

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

# APPLICATIONS OF DEEP LEARNING FOR PHYSICAL LAYER OF WIRELESS COMMUNICATION SYSTEMS

XIE YIHANG

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

**2024**

# APPLICATIONS OF DEEP LEARNING FOR PHYSICAL LAYER OF WIRELESS COMMUNICATION SYSTEMS

**XIE YIHANG**

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy

**2024**

# Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

10-04-2024

Date



XIE YIHANG

# Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

10-04-2024

Date



Prof. Teh Kah Chan

# Authorship Attribution Statement

This thesis contains material from 4 paper(s) published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Chapter 2 is published as Y. Xie, K. C. Teh and A. C. Kot, "Deep learning-based joint detection for OFDM-NOMA scheme," in IEEE Communications Letters, vol. 25, no. 8, pp. 2609-2613, Aug. 2021, doi: 10.1109/LCOMM.2021.3077878.

The contributions of the co-authors are as follows:

- A/Prof Teh provided the initial research direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by A/Prof Teh and Prof Kot.
- I co-designed the communication system structure and deep learning network with A/Prof Teh and Prof Kot.
- I performed all the numerical experiments with code programming and data analysis at the Center of Information Sciences and Systems (CISS).
- I co-derived the theoretical contribution with A/Prof Teh for performing optimization of the system design.

Chapter 3 is published as Y. Xie, X. Liu, K. C. Teh and Y. L. Guan, "Robust deep learning-based end-to-end receiver for OFDM system with non-linear distortion," in IEEE Communications Letters, vol. 26, no. 2, pp. 340-344, Feb. 2022, doi: 10.1109/LCOMM.2021.3132326.

The contributions of the co-authors are as follows:

- Dr. Liu and A/Prof Teh provided the initial research direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by Dr. Liu, A/Prof Teh and Prof Guan.
- I co-designed the communication system structure and deep learning network with Dr. Liu. Dr. Liu provided the preliminary study of the non-linear distorted communication system.
- I performed all the numerical experiments with code programming and data analysis at the Center of Information Sciences and Systems (CISS).

- I co-derived the theoretical contribution with Dr. Liu for performing optimization of the system design.

Chapter 5 is published as Y. Xie, K. C. Teh and A. C. Kot, "Comm-Transformer: A Robust Deep Learning Based Receiver for OFDM System under TDL Channel," in IEEE Transactions on Communications, doi: 10.1109/TCOMM.2023.3343787.

The contributions of the co-authors are as follows:

- A/Prof Teh provided the initial research direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by A/Prof Teh and Prof Kot.
- I co-designed the communication system structure and deep learning network with A/Prof Teh and Prof Kot.
- I performed all the numerical experiments with code programming and data analysis at the Center of Information Sciences and Systems (CISS).
- I co-derived the theoretical contribution with A/Prof Teh for performing optimization of the system design.

Chapter 6 is under review as Y. Xie, K. C. Teh and A. C. Kot, "Generative semantic communication system: Transmission beyond bits with diffusion model," under review in IEEE Communication Letters.

The contributions of the co-authors are as follows:

- A/Prof Teh provided the initial research direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by A/Prof Teh and Prof Kot.
- I co-designed the communication system structure and deep learning network with A/Prof Teh and Prof Kot.
- I performed all the numerical experiments with code programming and data analysis at the Center of Information Sciences and Systems (CISS).
- I co-derived the theoretical contribution with A/Prof Teh for performing optimization of the system design.

10-04-2024

xie yihang

Date

XIE YIHANG

# Acknowledgements

This research is supported by the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. It would not have been possible without the kind support and help of my supervisor Prof. Teh Kah Chan, who gives me a lot of encouragement and technical support. Also, it would be a great opportunity for me to meet Prof. Alex C. Kot and Prof. Guan Yong Liang, who give me kind support and suggestions from different perspectives. Again, I would like to extend my sincere thanks to all of them.

I would like to express my gratitude to members of the Centre for Information Sciences and Systems for their kind cooperation and encouragement which help me in the completion of this report.

I would like to thank my family for their constant love, encouragement, and understanding. Their unwavering support and sacrifices have been the cornerstone of my academic success, and I am grateful for their patience and belief in me.

I would also like to express my gratitude to my friends and colleagues for their camaraderie, motivation, and intellectual stimulation. Their diverse interests, experiences, and perspectives have enriched my research and broadened my horizons, and I am fortunate to have had the opportunity to work alongside them.

Last but not least, I would like to express my sincere gratitude to every individual I have met during the Ph.D. journey for the warmth and support extended to me during my baptism. It was an immensely meaningful and profound experience that deepened my faith and enriched my life.

Finally, I am deeply grateful to all those who have supported me throughout my Ph.D. journey, and I will always cherish the memories and lessons learned during this transformative experience.

*"The only true wisdom is in knowing you know nothing."*

—Socrates

# Abstract

The objective of the thesis is to exploit the recent advances in deep learning-based techniques for wireless communication systems. The proposed methods aim to improve the performance and robustness of communication systems in different scenarios. The first research work proposes a deep learning-based receiver for non-orthogonal multiple access (NOMA) joint signal detection. The receiver combines channel estimation, equalization, and demodulation into an end-to-end process, which results in improved performance and robustness compared to traditional methods. The proposed method is tested with a tapped-delay line (TDL) channel model, which is commonly used in fifth generation (5G) communication systems. The second research work presents a deep learning-based receiver for orthogonal frequency-division multiplexing (OFDM) systems. The proposed receiver, named one-dimensional transmit and recovery network (1D-TRNet), is designed to be robust to non-linear clipping distortion. The 1D-TRNet receiver outperforms traditional OFDM receivers and state-of-the-art generalized approximate message passing (GAMP) receivers in terms of bit-error rate (BER) performance. The third research work proposes a deep learning-based 5G receiver, named communication transformer network (Comm-Trans Net), which is designed to be robust for different sub-types of TDL channels. The proposed receiver uses an attention mechanism to compensate for the multi-path fading effect of different OFDM subcarriers. The proposed positional encoding method for each OFDM subcarrier offsets the deep fading effect, resulting in improved performance compared to traditional MMSE channel estimation with GAMP receivers and state-of-the-art deep learning-based receivers. The fourth research work introduces a novel communication system called the diffusion model-based generative semantic communication (DM-GSC)

system. The system utilizes the diffusion model and semantic embedding to enable efficient and robust communication in wireless networks. The proposed DM-GSC system shows promising results in robustness for different channel conditions. Overall, the thesis demonstrates the potential of deep learning-based techniques in improving the performance and robustness of wireless communication systems. The proposed methods have the potential to enhance the quality of service and enable novel communication scenarios in the future.

# Contents

**Acknowledgements** **ix**

**Abstract** **xiii**

**List of Figures** **xix**

**List of Tables** **xxi**

**Symbols and Acronyms** **xxiii**

**1 Introduction** **1**
1.1 Background Overview . . . . . 1
1.2 Literature Review . . . . . 3
1.2.1 DL-based End-to-End Wireless Communication Systems . . 4
1.2.2 DL-based CSI Feedback . . . . . 6
1.2.3 DL-based Channel Equalization . . . . . 8
1.2.4 DL-based Channel Estimation . . . . . 9
1.2.5 DL-based Semantic Communication Systems: Transmission Beyond Bits . . . . . 11
1.3 Motivation and Novelty . . . . . 13
1.3.1 Motivation of the thesis . . . . . 13
1.3.2 Novelty of the thesis . . . . . 14
1.4 Contribution of the Thesis . . . . . 15

**2 Deep Learning Based Joint Detection for OFDM-NOMA Scheme** **17**
2.1 Introduction . . . . . 17
2.2 System model and Problem Illustration . . . . . 19
2.2.1 System Architecture . . . . . 20
2.2.1.1 Signal structure . . . . . 20
2.2.1.2 Power domain NOMA scheme . . . . . 21
2.2.1.3 Tapped-delay line channel model . . . . . 22
2.2.2 Problem illustration . . . . . 22
2.3 Network design and Training Procedure . . . . . 22
2.3.1 Network Design . . . . . 22

2.3.2 Training Procedure . . . 25
2.4 Simulation Results and Discussion . . . 26
2.4.1 simulation results for different receivers . . . 28
2.4.2 Simulation results for three users . . . 29
2.4.3 Impact of Pilot Size . . . 31
2.4.4 Complexity Analysis . . . 32
2.5 Conclusion . . . 33

**3 Autoencoder-decoder based CSI Feedback for Massive MIMO Communication 35**
3.1 Introduction . . . 35
3.2 System model . . . 36
3.3 Problem Illustration . . . 37
3.4 Autoencoder-decoder based channel state information feedback process . . . 38
3.4.1 Autoencoder compressing process . . . 38
3.4.2 Autodecoder decompressing process . . . 40
3.5 Experiments Results . . . 43
3.5.1 Ablation Experiment . . . 43
3.5.2 NAIC Practical 5G CSI dataset Results . . . 46
3.5.3 Performance evaluation for different compression rates . . . 47
3.6 Conclusion . . . 50

**4 1D-TRNet Based Receiver for the OFDM System with Non-linear Distortion 51**
4.1 Introduction . . . 51
4.2 System Model . . . 53
4.3 Proposed 1D-TRNet Receiver . . . 54
4.3.1 Problem Formulation . . . 55
4.3.2 The Proposed 1D-TRNet Structure . . . 55
4.3.2.1 Dual 1D convolution operation (Yellow Block) . . . 56
4.3.2.2 Max-pooling operation (Green Block) . . . 57
4.3.2.3 Transpose convolution operation (Red Block) . . . 58
4.3.2.4 Concatenating operation (Grey Block) . . . 58
4.3.2.5 Sigmoid and classification . . . 58
4.3.3 Dataset Generation and Training Procedure . . . 59
4.4 Computational Complexity Analysis . . . 60
4.5 Simulation Results and Discussions . . . 61
4.5.1 Simulation Results without OOB Filter . . . 61
4.5.2 Simulation Results with OOB Filter . . . 63
4.5.3 Simulation Results for different numbers of Multipath . . . 64
4.5.4 Ablation Experiment . . . 66
4.5.5 Comparisons of Online Testing Computational Complexity . . . 69
4.6 Conclusion . . . 70

**5 Comm-Transformer: A Robust Deep Learning Based Receiver for 5G Communication under TDL Channel 71**
5.1 Introduction . . . 71
5.1.1 Motivation and Related Work . . . 74
5.1.2 Contribution . . . 75
5.2 System Model and Channel State Information . . . 76
5.3 Comm-Trans Net based End-to-End communication . . . 78
5.3.1 Attention Mechanism and Channel Positional Encoding . . . 78
5.3.2 Comm-Transformer Network . . . 81
5.3.3 Training and Testing . . . 85
5.3.3.1 Problem Formulation . . . 85
5.3.3.2 Loss Design . . . 85
5.4 Computational Complexity Analysis . . . 86
5.5 Simulation Results and Discussions . . . 88
5.5.1 Channel estimation results . . . 91
5.5.2 Robustness of the proposed Comm-Trans Net under different TDL sub-types . . . 95
5.5.3 Impact of different pilot conditions . . . 100
5.5.4 Ablation Experiment . . . 101
5.5.5 Comparisons of Online Testing Computational Complexity . 104
5.6 Conclusion . . . 106

**6 Generative Semantic Communication System: Transmission Beyond Bits with Diffusion Model 107**
6.1 Introduction . . . 107
6.2 System Model . . . 110
6.3 Proposed Diffusion Model-Based Receiver . . . 111
6.3.1 The Proposed DM-GSC Structure . . . 111
6.3.2 Diffusion Process Semantic Encoder . . . 112
6.3.3 Reverse Denoise Semantic Decoder . . . 115
6.3.4 Problem Formulation . . . 116
6.4 Training Procedure and Implementation Setting . . . 117
6.4.1 Forward diffusion semantic encoder . . . 117
6.4.2 Reverse denoise semantic decoder . . . 117
6.5 Numerical Results and Discussions . . . 118
6.5.1 Simulation Results for Tokenzied Feature Map Through Different Channels . . . 119
6.5.2 Generative Content Results for Diffusion Model-Based Semantic Communication . . . 120
6.6 Conclusion . . . 122

**7 Conclusion and Future Works 123**
7.1 Conclusion . . . 123
7.2 Future Work . . . 125

**List of Author's Awards, Patents, and Publications** **129**

**Bibliography** **133**

# List of Figures


2.1 System block diagram . . . . . 20
2.2 TR 38.901 TDL-E channel model . . . . . 27
2.3 Simulation results with 64 pilots and 16 cyclic prefixes. . . . . 28
2.4 Simulation results for three-user NOMA system with 16 pilots, CP length 16, and 64 subcarriers. . . . . 29
2.5 Simulation results with 16 pilots and 16 cyclic prefixes. . . . . 31

3.1 CSI feedback block diagram. . . . . 36
3.2 Autoencoder-decoder based CSI feedback structure. . . . . 38
3.3 Encoder block diagram . . . . . 39
3.4 Decoder block diagram . . . . . 40
3.5 Network hyperparameters for the CSI Encoder . . . . . 42
3.6 Network hyperparameters for the CSI Decoder . . . . . 43
3.7 Hyperparameters for the CRBlock . . . . . 44

4.1 The structure of the proposed 1D-TRNet receiver. . . . . 55
4.2 Simulation results without OOB filter. . . . . 62
4.3 Simulation results with different CRs ($E_b/N_0 = 25$ dB). . . . . 63
4.4 Simulation results with OOB filter. . . . . 64
4.5 Simulation results for different numbers of multipath. . . . . 65
4.6 Ablation experiment of block size. . . . . 68
4.7 Ablation experiment of kernel size. . . . . 68

5.1 System block diagram of the proposed Comm-Trans Net receiver. . 79
5.2 Attention block diagram. . . . . 80
5.3 Comm-Trans Net structure. . . . . 81
5.4 Different TDL sub-type channels. . . . . 89
5.5 TDL-A channel estimation result under SNR = $20dB$, NMSE = -15.7 dB. . . . . 92
5.6 TDL-B channel estimation result under SNR= $20dB$, NMSE = -16.2 dB. . . . . 92
5.7 TDL-C channel estimation result under SNR= $20dB$, NMSE = -17.5 dB. . . . . 92
5.8 TDL-D channel estimation result under SNR= $20dB$, NMSE = -19.6 dB. . . . . 93

5.9 TDL-E channel estimation result under SNR = $20dB$, NMSE = -19.0 dB. . . . . 93
5.10 NMSE simulation results for TDL-A channel. . . . . 94
5.11 NMSE simulation results for TDL-E channel. . . . . 95
5.12 BER simulation results for TDL-A channel. . . . . 96
5.13 BER simulation results for TDL-B channel. . . . . 97
5.14 BER simulation results for TDL-C channel. . . . . 97
5.15 BER simulation results for TDL-D channel. . . . . 98
5.16 BER simulation results for TDL-E channel. . . . . 98
5.17 BER simulation results for cross domain validation of TDL-E channel. 99
5.18 BER simulation results for TDL-A channel with different number of pilots. . . . . 100
5.19 BER simulation results for TDL-E channel with different number of pilots. . . . . 101
5.20 Ablation experiment of kernel size. . . . . 103
5.21 Ablation experiment of block size. . . . . 104

6.1 The structure of the proposed DM-GSC system. . . . . 112
6.2 The example for forward diffusion process and reverse denoise process. 112
6.3 The network structure of the proposed DM-GSC receiver. . . . . 113
6.4 MSE results for the transmission loss under different channels. . . . 119
6.5 DM-GSC generative examples for the semantic receiver. . . . . 120
6.6 SSIM results for DM-GSC system under different channels. . . . . 121

# List of Tables


2.1 Channel Model Parameters . . . . . 26
2.2 Computational Complexity Comparison . . . . . 32

3.1 Ablation experiments for the proposed AED . . . . . 46
3.2 Comparisons of NMSE Loss and Score . . . . . 48
3.3 Comparisons of NMSE Loss and Score under different compression rate . . . . . 49

4.1 Parameters for generating the dataset . . . . . 60
4.2 Comparisons of Online Testing Computation Complexities . . . . . 69

5.1 Computational Complexity Comparison . . . . . 88
5.2 System Model Parameters . . . . . 90
5.3 Channel Model Parameters . . . . . 91
5.4 Comparisons of Online Testing Computation Complexities . . . . . 105

6.1 Parameters for generating the channel conditions . . . . . 118

# Symbols and Acronyms

## Symbols

| | |
|---|---|
| $\mathcal{R}^n$ | the $n$-dimensional Euclidean space |
| $\mathcal{H}$ | the Euclidean space |
| $\|\cdot\|_2$ | the 2-norm of a vector or matrix in Euclidean space |
| $\|\cdot\|_G$ | the induced norm of a vector in G-space |
| $\|\cdot\|_E$ | the induced norm of a vector or matrix in probabilistic space |
| $\odot$ | the Hadamard (component-wise) product |
| $\otimes$ | the Kronecker product |
| $\langle\cdot,\cdot\rangle$ | the inner product of two vectors |
| $f(\cdot)$ | the composition of functions |
| $\nabla f$ | the gradient vector |
| $\mathcal{C}^k$ | the function with continuous partial derivatives up to $k$ orders |
| $x_i$ | the $i$-th component of a vector $x$ |
| $x_{i,t}$ | the $i$-th component of a vector $x$ at time $t$ |
| $\bar{x}$ | the vector with the average of all components of $x$ as each element |
| $\mathbf{1}$ | all-ones column vector with proper dimension |
| $\mathcal{C}$ | the average space, i.e., $span\{\mathbf{1}\}$ |
| $\mathcal{C}^\perp$ | the disagreement space, i.e., $span^\perp\{\mathbf{1}\}$ |
| $\Pi_\parallel$ | the projection matrix to the average space $\mathcal{C}$ |
| $\Pi_\perp$ | the projection matrix to the disagreement space $\mathcal{C}^\perp$ |
| $O(\cdot)$ | order of magnitude or ergodic convergence rate (running average) |
| $o(\cdot)$ | non-ergodic convergence rate |

| | |
|---|---|
| $\mathcal{N}_i$ | the index set of the neighbors of agent $i$ |

# Acronyms

| | |
|---|---|
| 3GPP | The 3rd Generation Partnership Project |
| 4G | Fourth Generation |
| 5G | Fifth Generation |
| ADC | Analog To Digital Converter |
| AI | Artificial Intelligence |
| AIGC | Artificial Intelligence Generative Content |
| ANN | Artificial Neural Network |
| BER | Bit Error Rate |
| B5G | Beyond Fifth Generation |
| CDL | Clustered Delay Line |
| CNN | Convolutional Neural Network |
| CSI | Channel State Information |
| DAC | Digital-To-Analog Converter |
| DL | Deep Learning |
| DM | Diffusion Model |
| DNN | Deep Neural Network |
| F-GAMP | Fast-Generalized Approximate Message Passing |
| FDD | Frequency-Division Duplex |
| GAMP | Generalized Approximate Message Passing |
| GRU | Gate Recurrent Unit |
| HPA | High Power Amplifier |
| ICI | Inter-Carrier Interference |
| IRS | Intelligent Reflecting Surfaces |
| ISI | Inter-Symbol Interference |
| LSTM | Long Short Term Memory Network |
| LS | Least Squares |

| | |
|---|---|
| LMMSE | Linear Minimum Mean-Square Error |
| MMSE | Minimum Mean-Square Error |
| NOMA | Non-Orthogonal Multiple Access |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| PAPR | Peak-To-Average Power Ratio |
| RNN | Recurrent Neural Network |
| SER | Symbol Error Rate |
| SINR | Signal-To-Noise-Plus-Interference Ratio |
| SNR | Signal-To-Noise Ratio |
| TDL | Tapped Delay Line |
| URLLC | Ultra-Reliable And Low Latency Communications |
| GSC | Generative Semantic Communication |
| i.i.d. | independent and identically distributed |
| $a.s.$ | almost sure convergence of a random sequence |

# Chapter 1

# Introduction

## 1.1 Background Overview

The demand for highly efficient communication systems, especially fifth-generation (5G) communication, has prompted extensive research. Traditional communication theories used in current systems have imposed significant limitations on further performance enhancements. To address this, recent attention has turned to deep learning (DL) techniques, which offer promising potential for optimizing wireless communications in complex scenarios. In this thesis, we delve into the development of DL solutions for 5G communication, proposing efficient schemes tailored to these scenarios. We present key ideas behind various DL-based communication methods, exploring both the opportunities and challenges they present. By envisioning the adoption of these innovative DL-based wireless physical layer frameworks, we anticipate a new direction in communication theories, propelling us forward in this realm.

