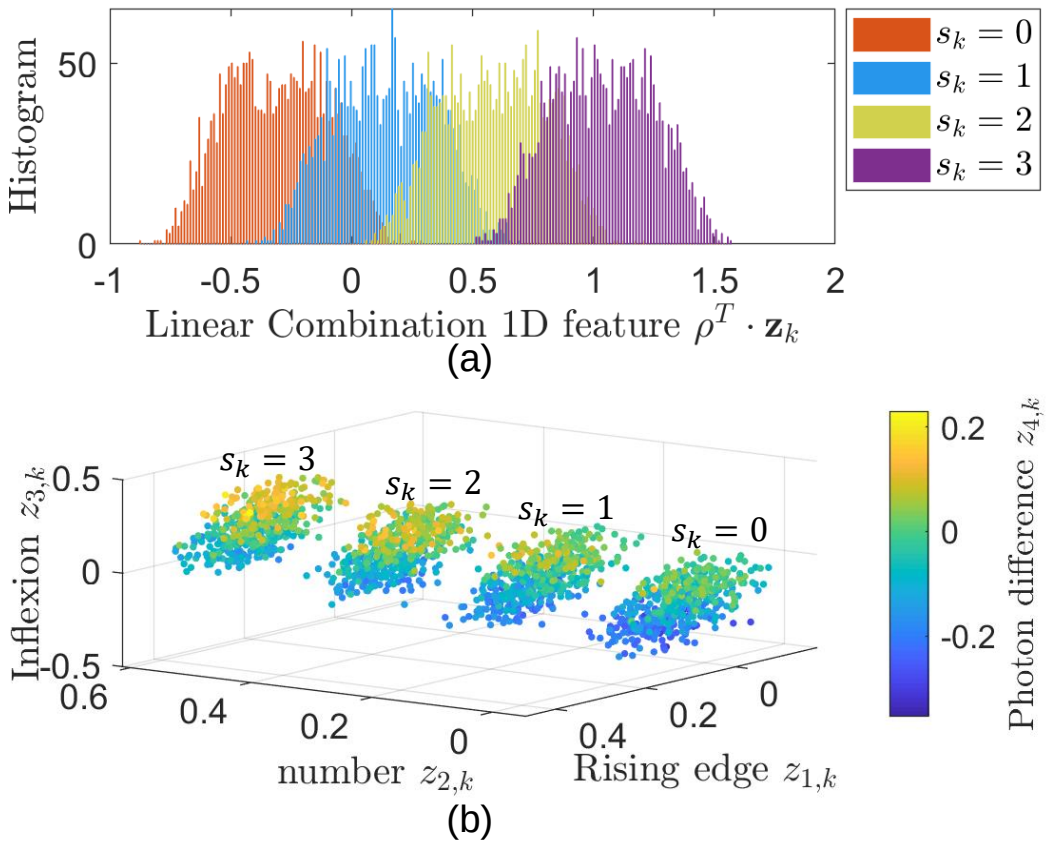


Figure 5.7: Illustration of non-coherent schemes that transform the ISI-caused sequential detection space into an ISI-released multi-hypothesis detection framework. Under the same received UV signals, (a) shows the poor detection accuracy from 1D non-coherent schemes, (b) gives the higher detection accuracy of the proposed HD non-coherent scheme.

can be re-written as:

$$\begin{aligned} &-\sum_{k=1}^{L'}\sum_{j=0}^{J-1} w_{j,k}\left(\frac{1}{2}(\mathbf{z}_k-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}_k-\boldsymbol{\mu}_j)\right) \\ =&-\sum_{k=1}^{L'} tr\left(\mathbf{W}_k\left(\mathbf{Z}_k-\mathbf{U}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{Z}_k-\mathbf{U}\right)\right), \end{aligned} \tag{5.29}$$

where $\mathbf{W}_k \triangleq diag([w_{0,k},\cdots,w_{J-1,k}]^T)$ is the diagonal matrix from vector $[w_{0,k},\cdots,w_{J-1,k}]^T$, $\mathbf{Z}_k \triangleq [\mathbf{z}_k,\cdots,\mathbf{z}_k]$ reproduces J times the vector $\mathbf{z}_k$, $\mathbf{U} \triangleq [\boldsymbol{\mu}_0,\cdots,\boldsymbol{\mu}_{J-1}]$, and $tr(\cdot)$ gives the trace operator of a matrix. Then, the maxi-

mization problem over $\boldsymbol{\mu}_{0:J-1}$ in Eq. (5.28b) can be converted to the following convex problem with UV signal related constraints:

$$\min_{\mathbf{U}} \sum_{k=1}^{L'} tr\left(\mathbf{W}_k \left(\mathbf{Z}_k - \mathbf{U}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{Z}_k - \mathbf{U}\right)\right) \tag{5.30a}$$

$$\text{s.t. } \|\mathbf{U}\boldsymbol{\Psi}\|_F^2 < \varepsilon, \tag{5.30b}$$

$$\mathbf{c}_d^T \mathbf{U}\mathbf{b} = \mu_{d,J-1} - \mu_{d,0} > 0, \ \forall d \in \{1, \cdots, D\}, \tag{5.30c}$$

$$\mathbf{c}_1^T \mathbf{U}\mathbf{b} = \mu_{1,J-1} - \mu_{1,0} < \max_{1 \le k,k' \le I} \{z_{1,k} - z_{1,k'}\}, \tag{5.30d}$$

$$\mathbf{c}_1^T \mathbf{U}\mathbf{b} = \mu_{1,J-1} - \mu_{1,0} > 0.5 \left(\mu_{1,J-1}^{(0)} - \mu_{1,0}^{(0)}\right) \tag{5.30e}$$

where $\|\cdot\|_F$ represents the Frobenius norm. Eq. (5.30a) is the objective function. Eq. (5.30b) is to ensure the equal difference expressed in Eq. (5.23), by making the squared Frobenius norm of $\mathbf{U}\boldsymbol{\Psi}$, i.e., $\|\mathbf{U}\boldsymbol{\Psi}\|_F^2$, less than some small ϵ. Eqs. (5.30c)-(5.30e) are to confine the differences between the expectations (clustering centres), where Eq. (5.30c) is to assign $\boldsymbol{\mu}_{J-1}$ and $\boldsymbol{\mu}_0$ corresponding to the largest and smallest symbols, respectively. Eqs. (5.30d)-(5.30e) are to confine the difference of the expectation of the rising-edge with respect to the largest and the smallest symbols. The convex optimisation problem can be solved using interior point method [81, 82], or CVX toolbox [83].

After the derivation of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$, the signal detection process can be pursued by computing and comparing the logarithm-form of the conditioned likelihood PDF, i.e.,

$$\hat{s}_k = \underset{j \in \{0,\cdots,J-1\}}{\text{argmax}} \ -(\mathbf{z}_k - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_j). \tag{5.31}$$

5.5.2 Overall Non-Coherent Detection Algorithm

After the elaboration of the designed HD non-coherent detection scheme, the algorithm flow is provided in Algo. 3. The inputs are the received UV sig-

Algorithm 3 HD Non-Coherent Signal Detection Algorithm

Input: Received UV signals $r_1, \cdots r_{KN}$, modulation type J-PAM, and number of features L' for unsupervised training.

1: $\forall k \in \{1, \cdots, K\}$, compute HD UV feature $\mathbf{z}_k$ via Eqs. (5.1)-(5.2).
2: Initialize $\boldsymbol{\mu}_m^{(0)}$ via Eq. (5.25).
3: Assign $i = 0$ and $\Delta^{(i)} = \|[\boldsymbol{\mu}_1^{(i)}, \cdots, \boldsymbol{\mu}_{M-1}^{(i)}]\|_F$.
4: **while** $\Delta^{(i)} > \epsilon$ **do**
5: Assign $i = i + 1$.
6: Update $w_{j,k}^{(i)}$ via Eq. (5.28a).
7: Update $\boldsymbol{\mu}_j$ by solving convex problem in Eq. (5.30).
8: Update $\boldsymbol{\Sigma}^{(i)}$ via Eq. (5.28c).
9: $\Delta^{(i)} = \|[\boldsymbol{\mu}_1^{(i)}, \cdots, \boldsymbol{\mu}_{J-1}^{(i)}] - [\boldsymbol{\mu}_1^{(i-1)}, \cdots, \boldsymbol{\mu}_{J-1}^{(i-1)}]\|_F$
10: **end while**
11: Assign $\boldsymbol{\mu}_j = \boldsymbol{\mu}_j^{(i)}$, and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{(i)}$.
12: **for** $k \in \{1, \cdots, K\}$ **do**
13: Determine the J-PAM symbol $\hat{s}_k$ via Eq. (5.31).
14: **end for**

Output: Return the detected symbol $\hat{s}_1, \cdots, \hat{s}_K$.

nals, the modulation type J, and the number of features L' for unsupervised learning. Step 1 is to compute the HD feature $\mathbf{z}_k$ from the received signals for each symbol. Steps 2-11 are to learn the expectations $\boldsymbol{\mu}_j$ and the covariance matrix $\boldsymbol{\Sigma}$ for the conditioned likelihood PDF of $\mathbf{z}_k$. Steps 12-14 are to pursue the signal detection for J-PAM symbol.

5.5.3 Convergence Analysis

The convergence of Algo. 3 can be proved via the non-decreasing likelihood from iterations i to $i+1$, i.e., $\mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1}^{(i)}, \boldsymbol{\Sigma}^{(i)}) \leq \mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1}^{(i+1)}, \boldsymbol{\Sigma}^{(i+1)})$:

$$
\begin{aligned}
&\mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1}^{(i)}, \boldsymbol{\Sigma}^{(i)}) = \sum_{k=1}^{L'} \log \sum_{j=0}^{J-1} \mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i)}, \boldsymbol{\Sigma}^{(i)}) \\
=&\sum_{k=1}^{L'} \log \sum_{j=0}^{J-1} w_{j,k}^{(i)} \frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i)}, \boldsymbol{\Sigma}^{(i)})}{w_{j,k}^{(i)}} && \text{(5.32a)} \\
=&\sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \log \frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i)}, \boldsymbol{\Sigma}^{(i)})}{w_{j,k}^{(i)}} && \text{(5.32b)} \\
=&\sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(-\log(w_{j,k}^{(i)}) - \frac{1}{2}\log(\boldsymbol{\Sigma}^{(i)}) \right) \\
&- \sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(\frac{1}{2}(\mathbf{z}_k - \boldsymbol{\mu}_j^{(i)})^T (\boldsymbol{\Sigma}^{(i)})^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_j^{(i)}) \right) && \text{(5.32c)} \\
\leq&\sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(-\log(w_{j,k}^{(i)}) - \frac{1}{2}\log(\boldsymbol{\Sigma}^{(i)}) \right) \\
&- \sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(\frac{1}{2}(\mathbf{z}_k - \boldsymbol{\mu}_j^{(i+1)})^T (\boldsymbol{\Sigma}^{(i)})^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_j^{(i+1)}) \right) && \text{(5.32d)} \\
\leq&\sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(-\log(w_{j,k}^{(i)}) - \frac{1}{2}\log(\boldsymbol{\Sigma}^{(i+1)}) \right) \\
&- \sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \left(\frac{1}{2}(\mathbf{z}_k - \boldsymbol{\mu}_j^{(i+1)})^T (\boldsymbol{\Sigma}^{(i+1)})^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_j^{(i+1)}) \right) && \text{(5.32e)} \\
=&\sum_{k=1}^{L'} \sum_{j=0}^{J-1} w_{j,k}^{(i)} \log \frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i+1)}, \boldsymbol{\Sigma}^{(i+1)})}{w_{j,k}^{(i)}} && \text{(5.32f)} \\
\leq&\sum_{k=1}^{L'} \log \sum_{j=0}^{J-1} w_{j,k}^{(i)} \frac{\mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i+1)}, \boldsymbol{\Sigma}^{(i+1)})}{w_{j,k}^{(i)}} && \text{(5.32g)} \\
=&\sum_{k=1}^{L'} \log \sum_{j=0}^{J-1} \mathcal{N}(\mathbf{z}_k|\boldsymbol{\mu}_j^{(i+1)}, \boldsymbol{\Sigma}^{(i+1)}) = \mathcal{L}(\mathbf{z}_{1:L'}|\boldsymbol{\mu}_{0:J-1}^{(i+1)}, \boldsymbol{\Sigma}^{(i+1)}). && \text{(5.32h)}
\end{aligned}
$$

Eq. (5.32a) describes the computation of the log-likelihood objective function. Eq. (5.32b) holds for the equality condition in Jensen's inequality, i.e., $w_{j,k}^{(i)} = \mathcal{N}(\mathbf{z}_k; \boldsymbol{\mu}_j^{(i)}, \boldsymbol{\Sigma}^{(i)}) / \sum_{j=0}^{J-1} \mathcal{N}(\mathbf{z}_k; \boldsymbol{\mu}_j^{(i)}, \boldsymbol{\Sigma}^{(i)})$ (corresponding to step 5 in Algo. 3). Eq. (5.32c) is the expansion of Eq. (5.32b). Eq. (5.32d) holds since $\boldsymbol{\mu}_{0:J-1}^{(i+1)}$ is the optimal solution of the proposed convex problem, which maximizes Eq. (5.32c) given fixed $w_{j,k}^{(i)}$ and $\boldsymbol{\Sigma}^{(i)}$ (corresponding to step 7 in Algo. 3). Eq. (5.32e) holds since $\boldsymbol{\Sigma}^{(i+1)}$ is the optimal solution that maximizes Eq. (5.32d), given fixed $w_{j,k}^{(i)}$ and $\boldsymbol{\mu}_{0:J-1}^{(i+1)}$ (corresponding to step 8 in Algo. 3). Eq. (5.32f) is the direct computation from Eq. (5.32e). Eq. (5.32g) arises from Jensen's inequality, which then gives Eq. (5.32h). Meanwhile, it is noted that there always exists a maximum of the log-likelihood in the GMM clustering. As a consequence, combined with Eqs. (5.32a)-(5.32h), the convergence of the proposed non-coherent detection algorithm can be proved.

5.6 Numerical Simulations

In this section, the performance of the proposed HD non-coherent detection scheme is evaluated. Firstly, an analysis of the performance of GMM clustering will be conducted. Then, the detection accuracy will be measured via the BER. The computational complexity is represented by the number of multiplication operations required.

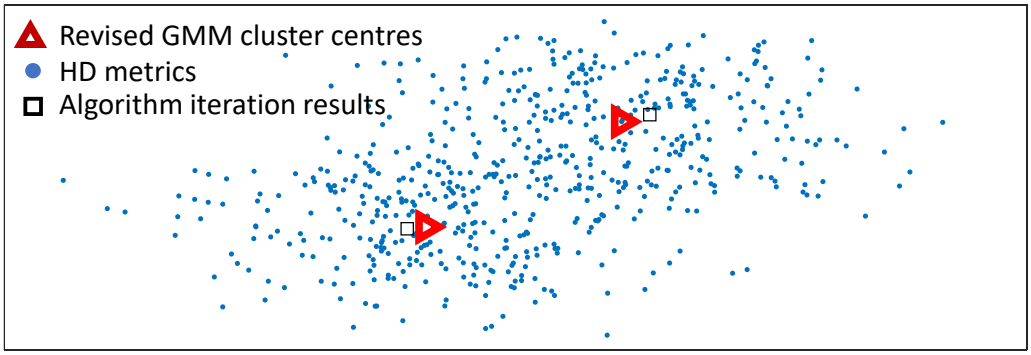
For the UV channel modelling, Monte Carlo techniques are adopted proposed by [29, 30], to obtain the discrete CIR coefficients in Eq. (3.4). The relevant channel parameters are configured according to Table 3.1.