This thesis introduces the potential of DL to revolutionize communication systems and provides an overview of recent advances in DL-based physical layer communications for wireless communication systems. DL has the ability to enhance the

performance of each individual block in communication systems, as well as optimize the entire transmitter/receiver system. For DL-based communication systems with block structures, we demonstrate DL's ability in signal compression and detection. We also explore recent developments in DL-based end-to-end communication systems.

The need for high-reliability and ultra-high capacity wireless communication has driven extensive research into 5G communications. However, conventional communication theories that underpin current communication systems significantly restrict further performance improvements and impose severe limitations. Recently, deep learning techniques have emerged as a promising tool for handling complex communication systems, with demonstrated potential for optimizing wireless communications.

Within this thesis, a comprehensive examination is undertaken to explore the advancements made in deep learning-based algorithms for various wireless communication scenarios. Beginning with the pioneering applications of deep learning in the Orthogonal Frequency Division Multiplexing (OFDM) system, the research progresses toward addressing more intricate and demanding problems. These challenges encompass severe doubly selective fading channels, non-linear distortion issues, and even transmission beyond traditional bit-based methods. The primary objective is to investigate the potential of deep learning techniques in effectively overcoming obstacles and providing solutions for the realm of wireless communication.

We envision that deep learning-based wireless physical layer frameworks will pave the way for a new direction in communication theories, and this work will contribute to advancing research in this area.

Overall, we have investigated various perspectives for physical layer communication problems for beyond 5G communication systems including but not limited to DL-based joint detection for imperfect successive interference cancellation (SIC) NOMA system, DL-based receiver for doubly selective fading channel, DL-based

massive massive multiple-input multiple-output (MIMO) channel state information (CSI) feedback, DL-based channel estimation for 5G tapped delay line (TDL) channel, and DL-based semantic communication system transmission beyond bits, etc.

## 1.2 Literature Review

Wireless communication systems are constantly evolving to meet the increasing demands for higher data rates, improved reliability, and enhanced spectral efficiency. The physical layer of these systems plays a crucial role in ensuring efficient and reliable transmission of data. In recent years, deep learning techniques have emerged as a promising approach for addressing various challenges and optimizing the physical layer of wireless communication systems.

DL-based end-to-end communication refers to the utilization of DL algorithms to optimize the entire transmission chain, from source encoding to channel decoding. By integrating DL models into the communication system, it becomes possible to learn and adapt to the complex characteristics of wireless channels, resulting in improved performance and enhanced end-to-end communication quality.

DL-based equalizers focus on mitigating the effects of channel distortions, such as multipath fading and interference. By training DL models to effectively equalize the received signals, these techniques can combat the impact of non-linear distortions and enhance the overall system performance.

DL-based channel estimation techniques aim to accurately estimate channel parameters, which are essential for efficient signal processing and resource allocation. By leveraging DL algorithms, these techniques improve the accuracy and efficiency of channel estimation in various channel conditions, including time-varying and frequency-selective fading channels.

DL-based CSI feedback addresses the challenge of obtaining accurate and timely CSI in communication systems, particularly in massive MIMO setups. By employing DL techniques, the efficiency and reliability of CSI feedback are enhanced, enabling improved beamforming, resource allocation, and interference management.

DL-based semantic communication explores the integration of DL models into the source coding process, enabling more efficient data compression and transmission. By capturing the semantic meaning or context of the data, DL-based techniques offer higher transmission rates and improved privacy, paving the way for advanced communication systems.

Throughout this thesis, we delve into these DL-based applications for the physical layer of wireless communication systems. By harnessing the power of DL algorithms, we aim to address various challenges, optimize system performance, and contribute to the advancement of wireless communication technologies. The subsequent chapters will delve into each application area, exploring the underlying techniques, methodologies, and their potential impact on future wireless communication systems.

### 1.2.1 DL-based End-to-End Wireless Communication Systems

Non-orthogonal multiple access (NOMA) technique has drawn much attention in recent years. It has also been a promising technique for 5G wireless communication system and beyond. In this thesis, we develop a novel DL-aided receiver for NOMA joint signal detection. The DL-based receiver serves as an end-to-end mode, which simultaneously fulfills the function of channel estimation, equalization, and demodulation. Compared with the traditional signal detection method for the NOMA scheme, the proposed deep learning method shows feasible improvement in performance and robustness with the tapped-delay line (TDL) channel model, which is adapted for the 5G communication environment.

NOMA is regarded as a promising technique for the next-generation wireless communication system to achieve higher spectral efficiency. Over the past several decades, the explosively growing demand for transmission data and mobile devices requires a higher transmission rate, low latency, and massive connectivity. The NOMA technique has the crucial concept to share the same frequency spectrum, which highly increases the efficiency, transmission rate, and connectivity [1]. Thus, NOMA shows great potential for the fifth-generation (5G) wireless communication system. Multi-user superposition transmission (MUST) technique is widely used for NOMA scheme in the power domain [2]. In traditional approaches, the SIC method is mainly used in NOMA signal detection for both uplink and downlink scenarios. However, the 5G communication scenario with a doubly-selective channel will have inter-carrier interference (ICI) and inter-symbol interference (ISI), which will increase the complexity and cause a disturbance on the receiver side.

Recently, DL has been applied to the physical layer of a wireless communication system for channel estimation and signal detection. Some works considered the deep learning-based receiver as an end-to-end approach, and other works considered a model-driven approach. In [3], the authors presented the use of emerging deep learning techniques in future wireless communication networks. A thorough survey of the literature on deep learning for wireless communication networks was also provided. In [4], the authors proposed two deep learning approaches based on artificial neural networks (ANNs) and they have been shown to provide good performance. In [5], the deep neural network (DNN) was first applied in the OFDM system for channel estimation and signal detection. It has been shown to provide significant improvement in symbol-error rate (SER) performance. Moreover, in [6], DL-based channel estimation for a time-varying Rayleigh fading channel was studied. In [7], the recurrent neural network was applied to a time-selective fading channel in the OFDM system, however, the multi-path effect is not considered.

Based on preliminary studies in the OFDM systems, the extension for DL-based application to the NOMA scheme has great potential. In [8], a fully-connected DNN was designed for signal detection in the NOMA system, which considered

the same neural network structure proposed in [5]. In [9], a model-driven deep-learning method was proposed for physical layer communications, which provided more learnable parameters to achieve better performance. The OFDM-NOMA transmission is also used to increase the system capacity. In [10], it was shown that the proposed scheme can achieve the optimal diversity gain based on an OFDM-based NOMA system. However, most research works assume that the stronger user can be perfectly detected and canceled, which may not be feasible for practical implementation. Meanwhile, the ideal Rayleigh fading channel with additive white Gaussian noise (AWGN) may not model accurately in a practical environment for the 5G wireless communication scenario.

### 1.2.2 DL-based CSI Feedback

MIMO techniques are considered a crucial technique for 5G wireless communications. MIMO-aided communication systems leverage downlink CSI to enable base stations (BS) with massive antennas to employ channel-adaptive equalization, channel-adaptive beamforming, enhancing channel capacity, and eliminating inter-user interference. However, in frequency-division duplex (FDD) systems, downlink CSI can only be estimated at user equipment (UE) and then sent back to the BS. This process incurs significant overhead, leading to the exploration of feedback reduction techniques like vector quantization and codebook-based approaches to mitigate this issue [11]. However, challenges arise due to quantization errors impacting CSI-sensitive applications, and the complexity of codebook design escalates with the increasing number of antennas.

Recently, compressive sensing (CS)-based CSI feedback approaches have emerged, addressing the mentioned problems by exploiting spatial and temporal CSI correlations. These methods employ CS techniques for feedback and reconstruction [12] or distributed compressive channel estimation [13]. However, in practice, CSI demonstrates only approximate sparsity, which introduces complexities in compressing and recovering CSI, primarily due to possible model mismatches.

To address the challenge of reducing feedback overhead, certain approaches take advantage of the time correlation of slow-varying channels [12]. These methods leverage the fact that the channel information changes slowly over time, allowing them to reuse previously learned CSI for subsequent CSI recovery, as long as the error remains below a certain threshold. This reuse of past information helps to alleviate the need for continuous feedback, thus reducing the overhead.

The literature on deep learning-based CSI feedback methodologies reveals a diverse array of techniques aimed at enhancing the performance and efficiency of convolutional neural networks (CNNs) in wireless communication systems. One prominent strategy focuses on increasing the receptive field of CNNs, thereby improving their ability to capture spatial dependencies within CSI data. For instance, CsiNet+ replaces the original 3x3 convolutional layer with two 7x7 convolutional layers in the encoder, while RefineNet sets the initial convolutional layers as 7x7 and 5x5 [14]. Additionally, CsiNet+DNN embeds fully connected layers after convolutional layers and utilizes more RefineNet blocks, whereas MRNet employs larger convolutional sizes in both the encoder and decoder [15] [16]. Another approach involves leveraging multiple resolutions to process CSI data, as demonstrated by CRNet, which utilizes parallel neural network paths with different kernel sizes at the encoder [17]. Furthermore, methods like BCsiNet stack multiple convolutional layers to enhance CSI feature quality [18]. Finally, the adoption of fully convolutional layers, as seen in ConvCsiNet, DeepCMC, ACCsiNet, and FullyConv, facilitates dimension reduction and interpolation operations, thereby improving feature extraction and reconstruction efficiency [14] [19] [20]. These advancements collectively contribute to the development of robust and efficient CSI feedback mechanisms in wireless communication systems.

However, a limitation of these methods is that the reused channel information is only an estimate and may not reflect real-time changes accurately. As a result, the resolution of the CSI may be compromised, particularly in fast-changing channels, leading to potential degradation in overall system performance. Despite this

drawback, these techniques still offer a practical solution for mitigating feedback overhead and are well-suited for scenarios with relatively slow channel variations.

In recent times, deep learning (DL) methods have shown success in wireless communications [21], [22], [5]. One notable DL-based CSI feedback network called CsiNet [23] utilizes an autoencoder-like architecture to mimic the CS and reconstruction processes. The encoder learns the channel response directly from training data to obtain compressed representations (codewords), while the decoder recovers CSI through feedforward multiplication. CsiNet outperforms conventional CS-based methods significantly. However, it neglects the time correlation present in time-varying channels and independently reconstructs CSI.

### 1.2.3 DL-based Channel Equalization

Multi-carrier techniques such as OFDM systems are widely used in wireless communications [24], [25]. Compared with single-carrier techniques, multi-carrier techniques have higher spectrum efficiency, and are more resistant to deep fading effects under multi-path channel. However, the peak-to-average power ratio (PAPR) is a persistent problem inherent in the multi-carrier system [26]. The design of high power amplifier (HPA) in the transmitter can reduce the PAPR at the expense of lower power efficiently [26], [27].

Traditionally, it has been shown that clipping of transmitted signals will cause BER performance degradation [28]. With the development of technology, it was noted that although clipping of transmitted signals causes ICI, it also introduces frequency diversity, which has the potential to be taken advantage of to improve the BER performance [29]. In [30], a generalized approximate message passing (GAMP) algorithm was proposed for recovering OFDM signals with severe non-linear distortion. It was shown in [30] that by using GAMP, the BER performance of OFDM signals with non-linear distortion is even better than OFDM signals without non-linear distortion. However, the GAMP algorithm has some limitations

due to its high computational complexity. To reduce the complexity of GAMP, a fast GAMP (F-GAMP) algorithm has been proposed in [31].

Recently, deep learning becomes an attractive tool to solve the physical layer problems in wireless communication, including this non-linear distortion in the OFDM system. Some preliminary results have been presented in [5]. However, in [5], the authors only considered OFDM signals with non-linear distortion which is not very severe. Using deep learning to recover OFDM signals with severe non-linear distortion is still an open problem.

Inspired by the GAMP theoretical analysis in [32] as well as the research work presented in [8], in this thesis, the design of a deep learning-based receiver that is robust to OFDM signals with severe non-linear clipping distortion is investigated. To exploit frequency diversity and combat ICI caused by non-linear distortion in OFDM signals, we propose a deep neural network named one-dimensional transmit and recovery net (1D-TRNet), which consists of 1D U-shape feature extractor with multiple gated recurrent units (GRUs). Our proposed 1D-TRNet extracts both low-level (local) and high-level (global) features from the received signal.

Specifically, the low-level feature denotes the information in a subcarrier and its adjacent subcarriers, and the high-level feature denotes the information in the whole OFDM block. Since clipping causes the energy on a subcarrier to spread into not only its adjacent subcarriers, but also its non-adjacent subcarriers, both the low-level and high-level features need to be exploited to preserve useful information existing in the received signal.

### 1.2.4 DL-based Channel Estimation

DL-based channel estimation is an active research area within the field of wireless communication. Channel estimation is a crucial step in communication systems, as it enables the receiver to recover transmitted signals in the presence of noise, interference, and channel impairments. In recent years, DL techniques have been

applied to channel estimation to overcome some limitations of conventional methods, such as the least squares (LS) and minimum mean square error (MMSE) estimators.

Recently, the DL-based channel estimation techniques have drawn tremendous attention for 5G communication and beyond. The initial endeavor [5] to leverage deep learning (DL) methods aimed at discerning the intricacies of frequency-selective wireless channels marks a significant milestone in combating nonlinear distortion and interference within OFDM systems. Subsequent innovations, such as the pioneering framework proposed in [33], integrate DL techniques into MIMO systems, addressing challenges like direction-of-arrival (DoA) and channel estimation. Building upon this foundation, [19] extends DL-based channel estimation to doubly selective channels, demonstrating superior performance compared to conventional methods across various scenarios. In a parallel effort, [6] treats the channel matrix as an image and employs DL-based image super-resolution and denoising techniques for channel estimation. Moreover, [34] introduces a sparse complex-valued neural network structure tailored for channel estimation in massive MIMO systems. Diverging from piecemeal enhancements, another line of research endeavors to establish end-to-end deep neural network (DNN) architectures, aiming to replace all transmitter and receiver modules, as evidenced by works such as [35], [36].

In [22], the authors proposed a DL-based channel estimation method for beamspace millimeter-wave (mmWave) massive MIMO systems. They used a convolutional neural network (CNN) to learn the channel estimation, which resulted in significant performance improvement over traditional LS and MMSE-based methods. In [37], the authors explored deep learning techniques for channel estimation in ultra-reliable low-latency communication (URLLC) scenarios within 5G networks. The authors proposed a stacked autoencoder-based channel estimator and showed that it outperformed traditional methods in various channel conditions.

In [38], the authors focused on channel estimation for massive MIMO systems with mixed-resolution analog-to-digital converters (ADCs). They proposed a deep

residual network (ResNet) based approach that could effectively recover CSI from mixed-resolution ADCs, achieving better performance than conventional methods. In [39], the authors investigated channel estimation for intelligent reflecting surfaces (IRS) using deep learning. The authors proposed a deep unfolding method based on the approximate message passing (AMP) algorithm and demonstrated its effectiveness in various IRS scenarios. In [40], the authors explored the use of graph neural networks (GNNs) for channel estimation in wireless communication systems. The authors designed a GNN-based channel estimator and showed that it could achieve superior performance compared to conventional methods, particularly in scenarios with irregular network topologies.

In summary, DL-based channel estimation techniques have shown promising results in various wireless communication scenarios, including massive MIMO, mmWave, URLLC, and IRS. These techniques, including CNNs, stacked autoencoders, ResNets, deep unfolding methods, and GNNs, have been demonstrated to outperform traditional channel estimation methods, such as LS and MMSE. However, challenges remain in terms of computational complexity, robustness, and generalization to different channel conditions and network configurations. As research in this area progresses, it is expected that DL-based channel estimation techniques will continue to improve and become more widely adopted in future wireless communication systems.

### 1.2.5 DL-based Semantic Communication Systems: Transmission Beyond Bits

DL-based semantic communication systems have emerged as a promising direction in wireless communication research. These systems aim to go beyond traditional bit-centric communication by directly transmitting and recovering high-level semantic information. This approach can potentially lead to more efficient and robust communication systems, as it allows for exploiting the structure and redundancy in the semantic content.

A literature review of some notable works in the area of DL-based semantic communication systems could be summarized. In this pioneering work [41], the authors proposed an end-to-end communication system based on DL, where both the transmitter and receiver were implemented as neural networks. This system was trained to minimize the symbol error rate directly, without relying on a predefined channel model, leading to improved robustness and performance. In [42], the authors introduced a DL-based semantic communication system that directly transmits high-level semantic information instead of bits. The authors used autoencoders for encoding and decoding the semantic content and demonstrated that the proposed system can provide better performance than traditional bit-centric approaches under various channel conditions.

As 5G deployment progresses, speculation about the next wireless generation arises. Semantic communication, driven by AI, extracts meaning from data for interpretation, posing as a promising paradigm. This article reviews DL and E2E semantic communication, highlighting unresolved challenges [43]. In [44], the authors propose a reinforcement learning based adaptive semantic coding (RL-ASC) approach that encodes images beyond pixel level. The task-related information is preserved and reconstructed properly while less important data is discarded. The compressed representation was then transmitted using a traditional communication scheme, and the system was shown to provide superior performance compared to conventional methods. In [45], authors highlight how current communication systems neglect semantics, despite 5G's potential for semantic communication, proposing a federated edge intelligence architecture to address efficiency challenges. In [46], authors discuss how end-to-end communication systems integrate physical layers for joint optimization. They introduce DeepSC, a semantic communication system utilizing deep learning and Transformers, outperforming traditional systems. In [47], authors introduce DeepSC-S, a deep learning-based semantic communication system for speech signals. It outperforms traditional methods, emphasizing semantic-level error minimization and adaptability to dynamic channel conditions

without retraining, validated across telephone and multimedia transmission systems.

In summary, DL-based semantic communication systems have shown promising results in various aspects of wireless communication, including end-to-end learning, semantic compression, modulation classification and demodulation, and semantic index modulation. These methods have the potential to provide more efficient and robust communication by exploiting the structure and redundancy in semantic content. Future research directions in this area include investigating new semantic representations, exploring novel DL architectures and algorithms for encoding and decoding, and extending the approach to different communication settings and scenarios.

## 1.3 Motivation and Novelty

### 1.3.1 Motivation of the thesis

The motivation for the project is to explore the DL-based method for existing wireless communication systems. Based on the background information, several research works have shown that DL-based methods can improve the performance of wireless communication systems regardless of whether they are model-based or data-based systems. The rapidly increasing of computation capability ensures the deep learning method can be implemented in the practical system. Since the study of this approach is in the elementary stage, there are extensive interesting problems to be studied. Generally, the objectives for each section can be summarized as follows.

### 1.3.2 Novelty of the thesis

The thesis represents a comprehensive exploration of advancements in the field of DL-based approaches for the physical layer of wireless communication, juxtaposed with established methodologies. A thorough examination of prominent deep learning-based receivers forms the cornerstone of this analysis. For instance, the DNN model [5], despite its widespread use, confronts challenges in adapting to the intricacies of complex channel environments. Similarly, CNN-based receivers [48] face hurdles in effectively processing time-sequence data, which limits their efficacy in specific scenarios. Additionally, LSTM-based receivers [49] exhibit deficiencies in feature extraction, impeding their capacity to capture nuanced patterns within the data. Furthermore, the deepRX receiver [50] grapples with non-linear distortion, presenting obstacles to its practical application in real-world settings.

In addition to assessing deep learning-based receivers, the thesis meticulously evaluates state-of-the-art conventional methods. For example, GAMP [51], while efficient, demands substantial computational resources, which may restrict its practical utility. Similarly, F-GMAP [31] showcases impressive performance but relies heavily on computational power, posing challenges in real-time applications.

The thesis's innovative contributions are underscored through a systematic comparison with existing methods. By addressing the limitations of both DL-based and conventional approaches, the thesis aims to bridge existing gaps in DL-based communication systems. Through meticulous evaluation and comparison, it not only identifies shortcomings but also introduces novel techniques capable of effectively mitigating these challenges. Ultimately, the thesis propels the field forward by offering detailed insights into enhanced methodologies for DL-based communication systems, paving the way for further advancements and applications in the realm of wireless communication.

## 1.4 Contribution of the Thesis

This thesis contributes to the advancement of deep learning-based techniques for physical layer wireless communication systems, addressing practical challenges and improving performance in various scenarios. In detail, we study the crucial problems for deep learing applications in beyond 5G (B5G) communication systems as follows:

1. DL-based joint detection for imperfect-SIC NOMA system: NOMA is a key technology for beyond 5G systems. However, imperfect SIC in NOMA systems can limit their performance. DL-based joint detection can be used to mitigate the impact of imperfect SIC on the performance of the NOMA system.

2. DL-based receiver for doubly selective fading channel: The performance of communication systems operating in doubly selective fading channels can be significantly limited by channel estimation errors. DL-based receivers can be used to improve the performance of communication systems in such channels.

3. DL-based massive MIMO CSI feedback: Massive multiple-input multiple-output is another key technology for beyond 5G systems. However, obtaining accurate channel state information (CSI) for massive MIMO systems can be challenging. DL-based techniques can be used to improve the accuracy and efficiency of CSI feedback for massive MIMO systems.

4. DL-based channel estimation for 5G TDL channel: Time-division duplex (TDD) and time-division multiplexing (TDM) channels are important components of 5G systems. Accurate channel estimation is critical for the performance of such channels. DL-based techniques can be used to improve the accuracy of channel estimation for TDL channels.

5. DL-based semantic communication system transmission beyond bits: Semantic communication systems can provide more efficient and reliable transmission than traditional bit-based systems. DL-based techniques can be used to develop and optimize semantic communication systems beyond 5G systems.

# Chapter 2

# Deep Learning Based Joint Detection for OFDM-NOMA Scheme

## 2.1 Introduction

In this chapter, the focus is on studying deep learning (DL) techniques for end-to-end communication systems. Specifically, the research centers around DL-based joint detection for the Non-Orthogonal Multiple Access (NOMA) system in the presence of imperfect SIC. The NOMA system is a key technology in 5G communication systems that enables multiple users to share the same frequency resources. However, the performance of NOMA can be affected by imperfect SIC, where residual interference from decoded signals may degrade the detection accuracy. To address this challenge, DL-based joint detection techniques are explored. These techniques leverage the power of DL algorithms to jointly detect multiple users' signals in NOMA systems, even in the presence of imperfect SIC. The aim is to enhance the overall detection accuracy and system performance. By applying DL techniques, the chapter aims to improve the performance of end-to-end communication systems, with a specific focus on the NOMA system and its joint detection

challenges. The research conducted in this chapter contributes to advancing the understanding and potential solutions for optimizing the performance of NOMA systems using DL-based approaches.

NOMA is a promising technique for achieving higher spectral efficiency in the next-generation wireless communication system, driven by the rapidly growing demand for high data transmission rates, low latency, and massive connectivity. Sharing the same frequency spectrum is a crucial concept in NOMA, significantly increasing efficiency, transmission rates, and connectivity. As a result, NOMA holds great potential for the fifth-generation (5G) wireless communication system.

In the context of 5G communication scenarios with doubly-selective channels, traditional approaches such as SIC for NOMA signal detection may face challenges due to ICI and ISI. These challenges can increase complexity and cause disturbances on the receiver side.

Recently, deep learning has been applied to the physical layer of wireless communication systems for channel estimation and signal detection. Various works have explored deep learning-based receivers, considering both end-to-end and model-driven approaches. For instance, deep neural networks (DNNs) have been applied to OFDM systems for channel estimation and signal detection, leading to significant improvements in symbol-error rate (SER) performance.

Based on preliminary studies in the OFDM systems, the extension for DL-based application to the NOMA scheme has great potential. In [8], a fully-connected DNN was designed for signal detection in the NOMA system, which considered the same neural network structure proposed in [5]. In [9], a model-driven deep-learning method was proposed for physical layer communications, which provided more learnable parameters to achieve better performance. The OFDM-NOMA transmission is also used to increase the system's capacity. In [10], it was shown that the proposed scheme can achieve the optimal diversity gain based on OFDM-based NOMA system. However, most research works assume that the stronger user can be perfectly detected and canceled, which may not be feasible for practical

implementation. Meanwhile, the ideal Rayleigh fading channel with AWGN may not model accurately in a practical environment for the 5G wireless communication scenario.

Inspired by these works, we design a novel convolutional neural network (CNN)-based feature extractor that considers the NOMA-OFDM frame's characteristics and modifies the network type for the doubly-selective fading channel under 5G communication scenarios. Our main contribution is to design the combined CNN feature extractor and time-series long short-term memory (LSTM) layer to solve the NOMA-OFDM problem. From the perspective of inter-carrier-interference and inter-symbol-interference, we use a CNN-based feature extractor for segmenting the information of the received signals. The CNN-based feature extractor imitates the conventional process to compensate for the severe ICI caused by the Doppler shift, and the LSTM layers are proposed to handle serious ISI caused by the multipath effect.

We proposed a deep-learning (DL) based receiver for the 5G communication scenario. In a practical 5G scenario which can be modeled by the tapped-delay line (TDL) doubly-selective fading channel model with severe multi-path and Doppler shift, the conventional algorithms will not perform well in symbol-error rate performance and may also incur high computational complexity. As such, our proposed DL-based receiver is designed for solving this problem in a more efficient way.

## 2.2 System model and Problem Illustration

In this section, we first present the power-domain superposition coding system with deep-learning based joint detection. Following that, the learning problem for neural networks is illustrated.

### 2.2.1 System Architecture

A power-domain NOMA with OFDM modulation scheme is considered in this thesis. Multi-user superposition coding technique is applied for the downlink NOMA scenario. The system block diagram is shown in Fig. 2.1. The upper portion of Figure 2.1 depicts a power NOMA system with QPSK modulation involving two users. The lower segment illustrates the proposed DL-based receiver designed for joint detection.

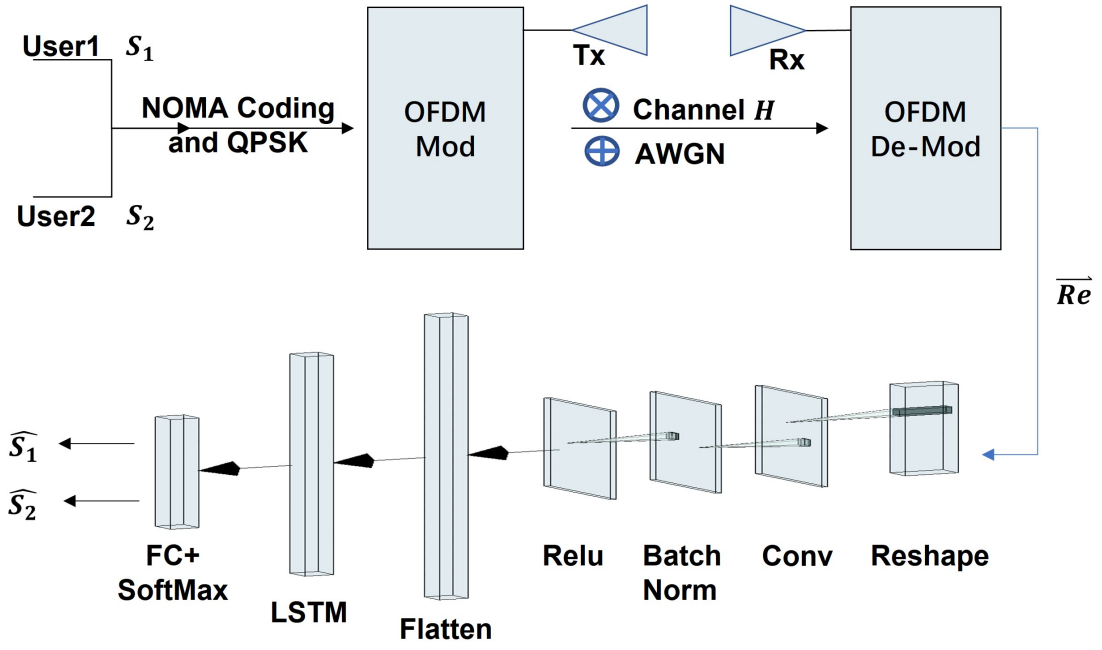



FIGURE 2.1: System block diagram

#### 2.2.1.1 Signal structure

For each user, the data frame consists of pilots and transmitted data. In a doubly-selective fading channel, the transmitted package requires cyclic prefix (CP) as protective interval to overcome the inter-carrier interference. Each frame has $N_s$ pilot symbols with $L_s$ cyclic prefix. The pilots in each frame are fixed both in magnitude and location.

#### 2.2.1.2 Power domain NOMA scheme

Multi-user superposition transmission technique is considered for $M$ users [52]. Each user is allocated with different power coefficient $\beta_i$, and $P$ denotes the total power. Let $s_i(n)$ denote the $n^{\text{th}}$ transmitted symbol of user $i$, where $n = 0, 1, 2, ..., N$. The total power allocation coefficient is summed up to 1, i.e., $\sum_{i=1}^{M} \beta_i = 1$. Therefore, the transmission sequence can be represented as

$$x(n) = \sum_{i=1}^{M} \sqrt{P\beta_i} s_i(n). \tag{2.1}$$

Following the power domain NOMA structure in [8], the received signal can be presented as

$$y(n) = h(n) * \sum_{i=1}^{M} \sqrt{P\beta_i} s_i(n) + N(n), \tag{2.2}$$

where $N(n)$ is the AWGN and $y(n)$ denotes the received signal. Furthermore, $h(n)$ denotes the TDL time-varying multi-path fading channel with $*$ denoting the convolution operation.

As a comparison, the traditional successive interference cancellation (SIC) with the least square (LS) and ideal minimum mean-square error (MMSE) methods will be applied to estimate the channel state information and detect the signal [7]. The correlation coefficient $R_{hh}$ for MMSE estimation will be calculated in advance. Overall, the process consists of two steps. Based on the received pilots, the channel state information can be estimated in the time domain. We reconstruct the first user's signal and then apply the SIC technique. After that, the second user's signal can be recovered iteratively. Let $y_2{}'(n)$ denote the estimation of the second user's signal, and it can be expressed as

$$y_2{}'(n) = y(n) - \sqrt{P\beta_1}\hat{h}(n)\hat{s_1}(n). \tag{2.3}$$

#### 2.2.1.3 Tapped-delay line channel model

In this section, the TDL channel model is considered for the 5G communication scenario. This model supports the frequency range from 0 to 100GHz with a maximum bandwidth of 2 GHz. The model covers different sub scenarios, including indoor and outdoor, consisting of line-of-sight (LoS) and non-LoS (NLoS) components. For channel coefficient generation, the fading model equation follows the TR38.901 release [53].

### 2.2.2 Problem illustration

For the proposed deep learning method, we aim to develop a novel approach to estimate the channel and detect the multi-users' signals simultaneously. Conventionally, different users' signals are considered as interference to others. Even though the SIC technique is applied to cancel the stronger signal, the cancellation is imperfect. Therefore, we propose a joint detection network to learn the TDL channel and detect the multi-user signal as an end-to-end model. In this case, the transmitted symbols are treated as the ground truth, and the received signals are considered as the input to the training network.

## 2.3 Network design and Training Procedure

### 2.3.1 Network Design

In this section, we present the structure of the proposed joint detection network. We first reshape the received signal after OFDM demodulation, and it can be expressed as

$$\mathbf{S}_{sq} = \begin{bmatrix} S^{Q}_{sq1,1} & S^{Q}_{sq1,2} & S^{Q}_{sq1,3} & \cdots & S^{Q}_{sq1,m} \\ S^{I}_{sq1,1} & S^{I}_{sq1,2} & S^{I}_{sq1,3} & \cdots & S^{I}_{sq1,m} \\ S^{I}_{sq2,1} & S^{I}_{sq2,2} & S^{I}_{sq2,3} & & S^{I}_{sq2,m} \end{bmatrix}^{T}, \tag{2.4}$$

where $\mathbf{S}_{sq}$ is the tensor form of the received sequence with dimensions of $m \times 2 \times 2$ (it can be considered as a 3D-matrix, where $m$ denotes the half size of the received signal), which contains the in-phase (I) and quadrature (Q) components. Normally, we treat the depth of tensor as the concept of channel in deep learning field, where the I and Q components are two different channels of the tensor $\mathbf{S}_{sq}$. In each row, we place half of the received information according to corresponding subcarrier for pilots and transmitted signals. In the following deduction, the superscript $l$ denotes the $l^{th}$ channels.

We use the two-dimension convolutional layer to extract the feature from $\mathbf{S}_{sq}$. The weights for the convolutional kernel $\mathbf{C}^{l}_{f(i,l)}$ can be expressed as

$$\mathbf{C}^{l}_{f(i,l)} \in \mathbb{R}^{2\times 2\times l} = \begin{bmatrix} w^{l}_{(1,1)} & w^{l}_{(1,2)} \\ w^{l}_{(2,1)} & w^{l}_{(2,2)} \end{bmatrix}^{T}. \tag{2.5}$$

The convolution operation [54] is processed for each kernel and different channel of the input tensor, which can be formulated as

$$S^{l}_{con\ i,j} = \sum_{m=0}^{1}\sum_{n=0}^{1} w^{l}_{2-m,2-n} S_{sq\ m+i,n+j} + b^{l}_{i,j}, \tag{2.6}$$

where $b^{l}_{i,j}$ denotes the bias in the convolution operation for training proposes. The output of the convoultion operation is $\mathbf{S}_{con} = S^{l}_{con\ i,j}$. We use the batch normalization layer [55] to normalize the $g^{th}$ $\mathbf{S}_{con}$ with the batch size $G$, which can be expressed as

$$\mu_B = \frac{1}{IJLG}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{g=1}^{G}(S^{l}_{con\ i,j})_g, \tag{2.7}$$

$$\sigma_B^2 = \frac{1}{IJLG}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{g=1}^{G}((S^l_{con\ i,j})_g - \mu_B)^2, \tag{2.8}$$

$$\widehat{S}^l_{con\ i,j} = \frac{S^l_{con\ i,j} - \mu_B}{\sqrt{{\sigma_B}^2 + \varepsilon}}, \tag{2.9}$$

$$S^l_{B\ i,j} = \gamma \widehat{S}^l_{con\ i,j} + \beta, \tag{2.10}$$

where $\mu_B$ and $\sigma_B^2$ denote the mean and variance of $S^l_{con\ i,j}$, respectively. We denote $I, J$, and $L$ as the sizes of the tensors to calculate the mean and variance. The learnable parameters $\gamma$ and $\beta$ will be updated in the training process. $\varepsilon$ is a small constant for numerical stability or to avoid division by zero). Note that the learnable parameters indicate the whole weights in convolutional layers, $\gamma$ and $\beta$ in batch normalization layer, as well as the weights in the LSTM layer and the fully-connected layer.

The Relu function is considered as an active function, which can be formulated as

$$f(x) = \begin{cases} 0\ for\ x < 0, \\ x\ for\ x \geq 0. \end{cases} \tag{2.11}$$

We use the Relu function to activate the output signal $S^l_{B\ i,j}$, which can be described as

$$S^l_{R\ i,j} = f(S^l_{B\ i,j}). \tag{2.12}$$

Next, we use flatten layer to reshape $S^l_{R\ i,j}$ to sequence for the LSTM layer, which can be described as

$$\mathbf{S}_F = [S_{F\ 1}\ S_{F\ 2} \cdots\ S_{F\ k}]^T, \tag{2.13}$$

where $\mathbf{S}_F$ denotes the output of flatten layer. The LSTM cell consists of input gate, output gate and forget gate. These three gates regulate the sequence into

and out of the cell for each time slot [56]. The fully-connected layer compresses the signal $S_{LSTM}$ from LSTM cell to $\mathbf{S}_{fc} \in \mathbb{R}^{1\times K}$ with the formulation of

$$\mathbf{S}_{fc} = \mathbf{S}_{LSTM}\mathbf{w}_{fc} + \mathbf{b}_{fc}, \tag{2.14}$$

where $\mathbf{w}_{fc} \in \mathbb{R}^{J\times K}$ and $\mathbf{b}_{fc} \in \mathbb{R}^{1\times K}$ denote the weights and bias for the fully connected layer, respectively.

Furthermore, we need to use the Softmax layer to calculate the possibility $S_i$ of each category which is given by

$$Softmax(S_i) = \frac{\exp(S_{fc\ i})}{\sum_{k=1}^{K}\exp(S_{fc\ k})}. \tag{2.15}$$

Finally, the cross-entropy loss denotes the cost function for the overall network together with the L2 regularization which is presented as

$$\mathcal{L}_{ce} = -\frac{1}{K}\sum_{i=1}^{K} T_i log(S_i) + \frac{1}{\lambda}\|w\|_2^2, \tag{2.16}$$

where $K$ represents the number of categories, and $T_i$ represents the possibility of the known target. The parameters $\lambda$ and $w$ are the regularization factor and the overall learnable weights, respectively. We use the well-known ADAM optimizer [57] to minimize the cross-entropy loss. Accordingly, the optimal weights, bias and other learnable factors can be obtained.

### 2.3.2 Training Procedure

We generate the bit sequence randomly following the binomial distribution. We choose quadrature phase-shift keying (QPSK) modulation for mapping the bits to symbols. In each simulation, we follow the system model to generate the NOMA scheme's transmission sequence before passing through the OFDM modulation. Then, we randomly transmit the signal over the TDL channel with AWGN noise

under different SNR conditions. Each received package of OFDM modulation, which contains the transmitted information and pilots, is treated as a cell for the joint detection network. Considering the maximum Doppler shift, the channel coefficient for each OFDM package's length can be approximately treated as a constant [5]. The ground truth is the same as the QPSK symbols. During the offline training stage, each subcarrier is trained separately under high SNR condition. In the testing stage, the channel state information is not required for the trained network.

TABLE 2.1: Channel Model Parameters

| **Parameters** | **Value** |
|---|---|
| Carrier Frequency | 4 GHz |
| Sampling Rate | 32.75 MHz |
| Maximum Doppler Shift | 148.18Hz |
| Mobile Speed | 40 km/h |
| Delay Profile | TDL - E |
| Number of Paths | 14 |
| Length of Cyclic Prefix | 16 |
| Number of Subcarriers | 64 |
| Number of Pilots | 64,16 |
| Block Length | 128 |
| Power Allocation Factor | UE1: 0.8; UE2: 0.2 |

## 2.4 Simulation Results and Discussion

We generate the TDL channel coefficients following the distribution property. The channel state information and channel response can be generated according to the parameters in Table 2.1. To visualize the channel model, the magnitude of different path for each frame is plotted in Fig. 2.2. The dataset contains 80,000 training samples, 10,000 validation samples, and 10,000 testing samples. Each frame of the OFDM package is randomly transmitted through the channel coefficient. The TDL channel response can be generated by the 5G Toolbox in MATLAB 2019a.

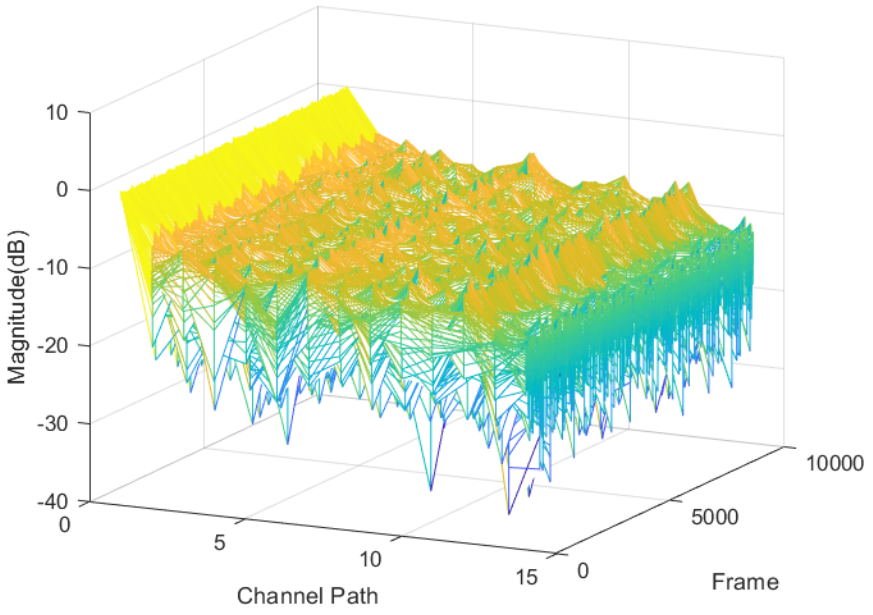



FIGURE 2.2: TR 38.901 TDL-E channel model

The training dataset and testing dataset are generated following the system model presented in Section 2.2.1. The parameters to generate the TDL channel response are shown in Table 2.1. During offline training, the SNR is set to 50 dB. We have conducted several experiments to illustrate the performance and robustness of the proposed method with different number of pilots. The LS-SIC and MMSE-SIC methods are considered as the baseline for comparisons. To be consistent with [8], we set the number of users to be two for the subsequent numerical results. Note that our proposed method can be extended to more than two users based on similar system setting. From [5] and [8], it has been concluded that the MMSE-SIC results are not optimal for the distorted practical scenarios as the ICI and ISI cannot be removed completely. Furthermore, the SIC considered in this thesis is not perfect, which imitates the practical situation. Thus, the maximum-likelihood (ML) joint detection with perfect channel state information condition is considered as the lower-bound in the ideal case.