For comparison, the MMSE-based linear equalizer (a 1D non-coherent scheme) as a benchmark and the state-of-the-art coherent MLSD using least square channel estimator and Viterbi detection algorithm [1]. The aim is to demonstrate that the BER here is close to that of MLSD, but at a lower com-

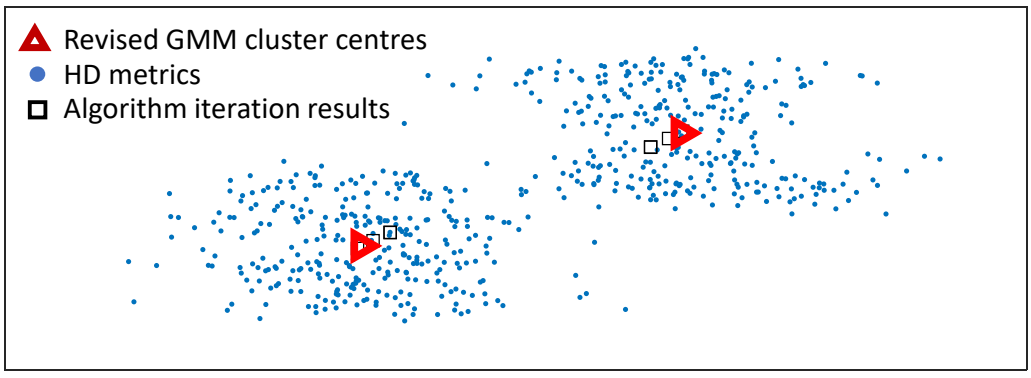
putational complexity given the exploitation of the HD ISI-intensive features. For fairness, the number of samples during one symbol interval T_b is assigned as $M = 20$ for all three schemes.

5.6.1 GMM Clustering accuracy

The analysis on the performance of GMM clustering will be conducted in this section, by showing the achieved clustering centres of the HD metrics, with various SNR, under different modulation modes (i.e., OOK, 4-PAM and 8-PAM), seen in Figs. 5.8 - 5.10.



(a)



(b)

Figure 5.8: Illustration of revised GMM on OOK signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=8dB, (b) SNR=15dB.

Firstly, the clustering accuracy based on OOK modulation is evaluated

using SNR values of 8dB and 15dB, as shown in Fig. 5.8.

Fig. 5.8 illustrates the revised GMM on OOK signal-based HD metrics, showing the ability to achieve the correct clustering centres. It can be seen, the blue dots represent every individual HD metrics which should be classified into two types. The black squares are the iteration results of the revised GMM algorithm for pursuing the centres of the clustering, while the red triangle is the achieved cluster centres by the revised GMM. It may be observed that in Fig. 5.8 (b) the blue dots are easy to be divided into two clusters, while in Fig. 5.8 (a) the dots gather into a mass without a boundary to distinguish them from each other. That means the transmitted signal under a higher SNR is easier to be classified into two types, compared with the signal delivered under a lower SNR. It is noteworthy that the centres for the two clusters could be figured out by the revised GMM clustering algorithm, in both of these two scenarios, although in Fig. 5.8 (a) the signal contamination seriously leads to boundary unclear.

Now, the clustering accuracy of the revised GMM algorithm, based on the 4-PAM modulation is evaluated, respectively under SNR=15dB and SNR=25dB, as shown in Fig. 5.9.

In Fig. 5.9, there is a demonstration of the revised GMM algorithm employed in 4-PAM modulation, showing its ability to achieve the correct clustering centres. As may be seen, the blue dots represent every individual HD metric, belonging to four different kinds of signal. The black squares and the red triangle are the iteration results and the finally achieved results respectively, for pursuing the centres of the clustering by the revised GMM algorithm. Same with the previous OOK scenario, in conditions of a higher SNR the blue dots are easy to divide into four clusters as seen in Fig. 5.9 (b), while in Fig. 5.9 (a) under a lower SNR the dots gather into a mass without clear boundaries to distinguish them from each other. Moreover, the revised

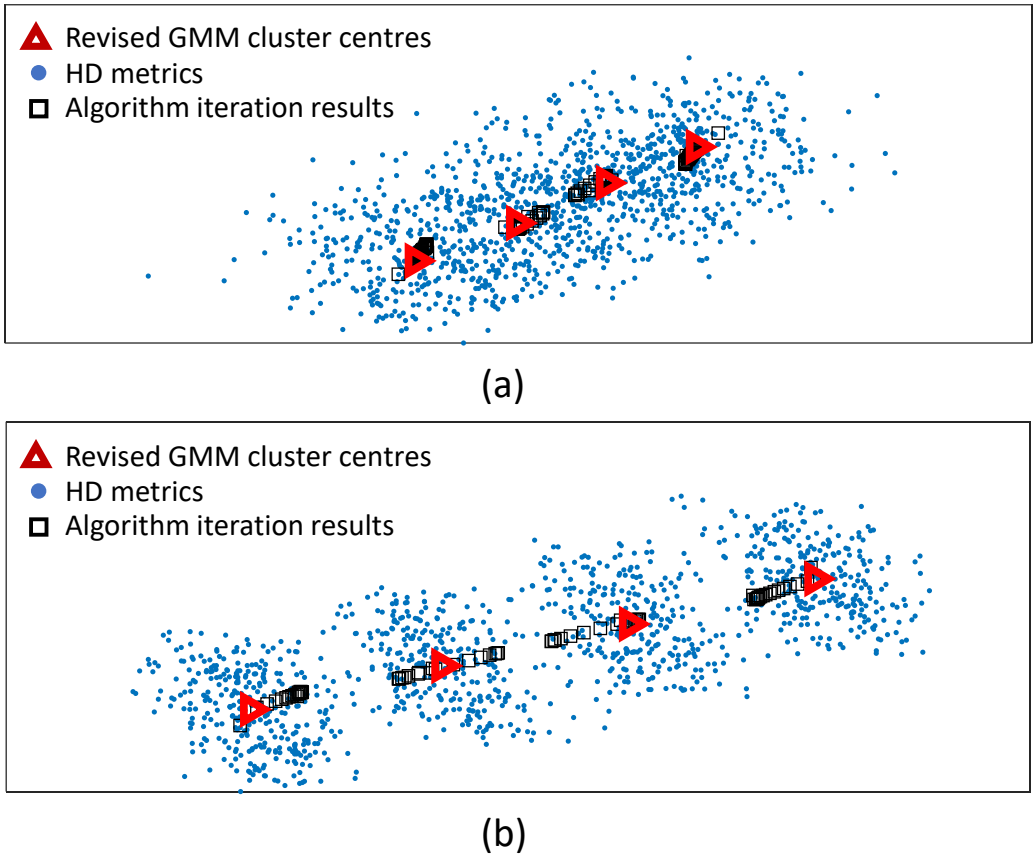


Figure 5.9: Illustration of revised GMM on 4-PAM signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=15dB, (b) SNR=25dB.

GMM clustering algorithm still can figure out four centres for the clustering, in cases of these two SNR scenarios. According to 5.9 (b), the transmission under a high SNR, it is obvious the achieved centres are reasonable by human experience.

Finally, the clustering performance of the revised GMM algorithm in the case of the 8-PAM modulation is evaluated, respectively under SNR=20dB and SNR=30dB, as shown in Fig. 5.10.

There is an illustration of revised GMM on 8-PAM signal-based HD metrics, showing the ability to achieve the correct clustering centres. It may be observed that when the SNR is 20dB, blue dots are gathered to a mass,

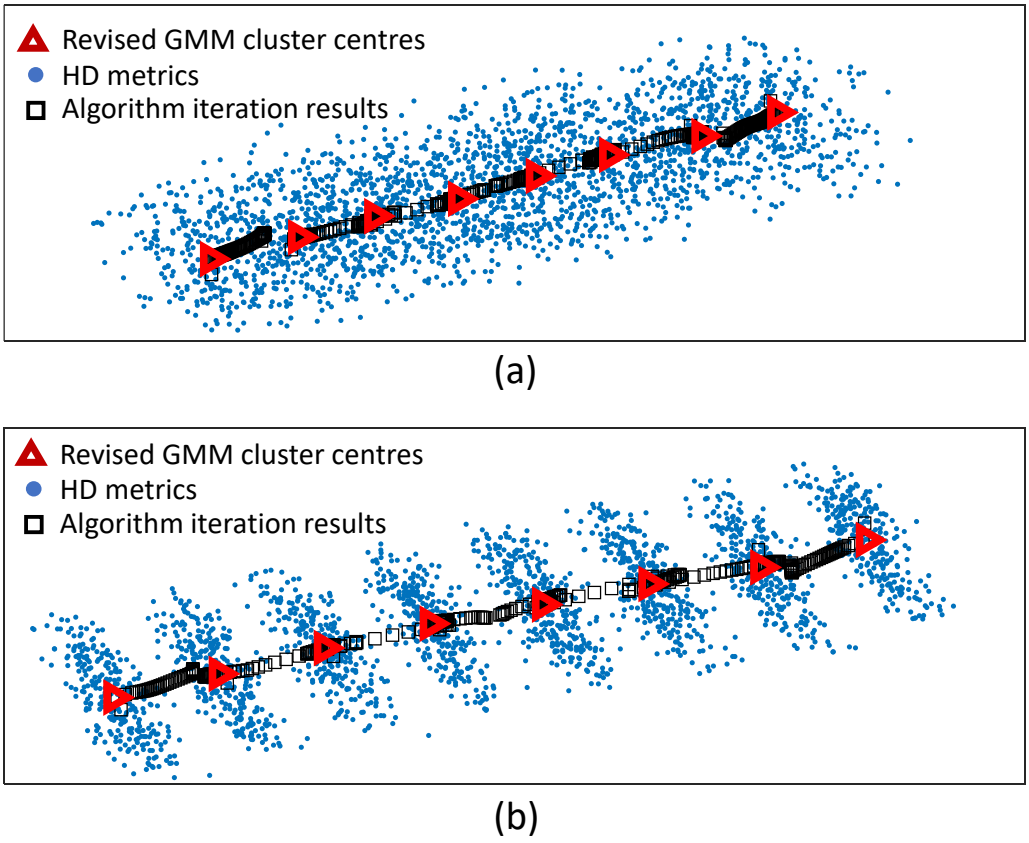


Figure 5.10: Illustration of revised GMM on 8-PAM signal-based HD metrics, showing the ability to achieve the correct clustering centres: (a) SNR=20dB, (b) SNR=30dB.

difficult to classify. Then, increasing the SNR to 30dB, the blue dots distribute in eight parts naturally. It can be seen, in both of these two cases, the revised GMM algorithm has a good performance in obtaining the clustering centres.

In Fig. 5.8 - Fig. 5.10, the iteration results show the traces of pursuing the centres, represented by black squares. The OOK needs the least iterative steps to figure out the results of clustering centres, which is followed by the 4-PAM modulation. Meanwhile, the 8-PAM modulation has the most iterative steps to get the results. Besides, in terms of the SNR, the OOK modulation can be clearly divided 0-bit from 1-bit utilizing the lowest SNR, while 8-PAM modulation requires the highest SNR to distinguish them from each other

distinctly. For instance, when SNR is 15dB, in Fig. 5.8 (b) the OOK signals are classified clearly, on the contrary, the 4-PAM signals are in a mass as seen in Fig. 5.9. The reason for this phenomenon is that operatng at the same SNR, means that a 4-PAM signal has a smaller voltage gap between two adjacent Amplitude levels, compared with OOK.

5.6.2 Communication Performance

The detection accuracy of the proposed HD non-coherent detection scheme is determined in Figs. 5.11-5.16, versus different levels of SNR and various symbol rates. This section divides into three parts according to the modulation modes, i.e., OOK, 4-PAM, and 8-PAM.

A) HD method for OOK modulation

Fig. 5.11 provides the BER comparison of the four schemes versus different levels of SNR, under OOK modulation (given that the compared scheme in [1] and 1D non-coherent methods are designed only for OOK).

It is observed from Fig. 5.11 that, the proposed HD non-coherent detection scheme outperforms the previous 1D non-coherent scheme (a gain of 5dB in SNR). This matches the theoretical error analysis in section 5.4.1. The theoretical BER of the proposed scheme is also provided, which gives a tight lower bound on numerical simulations. A small gap can be found between the theoretical and the simulated curves, especially in the low SNR region. This is because the training errors in the designed clustering method become worse as the SNR decreases.

Besides, it is seen from Fig. 5.11 that the BER of the proposed HD non-coherent scheme is lower than that of the MMSE equalizer benchmark, and is close to the state-of-the-art MLSD. For example, a 1dB gap can be achieved when the BER reaches the 7% overhead hard-decision forward error correction

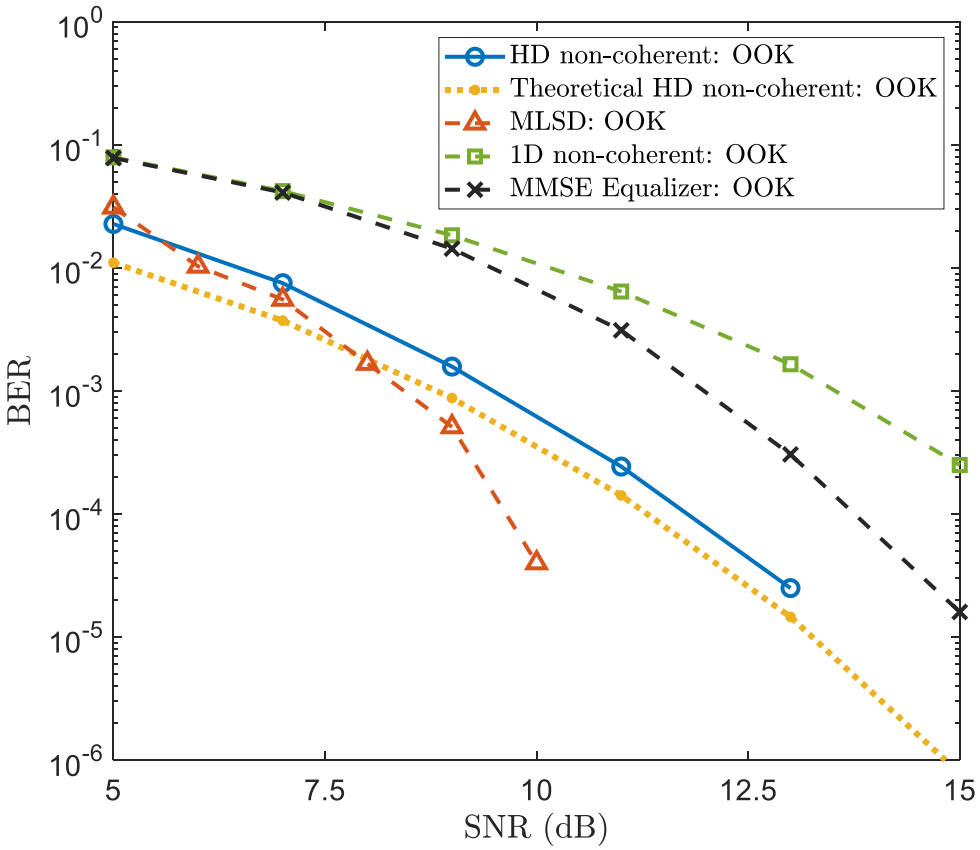


Figure 5.11: BER comparison versus SNR under OOK modulation. the proposed HD non-coherent detection outperforms the MMSE equalizer and the previous 1D non-coherent scheme, and approaches that of the MLSD method.