### 2.4.1 simulation results for different receivers

Simulation results of various schemes are presented in Fig. 2.3, where each transmitted package has 64 pilots, and the length of CP is 16. The curves represent the performance of the proposed joint detection network, LS-SIC, MMSE-SIC, and DNN methods for User1 and User2, respectively. Also, the maximum-likelihood method with perfect CSI conditions is considered the ideal case. The SER versus SNR curves show that for User1 with a higher power factor, the performance is quite similar for both the traditional methods and the proposed joint detection network. Compared with the MMSE-SIC method, our proposed DL method has 1.1 dB gain to achieve the same SER of $10^{-3}$. Moreover, in the range of high SNR, our proposed method outperforms the existing DNN method [5].

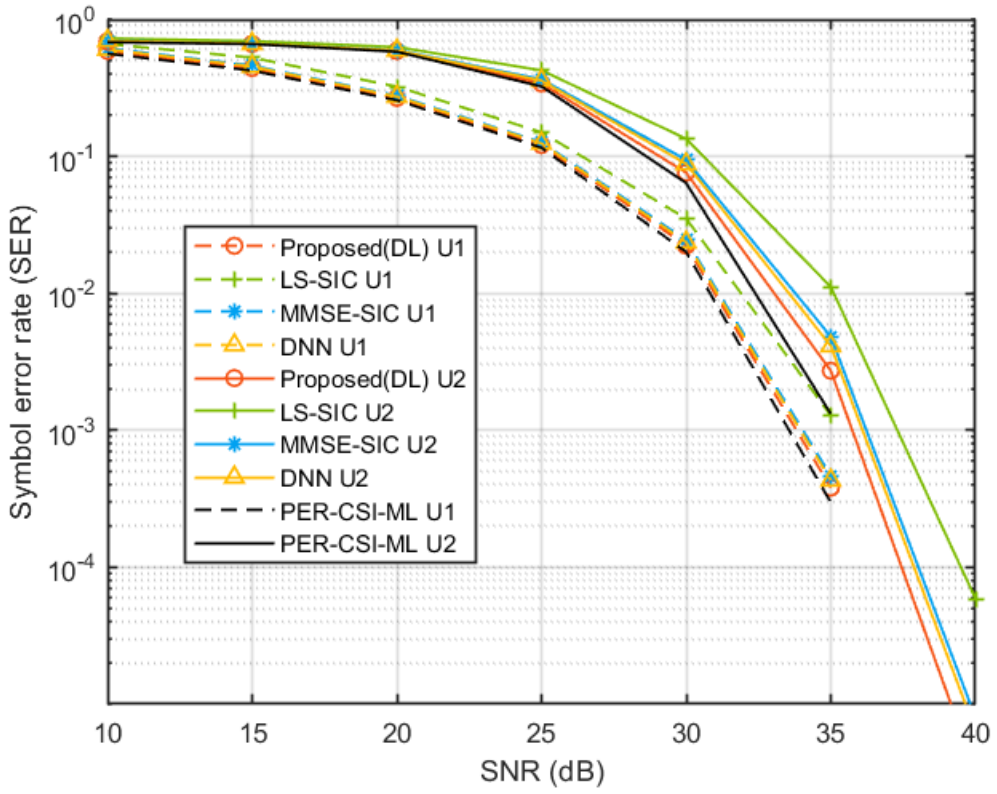



FIGURE 2.3: Simulation results with 64 pilots and 16 cyclic prefixes.

### 2.4.2 Simulation results for three users

The simulation results for three users are shown in Fig. 2.4. With 16 pilots, the SER performance of MMSE-SIC and LS-SIC methods is unsatisfactory for weak users, i.e., UE2 and UE3. The channel estimations for LS and MMSE methods are not accurate because of the insufficient pilots. Moreover, imperfect SIC exists in the practical wireless communication situation, which also degrades the performance of UE2 and UE3. Compared with MMSE-SIC and LS-SIC methods, it has been shown that our proposed DL-based receiver can outperform the conventional methods which is consistent with the conclusion for the two-user case.

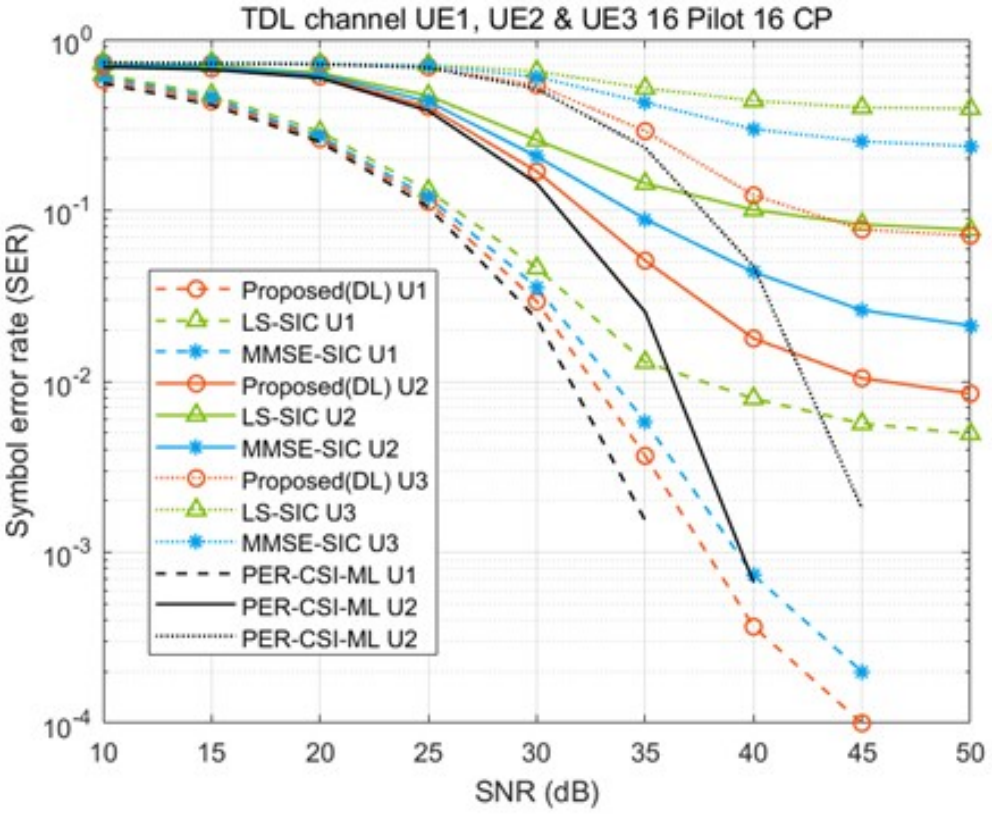



FIGURE 2.4: Simulation results for three-user NOMA system with 16 pilots, CP length 16, and 64 subcarriers.

The proposed DL-based receiver can handle inter-carrier interference and inter-symbol interference by CNN-based feature extractor and LSTM time-series layers. With the convolution operation between the input sequence and different kernels, the information about the interconnection between each subcarrier can be

extracted. The optimal weights for each kernel are learned in the training process to minimize the cross-entropy loss function, which is equivalent to maximize the posterior probability of the recovered symbol.

### 2.4.3 Impact of Pilot Size

In Fig. 2.5, the SER results of various schemes with 16 pilots are presented. As compared with Fig. 2.3, we observe that in general, the overall performance of various schemes degrades as the number of pilots decreases from 64 to 16. In Fig. 2.5, the performance of User1 degrades slightly, however, the performance of User2 degrades more severely as the channel estimation error increases. We also observe that our proposed DL-based method has around 4 dB and 2.4 dB gain in SER performance for the second user's detection as compared with MMSE-SIC and DNN [5] to achieve the same SER of $10^{-3}$, respectively. These simulation results for different number of pilots show the robustness of our proposed deep-learning-based method. Moreover, we observe that our proposed DL-based method significantly reduces the error floor under the condition of fewer pilots.

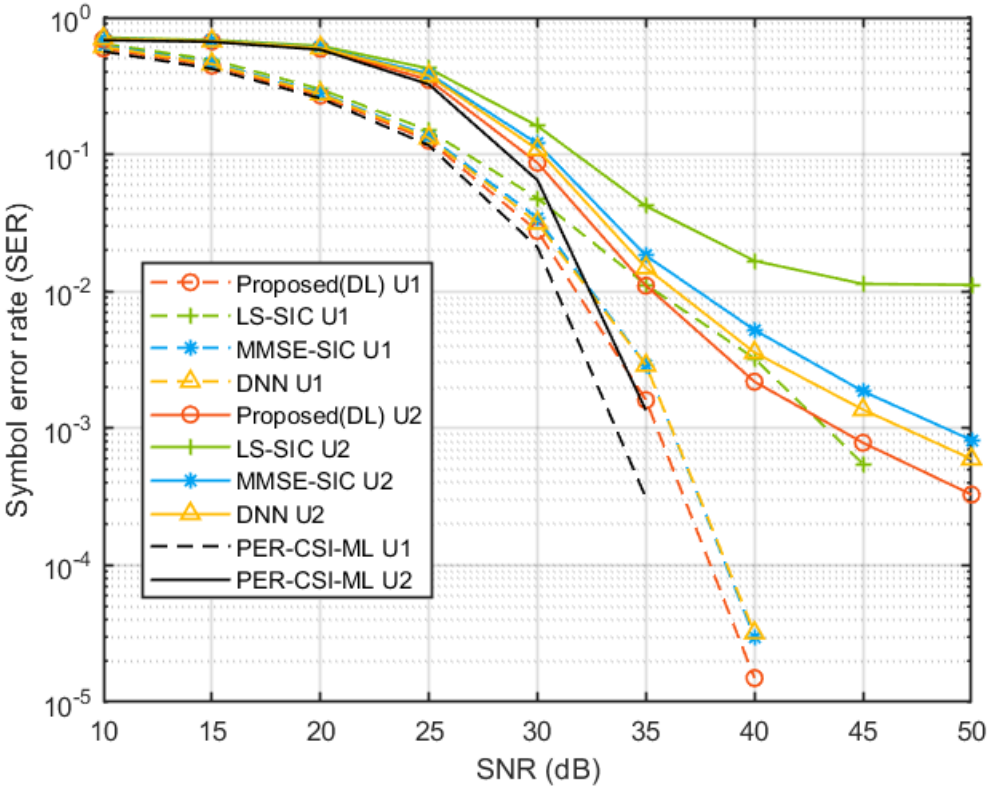



FIGURE 2.5: Simulation results with 16 pilots and 16 cyclic prefixes.

### 2.4.4 Complexity Analysis

In this section, the complexity analysis is presented to compare the complexities of the proposed DL-based method and traditional methods. The number of floating-point operations (FLOPs) is considered as the measurement of complexity. In general, the complexity of the joint detection network can be written as $C_{Net} \sim O(LW_L(4h_s^2 + h_s i_s + h_s) + L(i_d h_d + h_d))$, where $L$ and $W_L$ represent the length of received sequence and length of each OFDM block, respectively. The symbols $i_s$ and $h_s$ denote the input and hidden size of LSTM layer, respectively. Moreover, we denote $i_d$ and $h_d$ as the input and the hidden layer's neuron size, respectively. Also, we denote $n_l$ and $n_c$ as the number of the LSTM layers and hidden layers. By ignoring the first term which is not significant, the equation can be approximately simplified as $C_{Net} \sim O(L(4W_L h_s^2 n_l + i_d h_d n_c))$. As for the traditional method, we denote $M$ as the modulation order, and the complexity of the LS-SIC method is $O(L(2M+1))$. The complexity of the MMSE-SIC method is larger than $O(L^{2.37}(2M+1))$ due to the matrix inversion operation [58]. On the other hand, the ML method requires the complexity of $O(L^{W_L/2}M)$, which is much larger than all other methods. Since the length of signal $L$ is much larger than the length of block $W_L$, generally, the complexity of our proposed deep learning based method is less than the MMSE-SIC method but more than the LS-SIC method.

In summary, the computational complexities of different conventional and DL-based receivers are compared in Table 2.2.

Table 2.2: Computational Complexity Comparison

| **Algorithm** | **Computational Complexity** |
|---|---|
| LS-SIC | $\mathcal{O} \sim (L(2M+1))$ |
| MMSE-SIC | $\mathcal{O} \sim (L^{2.37}(2M+1)))$ |
| Proposed DL-based joint receiver | $\mathcal{O} \sim (L(4W_L h_s^2 n_l + i_d h_d n_c))$ |
| DNN Receiver [5] | $\mathcal{O} \sim (L\sum_{d=1}^{D} J_d I_d)$ |

[a] $J_d$ and $I_d$ denote the input and output dimension of the $d^{th}$ fully connected layer of DNN receiver, respectively.

## 2.5 Conclusion

In this chapter, a deep learning based joint detection receiver for the orthogonal frequency-division multiplexing non-orthogonal multiple-access scheme is designed under a doubly-selective tapped-delay line channel for 5G scenarios. It has been shown that under the condition of insufficient pilots, the proposed DL-based algorithm significantly reduces the error floor, and it can achieve around 4 dB gain in SER performance compared with the conventional MMSE-SIC method. Also, our proposed DL-based receiver is an end-to-end model, which integrates the function of channel estimation, equalization, and demodulation. The proposed DL-based algorithm has less computational complexity as compared with the MMSE-SIC method as the matrix inversion operation is required for the MMSE-SIC method. Moreover, as compared with the existing methods, the proposed DL-based method is more robust against reducing the number of pilots. In the next chapter, we delve into another essential aspect of deep learning based wireless communication systems: channel state information feedback for massive MIMO communication. We propose a novel approach based on autoencoder-decoder (AED) techniques to optimize the process of transmitting and receiving CSI information. By leveraging deep learning for CSI feedback, we aim to further enhance the efficiency and accuracy of massive MIMO communication systems, contributing to the advancement of wireless communication technologies.

# Chapter 3

# Autoencoder-decoder based CSI Feedback for Massive MIMO Communication

## 3.1 Introduction

In this chapter, the focus is on studying deep learning -based channel state Information feedback for frequency division duplexing (FDD) massive MIMO communication systems. FDD systems are known for their ability to provide cognitive compressing for both the base station and user side, making them an important and widely used technology in practical communication systems.

CSI feedback plays a crucial role in FDD massive MIMO systems as it enables the base station to have accurate knowledge of the channel conditions between the base station and user devices. This information is essential for optimizing the transmission parameters and improving system performance.

To address the challenges associated with CSI feedback in FDD massive MIMO systems, DL-based techniques are studied. DL algorithms are employed to learn

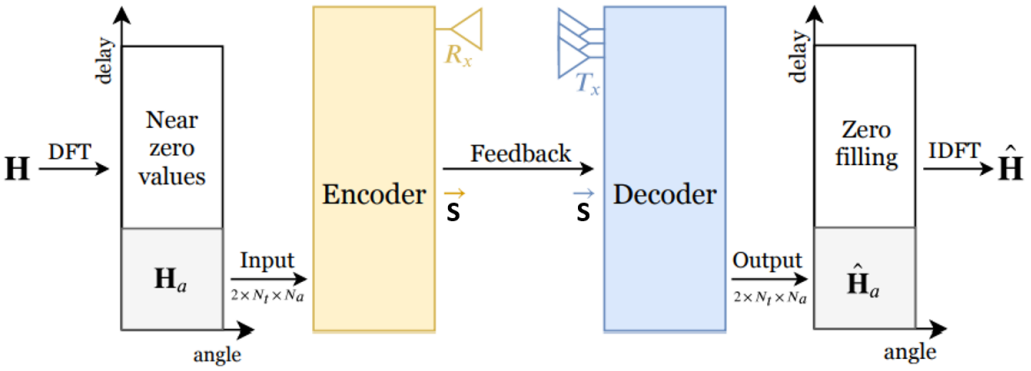



FIGURE 3.1: CSI feedback block diagram.

and predict the CSI feedback more efficiently and accurately. By utilizing DL models, the aim is to enhance the performance of CSI feedback mechanisms, resulting in improved system capacity, throughput, and overall communication quality.

The research conducted in this chapter contributes to the advancement of DL-based solutions for CSI feedback in FDD massive MIMO communication systems. By leveraging DL techniques, the chapter aims to optimize the cognitive compressing capabilities of FDD systems and demonstrate the wide-ranging practical applications of DL-based CSI feedback in real-world communication systems.

## 3.2 System model

A frequency-division duplex (FDD) downlink massive MIMO-orthogonal frequency division multiplexing (OFDM) system with $N_c$ subcarriers is considered. A general deep learning based CSI feedback system is shown in Fig. 3.1.

The BS deploys $N_t$ transmit antennas as uniform linear array (ULA). In a time-varying channel caused by UE mobility, the received signal at time $t$ on the $n^{th}$ subcarrier for UE with a single receiver antenna can be modeled as

$$y_{n,t} = \mathbf{h}_{n,t}^T \mathbf{v}_{n,t} x_{n,t} + z_{n,t}, \tag{3.1}$$

where $\mathbf{h}_{n,t} \in \mathcal{L}^{N_t \times 1}$, $x_{n,t} \in \mathcal{L}$, and$z_{n,t} \in \mathcal{L}$ denote time-varying channel vector in the frequency domain, transmitted sequence, and additive white Gaussian noise (AWGN), respectively. Then, we denote the channel state information (CSI) matrix at time $t$ in the spatial-frequency domain as $\mathbf{H_t} = [h_{1,t}, ..., h_{N_a,t}] \in \mathcal{L}^{N_a \times N_t}$. Note that $\mathbf{v}_{n,t}$ is the beamforming vector. In a practical FDD system, the UE should return $\mathbf{h}_{n,t}$ to the base station (BS) through feedback links. However, the total number of feedback information is $N_a \times N_t$, which is too large to transmit in a practical massive MIMO system. In this case, this work studies the deep learning based method to reduce the feedback overhead. First, we use 2D discrete Fourier transform (DFT) to transform the $\mathbf{H}_t$ into an approximately sparsified matrix $\widetilde{\mathbf{H}_t}$ in angular-delay domain, i.e.,

$$\widetilde{\mathbf{H}_a} = \mathbf{F_d} \mathbf{H}_t \mathbf{F_a}, \tag{3.2}$$

where $F_d \in \mathcal{L}^{N_a \times N_a}$ and $F_a \in \mathcal{L}^{N_t \times N_t}$ are the DFT matrices. The CSI dataset is provided by National Artificial Intelligence Competition (NAIC) as an open-source dataset. Generally, $\widetilde{\mathbf{H}_a}$ is only approximately sparse for finite $N_t$, which may not be applicable to conventional CS methods. Therefore, we will propose a DL-based feedback architecture without sparsity prior constraint. We perform sparsity transformation to decrease parameter overhead and training complexity.

## 3.3 Problem Illustration

The CSI feedback approach can be summarized as follows. Once the channel matrix $\mathbf{H}_t$ is acquired at the UE side, we perform 2D-DFT to obtain the truncated matrix $\widetilde{\mathbf{H}_a}$ and then use the encoder to generate a compressed-CSI $\mathbf{S}$. Next, $\mathbf{S}$ is returned to the BS, and the BS uses the decoder to obtain $\hat{\mathbf{H}}_a$. The final channel matrix in the spatial-frequency domain can be obtained by performing inverse DFT.