(FEC) limit (i.e., BER=4.5×10^{-3}) [57], which suggests a comparable error-free communication performance with MLSD method when adopting an appropriate FEC code. This is attributed to the exploration of the ISI-insensitive UV features (e.g., rising edge, inflection), which are able to inherently counteract the ISI effect, and thereby offer a remarkable detection performance.

Fig. 5.12 illustrates the BER comparisons versus the different intensities of the ISI effect. Here, a range of symbol rates are used (i.e., $R_s \in [10^6,\ 1.5\times10^7]$ bit/s) to reflect the different intensities of ISI, as a higher symbol rate will introduce a more severe ISI and will then negatively impact the performance of the communication system more significantly. Fig. 5.12

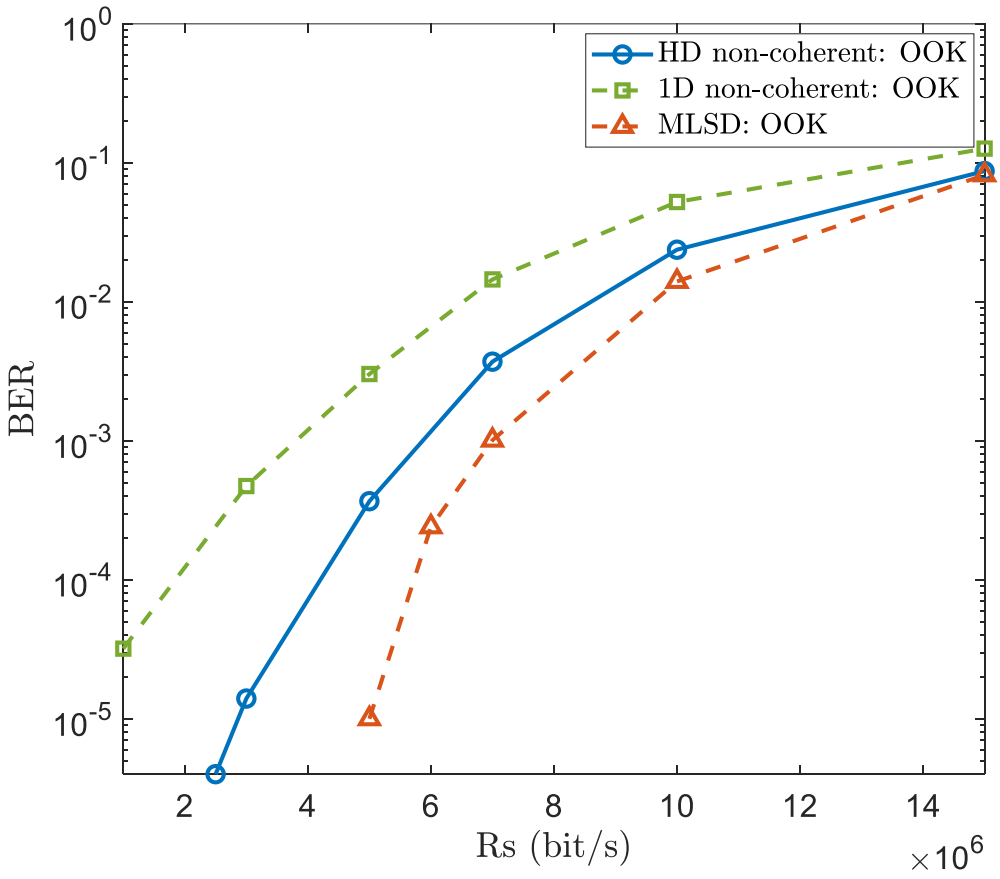


Figure 5.12: BER comparison versus ISI (rendered by different symbol rates R_s), under OOK modulation, with SNR=15dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method.

shows that with the growth of the ISI intensities (i.e., the increase of symbol rate R_s), the BERs of all schemes become larger, showing that the ISI induced by the NLoS UV scattering effect causes deterioration in the communication performance. For comparison among these schemes, two similar results are obtained. Firstly, the HD non-coherent scheme outperforms the 1D non-coherent scheme, which matches the proof in Section 5.4.1. Secondly, the proposed scheme can approach the state-of-the-art MLSD method, given its ability to transform (approximate) the ISI-caused sequential detection space into the ISI-released detection framework.

B) HD method under 4-PAM modulation

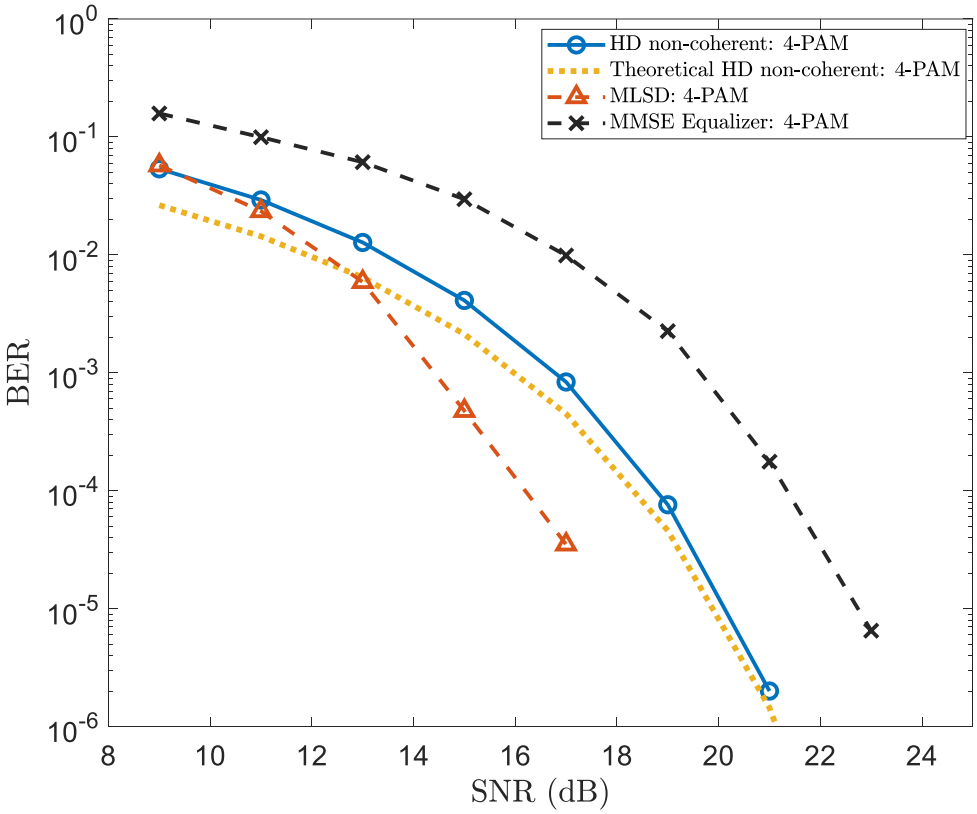


Figure 5.13: BER versus SNR under $J = 4$-PAM symbol. the proposed scheme shows comparable accuracy with the MLSD and MMSE equalizer, suggesting its suitability for multi-symbol UVC scenarios.

This part aims to evaluate the BER performance of the proposed HD method, in account of 4-PAM signals which suffer from a higher intensity of ISI compared to OOK modulating. Fig. 5.13 and Fig. 5.14 show the performance of the proposed scheme deployed in 4-PAM modulation system, via different SNR and symbol rates respectively. Here, it is only compared with the MMSE equalizer and the MLSD requiring a 4^L searching space, as the 1D non-coherent scheme cannot address the $(J > 2)$-PAM situation.

It can first be seen that, similar to the OOK case, the deduced theoretical BER serves as a very tight lower bound. The gap between them is

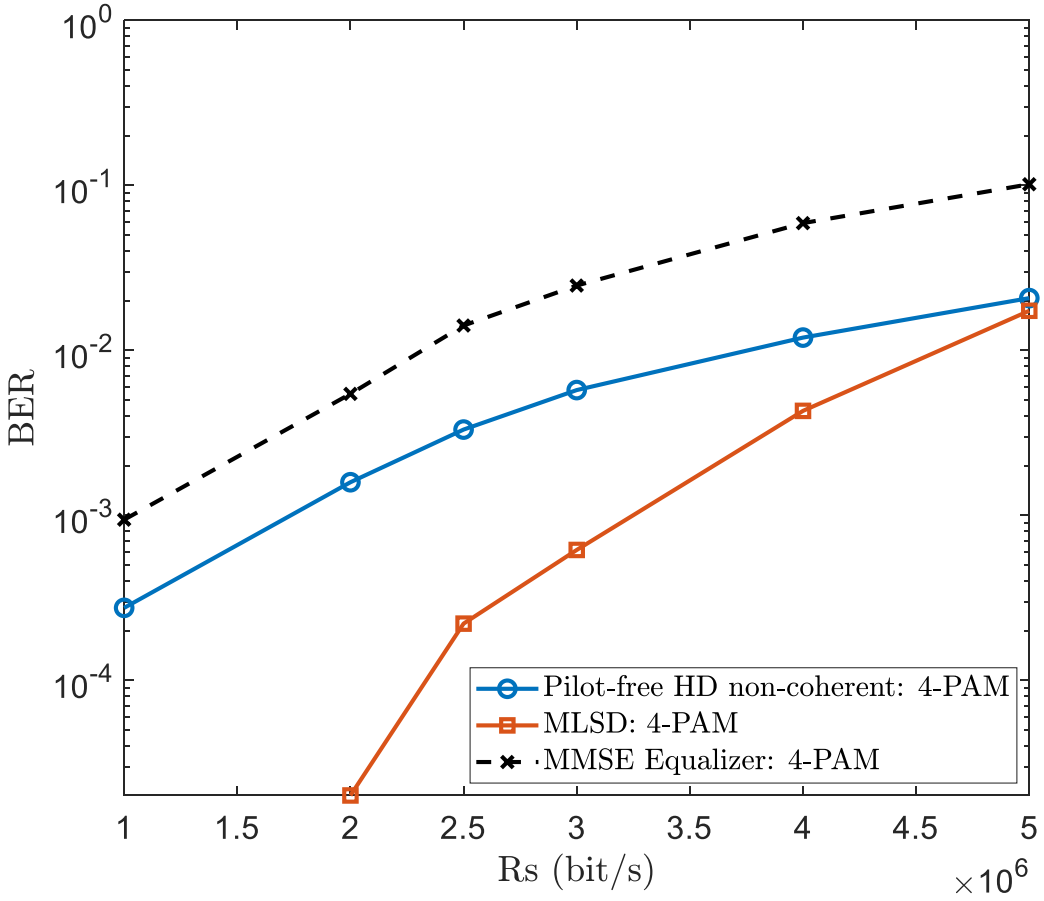


Figure 5.14: BER comparison versus ISI (rendered by different symbol rates R_s), under 4-PAM modulation, with SNR=20dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method.

induced by training errors, which are worse in the low SNR region, but tolerable with high SNR. Besides, in Fig. 5.13, it can be observed that for the 4-PAM symbol cases, the approach outperforms the benchmark provided by the MMSE equalizer with approximately 2dB, and shows a comparable BER accuracy with the MLSD method.

Fig. 5.14 demonstrates BER comparison versus different symbol rates R_s, under 4-PAM modulation, with fixed SNR=20dB. A higher symbol rate means a stronger ISI, consequently leading to a bad BER performance. Similar

to OOK situation, the BER performance of the HD method is better than the MMSE equalizer and gradually close to that of MLSD method in the high ISI domain. It indicates that the HD method is effective in dealing with the ISI.

C) HD method under 8-PAM modulation

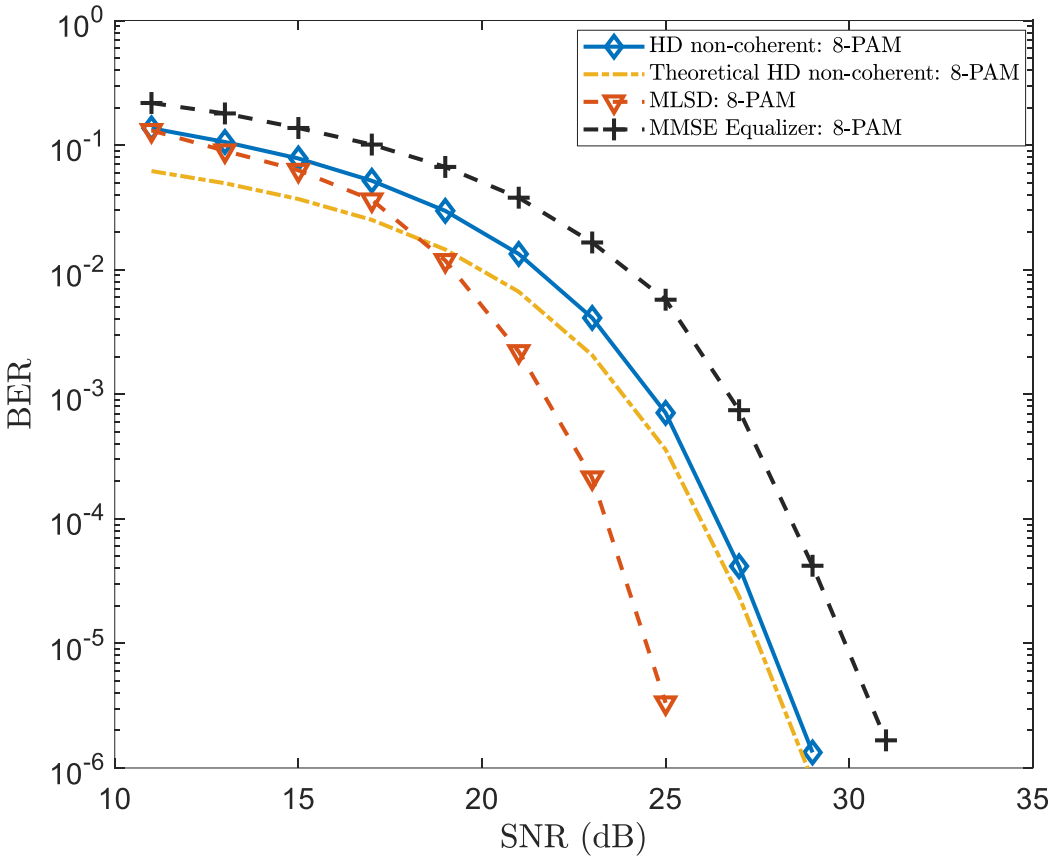


Figure 5.15: BER versus SNR under $J = 8$-PAM symbol. the proposed scheme shows comparable accuracy with the MLSD and MMSE equalizer, suggesting its suitability for multi-symbol UVC scenarios.

Here, the detection accuracy will be evaluated by 8-PAM modulation (i.e., 8 different levels) which means a more complex ISI situation. By Fig. 5.15 and Fig. 5.16, the approach demonstrates better BER accuracy compared to the benchmark MMSE equalizer for 8-PAM symbols and is comparable to the MLSD method.

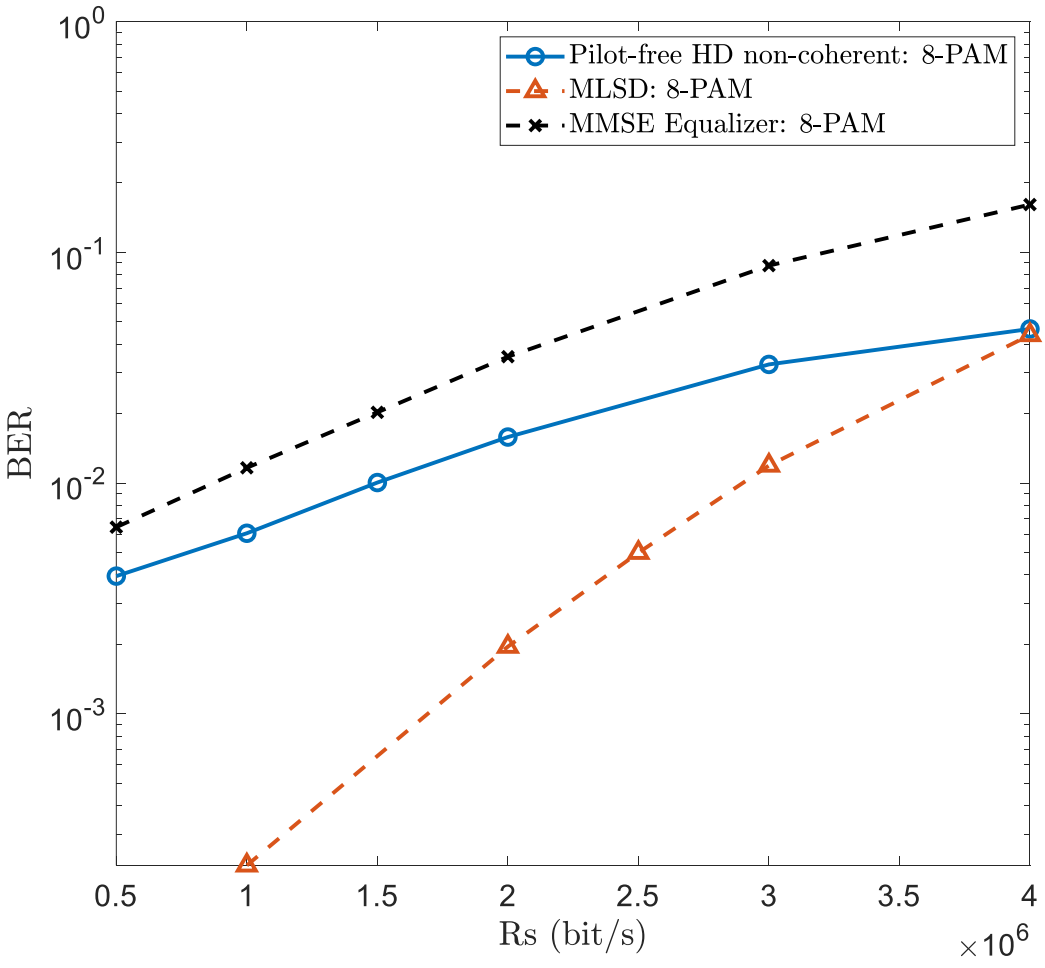


Figure 5.16: BER comparison versus ISI (rendered by different symbol rates R_s), under 8-PAM modulation, with SNR=25dB. The result suggests the robustness of the proposed HD non-coherent detection scheme to different ISI intensities and approaches that of the MLSD method.

Fig. 5.15 illustrates BER versus SNR and Fig. 5.16 demonstrates BER comparison versus ISI (rendered by different symbol rates R_s) with SNR=25dB, under 8-PAM modulation. In Fig. 5.15, Here, similar to the OOK and 4-PAM cases, the HD method obtains a close BER performance to the deduced theoretical BER, which indicates the clustering algorithm is also suitable for the 8-PAM signals. From both Fig. 5.15 and Fig. 5.16, it can be seen that the HD method has a better BER performance than the MMSE

equalizer, with a 2dB gain. Meanwhile, the HD method achieves a close BER performance with MLSD, especially in the high ISI region.

Then, it is observed that for the $J > 2$ symbol cases, the approach outperforms the benchmark provided by the MMSE equalizer, and shows a comparable BER accuracy with the MLSD method. This, combined with the previous OOK simulation results, demonstrates both the reliability and the robustness of the proposed scheme, which is suitable for different signal modulation schemes in UVC systems combating ISI distortion.

5.6.3 Complexity Comparison

After the demonstration of the comparable communication accuracy with the state-of-the-art MLSD, it remains to show that the proposed HD non-coherent scheme demands a smaller amount of computational complexity. The computational complexity is measured via the number of multiplications. For the proposed HD non-coherent scheme, the number of multiplications in the detection process in Eq. (5.31) is $M \cdot (D^2 + D) \cdot N = O(MD^2N)$. The detection process for MLSD has an order of $O(LM^LN)$ multiplications (L is the length of ISI). For the previous 1D non-coherent scheme, only ($J = 2$)-PAM (i.e., OOK) modulated signal can be detected and the number of multiplications has an order of $O(D^2N)$. To further compare the computational complexity in UV communications, Fig. 5.17 provides a comparison of the multiplication numbers for different schemes, under $D = 4$ sub-features, $N = 10^4$ transmitted symbols, and the length of ISI $L = 8$, for $J = 2, 4, 8$-PAM.

The proposed scheme has a very similar order of computational complexity to the previous 1D non-coherent scheme, but a significantly smaller order when compared to the MLSD. For OOK modulation, this gap is $2\times(4^2+4)\times10^4 = 4\times10^5 < 8\times2^8\times10^4 = 2\times10^7$, which, however, increases drastically for ($J = 4, 8$)-PAM modulation (i.e., $4\times(4^2+4)\times10^4 = 8\times10^5 < 8\times4^8\times10^4 =$

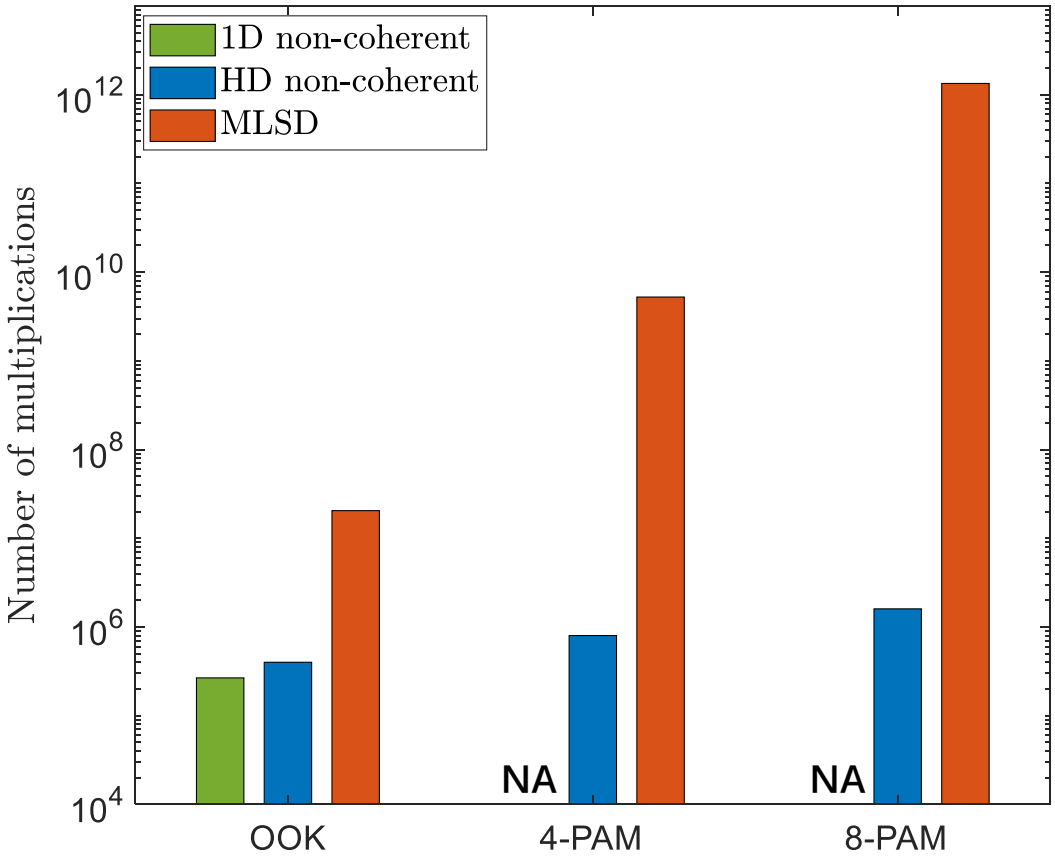


Figure 5.17: Computational complexity compared to the MLSD method, emphasizing the computational efficiency of the proposed HD non-coherent scheme.

5.2×10^9, and $8 \times (4^2 + 4) \times 10^4 = 1.6 \times 10^6 < 8 \times 8^8 \times 10^4 = 1.3 \times 10^{12}$). As such, by combining the results from Figs. 5.11-5.17, it may be concluded that the proposed HD non-coherent scheme can provide a low-complexity-based detection paradigm as well as comparable communication accuracy with the state-of-the-art coherent MLSD method.

5.7 Conclusions & Discussions

For most OWC applications, signal detection with intensive ISI contamination is of great difficulty. Current coherent schemes fall into high computational

consumption, and the non-coherent schemes cannot guarantee reliable detection accuracy. In this section, an HD feature-based non-coherent detection scheme has been proposed, relying on the HD construction of the four ISI-insensitive UV signals' sub-features, with the following two merits. Firstly, it transforms the ISI-caused sequential detection into the ISI-released multi-hypothesis detection framework, therefore avoiding the high computational CIR estimation and sequential likelihood computations required by the coherent schemes (a reduction of 4 magnitude orders of complexity is achieved). Secondly, the HD feature is able to provide higher SNRs for signal detection than current non-coherent schemes, which leads to a very-close detection accuracy with the state-of-the-art coherent MLSD ($< 1dB$ in SNR at FEC limit BER=4.5×10^{-3}). These thereby demonstrate the promising pathway of the proposed non-coherent scheme towards a low-computational but high-accurate detection paradigm for optical wireless communications with intensive ISI.

Chapter 6

Machine Learning based Nonlinear Feature for Non-coherent UV Signal Detection

In the previous chapters 4 and 5, UV signal detection features were designed manually via the geometry of ISI in the linear forms, which however does not fully exploit the properties of UV signals for symbol detection (e.g., whether and how the sequential nonlinearity can contribute to combating the ISI intensities and noises are not clear). In this chapter, to explore the nonlinear characteristic of UV signals, machine learning techniques are introduced to train the detector or feature construction process via the combinations of linear weights and bias of neurons and the nonlinear activation functions.

In essence, machine learning is a subfield of artificial intelligence focusing on the development of algorithms and models that can learn from and make predictions on data. To implement this, neural networks [84, 85], inspired by the structure and functionality of the human brain, aim to learn the

mapping/functions from a distribution of inputs to desired objective distribution of outputs. A neural network is composed of a graph of neurons, each of which performs a weighted sum of its inputs and applies an activation function. The outputs of the neurons in one layer are connected as inputs to the neurons in the next layer. A neural network can have one or multiple hidden layers between the input and output layers. In the context of signal detection, NNs can be created as a classifier (i.e., the outputs are the discrete symbols classes) [86–88].

For the structure of this chapter, the process and results of the supervised neural network (NN) framework are first provided in Section 6.1. Then, a half-supervised method embracing the previously constructed features is proposed in 6.2.

6.1 Supervised NN based UV Detector

Supervised learning is a type of machine learning where the algorithm learns from labelled training data to make predictions on new and unlabelled data [89,90]. In this section, the labelled data for training can be generated via the predefined pilot sequences. The NN is designed as a classifier to the received UV signals (as input data) into the transmitted symbol categories (e.g., binary OOK class as predefined categories), based on the relationships of their input and output distributions hidden reflected via the received signals and the corresponding symbols.

6.1.1 Implementation Details

For UV signal detection, a 4-layer fully connected NN is constructed. The input layer is the current kth and the previous $(k-1)$th received signals, i.e., the inputs are $2 \times M$ neurons, which is to count the main ISI from the adjacent

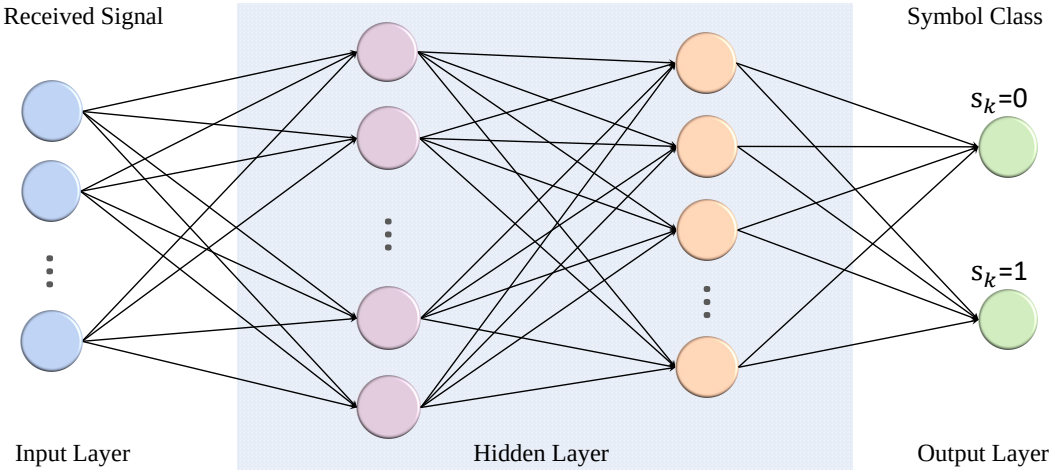


Figure 6.1: Illustration fully connected NN based detector, where the input layer is the received UV signals, and the output layer is the symbol classes (here the symbols have two classes, representing the OOK modulations for transmitter).

interval. The two hidden layers are with decreasing number of neurons and with ReLU activation functions, i.e.,

$$Activation(x) = max\{0, x\}. \tag{6.1}$$

The output layer is with neurons of which the number equals that of symbol categories (e.g., 2 for OOK), and whose outputs are the probabilities of belonging to current categories. As such, at each kth transmission interval, the forward function is to transform current and previous $2 \times M$ received signals into the probabilities of symbol categories.