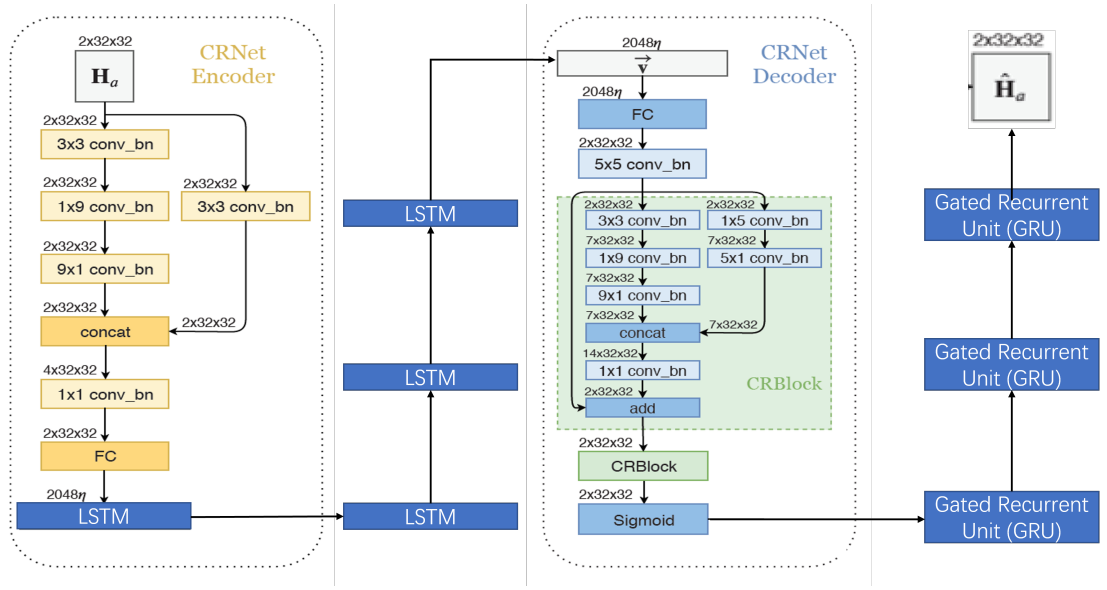



FIGURE 3.2: Autoencoder-decoder based CSI feedback structure.

## 3.4 Autoencoder-decoder based channel state information feedback process

In this section, we present a self-supervised autoencoder-decoder (AED). The compressing process and decompressing process are described in detail. The proposed AED-based network structure is shown in Fig. 3.2. We exploit the conventional neural networks (CNNs) for the encoder and decoder. They can exploit spatial local correlation by enforcing a local connectivity pattern among the neurons of adjacent layers. The overview of the proposed DL architecture consists of two major parts, one is the autoencoder compressing process at the user side, and the other is the decoder process at the base station.

### 3.4.1 Autoencoder compressing process

The encoder neural network is designed as shown in Fig. 3.3, which consists of multiple convolutional layers. If the CSI matrix is relatively dense, its feature granularity will be finer. Convolution with smaller kernel sizes can extract finer features better. Conversely, the larger kernel is preferred for a sparser CSI matrix.

Single-resolution networks can not adapt to the feature granularity change well, which may lead to the loss in network performance and robustness. Since the channel response in NAIC dataset [59] is relatively dense, we choose the small convolutional kernel to extract the feature. The autoencoder can not process the complex value, thus we separate the real part and the imaginary part. The total number of feedback information to the autoencoder is $N = 2 \times N_a \times N_t$.

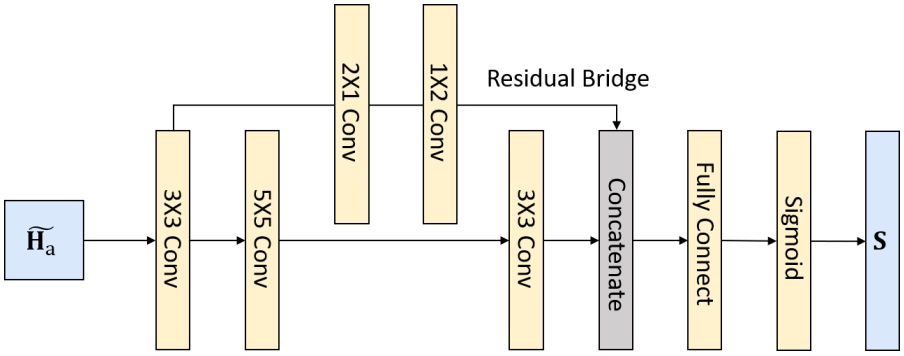



FIGURE 3.3: Encoder block diagram

The input to autoencoder is $\widetilde{\mathbf{H}}_a$ which is processed by 2D-DFT. First, convolutional layer with small kernel extracts the low-level features, which can be formulated as

$$S^l_{con\ (m,j)} = \sum_{i=0}^{I-1} w^l_{I-i} H_{a\ (m,j+i)} + b^l{}_{(m,j)}, \tag{3.3}$$

where $S^l_{con\ (m,j)}$ denotes the output of the convolutional layer, and $l$ denotes the number of kernels in each convolutional layer, as well as we define the number of kernel as channel number inside the autoencoder. Note that $w$ and $b$ are the weights for each kernel and bias, respectively. To maintain the feature size, zero padding is used at both ends. The residual bridge provides an extra path for feature extraction. The residual path has two convolutional layers with different kernel sizes. As the density of the channel matrix is uncertain, different size of kernel can extract different levels of features. This residual structure can enhance the integrity of extracted features, and concatenate different levels of features together. The Leaky-Relu [60] is used as an activation function in each convolutional layer, which is formulated as

$$L - Relu(x) = \begin{cases} S^l_{con\ (m,j)}, & S^l_{con\ (m,j)} > 0, \\ \beta S^l_{con\ (m,j)}, & S^l_{con\ (m,j)} \leq 0, \end{cases} \tag{3.4}$$

where $\beta$ is the coefficient for the negative slope, and we set it to 0.2 in our experiment follow the same setting as [23]. The fully-connected layer compresses the signal from the convolutional layer. Next, the *sigmoid* function centralized the signal in $(0, 1)$

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{3.5}$$

### 3.4.2 Autodecoder decompressing process

The feedback CSI $\mathbf{S}$ is decoded at the base station by the autodecoder, which is designed as shown in Fig. 3.4. The structure of the autodecoder is designed as the inverse of the autoencoder. The residual bridge is also considered in the autodecoder.

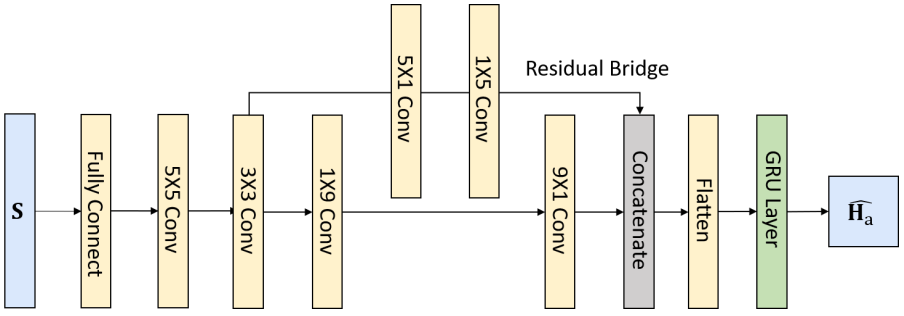



FIGURE 3.4: Decoder block diagram

To recover the compressed CSI, we choose gate recurrent unit (GRU) layer as the final output layer to get $\hat{\mathbf{H}}_a$. The GRU layer has the advantage to predict the sequence information in the regression problem, which can be formulated as

$$z_n = \sigma_g(W_z x_n + U_z p_{n-1} + b_z), \tag{3.6}$$

$$r_n = \sigma_g(W_r x_n + U_r p_{n-1} + b_r), \tag{3.7}$$

$$\hat{p_n} = \phi_p(W_h x_n + U_p(r_n \bigodot p_{n-1}) + b_p), \tag{3.8}$$

$$p_n = (1 - z_n) \bigodot (p_{n-1} + z_n \bigodot \hat{p_n}), \tag{3.9}$$

where $x_n$ and $p_n$ are the input information and output information, respectively. Note that $z_n$ and $r_n$ denote the update gate information and reset get information, which controls the candidate activation information $\hat{p_n}$. Furthermore, $W$, $U$, and $b$ are learnable parameters optimized in the training process by the ADAM optimizer. Finally, normalized mean-squared error (NMSE) is used as the loss function for each batch with size $M$, which is defined as

$$\mathcal{L}_{loss} = \frac{1}{M} \sum_{m=0}^{M} \left\| \hat{\mathbf{H}}_a - \mathbf{H}_a \right\|_2^2. \tag{3.10}$$

We also define the score $\rho$ to measure the accuracy, which can be formulated as

$$\rho = 1 - \mathcal{L}_{loss}. \tag{3.11}$$

The hyperparameters for the CSI encoder and decoder are shown in Fig. 3.5 and 3.6. We use the PyTorch library as the deep learning framework for network structure, and the code implementation is also shown in Fig. 3.5 and 3.6. We choose CRBlock as the backbone for both the CSI encoder and decoder, which is implemented as Fig. 3.7.

```
class Encoder(nn.Module):
    B = 2

    def __init__(self, feedback_bits, quantization=True):
        super(Encoder, self).__init__()
        self.encoder1 = nn.Sequential(OrderedDict([
            ("conv3x3_bn", ConvBN(2, 64, 3)),
            ("relu1", nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ("conv1x9_bn", ConvBN(64, 64, [5, 5])),
            ("relu2", nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ("conv9x1_bn", ConvBN(64, 64, [3, 3])),
        ]))
        self.encoder2 = ConvBN(2, 64, 3)
        self.encoder_conv = nn.Sequential(OrderedDict([
            ("relu1", nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ("conv1x1_bn", ConvBN(64 * 2, 2, 1)),
            ("relu2", nn.LeakyReLU(negative_slope=0.3, inplace=True)),
        ]))

        self.fc = nn.Linear(256, int(feedback_bits / self.B))
        self.LSTM = nn.LSTMCell(2048, 2048, 3)
        self.fc1 = nn.Linear(2048, 512)
        self.fc2 = nn.Linear(512, 256)
        self.sig = nn.Sigmoid()
        self.quantize = QuantizationLayer(self.B)
        self.quantization = quantization

        for m in self.modules():
            if isinstance(m, (nn.Conv2d, nn.Linear)):
                nn.init.xavier_uniform_(m.weight)
            elif isinstance(m, nn.BatchNorm2d):
                nn.init.constant_(m.weight, 1)
                nn.init.constant_(m.bias, 0)
```

FIGURE 3.5: Network hyperparameters for the CSI Encoder

```
class Decoder(nn.Module):
    B = 2

    def __init__(self, feedback_bits, quantization=True):
        super(Decoder, self).__init__()
        self.feedback_bits = feedback_bits
        self.dequantize = DequantizationLayer(self.B)
        self.fc = nn.Linear(int(feedback_bits / self.B), 256)
        self.fc2 = nn.Linear(256, 512)
        self.fc3 = nn.Linear(512, 2048)
        decoder = OrderedDict([
            ("conv5x5_bn", ConvBN(2, 64, 5)),
            ("relu", nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ("CRBlock1", CRBlock()),
            ("CRBlock2", CRBlock()),
        ])
        self.GRU = nn.GRU(2048, 2048, 4, batch_first=True)
        self.decoder_feature = nn.Sequential(decoder)
        self.out_cov = conv3x3(64, 2)
        self.sig = nn.Sigmoid()
        self.quantization = quantization

        for m in self.modules():
            if isinstance(m, (nn.Conv2d, nn.Linear)):
                nn.init.xavier_uniform_(m.weight)
            elif isinstance(m, nn.BatchNorm2d):
                nn.init.constant_(m.weight, 1)

                nn.init.constant_(m.bias, 0)
```

FIGURE 3.6: Network hyperparameters for the CSI Decoder

## 3.5 Experiments Results

### 3.5.1 Ablation Experiment

In this section, we conduct ablation experiments to comprehensively analyze the impact of various hyperparameters and components of the proposed AED model on the performance of CSI feedback in the practical NAIC 5G communication dataset.

```
class CRBlock(nn.Module):
    def __init__(self):
        super(CRBlock, self).__init__()
        self.path1 = nn.Sequential(OrderedDict([
            ('conv3x3', ConvBN(64, 64, 3)),
            ('relu1', nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ('conv1x9', ConvBN(64, 64, [1, 9])),
            ('relu2', nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ('conv9x1', ConvBN(64, 64, [9, 1])),
        ]))
        self.path2 = nn.Sequential(OrderedDict([
            ('conv1x5', ConvBN(64, 64, [1, 5])),
            ('relu', nn.LeakyReLU(negative_slope=0.3, inplace=True)),
            ('conv5x1', ConvBN(64, 64, [5, 1])),
        ]))
        self.conv1x1 = ConvBN(64 * 2, 64, 1)
        self.identity = nn.Identity()
        self.relu = nn.LeakyReLU(negative_slope=0.3, inplace=True)

    def forward(self, x):
        identity = self.identity(x)

        out1 = self.path1(x)
        out2 = self.path2(x)
        out = torch.cat((out1, out2), dim=1)
        out = self.relu(out)
        out = self.conv1x1(out)

        out = out + identity
        return out
```

FIGURE 3.7: Hyperparameters for the CRBlock

The AED model is designed to optimize the compression and reconstruction of CSI, which is crucial for efficient feedback in massive MIMO systems. We study the three main components of the proposed AED, which are the number of kernels, kernel size, and number of FC layers. The ablation experiment results are shown in Table. 3.1.

The convolution-Batch Normalization (ConvBN) layer plays a crucial role in feature

extraction from the CSI data. We evaluate the impact of varying the number of kernels in this layer on the AED model's performance. The results indicate that increasing the number of kernels improves the AED model's performance in terms of NMSE and score. A larger number of kernels allows the model to capture more complex CSI patterns, leading to better compression and reconstruction. However, too many kernels may lead to overfitting, especially when the training data is limited. On the other hand, too few kernels may limit the model's capacity to represent intricate features accurately.

The kernel size in the convolutional layers directly influences the receptive field and feature extraction capabilities of the AED model. We investigate the effect of varying kernel sizes on the model's performance. Table. 3.1 present the outcomes for different kernel dimensions. The results show that a kernel size of 5x5 achieves the best performance in terms of NMSE and score. A 5x5 kernel has a larger receptive field, allowing the model to capture spatial correlations and extract more meaningful features from the CSI data. Smaller kernel sizes, such as 1x1, may result in insufficient feature extraction, while larger ones, such as 7x7, may lead to increased computational complexity and reduced performance.

We investigate the impact of varying the number of FC layers on the AED model's performance. The fully connected (FC) layers in the decoder contribute to reconstructing the compressed CSI data. The analysis reveals that an AED model with three FC layers achieves the best overall performance. Adding more layers may not lead to substantial improvements, as it can introduce additional computational overhead without significant gains. On the other hand, reducing the number of FC layers could compromise the model's capacity to learn complex CSI representations, impacting performance.

The ablation experiments yield valuable insights into the AED model's hyperparameters, shedding light on their pivotal role in CSI feedback performance. A well-balanced configuration with 64 kernels in the ConvBN layer, a 5x5 kernel size,

and three FC layers emerge as the optimal choice for efficient and high-performance CSI feedback in 5G communication systems.

TABLE 3.1: Ablation experiments for the proposed AED

| **Hyperparameters** | **NMSE** | **Score $\rho$** |
|---|---|---|
| Number of Kernels in ConvBN layer | | |
| 64 | -39.9540 (dB) | 0.9816 |
| 32 | -18.3572 (dB) | 0.8405 |
| 16 | -15.2763 (dB) | 0.7829 |
| 128 | -31.2963 (dB) | 0.9572 |
| Kernel Size | | |
| 1x1 | -15.9061 (dB) | 0.7962 |
| 3x3 | -36.3514 (dB) | 0.9736 |
| 5x5 | -39.9540 (dB) | 0.9816 |
| 7x7 | -18.0970 (dB) | 0.8362 |
| Number of FC layers | | |
| 1 | -22.2688 (dB) | 0.8921 |
| 2 | -27.1055 (dB) | 0.9335 |
| 3 | -39.9540 (dB) | 0.9816 |
| 4 | -34.1530 (dB) | 0.9671 |

### 3.5.2 NAIC Practical 5G CSI dataset Results

In this section, we conducted a comprehensive analysis of the proposed AED for data compression in wireless communication systems. In Table 3.2, we present a comprehensive comparison of different algorithms for channel state information (CSI) estimation in practical 5G MIMO communication scenarios with non-standard noise. The evaluation metrics used for comparison are the Normalized Mean Squared Error (NMSE) and the Score ($\rho$), which measure the accuracy of the estimated CSI and the overall performance of the algorithms, respectively. The comparison between CsiNet, CsiNet-LSTM, AnciNet, CRNet from references [23],

[61], [62], [17] and our proposed AED network are shown in Table 3.2. Each algorithm's performance is assessed using the open-source dataset (which is available on https://naic.pcl.ac.cn/frame/1.) collected from practical 5G MIMO communication scenarios with a compression ratio of $\gamma = 1/16$. The dataset consists of 320,000 total samples with dimensions of $32 \times 16 \times 2$.

The results clearly demonstrate the superior performance of our proposed AED network. It achieves a significantly lower NMSE of -39.9540 dB, indicating highly accurate and reliable CSI estimation. Moreover, the AED network achieves an impressive Score of 0.9816, indicating excellent overall performance in capturing semantic information. Comparatively, other state-of-the-art algorithms, such as CSiNet, CsiNet-LSTM, AnciNet, and CRNet, also exhibit respectable performance with decreasing NMSE and increasing Score values. However, our proposed AED network outperforms all of them, showing the potential of the DL-based AED approach in achieving superior performance for CSI estimation in 5G MIMO communication scenarios. These results showcase the effectiveness of the AED network for data compression and CSI estimation in practical 5G communication systems with non-standard noise. The proposed AED network demonstrates its ability to achieve higher accuracy and better semantic information preservation compared to other existing algorithms.

The availability of an open-source dataset and the use of standardized evaluation metrics ensure a fair and rigorous comparison of the algorithms. The promising results of the proposed AED network suggest its potential application in real-world 5G communication systems, where accurate and efficient CSI estimation is essential for optimizing communication performance.

### 3.5.3 Performance evaluation for different compression rates

Moreover, we evaluate the performance based on the NMSE and Score metrics under different compression ratios ($\gamma$). Table. 3.3 presents the NMSE and Score values for four different compression ratios: 1/32, 1/16, 1/8, and 1/4. As expected, we

TABLE 3.2: Comparisons of NMSE Loss and Score

| **Algorithm** | **NMSE** | **Score $\rho$** |
|---|---|---|
| DNN-AE | -9.2963 (dB) | 0.6053 |
| CSiNet [23] | -14.5329 (dB) | 0.7662 |
| CsiNet-LSTM [61] | -16.8470 (dB) | 0.8145 |
| AnciNet [62] | -24.2250 (dB) | 0.9113 |
| CRNet [17] | -26.9563 (dB) | 0.9325 |
| Propused AED | -39.9540 (dB) | 0.9816 |

observe that as the compression ratio increases (i.e., the level of data compression decreases), the NMSE decreases, indicating improved data reconstruction accuracy. Additionally, the Score metric increases, suggesting enhanced overall performance in capturing semantic information. These results indicate that the proposed AED performs well under higher compression ratios, achieving higher data reconstruction accuracy and better preservation of semantic information. This is a crucial finding, as it highlights the Algorithm's effectiveness in achieving efficient data compression without significant loss of information.

Moreover, comparing the NMSE and Score values across different compression ratios provides insights into the trade-off between data compression and reconstruction accuracy. While higher compression ratios lead to greater data compression, they can potentially result in higher reconstruction errors and reduced performance in capturing semantic information. Therefore, careful consideration should be given to selecting an appropriate compression ratio based on the specific requirements of the communication system.

The results of this analysis demonstrate the potential of the proposed Algorithm for data compression in wireless communication systems. By achieving high data reconstruction accuracy and preserving semantic information, the Algorithm shows promise in enhancing the efficiency and performance of communication systems

with reduced data transmission requirements. It is essential to note that the performance of the proposed Algorithm should be further validated and compared against existing state-of-the-art methods or benchmarks. Additionally, the analysis should consider a broader range of compression ratios and evaluate the Algorithm's performance under various communication scenarios to ensure its robustness and applicability in real-world settings.

TABLE 3.3: Comparisons of NMSE Loss and Score under different compression rate

| **Algorithm** | **NMSE** | **Score $\rho$** | **Compression Ratio $\gamma$** |
|---|---|---|---|
| Propused AED | -21.27 (dB) | 0.8808 | 1/32 |
| Propused AED | -39.9540 (dB) | 0.9816 | 1/16 |
| Propused AED | -43.570 (dB) | 0.9871 | 1/8 |
| Propused AED | -48.124 (dB) | 0.99113 | 1/4 |

## 3.6 Conclusion

In conclusion, our study on deep learning-based network structures for channel state information feedback in 5G communication systems reveals that networks with a combination of convolutional feature extractors and recurrent layers. Our proposed AED outperforms the state-of-the-art deep learning based CSI feedback algorithms e.g., DNN, CRNet, CsiNet-LSTM, and CsiNet.

Furthermore, we compared our proposed AED network with a baseline DNN-AE that consists of fully connected layers. Through careful modification and fine-tuning, our AED network demonstrates superior performance with a higher score and lower normalized mean square error on the 5G NAIC CSI dataset.