The training process involves adjusting the weights and biases of the neurons in the network such that the output of the network matches the target classes in the training data via minimizing a loss function. To be specific, as Tx transmits a sequence of symbols that Rx is awareness of, the training data can be specified as the knowledge of pairs of received signals and their corresponding symbol categories. Then, the loss function is selected as cross entropy

loss, which measures the dissimilarity between the predicted probabilities of the symbol categories and the true symbol class labels.i.e.,

$$Loss_{CE} = -\frac{1}{|\mathcal{B}|}\sum_{k\in\mathcal{B}} s_k \cdot \log_2(y_k). \tag{6.2}$$

In Eq. (6.2), $\mathcal{B}$ is the batch set with size $|\mathcal{B}|$, which are randomly selected from total L pilot symbols transmitted at 1th, 2th, $\cdots$, Lth transmission intervals, s_k is the exact symbol transmitted at kth transmission interval, and y_k is the result of forward function of NN whose inputs are the corresponding received signals of kth and $k-1$th transmission interval, i.e., $r_{(k-1)M+1}, r_{(k-1)M+1}, \cdots, r_{kM}$ and $r_{(k-2)M+1}, r_{(k-2)M+1}, \cdots, r_{(k-1)M}$.

After the assignment of loss function with respect to the weights and bias of each neuron, the training process is pursued via back propagation based optimization algorithms, e.g., gradient descent, stochastic gradient descent, Adam, to minimize the loss function. Here, the Adam optimiser [91, 92] is selected, to derive the optimal/sub-optimal weights and bias for every neurons.

6.1.2 Simulation Results

In this part, the signal detection performance of the supervised NN-based UV signal detector is evaluated. The simulation setting is shown in Table 3.1. Here, only the OOK signal is provided for illustration. For the structure of the NN, aside from the setting of input and output layers and the activation functions, the number of neurons of the second and the third hidden layers are assigned to be 128 and 64, respectively. In Fig. 6.2, the x-coordinate of is the SNR in terms of dB, and the y-coordinate represents BER. Along with the BER curve of the supervised NN detector, the previously proposed HD non-coherent scheme and the state-of-the-art MLSD scheme are placed for comparison. It is first observed that with the increase of SNR, the noise effect

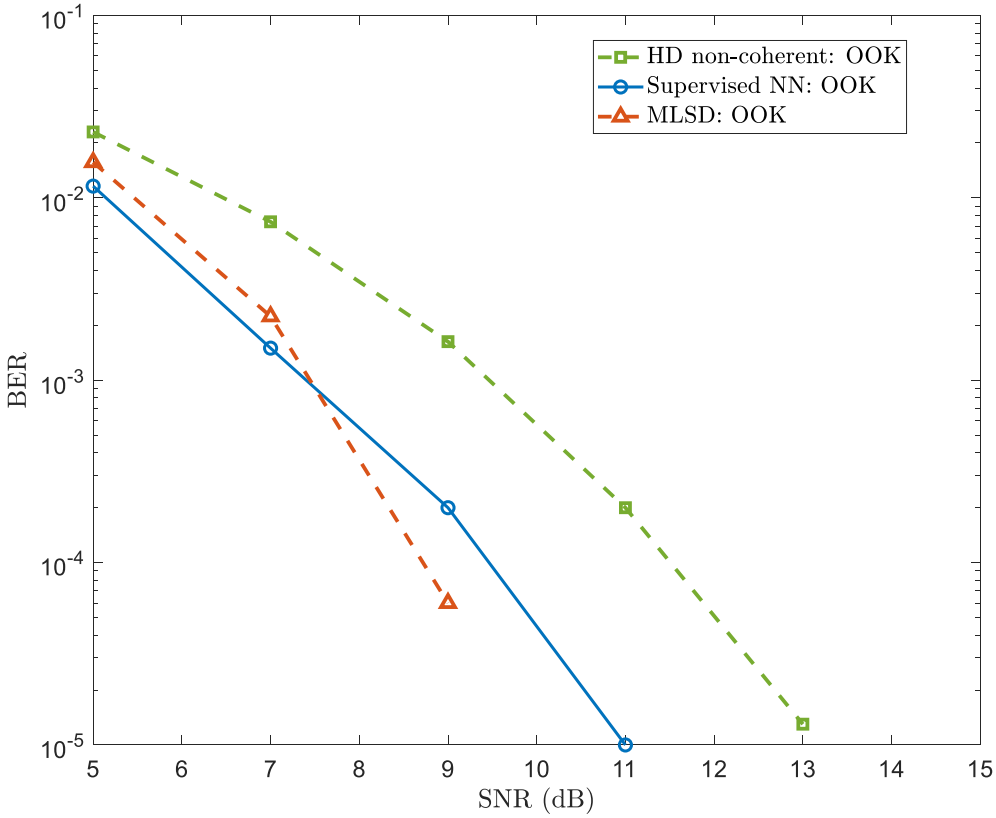


Figure 6.2: Performance of supervised NN based signal detector: BER versus SNR in dB.

decreases and the BER of all schemes is decreased. Second, it is seen that the supervised NN detector has better performance in the low SNR region. For example, when SNR= $5dB$, the BER of the supervised NN detector approaches 10^{-2}, which is lower than that of the MLSD and the non-coherent scheme. This is because in low SNR region, the performance of classical linear channel estimation, e.g., LS, (serving as the prerequisite for coherent detection such as MLSD) is compromised by the noise effect. In contrast, even if with the same noisy pilot sequences as training data, the introduction of non-linearity via the activation functions of NN paves its way to explore the hidden mapping from the received signals to its transmitted symbols, which renders the lower BER by a well trained supervised NN based detector. Next, it is shown in Fig. 6.2

that with the SNR increases, MLSD outperforms the supervised NN detector, consistent with the fact that MLSD with perfect CSI gives the optimal signal detection performance.

6.2 Half-Supervised NN UV Detector with Explainable AI

It is noteworthy that the supervised NN detector inevitably suffers from the heavy overhead burden of training data. This is especially worse when the channel is time-varying, as NN requires to be re-trained for different CIR. Moreover, fully connected NN does not have explainability given its black-box nature. In this section, a half-supervised NN detector is provided, which is able to adapt with the time-varying channel effect for UV communications. Also, this NN is explainable by approximating the likelihood-based decision surfaces in its core. The idea is to put the received UV signals as the input of a Parzen window-based neural network (PNN), which aims to approximate the likelihood PDF (conditioned on transmitted symbols). Then, a classifier can be easily pursued given the approximated likelihood PDF.

The advantages here have two-folds. First, compared to the fully supervised fully connected NN that requires data overhead for training and re-training in the time-varying channel scenarios, the training process and overhead of PNN is lightweight [93]. This is because the constructed feature is in the binary detection framework, irrelevant with ISI, which makes its likelihood PDF simple and therefore can be approximated by less samples (training data). Second, other than the black-box fully connected NN that is opaque and difficult for explanation, the essential of PNN is to approximate likelihood for signal detection purposes, which is straightforward and is easily explainable via mathematical deduction.

6.2.1 Parzen Window PDF Approximation

The Parzen-windowing method [94] aims to approximate the desired PDF via an assigned window and a function that is defined on it. The function here is referred to as the window tailed function. As such, for the total M received signals of kth symbol, i.e., $\mathbf{r}_k \triangleq [r_{(k-1)M+1}, r_{(k-1)M+2}, \cdots, r_{kM}]^T$ with the unknown likelihood PDF conditioned on its transmitted symbol s_k, i.e., $p(\mathbf{r}_k|s_k)$, Parzen-windowing can approximate $p(\mathbf{r}_k|s_k)$ via the samples extracted by the window function $W(\mathbf{r})$, i.e. [94, 95],

$$p(\mathbf{r}_k|s_k) \simeq \frac{1}{L}\sum_{l=1}^{L}\frac{1}{\eta^d}\cdot W\left(\frac{\mathbf{r}_k - \mathbf{r}^{(l)}}{\eta}\right). \tag{6.3}$$

In Eq. (6.3), $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \cdots, \mathbf{r}^{(L)}$ are the L samples of the $\mathbf{r}$. η represents the virtual variance which will be discussed later. From Eq. (6.3), it is noticed that the accuracy of the Parzen window approximation relies on how one designs the window function $W(\mathbf{r})$. Here, the general Gaussian Parzen window function is adopted as the stacked received signals $\mathbf{r}_k$ follows the high-dimensional Gaussian distribution (the received noise is assumed as independent Gaussian distributed). As such, the window function is expressed as:

$$W(\mathbf{r}) = \frac{1}{\sqrt{(2\pi)^d}}\cdot\exp\left(-\frac{1}{2}\mathbf{r}^T\mathbf{r}\right). \tag{6.4}$$

Then, by taking Eq. (6.4) into Eq. (6.3), the approximated likelihood PDF $p(\mathbf{r}_k|s_k)$ can be computed as:

$$p(\mathbf{r}_k|s_k) \backsimeq \frac{1}{L}\sum_{l=1}^{L}\frac{1}{\sqrt{(2\pi\eta^2)^d}}\cdot\exp\left[-\frac{\left(\mathbf{r}_k - \mathbf{r}^{(l)}\right)^T\cdot\left(\mathbf{r}_k - \mathbf{r}^{(l)}\right)}{2\eta^2}\right]. \tag{6.5}$$

After the approximation process of the likelihood PDF, it is shown in the following equation that such approximation is unbiased with the increase

of the number of samples. This is done by showing with the infinite number of samples (training data), the approximated PDF equals the real PDF, i.e.,

$$\begin{aligned}
&\lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L}\frac{1}{\sqrt{(2\pi\eta^2)^d}}\cdot\exp\left[-\frac{\left(\mathbf{r}_k-\mathbf{r}^{(l)}\right)^T\cdot\left(\mathbf{r}_k-\mathbf{r}^{(l)}\right)}{2\eta^2}\right]\\
=&\mathbb{E}_{r^*\sim p(r^*)}\left\{\frac{1}{\sqrt{(2\pi\eta^2)^d}}\cdot\exp\left[-\frac{\left(\mathbf{r}_k-\mathbf{r}^*\right)^T\cdot\left(\mathbf{r}_k-\mathbf{r}^*\right)}{2\eta^2}\right]\right\}\\
=&\int_{\mathbb{R}^d}p(\mathbf{r}^*)\cdot\frac{1}{\sqrt{(2\pi\eta^2)^d}}\cdot\exp\left[-\frac{\left(\mathbf{r}_k-\mathbf{r}^*\right)^T\cdot\left(\mathbf{r}_k-\mathbf{r}^*\right)}{2\eta^2}\right]d\mathbf{r}^*\\
\overset{(a)}{=}&\int_{\mathbb{R}^d}p(\mathbf{r}^*|s_k)\cdot p(\mathbf{r}_k|\mathbf{r}^*)d\mathbf{r}^*\\
\overset{(b)}{=}&p(\mathbf{r}_k|s_k).
\end{aligned}\tag{6.6}$$

In Eq. (6.6), (a) is due to the fact that the samples follow the distributions of the real likelihood PDF, i.e., $\mathbf{r}^*\sim p(\mathbf{r}^*|s_k)$, and $\mathbf{r}_k$ follows a Gaussian PDF conditioned on $\mathbf{r}^*$. (b) is because the Chapman-Kolmogorov function [96], which suggests the unbiased likelihood approximation from the Parzen window technique.

6.2.2 Parzen Window-based Neural Network Detector

After the description of the Parzen window technique, the implementation details for likelihood approximation and further signal detection process are elaborated. This is done by Parzen window-based probabilistic neural networks [97]. The structure of the PNN is shown in Fig. 6.3. The input layer is the M received UV signals of kth transmitted symbol s_k, which are then forwarded to the pattern layer. The pattern layer is composed of the previous L stacked received signals, i.e., $\mathbf{r}^{(L)}=\mathbf{r}_{k-1},\mathbf{r}^{(L-1)}=\mathbf{r}_{k-2},\cdots,\mathbf{r}^{(1)}=\mathbf{r}_{k-L}$, labelled by the corresponding symbols $s_{k-1},s_{k-2},\cdots,s_{k-L}$. Here, such symbols can be the pilot sequences or from the previous signal detection results from proposed non-coherent detection methods in Chapters 4-5. The output layer

is the symbol classes, which are connected by the pattern neurons that are labelled as belonging to those classes.

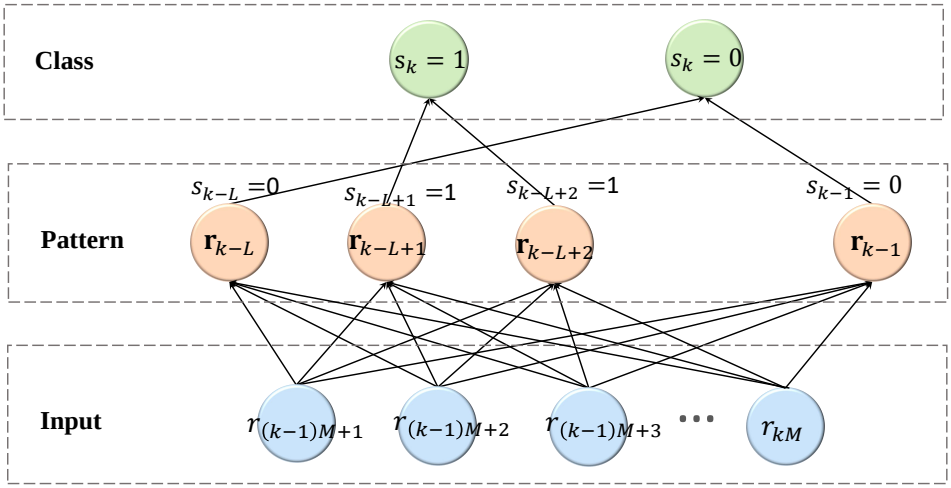


Figure 6.3: Illustration of Parzen-PNN that consists of three layers, i.e., the input layer fed by the designed metric $\mathbf{z}_k$, the pattern layer trained by the labels $s_{k-L}, s_{k-L+1}, \cdots, s_{k-1}$, and the class layer for classification.

Training Process & Signal Detection

The training process of the PNN is straightforward from the samples, i.e., to directly assign L previous stacked received signals $\mathbf{r}^{(L)} = \mathbf{r}_{k-1}, \mathbf{r}^{(L-1)} = \mathbf{r}_{k-2}, \cdots, \mathbf{r}^{(1)} = \mathbf{r}_{k-L}$ as the values of the pattern neurons. Then, the output of each pattern neuron $\mathbf{r}^l$ (with $1 \leq l \leq L$) is:

$$T(\mathbf{r}_k, \mathbf{r}^{(l)}) = \exp\left[-\frac{\left(\mathbf{r}_k - \mathbf{r}^{(l)}\right)^T \cdot \left(\mathbf{r}_k - \mathbf{r}^{(l)}\right)}{2\eta^2}\right]. \tag{6.7}$$

In the output layer, for each output neuron (i.e., one symbol class), only the pattern neurons that are labelled as that class will connect to it, which makes the output of each neuron j as:

$$\Delta_j(\mathbf{r}_k) = \sum_{\forall l,\ s.t.\ s_l^*=j} T(\mathbf{r}_k, \mathbf{r}^{(l)}). \tag{6.8}$$

Finally, the signal detection process can be pursued by the classification decision, as

$$\hat{s}_k = \begin{cases} 1, & \Delta_1(\mathbf{r}_k) \geq \Delta_0(\mathbf{r}_k), \\ 0, & \Delta_1(\mathbf{r}_k) < \Delta_0(\mathbf{r}_k). \end{cases} \tag{6.9}$$

In comparison with the fully connected NN detector with supervision, the proposed PNN detector does not require training data overhead. Instead, it directly extracts the previously received signals $\mathbf{r}_{k-1}, \mathbf{r}_{k-2}, \cdots, \mathbf{r}_{k-L}$ for further likelihood PDF approximation and detection process. Also, PNN is relatively simple compared with the other neural network architectures, e.g., recurrent neural network (RNN) [98], and the radial basis function networks (RBFN) [99].

Virtual Variance Assignment

One practical issue is how to assign the virtual variance η. The virtual variance η^2 serves as the variance-like parameter in the exponential expression, which will affect the accuracy of likelihood approximation and further the signal detection [100]. To be specific, a small η^2 will produce a spike-like PDF approximation, which leads to a curving decision surface and much extra false detection (missing 1 bits). On the other hand, an over-smoothing likelihood PDF would be provided with a large η^2, which makes it insensitive to the receiving and thermal noises. Here, two guides to assign an appropriate virtual variance are provided in the following.

Given the Gaussian distribution of the stacked received signals, an straightforward way to assign the virtual variance as the variance of receiving noise, i.e., $\eta^2 = \sigma^2$. The performance of this setting is provided in the following simulation part.

Whole Parzen-PNN Signal Detection Process

Parzen-PNN is a three-layer neural network. The first layer is the input layer with M received samples of kth time-slot symbol, which then fully connect to the second layer. The second layer is the pattern layer consisting of L neurons, with the activation function as the exponential expression in Eq. (6.7). The third layer is the output layer, whose neurons represent the classes. Recall that the training process is to directly assign the pattern neurons as the vector of the received signals of previous $k-l$ $(l = 1, \cdots, L)$ time interval, and to connect to the class neurons with the previous detection results. The process of Parzen-PNN for kth time-slot is to (i) input the M received samples, (ii) compute the exponential expression (pattern neuron activation function) by Eq. (6.7), (iii) compute the value of each class neuron by Eq. (6.8), and (iv) pursue symbol detection (classification) by comparing the computed values of the class neurons via Eq. (6.9).

6.2.3 Simulation Results

In this part, the PNN based UV signal detector will be evaluated with OOK modulation pattern for UV communications. The simulation configuration is listed in Table 3.1.

The BER performance versus the SNR of the received signals is shown in Fig. 6.4, where the x-coordinate of is the SNR in terms of dB, and the y-coordinate represents BER. Also, the previously proposed HD non-coherent schemes and the supervised NN detector are included for comparison. It is observed that with increasing SNR, the noise effect decreases and the BER of all schemes is decreased. Then, it is observed that the PNN scheme outperforms the HD non-coherent scheme, but is worse than the supervised NN-based detector. This is explained via the following two parts. First, when compared with HD non-coherent detection scheme, PNN actually uses the full likelihood

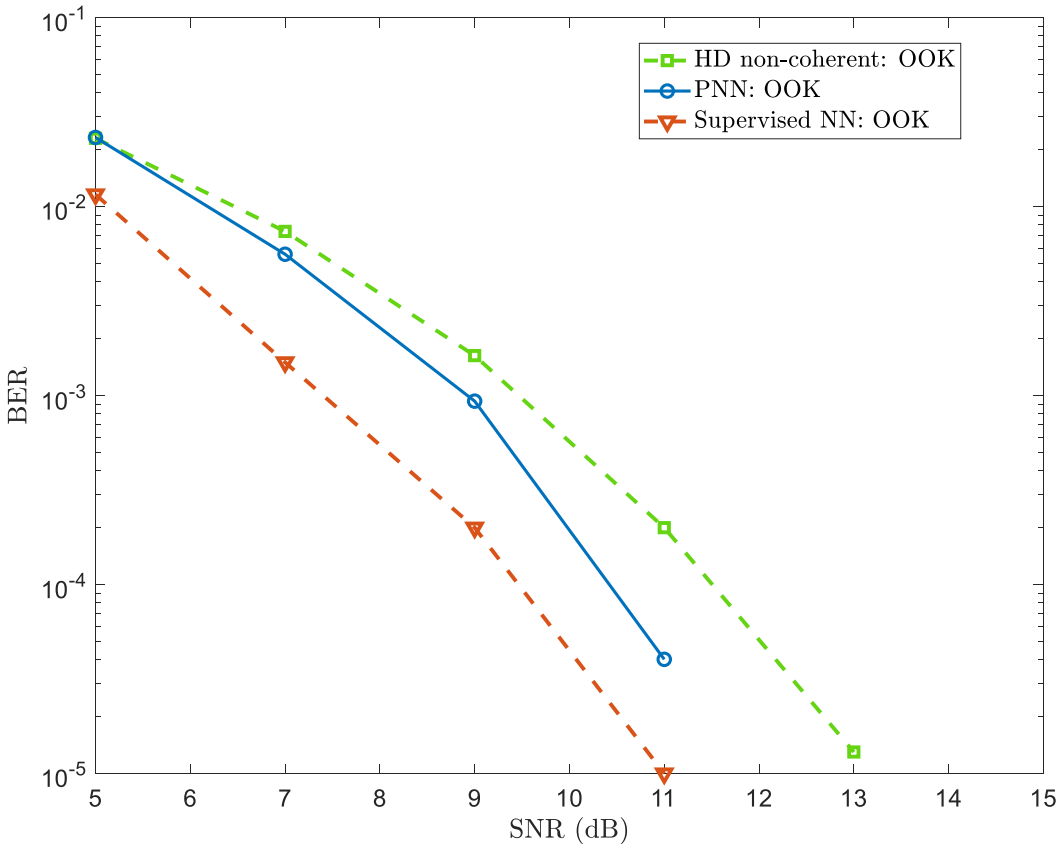


Figure 6.4: Performance of PNN based signal detector.

PDF, i.e., $p(\mathbf{r}_k|s_k)$, which contains more information than the compressed one (compressed from M to D dimensions) used by HD non-coherent, i.e., $p(\mathbf{z}_k = \mathbf{H}\mathbf{r}_k|s_k)$. From this point of view, the detection of PNN can rely on more signal-level features provided by the full likelihood, which leads to better signal detection as opposed to the HD non-coherent scheme. Second, when compared to the supervised NN-based detector, PNN has one layer less, which makes it less effective to capture the nonlinearity and hidden relations between the received signals and the transmitted symbols. In this view, as opposed to the fully connected NN-based detector that is opaque to explainability, PNN indeed sacrifices detection accuracy but has more explainable features given its aims to approximate the likelihood PDF for signal detection.

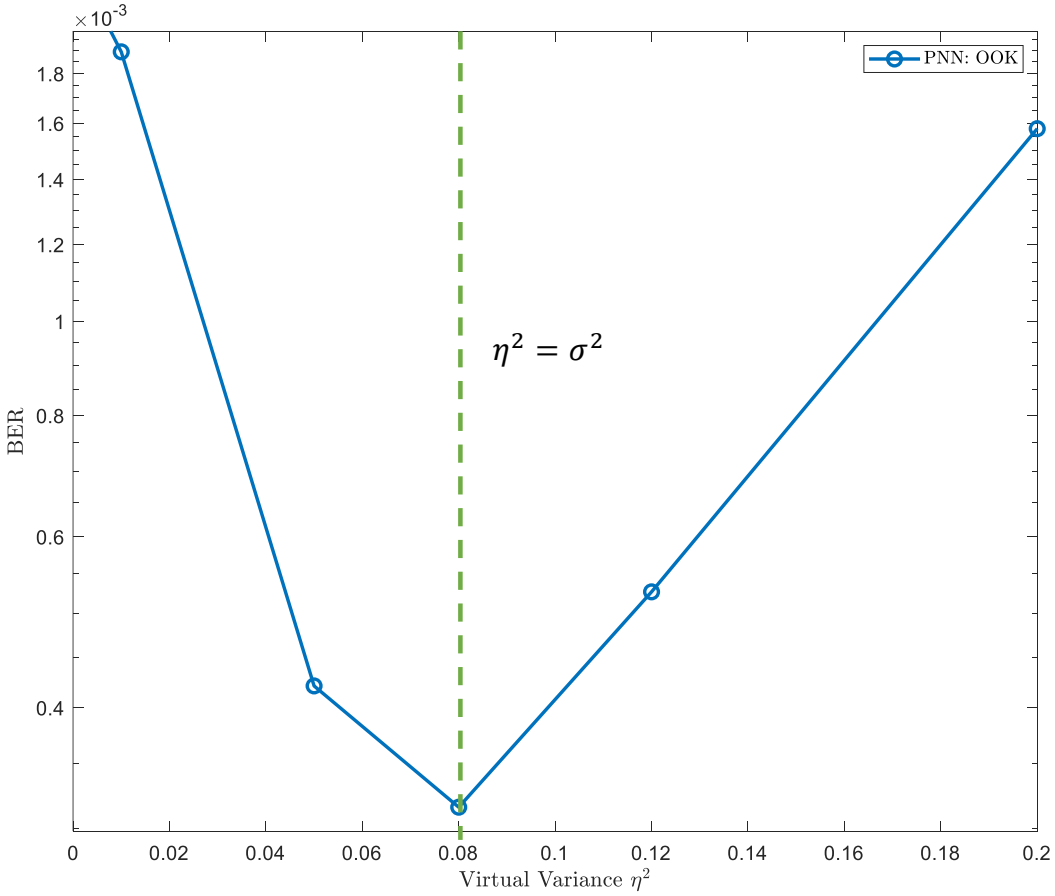


Figure 6.5: Selection of virtual variance η^2, in which the suggested virtual variance η^2 gives the lowest BER.

Next, the effect of virtual variance on detection performance is considered. This is shown by Fig. 6.5, where the x-coordinate represents the virtual variance η^2 and the y-coordinate is the corresponding BER. As can be seen from Fig. 6.5, with the increase of the virtual variance η^2, the BER at first goes down and then grows up. This can be explained by looking into Eq. (6.5), in which a spiky Gaussian likelihood PDF will be achieved with a small η^2, but a smooth likelihood PDF will be reached if η^2 is too large. The former will cause false detection by missing the amount of $1 - bit$ whereas the latter will make the detector insensitive to the receiving noise. Then, it is found that the suggested virtual variance setting, i.e., $\eta^2 = \sigma^2$ provides the optimal

BER, which demonstrates the completion of the proposed PNN-based signal detector.

6.2.4 Conclusion & Discussion

This chapter has deeply explored the hidden nonlinear characteristic of UV signals, in which machine learning techniques have been utilised to train the detector process via the combinations of linear weights and bias of neurons and the nonlinear activation functions. Firstly, a supervised fully connected NN has been used to map the received UV signals to their transmitted symbols. The neurons of the output layer of the trained NN present the nonlinear features for non-coherent detection, which outperforms the state-of-the-art MLSD in the low SNR region and gives a comparable BER (the gap is less than 1dB) in the high SNR region. The supervised NN-based detector requires data overhead for training and re-training in the time-varying channel scenarios and is difficult to be explainable due to its black-box nature. This further makes the NN-based features in-explainable. To solve this problem, a physical-aware Parzen window-based NN detector has been designed, which aims to approximate the likelihood PDF of the received signals, and thereby is easily explainable via mathematical deduction.

Chapter 7

Summary & Future Works

7.1 Summary of this Thesis

Wireless UV communication is regarded as a promising supplement for conventional wireless communication systems, due to its (i) strong scattering induced NLoS to make flexible device alignment, and (ii) the solar-blind property to prevent from contamination of solar beam. However, these characteristic also leads to two challenges that cause communication performance deterioration. Firstly, compared with other FSOs using different wavelength, the scattering phenomenon gives rise to more intensive ISI. Secondly, the flexible device deployment may produce a time-varying channel effect. Both are heavily deteriorating the estimation/detection of the transmitted symbols from the received signals. To deal with the these issues, the coherent detection schemes leveraging the CIR estimation have been well studied to compensate the ISI effect. However, the coherent schemes are less attractive when addressing the signal detection in the time-varying UV channel, as it will fall into either overhead burden of pilot sequences and resources or poor CIR acquisition that further hinders the detection performance.

In this thesis, a novel non-coherent detection framework has been pro-

posed and studied, which aims to convert the sequential signal detection process into a discrete binary detection framework, by the construction of ISI-resistant features. By doing so, the complex channel estimation process and the pilot sequence overhead issues for time-varying CIR scenarios have been solved. To implement this idea, features were designed by (i) manually extracting the signal geometric features (e.g., the rising edge, the inflection, and the energy difference) that are inherently resistant to the ISI effect, and (ii) using NN-based methods to learn more sophisticated nonlinear and hidden features for signal detection. Theoretical proof and simulation have been performed to demonstrate the feasibility and the performance of the proposed non-coherent idea and algorithms, which show comparable symbol detection accuracy with the state-of-the-art coherent detection scheme but with lower computational complexity. In this view, by casting the complex sequential detection into the concise discrete detection framework, and combining the manual and machine learning-based ISI-resistant feature construction, this work provides a novel idea, not only for UV communications but also can be applied to other communication and signal detection scenarios suffering from ISI.

In Chapter 2, a literature review on existing detection methods have been conducted. These methods have been classified into two genres: coherent detection and non-coherent detection. Several conventional methods have been introduced with detailed theories and implementations.

In Chapter 3, the basic ultraviolet communication system has been illustrated. Besides, the effect of ISI on the UV channel has been evaluated by numerical simulation, in terms of varying the symbol rate and the different modulation methods (i.e., J-PAM).

In Chapter 4, the 1D non-coherent method tries to find the optimal linear combination of the designed ISI-resistant features, by minimizing the theoretical BER. Firstly, four linear features of UV signals have been designed

and extracted based on the geometry shapes of the received signals. These include the intra-symbol rising edge, the intra-symbol energy, the inter-symbol minimum inflection, and the inter-symbol energy difference of the signal respectively. Such features are inherently insensitive to ISI contamination. Secondely, leveraging the extracted four features, two novel non-coherent schemes have been proposed (i.e., EWNCD and OWNCD). The proposed 1D non-coherent schemes are capable of converting the detection scenarios with ISI into a binary detection framework, which avoids the complex channel estimation process. Moreover, the proposed OWNCD method is able to adjust the optimal weights with respect to the different levels of the ISI intensities and therefore delivers a better detection performance. Thirdly, in order to automatically adjust the time-varying channel, a heuristic algorithm has been designed to adaptively achieve the optimal weights. This heuristic algorithm relies on the previous data that can reflect the continuous changes in the channel environment, thereby capable of addressing the time-varying UV channel challenges, but with lower computational complexity. Finally, the detection performance of the proposed EWNCD and OWNCD schemes have been numerically evaluated. Compared to the state-of-the-art coherent MLSD in the cases of static and time-varying CSI, the proposed OWNCD can gain $\sim$1 dB and $\sim$8 dB in SNR at the 7% overhead FEC limit (BER of 4.5×10^{-3}), respectively, and has also reduced the computational complexity by four orders of magnitude.