Moreover, we explored the impact of kernel size design on the network's performance. By using an appropriate kernel size, we can effectively extract more features that carry valuable information, especially in cases where the CSI matrix exhibits sparsity.

Overall, the findings from this chapter shed light on the potential of deep learning-based network structures for optimizing CSI feedback in 5G communication systems, offering promising avenues for future research and advancements in wireless communication technologies.

In the next chapter, we will delve into the exploration of deep learning-based algorithms to address the challenges posed by non-linear distortion in the OFDM system. Non-linear distortion is a common issue in OFDM systems and can lead to degradation in performance, particularly in high-power scenarios. The next chapter will present our novel deep learning-based algorithm, providing insights into the architecture, training process, and evaluation metrics used for performance assessment. We will also conduct extensive experiments on practical 5G OFDM communication scenarios with varying levels of non-linear distortion.

# Chapter 4

# 1D-TRNet Based Receiver for the OFDM System with Non-linear Distortion

## 4.1 Introduction

In this chapter, the focus is on studying the application of deep learning techniques for equalization in the presence of non-linear distortion under OFDM systems. The presence of non-linear distortions can significantly degrade the performance of OFDM systems, posing a challenge that needs to be addressed.

The objective of this chapter is to explore deep learning-based equalization techniques that can mitigate the effects of non-linear distortion in OFDM systems. By leveraging the power of deep learning algorithms, it is possible to learn and model the non-linear distortions, allowing for more effective equalization and improved system performance.

The study of DL-based equalization for non-linear distortion in OFDM systems presents a significant challenge due to the complexity and variability of non-linear distortions. However, by developing DL-based models and algorithms specifically

tailored to address these challenges, it is possible to enhance the equalization process and mitigate the impact of non-linear distortions on system performance.

By focusing on this specific problem, the chapter contributes to the understanding and development of DL-based solutions for equalization under non-linear distortion in OFDM systems. It demonstrates the potential of DL techniques to overcome challenges and improve the performance of wireless communication systems in the presence of non-linear distortions.

Multi-carrier techniques, particularly OFDM systems, are widely utilized in wireless communications due to their higher spectrum efficiency and resistance to deep fading effects [24], [25]. However, they suffer from high PAPR, which can be mitigated using high power amplifiers [26], [27]. Clipping transmitted signals can cause BER degradation, but it also introduces frequency diversity that improves BER performance [28], [29]. The generalized approximate message passing (GAMP) algorithm has been proposed for recovering OFDM signals with non-linear distortion, showing better BER performance than without non-linear distortion [30]. In [31], the authors introduce a fast-GAMP algorithm to recover OFDM signals with non-linear distortion, which has less computational complexity than the GAMP algorithm.

Recently, deep learning has been applied to wireless communication systems. In [5], the authors only considered OFDM signals with slight non-linear distortion, and in [63], the authors considered the non-linear distortion caused by the terminal power amplifier. In [64], the authors propose a CNN-based equalizer for the OFDM system. Using deep learning to recover OFDM signals with severe non-linear distortion is still an open problem.

Inspired by the GAMP theoretical analysis in [32] as well as the research work presented in [8], we propose a deep neural network named one-dimensional transmit and recovery net (1D-TRNet) to exploit frequency diversity and combat ICI caused by severe non-linear distortion in OFDM signals. Our proposed 1D-TRNet consists of 1D U-shape feature extractor with multiple gated recurrent units (GRUs), which

can extract both low-level (local) and high-level (global) features from the received signal. Specifically, the low-level feature contains the information in a subcarrier and its adjacent subcarriers, and the high-level feature contains the information in the whole OFDM block. Since clipping causes the energy on a subcarrier to spread into both adjacent and non-adjacent subcarriers, both the low-level and high-level features need to be exploited to preserve useful information contained in the received signal.

The contribution of this chapter can be summarized as follows. We design a 1D-TRNet based receiver for the non-linearly distorted OFDM system. This 1D-TRNet receiver serves as an end-to-end model and it integrates the functions of demodulation as well as removing inter-symbol interference and inter-carrier interference. The proposed 1D-TRNet has been shown to outperform existing methods such as the GAMP algorithm with better BER and less computational complexity. Note that we have considered an uncoded system though a similar observation can be applied to a coded system.

The rest of this chapter is organized as follows. In Section 4.2, the system model is introduced. In Section 4.3, the proposed 1D-TRNet receiver is described in detail. In Section 4.4, the computational complexity of the proposed 1D-TRNet is analyzed. Simulation results are presented in Section 4.5 and conclusion is drawn in Section 4.6.

## 4.2 System Model

In this section, the system model for OFDM signals with clipping is described. The randomly generated bit sequence $b_k$ is mapped by quadrature phase-shift keying (QPSK) modulation to produce the symbol sequence $a_k$, where $k = 1, 2, ..., K$ denotes the index of the sequence. The time-domain transmitted OFDM signals

can be expressed as

$$x_n = \sum_{k=1}^{N\zeta} F_{n,k} b_k, \qquad n = 1, \cdots, N\zeta, \tag{4.1}$$

where $F_{n,k} = \frac{1}{\sqrt{N\zeta}} e^{j\frac{2\pi kn}{N\zeta}}$, $N$ denotes the number of subcarriers, and $\zeta$ is the oversampling factor. The clipped signal can be formulated as

$$f_{cr}(x_n) = \begin{cases} x_n, & |x_n| \leq \Psi_{cr}, \\ \Psi_{cr} e^{j\angle(x_n)}, & |x_n| > \Psi_{cr}, \end{cases} \tag{4.2}$$

where $\Psi_{cr}$ is the clipping threshold and $\angle(x_n)$ denotes the phase angle of $x_n$. The clipping ratio (CR) can be expressed as $\Psi_{cr}/\sigma$, where $\sigma^2$ denotes the signal power. The received time-domain OFDM signals can be formulated as

$$r_n = \mathbf{f}_{cr}(x_n) * \mathbf{h}_n * \mathbf{f}_{oob} + \alpha_n, \tag{4.3}$$

where $*$ denotes the convolution operation, $\mathbf{h}_n$ is the multi-path channel response, $\mathbf{f}_{oob}$ is the out-of-band (OOB) filter which is applied according to the setting in [31], and $\alpha_n$ denotes the additive white Gaussian noise (AWGN) with power $\sigma_\alpha^2$.

## 4.3 Proposed 1D-TRNet Receiver

The proposed 1D-TRNet consists of a 1D U-shape feature extractor and multiple GRUs to recover the signal. In this section, the proposed 1D-TRNet will be described in details.

### 4.3.1 Problem Formulation

Before passing through the 1D-TRNet neural network, the received signal $r_n$ will firstly pass through a pre-processing block, which contains cyclic prefix removal, frequency domain conversion using FFT, and one-tap frequency-domain equalization (FDE) to obtain the output $y_n$. The detection of the transmitted bits $b_k$ can be formulated as a classification problem. The input is the received frequency-domain signal $\mathbf{y} = \{y_n,\ n = 1, ..., N\zeta\}$. The ground truth is the transmitted frequency-domain bits $b_k,\ k = 1, ..., N$. The target is the estimated frequency-domain bits $\hat{b}_k,\ k = 1, ..., N$. We maximize the posterior probability of the estimated bits $\hat{b}_k$ given by

$$\hat{b}_k = \arg\max_{b_k} P(b_k|\mathbf{y}). \tag{4.4}$$

### 4.3.2 The Proposed 1D-TRNet Structure

Overall, the proposed 1D-TRNet based receiver consists of two different components, i.e., the 1D U-shape feature extractor and the GRU layers. As shown in Fig. 4.1, different colour blocks in 1D U-shape feature extractor represent different operations. Subsequently, the received signals are processed by 1D-convolution, max-pooling, transpose convolution, concatenating operation and GRU layers. For each operation, the function is described in the following subsections in detail.

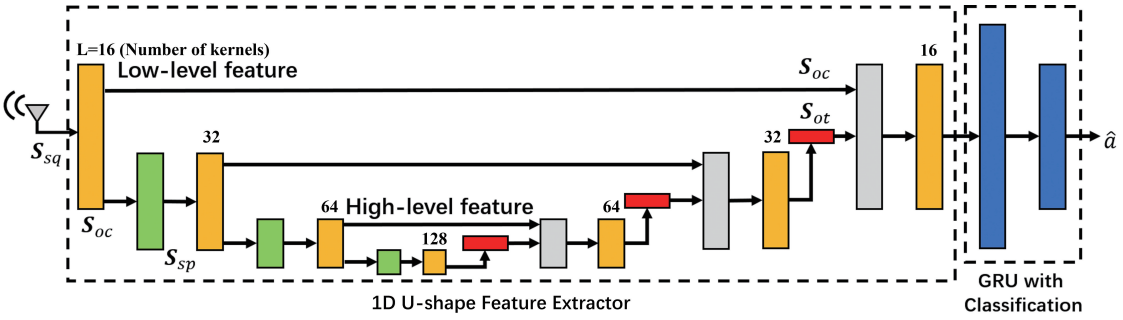



FIGURE 4.1: The structure of the proposed 1D-TRNet receiver.

The proposed 1D-TRNet serves as an end-to-end model. The signal $y_n$ is firstly reshaped into $\mathbf{S}_{sq}$ with batch size $M$ and block length $J = N\zeta$, as given by

$$\mathbf{S}_{sq} = \begin{bmatrix} y^Q{}_{(1,1)} & y^Q{}_{(1,2)} & \cdots & y^Q{}_{(1,J)} \\ y^I{}_{(1,1)} & y^I{}_{(1,2)} & \cdots & y^I{}_{(1,J)} \\ y^I{}_{(2,1)} & y^I{}_{(2,2)} & & y^I{}_{(2,J)} \\ \vdots & & \ddots & \\ y^I{}_{(M,1)} & & & y^I{}_{(M,J)} \end{bmatrix}^T. \tag{4.5}$$

In (4.5), $y^I{}_{(i,j)}$ and $y^Q{}_{(i,j)}$ denote the real and imaginary parts of the reshaped signal $y_n$, respectively.

#### 4.3.2.1 Dual 1D convolution operation (Yellow Block)

The dual 1D convolution operation follows the following processes: 1D-convolution $\rightarrow$ batch normalization $\rightarrow$ *Relu* activation function $\rightarrow$ 1D-convolution. Multiple 1D-kernels are used in the 1D convolution layer. Each kernel $\mathbf{K}$ can be represented as

$$\mathbf{K} = [w_1 \ w_2 \ ... \ w_I]^T, \tag{4.6}$$

where $w_i$ denotes the weight (learnable parameters) for each kernel, and $i = 1, 2, .., I$ denotes the $i^{\text{th}}$ weight. The convolution layer performs convolution operation between the kernel and the input sequence, which can be formulated as

$$S^l_{con\ (m,j)} = \sum_{i=0}^{I-1} w^l_{I-i} S_{sq\ (m,j+i)} + b^l{}_{(m,j)}, \tag{4.7}$$

where the output of convolution can be expressed as $\mathbf{S}_{con} = S^l_{con(m,j)}$, and $l$ denotes the $l^{\text{th}}$ kernel. Note that $b^l{}_{(m,j)}$ denotes the bias, and zero padding is added at both ends of input signal to maintain the feature size. In the second step, the batch

normalization layer can be expressed as

$$\mu_B = \frac{1}{MJL}\sum_{m=1}^{M}\sum_{j=1}^{J}\sum_{l=1}^{L} S^l_{con\ (m,j)}, \tag{4.8}$$

$$\sigma_B^2 = \frac{1}{MJL}\sum_{m=1}^{M}\sum_{j=1}^{J}\sum_{l=1}^{L} (S^l_{con\ (m,j)} - \mu_B)^2, \tag{4.9}$$

$$\widehat{S}^l_{con\ (m,j)} = \frac{S^l_{con\ (m,j)} - \mu_B}{\sqrt{{\sigma_B}^2 + \varepsilon}}, \tag{4.10}$$

$$S^l_{oc\ (m,j)} = \gamma \widehat{S}^l_{con\ (m,j)} + \beta, \tag{4.11}$$

where $\mu_B$ and $\sigma_B^2$ denote the mean and variance of $S^l_{con\ (m,j)}$, respectively, and $\varepsilon$ is added in the denominator for numerical stability with an arbitrarily small constant. Note that $M$, $J$, and $L$ are the sizes of tensor to calculate the mean and variance. In addition, $\gamma$ and $\beta$ are learnable parameters, which are updated in the training process. The output is denoted as $\mathbf{S}_{oc}$.

#### 4.3.2.2 Max-pooling operation (Green Block)

The max-pooling operation is to compress the input sequence. By passing through the max-pooling operation, the convolution kernels extract the information gradually from local to global (i.e., from low level to high level), which can be formulated as

$$S^l_{sp\ (m,n)} = \max\left(S^l_{con\ (m,2n-1)}, S^l_{con\ (m,2n)}\right), \tag{4.12}$$

$$\mathbf{S}_{sp} = \left[\begin{array}{cccc} S^L & S^L & \cdots & S^L \\ S^2 & S^2 & \cdots & S^2 \\ S^1_{sp\ (m,1)} & S^1_{sp\ (m.2)} & \cdots & S^1_{sp\ (m,J/2)} \end{array}\right]^T, \tag{4.13}$$

where $\mathbf{S}_{sp}$ denotes the output of the max-pooling operation in tensor form.

#### 4.3.2.3 Transpose convolution operation (Red Block)

The transpose convolution operation is used as an up-sampling process [65]. We denote the output of the transpose convolution operation as $\mathbf{S}_{ot}$.

#### 4.3.2.4 Concatenating operation (Grey Block)

After each transpose convolution, the output will be concatenated with the corresponding cropped feature as illustrated by grey block in Fig. 4.1. We define the information before and after the max-pooling operation as low-level and high-level features, respectively. This process will prevent information loss in the max-pooling operation. The concatenated information is given by

$$Cat(\mathbf{S}_{oc}, \mathbf{S}_{ot}) =$$

$$\begin{bmatrix} S^L \cdots S^L & S^L \cdots S^L \\ S^2 \cdots S^2 & S^2 \cdots S^2 \\ S^1_{oc\ (m,1)} \cdots S^1_{oc\ (m,J)} & S^1_{ot\ (m,1)} \cdots S^1_{ot\ (m,J)} \end{bmatrix}^T \quad . \tag{4.14}$$

After concatenation, both the low-level and high-level features are completely extracted and flattened before sending to the GRU layers [66]. The GRU layers are used in the end due to its advantage in the processing of time series sequence. We denote $\{g_k\}$ as the output of the GRU layers. The GRU can properly calculate the influence for the sequence of feature maps. Since the range of influenced subcarriers caused by clipping is uncertain, it is necessary to use GRU to process the feature map ergodically.

#### 4.3.2.5 Sigmoid and classification

In our proposed 1D-TRNet receiver, we use the *Sigmoid* function to normalize the probability of the estimated bits $\hat{a}_k$ in the range of $(0, 1)$ given by

$$P(\hat{a}_k) = \frac{1}{1 + \exp(-g_k)}, \tag{4.15}$$

where $P(\hat{a}_k)$ is defined as the probability of estimated bits. The binary cross-entropy loss function can be formulated as

$$\mathcal{L}_{loss} = -\sum_{k=0}^{N-1} P(a_k) \log P(\hat{a}_k) + (1 - P(a_k)) \log(1 - P(\hat{a}_k)), \tag{4.16}$$

where $P(a_k)$ is the *a priori* probability of the transmitted bits.

### 4.3.3 Dataset Generation and Training Procedure

All the training data and test data are generated by MATLAB 2020 following the system model setting. The multi-path Rayleigh time-varying fading channel is considered and the parameters are shown in Table 4.1 [31]. The one-tap frequency-domain equalization is adopted for both conventional methods and deep learning based methods [31]. The transmitted bits are treated as labels in the training process. Training and testing datasets have 50,000 OFDM frames and 5,000 OFDM frames with 128 subcarriers, respectively.

In the training procedure, we set the batch size to $M = 128$ and the learning rate to 0.0001 with 500 epochs. The well-knwon ADAM optimizer is used for training proposes to find the global minimum. Our 1D-TRNet has 7 dual 1D convolution operations with kernel size of $1 \times 3$. The number of kernels is shown in Fig. 4.1. Since $E_b/N_0$ is considered as a known parameter in the GAMP and F-GAMP algorithms, we train the network under different $E_b/N_0$ separately, and assume that the network knows the $E_b/N_0$ information. Following the training procedure presented in [5], 16 bits are predicted for every block. The training procedure is based on the Python 3.8 environment with the Pytorch deep learning framework deploying the NVIDIA RTX 3080 GPU.

TABLE 4.1: Parameters for generating the dataset

| Parameters | Value |
|---|---|
| Number of subcarriers $N$ | 128 |
| Number of multi-path | 4 |
| Normalized path delays | [0, 0.001, 0.002, 0.003] |
| Average path gain (dB) | [0, -3, -6, -9] |
| Oversampling factor $\zeta$ | 4 |
| Fading model | Rayleigh fading |

## 4.4 Computational Complexity Analysis

In this section, we analyze the computational complexity of the proposed scheme for online testing. The number of multiplications is considered to measure the computational complexity. The complexity of the proposed network consists of two parts, one is from the 1D U-shape based feature extractor, and the other is from the GRU recurrent layer. For each dual 1D-convolutional layer, the number of multiplications can be expressed as $\mathcal{O}(IJC_{out}(C_{in}+C_{out}))$, where $I$ and $J$ denote the length of each kernel and input feature, respectively. Note that $C_{in}$ and $C_{out}$ denote the number of input kernels and output kernels, respectively. Thus, the computational complexity of the U-shape feature extractor can be expressed as $\mathcal{O}(\sum_{k=1}^{K} I_k J_k C_{out}^k (C_{in}^k + C_{out}^k))$ where $k$ is the $k^{\text{th}}$ convolutional layer. Moreover, considering the recurrent component, the complexity of GRU can be expressed as $\mathcal{O}(3h_g(h_g + J) + h_g)$, where $h_g$ denotes the number of hidden neurons of the GRU layer. The computational complexity of the proposed 1D-TRNet receiver can be expressed as

$$\mathcal{O} \sim (\sum_{k=1}^{K} I_k J_k C_{out}^k (C_{in}^k + C_{out}^k) + \sum_{s=1}^{S} (h_g^s (h_g^s + J_s) + h_g^s)), \tag{4.17}$$

where $s$ denotes the $s^{\text{th}}$ GRU layer. In our design, since only two GRU layers are used, most of the complexity comes from the 1D-convolutional layer feature

extractor.

## 4.5 Simulation Results and Discussions

In this section, the BER performance of the proposed 1D-TRNet receiver is compared with some existing receivers developed for OFDM signals with non-linear distortion. The experiments consider both the situations with and without out-of-band (OOB) filter, which correspond to practical and ideal scenarios, respectively. QPSK modulation, 16 quadrature amplitude modulation (QAM) and 64 QAM are applied in each subcarrier. Following [30] and [51], we use an oversampling factor of $\zeta = 4$. The channel is a 4-path frequency-selective Rayleigh fading channel with delays of 0, 1, 2, and 3 symbol periods, respectively [31]. Our proposed 1D-TRNet can also be extended to other soft nonlinear distortion problems.

### 4.5.1 Simulation Results without OOB Filter

In Fig. 4.2, we present the BER performance of the proposed 1D-TRNet based receiver, the DNN receiver [5], the F-GAMP receiver proposed in [31] for clipped OFDM signal, as well as the asymptotic performance bound in [51]. To be consistent with the results presented in [30] and [67], we set the number of iterations in F-GAMP to be 5. From Fig. 4.2, it can be observed that the average BER performance gain for the proposed 1D-TRNet receiver is about 4.3 dB at BER of $10^{-3}$ compared with the F-GAMP algorithm. The clipping effect will introduce frequency diversity and inter-carrier interference. The frequency diversity can bring performance gain, if inter-carrier interference can be properly removed, which has been achieved by our proposed 1D-TRNet. In this case, the high-level feature becomes important, and since our proposed 1D-TRNet exploits both the low and high-level features, it outperforms the DNN scheme proposed in [5].

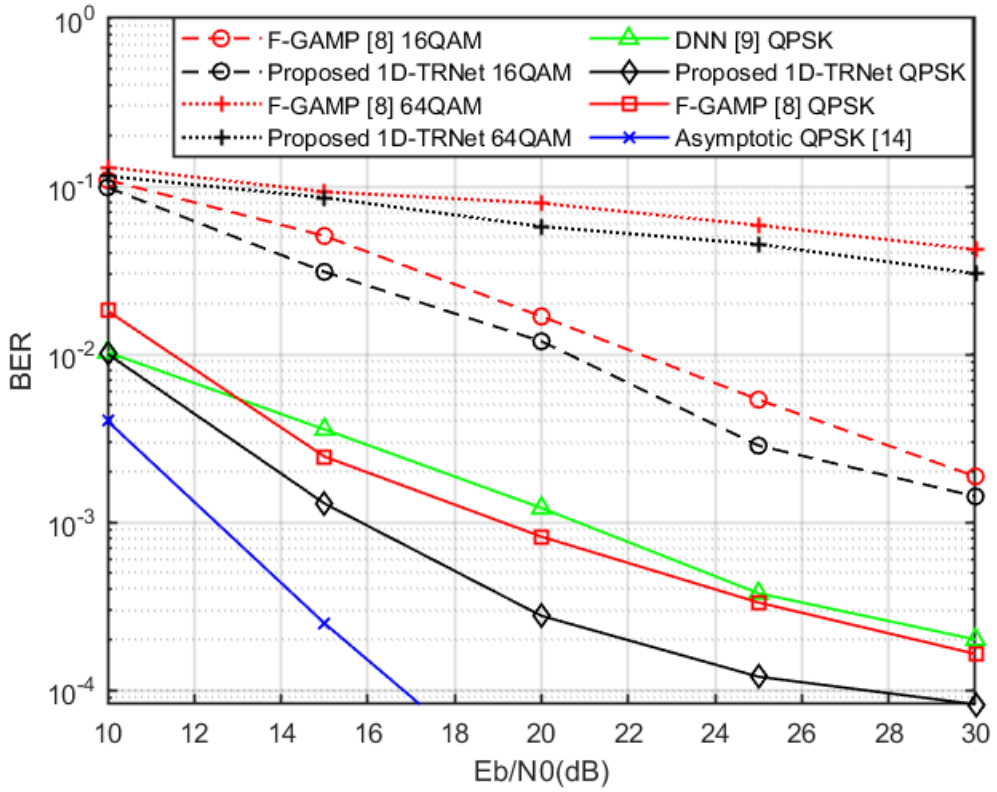



FIGURE 4.2: Simulation results without OOB filter.

The BER performance of the proposed 1D-TRNet receiver, the DNN receiver, and the F-GAMP receiver is compared in Fig. 4.3 for different clipping ratios with QPSK modulation and $E_b/N_0 = 25$ dB. It can be observed that our proposed 1D-TRNet based receiver outperforms the DNN and F-GAMP receivers for all CR levels. In addition, there is an optimum CR level, which is at around 0.6-0.8. This is because the non-linear clipping operation will introduce two opposite effects, i.e., frequency diversity which has positive effect and ICI which has negative effect [68]. In general, both the frequency diversity and ICI will increase as CR decreases. At the very low CR region, the ICI effect dominates and degrades the BER performance substantially. When CR increases, the frequency diversity effect becomes stronger than the ICI effect, the BER performance improves, and the optimum BER performance is achieved at CR level of around 0.6-0.8. Beyond this optimum CR level, the BER will increase slightly.

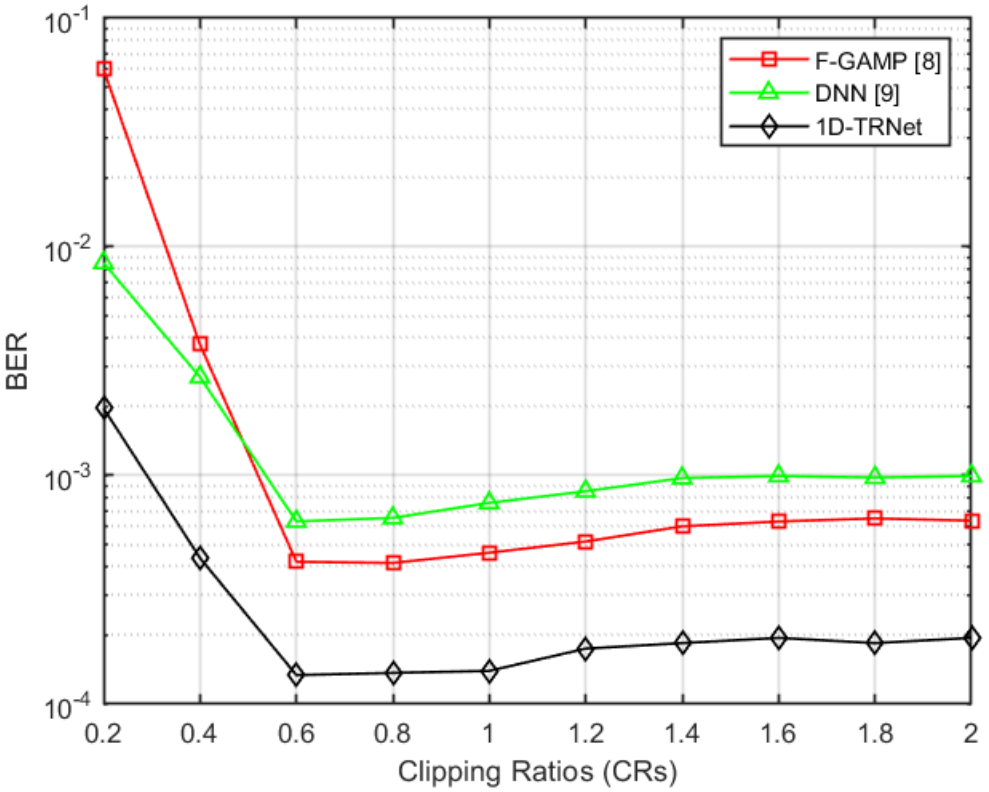



FIGURE 4.3: Simulation results with different CRs ($E_b/N_0 = 25$ dB).

### 4.5.2 Simulation Results with OOB Filter

In Fig. 4.4, we compare the BER performance of the proposed 1D-TRNet, GAMP, F-GAMP, and DNN-based receivers with OOB filter. Here, we apply the OOB filter on both the transmitter side and receiver side to simulate more practical situations. We follow [31] and [51] to set the OOB filter parameters. Moreover, we assume that for every eight subcarriers, a pilot is inserted to estimate the channel state information for frequency-domain equalization. For the GAMP method, we use step sizes of $N_x = 16$ and $N_x = 64$ denoted as GAMP 16 and GAMP 64, respectively. The number of iterations for the GAMP and F-GAMP algorithms is set at 5. From Fig. 4.4, we observe that the proposed receiver gives the best BER performance, which has around 3 dB gain as compared to F-GAMP and GAMP 64 algorithms. This is because the proposed receiver can extract both the low-level and high-level features and preserve all the useful information in the received

signal. In contrast, in the high $E_b/N_0$ range, i.e., $E_b/N_0 > 23$ dB, the DNN-based receiver is worse than both F-GAMP and GAMP 64 algorithms. This is due to the limitation of DNN receiver [5] where the high-level features (information on non-adjacent subcarriers) are not extracted.

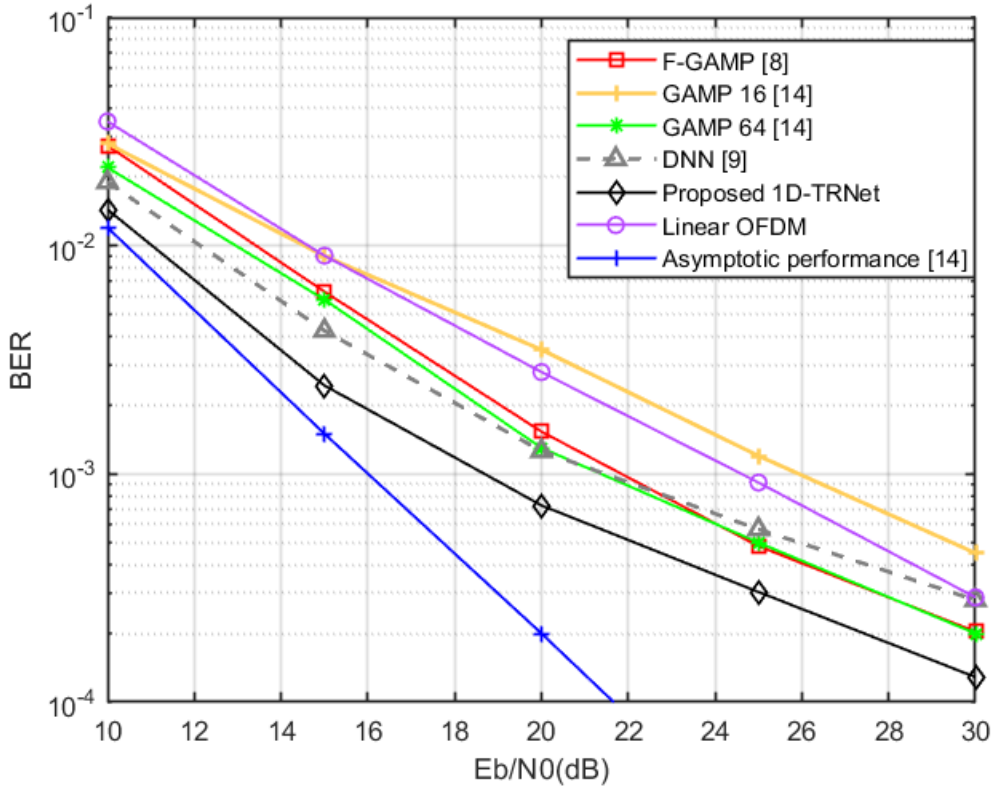



FIGURE 4.4: Simulation results with OOB filter.

### 4.5.3 Simulation Results for different numbers of Multi-path

As mentioned in Section 4.3, we apply the one-tap frequency-domain equalization for the received signal, and we assume that the channel estimation is perfect, which follows the same system setting and assumption as [31]. Firstly, if there is no nonlinear distortion, the increase of multipath will not bring a larger gain because OFDM converts frequency-selective channels into a number of flat fading channels. Unless the channel is time-varying (for each OFDM frame), otherwise, the channel

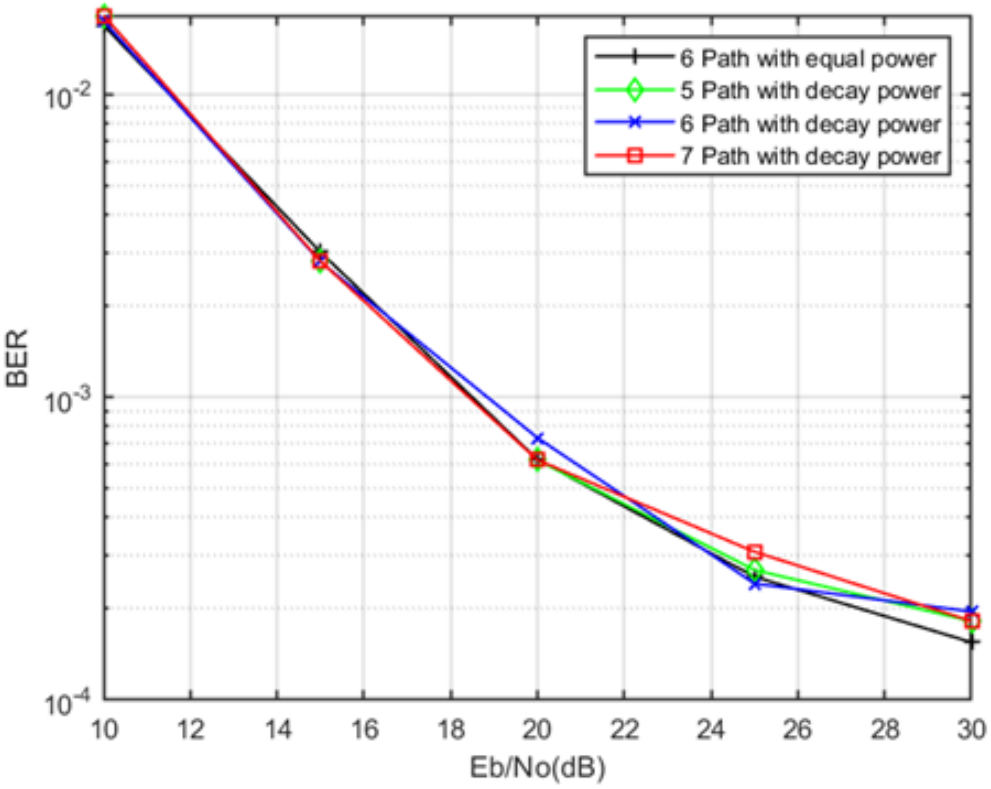



FIGURE 4.5: Simulation results for different numbers of multipath.

matrix is a diagonal matrix, so there is no multipath diversity gain. Secondly, since we consider nonlinear distortion, the orthogonality between subcarriers is distorted, and the channel matrix will not be a diagonal matrix anymore, so there should have diversity gain. However, just like the effect of clipping, if increasing the number of multipath brings more frequency diversity gain, it may also incur more ICI. The final BER depends on which of the two factors is dominant. Thus, even with the increase of multipath components, the gain will not increase significantly as shown in Fig. 4.5.

### 4.5.4 Ablation Experiment

In this section, we conduct an ablation experiment focused on network structure fine-tuning. The objective is to evaluate the impact of modifying specific components or layers of the network architecture on the overall performance of the model.

To conduct the ablation experiment, we start with a baseline network architecture that has been previously trained and optimized for a specific task or dataset. We then systematically modify or remove specific components, such as layers, connections, or modules, and observe the resulting changes in performance metrics. we present the results of two ablation experiments conducted to analyze the impact of different network structures on the performance of the 1D-TRNet model in practical 5G communication scenarios. The experiments focus on fine-tuning the network architecture by systematically modifying specific components and evaluating their effect on the BER metric at various SNR levels.

In the first ablation experiment, we investigate the influence of different convolutional block sizes on the performance of the TRNet model. Four configurations are compared, varying the number of Convolutional Blocks (2, 3, 4, and 5), to understand how it affects the BER values at different SNR levels (10 dB to 30 dB). As shown in Fig. 4.6, we observe that increasing the number of convolutional blocks generally leads to a decrease in BER performance. This suggests that deeper network architectures with more convolutional blocks have the potential to capture and exploit underlying features more effectively, resulting in improved channel state information estimation. However, we also notice that beyond a certain block size, the performance gains become less significant, indicating the presence of diminishing returns. Hence, there exists a trade-off between model complexity and performance enhancement.

In the second ablation experiment, we study the impact of different kernel sizes on the performance of the 1D-TRNet model. Four configurations with varying kernel sizes (1x1, 1x3, 1x5, and 1x7) were compared, and the BER values are analyzed at

different SNR levels. As shown in Fig. 4.7, the choice of kernel size significantly influences the model's performance. Smaller kernel sizes (1x1 and 3x3) tend to perform well in capturing local patterns and details, leading to lower BER values. Conversely, larger kernel sizes (1x5 and 1x7) capture broader spatial information but exhibited slightly higher BER values. The optimal kernel size depends on the communication scenario and the desired balance between local and global feature representations.

The ablation experiments provide valuable insights into the importance of specific network components in the 1D-TRNet model for handling non-linear distortion. Deeper network architectures and smaller kernel sizes generally result in better performance. However, it is essential to consider the trade-offs between model complexity and performance gain, as well as the specific requirements of the communication scenario. These findings offer guidance for fine-tuning the 1D-TRNet model and customizing it to suit different communication scenarios. By making informed decisions about the network architecture and hyperparameters, researchers can optimize the model's performance and achieve improved channel state information estimation in practical 5G communication systems.

It is important to note that the results presented in this section are based on specific experimental settings and datasets. Further evaluation on diverse datasets and communication scenarios will be necessary to validate the robustness and generalizability of the proposed ablation experiments.

To ensure the validity of the experiment, it is important to carefully define evaluation metrics and perform multiple runs to account for variability. Additionally, appropriate control experiments should be conducted to compare against randomly modified or shuffled networks to establish the significance of the observed performance changes.

The results of the ablation experiment can provide valuable insights into the network structure, identifying critical components and potentially highlighting areas where further optimization or modification may be beneficial.

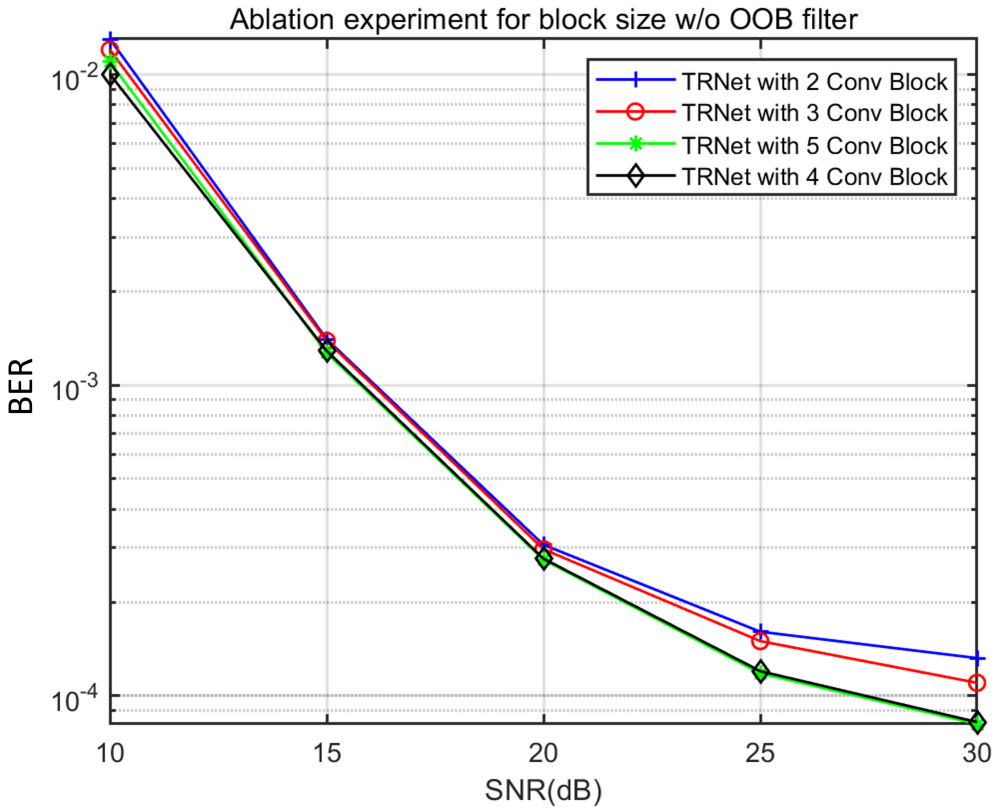



FIGURE 4.6: Ablation experiment of block size.

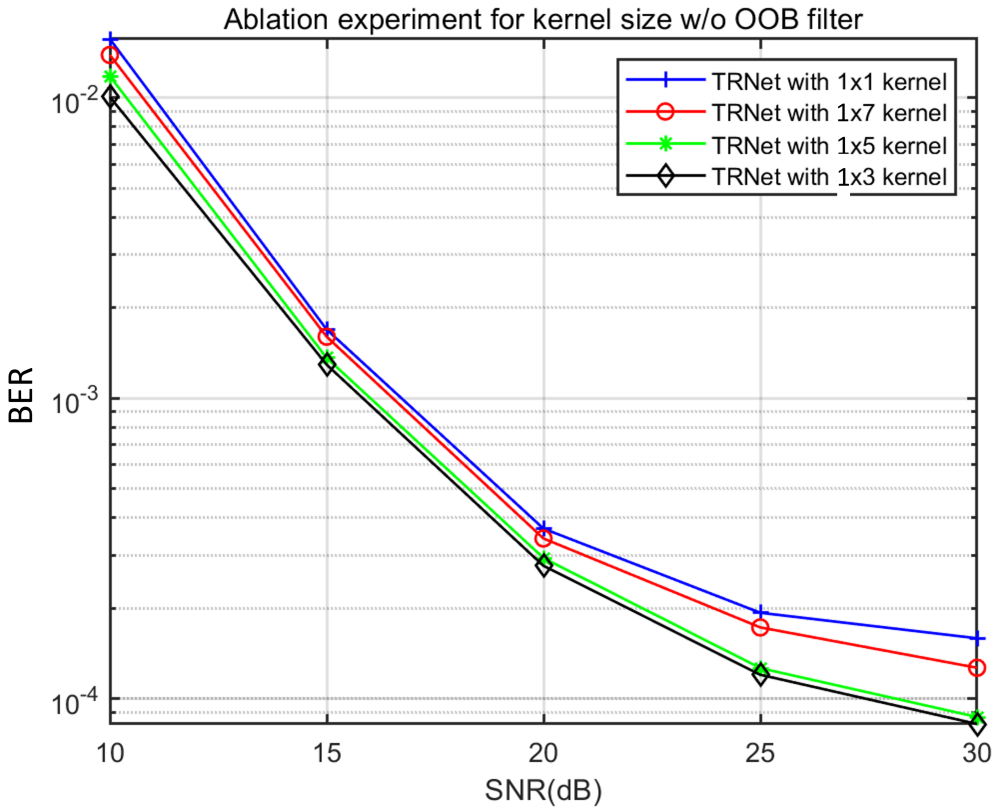



FIGURE 4.7: Ablation experiment of kernel size.