In Chapter 5, a high-dimensional non-coherent detection method has been illustrated, which is capable of dealing with the multi-level PAM modulations. To be specific, a high-dimensional combination of the designed geometrical features has been constructed, which leads to a high-dimensional feature space. Leveraging these extracted ISI-resistant features, the HD detection method transforms the ISI-caused sequential detection into the ISI-released

multi-hypothesis detection framework, which, therefore, avoids the high computational CIR estimation and sequential likelihood computations required by the coherent schemes. Then, the BER performance of the HD method has been analyzed, by deriving the theoretical BER of the HD feature-based non-coherent detection. It mathematically shows it is better than the previously proposed one-dimensional non-coherent schemes, in terms of the BER performance. Next, a clustering algorithm has been designed to conduct the classification of the HD features for signal determination, based on the GMM clustering framework. Finally, numerical simulation has been conducted to evaluate the HD method. This has illustrated that the proposed HD method is able to provide higher SNR for signal detection than current non-coherent schemes, which leads to a very-close detection accuracy with the state-of-the-art coherent MLSD ($< 1dB$ in SNR at the FEC limit BER=4.5×10^{-3}). These thereby have demonstrated the promising pathway of the proposed non-coherent scheme towards a low-computational but high-accurate detection paradigm for optical wireless communications with intensive ISI.

In Chapter 6, to deeply explore the hidden nonlinear characteristic of UV signals, machine learning techniques have been utilised to train the detector or feature construction process via the combinations of linear weights and bias of neurons and the nonlinear activation functions. Two different machine learning technologies have been investigated. First, a supervised fully connected NN has been used to map the received UV signals to their transmitted symbols. The neurons of the output layer of the trained NN present the nonlinear features for non-coherent detection, which outperforms the state-of-the-art MLSD in the low SNR region and gives a comparable BER (the gap is less than 1dB) in the high SNR region. However,the supervised NN-based detector requires data overhead for training and re-training in the time-varying channel scenarios and is difficult to be explainable due to its black-box nature. This

further makes the NN-based features in-explainable. To solve this problem, a physical-aware Parzen window-based NN detector has been designed, which aims to approximate the likelihood PDF of the received signals, and thereby is easily explainable via mathematical deduction.

7.2 Future Works

The future works will mainly focus on pursuing deep machine learning method to extract the hidden feature for accuracy detection, and try to build a reliable and explainable artificial intelligence (X-AI) [101] framework for signal detection combating ISI.

7.2.1 Reinforcement Learning the Optimal Feature Construction

In the Chapter 6, NN-based algorithms have been used as classifiers to map the input received UV signals to the output symbol category for detection purposes. The disadvantage is their reliance on labelled pilot sequences. In the future, I would like to use reinforcement learning to learn the features that can distinguish the transmitted symbols, which does not require pilot sequences. The aim is to learn the mapping from the received signals to detection features, which will then be used for classic signal detection method (e.g., GMM in Chapter 5).

Given the continuous space of the features, the deep deterministic policy gradient (DDPG) based feature construction scheme will be used. This is different from other popular reinforcement learning framework (e.g., deep Q network, and Proximal Policy Optimization), where only discrete actions or stochastic actions can be achieved [102–104]. In the following, the DDPG based feature construction method will be provided, with preliminary schematic

flow.

DDPG based Feature Construction

DDPG framework contains the actor and critic NNs. As is shown in Fig. 7.1, the input of the actor NN is the state from environment, and the output is the action corresponding to the input state. For the critic NN, the input is the combination of the state from environment, and the action provided by the actor NN. The output of the critic NN is the q value to evaluate the current inputted state and action. Here, both actor and critic NNs contain the real-time ones, denoted as $\pi_\theta(state)$ and $Q_\varpi(state, action)$, and the targeted ones, denoted as $\pi_{\theta^-}(state)$ and $Q_{\varpi^-}(state, action)$, where the aim of the targeted ones are to store the NN parameters θ and ϖ from speed updating (referred to as the off-policy).

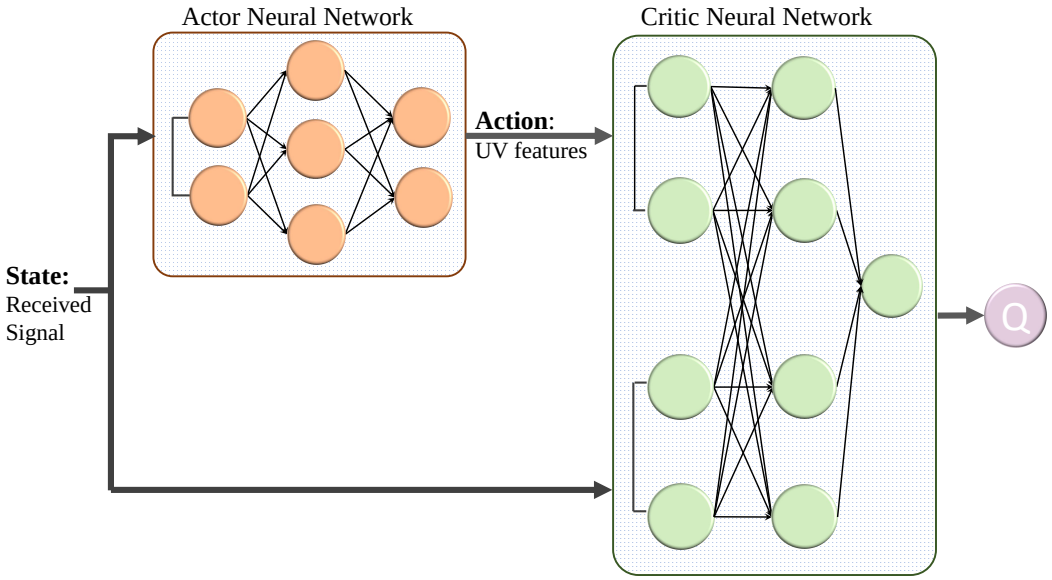


Figure 7.1: Illustration fully connected NN-based detector, where the input layer is the received UV signals, and the output layer is the symbol classes (here the symbols have two classes, representing the OOK modulations for transmitter).

To implement the DDPG framework, the first step is to define the state and action spaces, and the corresponding reward function. In the context of the UV communication scenario, the state at kth time step, denoted as $\mathbf{u}_k$, is set to be the received signals of $(k-1)th$ and kth transmitted symbols, i.e.,

$$\mathbf{u}_k = \left[r_{(k-2)M+1}, ..., r_{kM}\right]^T \tag{7.1}$$

Here, the inclusion of $(k-1)$th symbol is to compensate main part of ISI components. Then, the action for kth transmitted symbol is the features for further detection purpose, denoted as $\tilde{z}_k$, which is assigned as the output of the actor network, i.e.,

$$\tilde{z}_k = \pi_\theta(\mathbf{u}_k). \tag{7.2}$$

After the description of the state and action pair, the reward is set as by the GMM clustering results of $\tilde{z}_1, \tilde{z}_2, \cdots, \tilde{z}_k$. To be specific, by denoting the two (for OOK symbols) clustering centres as μ_0 and μ_1, and the variance of clusters as σ_0^2 and σ_1^2, the reward is set as:

$$reward = c_1|\mu_1 - \mu_0|^2 - c_2(\sigma_0^2 + \sigma_1^2) \tag{7.3}$$

where c_1 and c_2 are the weights for the combination. In Eq. (7.3), the first term is to make the gap between two clustering centres be huge to distinguish the different symbols, and the second term is to make the variance of each cluster be small to avoid the noise effect.

Training Process

The training process of DDPG can be revealed in Fig. 7.1. For the critic network, the training process follows the Bellman function. To be specific, given a batch set $\mathcal{B}$, for $\forall i \in \mathcal{B}$ state $\mathbf{u}_i$, the wanted q value can be computed

as:

$$y_i = reward + \gamma Q_{\varpi^-}(\mathbf{u}_i, \pi_{\theta^-}(\mathbf{u}_i)), \tag{7.4}$$

where γ is the discount, and $\pi_{\theta^-}(\mathbf{u}_i)$ represents the current action computed by the target actor NN. Then, the aim is to make the q value from current critic network being close to the wanted q value, i.e.,

$$Loss_{critic} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \left(y_i - Q_{\varpi}(\mathbf{u}_i, z_i)\right)^2. \tag{7.5}$$

Leveraging Eq. (7.5), the parameters in critic NN, i.e., ω can be learned by the Adam optimiser via stochastic gradient descend.

Then, for the actor NN, the objective is to maximize the q values of current state and the action. As such, the loss function is expressed as:

$$Loss_{actor} = -\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} Q_{\omega}(\mathbf{u}_i, \pi_{\theta}(\mathbf{u}_i)). \tag{7.6}$$

As such, the parameters in the actor NN, i.e., θ, can be learned by the Adam optimiser with stochastic gradient descend.

Next, a soft strategy is employed to update the target actor and critic NNs. This is done by a small merging parameter τ to combine the current updated NNs with the target ones, i.e.,

$$\theta^- = \tau\theta + (1-\tau)\theta, \tag{7.7}$$

$$\varpi^- = \tau\varpi + (1-\tau)\varpi. \tag{7.8}$$

Detection Process

After the convergence of the training process, the feature can be derived via the trained actor NN. Then, the procedure to detect the transmitted symbol

is to input the constructed feature into the classic GMM clustering algorithm.

7.2.2 Expand to Explainable AI

Then, in the future works, industrial engineering and applications will be taken into account. This includes but is not limited to understanding their KPIs, and also developing a more straightforward, interpretable, and trustworthy approach is required. Machine learning is a useful tool in solving complex problems with little expertise, which has been considered in this work to explore better and hidden features for detecting signals in the ISI contaminated scenarios. However, there is still a gap between the manually designed features and the NN trained features, and how to make those human expert features guide the NN for better ones would be an interesting topic.

To explore better features that is explainable and can resist ISI better, relying solely on the NN-based feature construction is inadequate. This is because the black-box nature of NN lacks transparency and interpretability via its inexplainable weights and bias. From this point, this may also makes them being susceptible to malicious/unintended inputs that will result in unreliable results. Therefore, the next step involves enhancing the explainability of the NN based feature construction schemes. Cutting-edge techniques and metrics, such as neuron visualization and NN symbolic representation, will be employed to create a trustworthy and interpretable artificial intelligence (X-AI) framework for signal detection that addresses ISI and adversarial environments.

Bibliography

[1] C. Gong and Z. Xu, "Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 5326–5337, Oct. 2015.

[2] H. Xu, W. Yu, D. Griffith, and N. Golmie, "A survey on industrial internet of things: A cyber-physical systems perspective," *Ieee access*, vol. 6, pp. 78238–78259, 2018.

[3] J. Wan, H. Yan, Q. Liu, K. Zhou, R. Lu, and D. Li, "Enabling cyber–physical systems with machine–to–machine technologies," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 13, no. 3-4, pp. 187–196, 2013.

[4] L. Wang and G. Wang, "Big data in cyber-physical systems, digital manufacturing and industry 4.0," *International Journal of Engineering and Manufacturing (IJEM)*, vol. 6, no. 4, pp. 1–8, 2016.

[5] G. Staple and K. Werbach, "The end of spectrum scarcity [spectrum allocation and utilization]," *IEEE spectrum*, vol. 41, no. 3, pp. 48–52, 2004.

[6] W. J. Sabol, H. C. West, M. Cooper, *et al.*, "Prisoners in 2008," *NCJ*, vol. 228417, pp. 1–46, 2009.

[7] H. Elgala, R. Mesleh, and H. Haas, "Indoor optical wireless communication: potential and state-of-the-art," *IEEE Communications Magazine*, vol. 49, no. 9, pp. 56–62, 2011.

[8] H. Haas, J. Elmirghani, and I. White, "Optical wireless communication," 2020.

[9] M. Z. Chowdhury, M. T. Hossan, A. Islam, and Y. M. Jang, "A comparative survey of optical wireless technologies: Architectures and applications," *ieee Access*, vol. 6, pp. 9819–9840, 2018.

[10] N. L. Alpert, W. E. Keiser, and H. A. Szymanski, *IR: theory and practice of infrared spectroscopy*. Springer Science & Business Media, 2012.

[11] I. Siddique, M. Z. Awan, M. Y. Khan, and A. Mazhar, "Li-fi the next generation of wireless communication through visible light communication (vlc) technology," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 5, no. 1, pp. 30–37, 2019.

[12] R. J. Drost and B. M. Sadler, "Survey of ultraviolet non-line-of-sight communications," *Semicond. Sci. Tech.*, vol. 29, p. 084006, Jun. 2014.

[13] L. Feng, R. Q. Hu, J. Wang, P. Xu, and Y. Qian, "Applying vlc in 5g networks: Architectures and key technologies," *IEEE Network*, vol. 30, no. 6, pp. 77–83, 2016.

[14] F. Delgado-Rajo, A. Melian-Segura, V. Guerra, R. Perez-Jimenez, and D. Sanchez-Rodriguez, "Hybrid rf/vlc network architecture for the internet of things," *Sensors*, vol. 20, no. 2, p. 478, 2020.

[15] E. Poves, G. del Campo-Jimenez, and F. J. López-Hernández, "Vlc-based light-weight portable user interface for in-house applications,"

in *2011 IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 365–367, IEEE, 2011.

[16] X. Cao, J. Zhao, Z. Wang, and B. Xing, "New insight into the photo-transformation mechanisms of graphene oxide under uv-a, uv-b and uv-c lights," *Journal of Hazardous Materials*, vol. 403, p. 123683, 2021.

[17] E. Trakhovsky, A. Ben-Shalom, U. P. Oppenheim, A. D. Devir, L. S. Balfour, and M. Engel, "Contribution of oxygen to attenuation in the solar blind uv spectral region," *Applied optics*, vol. 28, no. 8, pp. 1588–1591, 1989.

[18] N. Alwadai, S. Mitra, M. N. Hedhili, H. Alamoudi, B. Xin, N. Alaal, and I. S. Roqan, "Enhanced-performance self-powered solar-blind uv-c photodetector based on n-zno quantum dots functionalized by p-cuo micro-pyramids," *ACS Applied Materials & Interfaces*, vol. 13, no. 28, pp. 33335–33344, 2021.

[19] Y. Tang, G.-q. Ni, Z.-l. Wu, L.-j. Zhang, and Y. Lin, "Research on channel character of solar blind uv communication," in *Advanced Materials and Devices for Sensing and Imaging III*, vol. 6829, pp. 21–30, SPIE, 2007.

[20] D. Kedar, "Multiaccess interference in a non-line-of-sight ultraviolet optical wireless sensor network," *Applied optics*, vol. 46, no. 23, pp. 5895–5901, 2007.

[21] A. Vavoulas, H. G. Sandalidis, N. D. Chatzidiamantis, Z. Xu, and G. K. Karagiannidis, "A survey on ultraviolet C-Band (UV-C) communications," *IEEE Commun. Surveys Tut.*, vol. 21, pp. 2111–2133, thirdquarter 2019.

[22] M. A. El-Shimy and S. Hranilovic, "Binary-input non-line-of-sight solar-blind UV channels: Modeling, capacity and coding," *J. Opt. Commun. Netw.*, vol. 4, pp. 1008–1017, Dec. 2012.

[23] D. Kedar and S. Arnon, "Subsea ultraviolet solar-blind broadband free-space optics communication," *Opt. Eng.*, vol. 48, p. 046001, Apr. 2009.

[24] Z. Xu and B. M. Sadler, "Ultraviolet communications: potential and state-of-the-art," *IEEE Commun. Mag.*, vol. 46, pp. 67–73, May 2008.

[25] G. A. Shaw, A. M. Siegel, and M. L. Nischan, "Demonstration system and applications for compact wireless ultraviolet communications," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement II*, vol. 5071, pp. 241–252, International Society for Optics and Photonics, Sep. 2003.

[26] K.-X. Sun, B. Allard, S. Buchman, S. Williams, and R. L. Byer, "LED deep UV source for charge management of gravitational reference sensors," *Classical Quant. Gravity*, vol. 23, p. S141, Mar. 2006.

[27] B. Albrecht, S. Kopta, O. John, M. Rütters, M. Kunzer, R. Driad, N. Marenco, K. Köhler, M. Walther, and O. Ambacher, "Improved AlGaN pin photodetectors for monitoring of ultraviolet radiation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, pp. 166–172, May 2014.

[28] H.-Y. Liu, W.-C. Hsu, B.-Y. Chou, and Y.-H. Wang, "Fabrication AlGaN/GaN MIS UV photodetector by H2O2 oxidation," *IEEE Photon. Technol. Lett.*, vol. 27, pp. 101–104, Oct. 2014.

[29] H. Ding, G. Chen, A. K. Majumdar, B. M. Sadler, and Z. Xu, "Modeling of non-line-of-sight ultraviolet scattering channels for communication," *IEEE J. Sel. Area Commun.*, vol. 27, pp. 1535–1544, Dec. 2009.

[30] R. J. Drost, T. J. Moore, and B. M. Sadler, "UV communications channel modeling incorporating multiple scattering interactions," *J. Opt. Soc. Am. A.*, vol. 28, pp. 686–695, Apr. 2011.

[31] M. R. Luettgen, J. H. Shapiro, and D. M. Reilly, "Non-line-of-sight single-scatter propagation model," *J. Opt. Soc. Am. A.*, vol. 8, pp. 1964–1972, Dec. 1991.

[32] R. J. Drost, T. J. Moore, and B. M. Sadler, "Ultraviolet scattering propagation modeling: analysis of path loss versus range," *J. Opt. Soc. Am. A.*, vol. 30, pp. 2259–2265, Nov. 2013.

[33] M. A. Elshimy and S. Hranilovic, "Impact of finite receiver-aperture size in a non-line-of-sight single-scatter propagation model," *J. Opt. Soc. Am. A.*, vol. 28, pp. 2568–2576, Dec. 2011.

[34] M. Wu, D. Han, X. Zhang, F. Zhang, M. Zhang, and G. Yue, "Experimental research and comparison of LDPC and RS channel coding in ultraviolet communication systems," *Opt. Express*, vol. 22, pp. 5422–5430, Mar 2014.

[35] H. Qin, Y. Zuo, F. Li, R. Cong, L. Meng, and J. Wu, "Analytical link bandwidth model based square array reception for non-line-of-sight ultraviolet communication," *Opt. Express*, vol. 25, pp. 22693–22703, Sep. 2017.

[36] W. Hu, M. Zhang, D. Han, Q. Chen, L. Ai, Q. Li, and Z. Ghassemlooy, "Research on channel-related polar code with an optimum code length for wireless ultraviolet communications," *Opt. Express*, vol. 25, pp. 28630–28642, Nov. 2017.

[37] X. Liang, M. Zhang, X. Li, L. Ai, and Z. Ghassemlooy, "Security performance of LDPC and polar codes in UV wireless communications," in

in Proc. 1st West Asian Colloquium Opt. Wireless Commun., Isfahan, Iran, pp. 1–6, Apr. 2018.

[38] N. D. Chatzidiamantis, G. K. Karagiannidis, and M. Uysal, "Generalized maximum-likelihood sequence detection for photon-counting free space optical systems," *IEEE Trans. Commun.*, vol. 58, pp. 3381–3385, Dec. 2010.

[39] Z. Wei, W. Hu, M. Zhang, D. Han, B. Li, and C. Zhao, "Viterbi estimation on the finite-state markov ultra-violet channels," in *Asia Communications and Photonics Conference (ACP)*, pp. 1–3, Nov. 2017.

[40] Z. Wei, W. Hu, D. Han, M. Zhang, B. Li, and C. Zhao, "Simultaneous channel estimation and signal detection in wireless ultraviolet communications combating inter-symbol-interference," *Opt. Express*, vol. 26, pp. 3260–3270, Feb. 2018.

[41] D. Zou, C. Gong, and Z. Xu, "Signal detection under short-interval sampling of continuous waveforms for optical wireless scattering communication," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 3431–3443, May 2018.

[42] J. G. Proakis and M. Salehi, *Digital communications*, vol. 4. McGraw-hill New York, 2001.

[43] X. Zhu and J. M. Kahn, "Free-space optical communication through atmospheric turbulence channels," *IEEE Trans. Commun.*, vol. 50, pp. 1293–1300, Nov. 2002.

[44] J. Li, J. Q. Liu, and D. P. Taylor, "Optical communication using subcarrier PSK intensity modulation through atmospheric turbulence channels," *IEEE Trans. Commun.*, vol. 55, pp. 1598–1606, Aug. 2007.

[45] M. L. B. Riediger, R. Schober, and L. Lampe, "Multiple-symbol detection for photon-counting MIMO free-space optical communications," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 5369–5379, Dec. 2008.

[46] Y. J. Cho, H. B. Yilmaz, W. Guo, and C. Chae, "Effective enzyme deployment for degradation of interference molecules in molecular communication," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2017.

[47] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Transactions on NanoBioscience*, vol. 13, no. 1, pp. 31–43, 2014.

[48] Z. Lin, B. Li, Z. Wei, Y. Huang, W. Guo, and C. Zhao, "Diversity-based non-coherent signal detector for molecular communication via reaction-diffusion," *IEEE Transactions on Communications*, vol. 71, no. 5, pp. 2618–2631, 2023.

[49] W. Hu, M. Zhang, D. Han, Q. Chen, L. Ai, Q. Li, and Z. Ghassemlooy, "Research on channel-related polar code with an optimum code length for wireless ultraviolet communications," *Opt. Express*, vol. 25, pp. 28630–28642, Nov 2017.

[50] Z. Wei, B. Li, and C. Zhao, "On the polar code for the 60-ghz millimeter-wave systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, pp. 1–11, 2015.

[51] L. Jin, Y. Li, B. Li, Z. Wei, and J. Shi, "Performance of polar coding for the power line communications in the presence of impulsive noise," *IET Communications*, vol. 9, no. 17, pp. 2101–2106, 2015.

[52] W. Hu, Z. Luo, D. Han, Q. Chen, L. Ai, and M. Zhang, "A scheme of ultraviolet communication system with polar channel coding," in *2017*

16th International Conference on Optical Communications and Networks (ICOCN), pp. 1–3, IEEE, 2017.

[53] A. Shokrollahi, "Ldpc codes: An introduction," in *Coding, cryptography and combinatorics*, pp. 85–110, Springer, 2004.

[54] Z. Wei, B. Li, W. Hu, W. Guo, and C. Zhao, "Hamming–luby rateless codes for molecular erasure channels," *Nano Communication Networks*, vol. 23, p. 100280, 2020.

[55] L. Jin, Y. Li, C. Zhao, Z. Wei, B. Li, and J. Shi, "Cascading polar coding and lt coding for radar and sonar networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, pp. 1–12, 2016.

[56] P. Song, Y. Tan, X. Geng, and T. Zhao, "Noise reduction on received signals in wireless ultraviolet communications using wavelet transform," *IEEE Access*, vol. 8, pp. 131626–131635, 2020.

[57] F. Chang, K. Onohara, and T. Mizuochi, "Forward error correction for 100G transport networks," *IEEE Commun. Mag.*, vol. 48, pp. S48–S55, Mar. 2010.

[58] W. Hu, Z. Wei, S. Popov, M. Leeson, M. Zhang, and T. Xu, "Non-coherent detection for ultraviolet communications with inter-symbol interference," *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4699–4707, 2020.

[59] W. Hu, M. Zhang, Z. Li, S. Popov, M. Leeson, and T. Xu, "High-dimensional feature based non-coherent detection for multi-intensity modulated ultraviolet communications," *Journal of Lightwave Technology*, vol. 40, no. 7, pp. 1879–1887, 2021.

[60] S. J. Lee, "On the training of mimo-ofdm channels with least square channel estimation and linear interpolation," *IEEE Communications Letters*, vol. 12, no. 2, pp. 100–102, 2008.

[61] Z. Wei, B. Li, C. Sun, and W. Guo, "Sampling and inference of networked dynamics using log-koopman nonlinear graph fourier transform," *IEEE Transactions on Signal Processing*, vol. 68, pp. 6187–6197, 2020.

[62] Å. Björck, "Least squares methods," *Handbook of numerical analysis*, vol. 1, pp. 465–652, 1990.

[63] D. Kilinc and O. B. Akan, "Receiver design for molecular communication," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 705–714, 2013.

[64] A. Klein, G. K. Kaleh, and P. W. Baier, "Zero forcing and minimum mean-square-error equalization for multiuser detection in code-division multiple-access channels," *IEEE transactions on Vehicular Technology*, vol. 45, no. 2, pp. 276–287, 1996.

[65] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum ber block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2410–2423, 2003.

[66] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "Mmse decision-feedback equalizers and coding. i. equalization results," *IEEE transactions on Communications*, vol. 43, no. 10, pp. 2582–2594, 1995.

[67] G. L. Stüber and G. L. Stèuber, *Principles of mobile communication*, vol. 2. Springer, 2001.

[68] T. Foggi, G. Colavolpe, E. Forestieri, and G. Prati, "Channel estimation algorithms for mlsd in optical communication systems," *IEEE Photonics Technology Letters*, vol. 18, no. 19, pp. 1984–1986, 2006.

[69] Z. Wei, B. Li, W. Guo, W. Hu, and C. Zhao, "Sequential bayesian detection of spike activities from fluorescence observations," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 5, no. 1, pp. 3–18, 2019.

[70] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE signal processing magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[71] A. Doucet, A. M. Johansen, *et al.*, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of nonlinear filtering*, vol. 12, no. 656-704, p. 3, 2009.

[72] B. Li, J. Hou, X. Li, Y. Nan, A. Nallanathan, and C. Zhao, "Deep sensing for space-time doubly selective channels: When a primary user is mobile and the channel is flat rayleigh fading," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3362–3375, 2016.

[73] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on bernoulli filters: theory, implementation and applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3406–3430, 2013.

[74] J. M. Borwein and P. B. Borwein, *Pi and the AGM*, vol. 9. Wiley New York, 1987.

[75] B. Li, C. Zhao, and W. Guo, "Non-linear signal detection for molecular communications," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–6, 2017.

[76] Z. Wei, W. Guo, B. Li, J. Charmet, and C. Zhao, "High-dimensional metric combining for non-coherent molecular signal detection," *IEEE Transactions on Communications*, vol. 68, pp. 1479–1493, Mar. 2020.

[77] Y. Zhang, M. Li, S. Wang, S. Dai, L. Luo, E. Zhu, H. Xu, X. Zhu, C. Yao, and H. Zhou, "Gaussian mixture model clustering with incomplete data," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 1s, pp. 1–14, 2021.

[78] Y. Li, J. Zhang, Z. Ma, and Y. Zhang, "Clustering analysis in the wireless propagation channel with a variational gaussian mixture model," *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 223–232, 2018.

[79] G. Gan, C. Ma, and J. Wu, *Data clustering: theory, algorithms, and applications*. SIAM, 2020.

[80] S. Borman, "The expectation maximization algorithm-a short tutorial," *Submitted for publication*, vol. 41, 2004.

[81] F. A. Potra and S. J. Wright, "Interior-point methods," *Journal of computational and applied mathematics*, vol. 124, no. 1-2, pp. 281–302, 2000.

[82] A. S. Nemirovski and M. J. Todd, "Interior-point methods for optimization," *Acta Numerica*, vol. 17, pp. 191–234, 2008.

[83] M. Grant, S. Boyd, and Y. Ye, "Cvx users' guide," *online: http://www. stanford. edu/boyd/software. html*, 2009.

[84] J. A. Anderson, *An introduction to neural networks*. MIT press, 1995.

[85] H. Abdi, D. Valentin, and B. Edelman, *Neural networks*. No. 124, Sage, 1999.

[86] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 30, no. 4, pp. 451–462, 2000.

[87] G. Rogova, "Combining the results of several neural network classifiers," *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pp. 683–692, 2008.

[88] T. Denoeux, "A neural network classifier based on dempster-shafer theory," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 30, no. 2, pp. 131–150, 2000.

[89] R. Reed and R. J. MarksII, *Neural smithing: supervised learning in feedforward artificial neural networks*. Mit Press, 1999.

[90] X. Wang, X. Lin, and X. Dang, "Supervised learning in spiking neural networks: A review of algorithms and evaluations," *Neural Networks*, vol. 125, pp. 258–280, 2020.

[91] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pp. 1–2, Ieee, 2018.

[92] S. Bock and M. Weiß, "A proof of local convergence for the adam optimizer," in *2019 international joint conference on neural networks (IJCNN)*, pp. 1–8, IEEE, 2019.

[93] B. Yadav and V. S. Devi, "Novelty detection applied to the classification problem using probabilistic neural network," in *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 265–272, IEEE, 2014.

[94] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[95] B. Li, Z. Zhou, D. Li, and W. Zou, "A novel parzen probabilistic neural network based noncoherent detection algorithm for distributed ultra-wideband sensors," *Journal of Network and Computer Applications*, vol. 34, no. 6, pp. 1894–1902, 2011.

[96] A. H. Jazwinski, *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[97] D. F. Specht, "Probabilistic neural networks," *Neural networks*, vol. 3, no. 1, pp. 109–118, 1990.

[98] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (indrnn): Building a longer and deeper rnn," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5457–5466, 2018.

[99] J. Ghosh and A. Nag, "An overview of radial basis function networks," *Radial basis function networks 2: new advances in design*, pp. 1–36, 2001.

[100] Z. Wei, W. Guo, B. Li, J. Charmet, and C. Zhao, "High-dimensional metric combining for non-coherent molecular signal detection," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1479–1493, 2019.

[101] B. Leichtmann, C. Humer, A. Hinterreiter, M. Streit, and M. Mara, "Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task," *Computers in Human Behavior*, vol. 139, p. 107539, 2023.

[102] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*, pp. 387–395, Pmlr, 2014.

[103] D. Wang and M. Hu, "Deep deterministic policy gradient with compatible critic network," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[104] P. Wang, H. Li, and C.-Y. Chan, "Continuous control for automated lane change behavior based on deep deterministic policy gradient algorithm," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1454–1460, IEEE, 2019.