### 4.5.5 Comparisons of Online Testing Computational Complexity

In this section, we compare the computational complexities of the GAMP algorithm with step size of $N_x = 16$ and $N_x = 64$, the F-GAMP algorithm and the proposed 1D-TRNet receiver. For the conventional method, the typical number of iterations required is set to $t_{max} = 5$ [31]. The required numbers of multiplications for various algorithms are shown in Table II. Our proposed 1D-TRNet based receiver requires less computational complexity than the conventional GAMP algorithm with $N_x = 64$ and $N_x = 16$. Moreover, our proposed 1D-TRNet based receiver can achieve around 3dB gain in BER performance as compared with the conventional GAMP algorithm with $N_x = 64$ and the F-GAMP algorithm. In addition, the proposed 1D-TRNet scheme requires storage of the learned parameters for each SNR though it is not very significant.

TABLE 4.2: Comparisons of Online Testing Computation Complexities

| **Algorithm** | **Number of multiplications** | **Memory** |
|---|---|---|
| F-GAMP [31] | $1.13 \times 10^5$ | – |
| GAMP 16 [51] | $2.37 \times 10^7$ | – |
| GAMP 64 [51] | $3.78 \times 10^8$ | – |
| Proposed 1D-TRNet | $8.64 \times 10^6$ | 3.6MBytes |

## 4.6 Conclusion

In this chapter, a 1D-TRNet based end-to-end receiver has been proposed to detect OFDM signals which are non-linearly distorted by clipping. Our proposed network extracts both low-level and high-level features of the nonlinearly distorted signal. Simulation results have shown that our proposed 1D-TRNet based receiver outperforms the conventional methods such as GAMP and F-GAMP, as well as the DNN receivers. Furthermore, we have also conducted computational complexity analysis and it has been shown that our proposed 1D-TRNet receiver requires less computational complexity as compared with the conventional GAMP method.

In the following chapter, we present a novel and robust deep learning-based receiver for 5G communication systems operating under time-variant tapped-delay line (TDL) channels. The time-variant nature of TDL channels poses significant challenges for conventional communication techniques, including inter-carrier interference (ICI) and inter-symbol interference (ISI), which can lead to performance degradation and increased complexity. We anticipate that the deep learning based receiver will demonstrate superior robustness and performance in handling TDL channel effects, leading to more reliable and efficient 5G communication systems.

# Chapter 5

# Comm-Transformer: A Robust Deep Learning Based Receiver for 5G Communication under TDL Channel

## 5.1 Introduction

In this chapter, the primary focus is on studying the application of deep learning techniques for channel estimation in 5G communication systems, specifically for challenging channel conditions such as doubly selective fading channels. Channel estimation plays a crucial role in 5G systems as it enables accurate characterization of the wireless channel, which is essential for optimizing transmission parameters and achieving reliable communication.

Doubly selective fading channels are characterized by time-varying and frequency-selective fading effects, which present significant challenges for traditional channel estimation techniques. DL-based approaches are explored as a potential solution to

enhance the accuracy and efficiency of channel estimation in such complex channel conditions.

By utilizing DL algorithms, the chapter aims to learn and predict the channel characteristics more effectively, improving the accuracy of channel estimation in 5G systems. The research conducted in this chapter contributes to the understanding and development of DL-based channel estimation techniques for 5G communication systems, emphasizing the importance of accurate channel estimation in achieving reliable and efficient communication in challenging channel conditions.

Overall, the chapter aligns with the thesis's objective of exploring the applications of deep learning in the physical layer of wireless communication systems. It highlights the significance of DL-based channel estimation in 5G systems and its potential to improve the overall performance of communication systems operating in challenging channel conditions.

With the evolution of communication systems beyond the fifth generation (B5G), we will encounter more and more requirements and challenges [3]. Specifically, physical layer (PHY) communication techniques play a vital role in the next-generation communication system. Considering the more complex channel conditions under the millimeter-wave (mmWave) and Terahertz (THz) communication, channel estimation techniques for B5G will highly affect the quality of service (QoS). Conventional channel estimation methods such as least square (LS), minimum mean square error (MMSE), and linear MMSE (LMMSE) could not fulfill the conditions under the severe multi-path and fast time-varying channel. Meanwhile, the development of deep learning (DL) techniques and computational power provides an alternative solution for the B5G challenges [9]. DL is making a profound technological revolution to the concepts, patterns, methods, and architectures of wireless communication systems [3], [69], [70]. There have been some interesting DL-based research works for PHY of communication systems, including channel estimation, channel state information (CSI) feedback compression [23], [71], signal detection [72], dynamic resource allocation [73], and beamforming [74] , etc. Under

the instruction of the $3^{rd}$ Generation Partnership Project (3GPP) release 17 [75] and release 18 [76], the incorporation of artificial intelligence (AI) and machine learning (ML) technologies are currently under study for potential inclusion in the evolution of 5G Advanced.

Channel estimation is a widely-studied case among all DL applications for wireless communication systems [77]. The first attempt of DL methods to learn the characteristics of CSI in orthogonal frequency division multiplexing (OFDM) systems for channel estimation problems was presented in [5]. In addition, in [6] and [22], the authors introduced a novel framework that combines DL methods and multiple-input multiple-output (MIMO) systems to address direction-of-arrival (DoA) estimation and channel estimation problems. Furthermore, in [19], the authors extended DL-based channel estimation to doubly selective fading channels and demonstrated superior performance compared to many conventional methods. Moreover, a sparse complex-valued neural network structure was proposed in [19] to tackle channel estimation in massive MIMO systems. Another line of research aimed to establish an end-to-end deep neural network (DNN) architecture that replaces all modules at the transmitter and the receiver, respectively [35].

Meanwhile, DL techniques have emerged as a promising alternative for high-dimensional channel estimation. For instance, in [35], the authors employed convolutional neural networks (CNNs) to interpolate and denoise 2-dimensional OFDM channels, while in [78], a specialized CNN was integrated as a denoiser into the approximate message passing (AMP) algorithm for beamspace channels. In [36], [79], the authors presented a novel end-to-end wireless communication system using DNN and a conditional generative adversarial net (GAN) to model variable CSI effects, demonstrating effective performance across various channel types and paving the way for data-driven DNNs in communication systems. Additionally, in [80], CNNs were utilized for 3-dimensional channels to exploit correlations in frequency, time, and space. Moreover, a DL-based receiver for non-linear distortion in the OFDM system was proposed in [81]. In [82], the authors used the DL approach for signal authentication and security in massive internet-of-things systems.

### 5.1.1 Motivation and Related Work

Recently, more and more researchers have demonstrated that DL-based algorithms could be applied to channel estimation problems. In [5], DL-based channel estimation has been proposed, which indicates that a fully-connected layer neural network could be adapted to the OFDM system and provides a better QoS compared with the MMSE receiver. Following that, the channel estimation problem drew more attention, from classic Rayleigh fading to complicated doubly selective fading such as TDL channel. Specifically, in [48] and [83], CNNs have been used to interpolate and denoise the 2-D OFDM channels. In [84], the authors provided a solid support to theoretically analyze the Relu-DNN for channel estimation problems. Also, some researchers applied DL methods to the doubly selective channel. In [7], the authors proposed a Bi-gate recurrent unit (GRU) based DL receiver for the multi-path Rayleigh fading channel. In [85], the authors proposed a DL-based receiver for high mobility channel conditions. In [50], the authors investigated the DL-based receiver with ResNet backbone under TDL channels without nonlinear distortions. However, the research for the 5G channel with severe doubly selective fading is still a critical problem, e.g., the TDL-A channel will be much more difficult to handle than the TDL-E channel. Besides, the robustness of the DL-based receiver also needs to be further studied.

What distinguishes this chapter from the existing techniques is that we apply the channel positional encoding with attention mechanisms to the OFDM subcarriers. With the channel positional weights compensating the subcarriers which have deep fading effects, our proposed Comm-Trans Net could provide a very precise channel estimation result. Furthermore, with two-phase training, Comm-Trans Net could serve as an end-to-end receiver for the 5G TDL channel. Moreover, our proposed Comm-Trans Net is robust against different communication scenarios. The average gain from the TDL-A channel to the TDL-E channel is around 3 dB. Additionally, we also study the impact of the insufficient pilot situation on the system performance.

The proposed Comm-Trans Net is motivated by the need to effectively handle the challenges inherent in wireless communication, such as deep fading and ICI. Each of the components—self-attention, 1D-U shape feature extractor, and GRU—contributes uniquely to this goal, equipping the network with the necessary tools to learn from and adapt to the dynamic nature of wireless communication channels.

### 5.1.2 Contribution

The main contribution of this chapter is to propose a Comm-Trans Net based receiver for the 5G doubly selective channel under different scenarios. This DL-based receiver integrates the function of channel estimation, equalization, and demodulation. Our proposed Comm-Trans Net receiver takes advantage of attention mechanism [86], [87], which gives each subcarrier of the OFDM system positional encoding for compensating the deep fading effect. The weights of the positional encoder are learned by the training phase based on the different sub-types of the TDL channel. Numerical results show that our proposed Comm-Trans Net receiver is robust against the different communication scenarios under the TDL channel. We demonstrate that the proposed Comm-Trans Net can serve as an end-to-end receiver for OFDM systems under the 5G channel condition and is robust to the 5G communication environment. Our contributions extend beyond the novel architecture, encompassing both theoretical and empirical aspects. Specifically, we have two key theoretical contributions. First, the attention mechanism is organically combined with the subcarriers of the OFDM system. Due to the multi-path fading and time-varying nature of the TDL channel, channel estimation is challenging for conventional methods. Thus, positional encoding is applied for each subcarrier to compensate for the path loss. Second, we investigate the robustness of the proposed Comm-Tran Net receiver for TDL channels with different sub-types according to different communication scenarios.

Our empirical results indicate that the proposed Comm-Trans Net receiver is capable of addressing the significant challenges associated with severe doubly selective

fading, the robustness of DL-based receivers, and insufficient pilots. To be more specific, we conducted a benchmark of our technique against the performance of conventional channel estimators as well as several well-known DL-based receivers. Note that we have considered an uncoded system though similar observation can be applied to a coded system. Our proposed Comm-Trans Net receiver can achieve a 3 dB gain in bit-error rate (BER) performance at SNR = 20 dB as compared to existing methods. Furthermore, it can be shown that the proposed Comm-Trans Net receiver outperforms the DNN receiver [5].

## 5.2 System Model and Channel State Information

In this section, the communication system for non-linear distorted OFDM signals is described. The baseband-equivalent model is presented for illustration purposes. First, the bit sequence $b_k$ is randomly generated and mapped by quadrature phase-shift keying (QPSK) modulation as $a_k$, where $k = 1, 2, ..., K$. Note that it also can be extended to high-order modulation schemes. The transmitted signals through the OFDM scheme can be expressed as

$$x_n = \frac{1}{\sqrt{N\zeta}} \sum_{k=1}^{N\zeta} F_{n,k} b_k, \qquad n = 1, \cdots, N\zeta, \tag{5.1}$$

where $F_{n,k} = e^{j\frac{2\pi kn}{N\zeta}}$, $N$ denotes the number of subcarriers, and $\zeta$ represents the oversampling factor. The clipping operation is performed as the non-linear distortion for the OFDM signal, which can be formulated as

$$\phi_{cr}(x_n) = \begin{cases} x_n, & |x_n| \leq \theta_{cr}, \\ \theta_{cr} e^{j2\pi \angle(x_n)}, & |x_n| > \theta_{cr}, \end{cases} \tag{5.2}$$

where $\theta_{cr}$ and $\angle(x_n)$ denote the clipping amplitude and phase angle of the transmitted sequence, respectively. The clipping ratio (CR) is defined as $\theta_{cr}/\sigma$, where $\sigma^2$ is the power of the transmitted signal $x_n$. Thus, the non-linearly distorted OFDM signals in time-domain can be expressed as

$$r_n = h_n * \phi_{cr}(x_n) + \alpha_n, \tag{5.3}$$

where $\alpha_n$ denotes the additive white Gaussian noise (AWGN) with power $\sigma_\alpha^2$, and $*$ is the convolution operation. Note that $h_n$ denotes the TDL channel response, which is considered for both outdoor and indoor 5G communication systems. This model supports the frequency range from 0 to 100 GHz with a maximum bandwidth of 2 GHz [53]. The TDL is constructed for the line-of-sight (LoS) and non-line-of-sight (NLoS) scenarios, and the Doppler spectrum for each tap is characterized by a classical (Jakes) spectrum shape, and the first tap follows the Rician fading distribution. To generate the channel coefficient, the following fast-fading model will be adopted:

$$h_{u,s,n,m}^{NLoS}(t) = \sqrt{\frac{P_n}{M}} \begin{bmatrix} F_{rx,u,\theta}(\theta_{n,m},\phi_{n,m}) \\ F_{rx,u,\phi}(\theta_{n,m},\phi_{n,m}) \end{bmatrix}^T \begin{bmatrix} \exp\left(j\Phi_{n,m}^{\theta\theta}\right) & \sqrt{K_{n,m}^{-1}}\exp\left(j\Phi_{n,m}^{\theta\phi}\right) \\ \sqrt{K_{n,m}^{-1}}exp\left(j\Phi_{n,m}^{\phi\theta}\right) & \exp\left(j\Phi_{n,m}^{\theta\theta}\right) \end{bmatrix}$$
$$\begin{bmatrix} F_{tx,s,\theta}(\theta_{n,m},\phi_{n,m}) \\ F_{tx,s,\phi}(\theta_{n,m},\phi_{n,m}) \end{bmatrix}^T \exp\left(j2\pi\frac{\hat{r}_{rx,n,m}^T.\bar{d}_{rx,u}}{\lambda_0}\right)\exp\left(j2\pi\frac{\hat{r}_{tx,n,m}^T.\bar{d}_{tx,s}}{\lambda_0}\right)\left(j2\pi\frac{\hat{r}_{rx,n,m}^T.\bar{v}}{\lambda_0}\right). \tag{5.4}$$

In the LoS case, the channel coefficient can be obtained as follows:

$$h_{u,s,n,m}^{LoS}(t) = \begin{bmatrix} F_{rx,u,\theta}(\theta_{LoS},\phi_{LoS}) \\ F_{rx,u,\phi}(\theta_{LoS},\phi_{LoS}) \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} F_{tx,s,\theta}(\theta_{LoS},\phi_{LoS}) \\ F_{tx,s,\phi}(\theta_{LoS},\phi_{LoS}) \end{bmatrix}$$
$$\exp\left(-j2\pi\frac{d_{3D}}{\lambda_0}t\right)\exp\left(j2\pi\frac{\hat{r}_{rx,n,m}^T.\bar{d}_{rx,u}}{\lambda_0}\right)\exp\left(j2\pi\frac{\hat{r}_{tx,n,m}^T.\bar{d}_{tx,s}}{\lambda_0}\right)\left(j2\pi\frac{\hat{r}_{rx,n,m}^T.\bar{v}}{\lambda_0}t\right). \tag{5.5}$$

where $F_{rx,u,\theta}$ and $F_{rx,u,\phi}$ are the field patterns of receive antenna element $u$ according to [53] and in the direction of the spherical basis vectors, $\theta$ and $\phi$, respectively. $F_{tx,s,\theta}$ and $F_{tx,s,\phi}$ are the field patterns of the transmit antenna element $s$ in the direction of the spherical basis vectors, $\theta$ and $\phi$, respectively. Also, $\bar{d}_{tx,u}$ is the location vector of the receive antenna element $u$ and $\bar{d}_{tx,s}$ is the location vector of the transmit antenna element s, $K_{n,m}$ is the cross polarisation power ratio in linear scale, and $\lambda_0$ is the wavelength of the carrier frequency. Note that $\hat{r}_{rx,n,m}$ and $\hat{r}_{tx,n,m}$ are the spherical unit vector, and $d_{3D}$ is the distance for oxygen absorption loss. The path loss and shadowing for the TDL model follow the TR 38.901 standard [53].

## 5.3 Comm-Trans Net based End-to-End communication

In this section, we present the attention mechanism-based end-to-end Comm-Trans Net for doubly-selective channels, in which the overall system block diagram is shown in Fig. 5.1. First, we describe the details of the attention mechanism, and how it works for communication systems. Second, we address the channel positional encoding for the OFDM system. Then, we explain the details of Comm-Trans Net for the end-to-end communication system. Following that, the data flow for the Comm-Trans Net and training procedure are discussed.

### 5.3.1 Attention Mechanism and Channel Positional Encoding

In this part, we describe the attention mechanism and channel positional encoding, which consider the fading difference between different subcarriers. In general, the attention block is shown in Fig. 5.2, where $\mathbf{S}_{sq}$ and $\mathbf{S}_{atten}$ denote the input and output of the attention block, respectively. Note that $\mathbf{B}$ and $\mathbf{J}$ denote the batch size

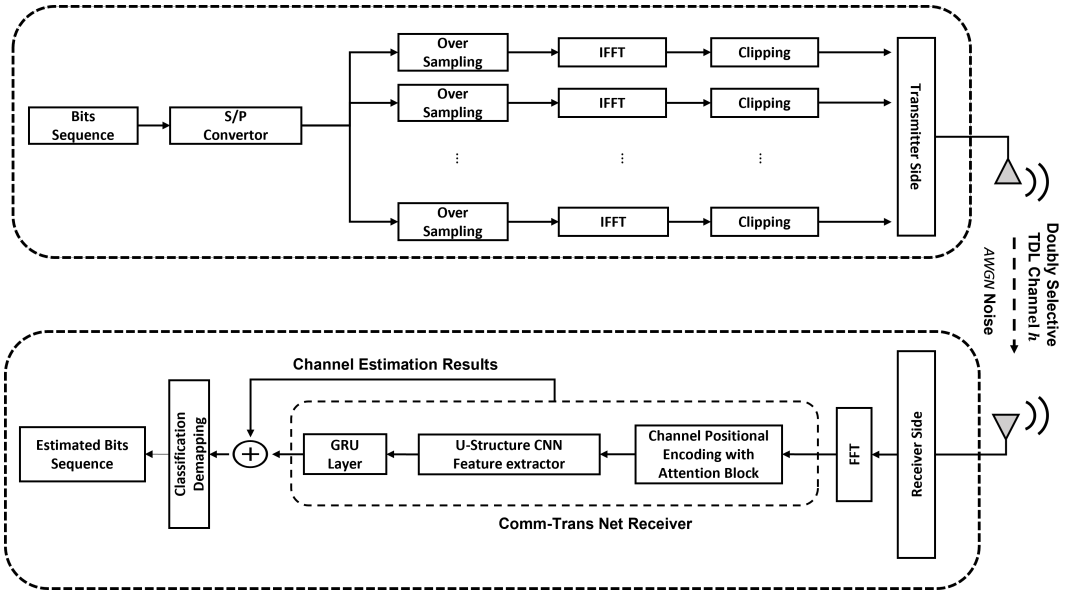



FIGURE 5.1: System block diagram of the proposed Comm-Trans Net receiver.

and dimension of subcarriers, respectively. The attention mechanism aims to focus on the subcarrier which has a severe fading effect. The channel positional encoder will be compensated by different weights accordingly which are calculated by the similarity function. First, we define a general attention operation in Comm-Trans Net as:

$$y_i = \psi(\frac{1}{C(\mathbf{x})} \sum_{\forall j} f(x_i, x_j) g(x_j)), \tag{5.6}$$

where $i$ is the index of an output position (subcarrier) whose response is to be computed and $j$ is the index that enumerates all possible positions. We denote $g(\cdot)$ as the linear mapping for the parameter $x_j$. Note that $\psi(\cdot)$ and $C(\mathbf{x})$ denote the activation function and normalized factor, respectively. In particular, the parameter $x_j$ denotes the subcarrier location information.

Following that, we choose the Gaussian function $f(\cdot)$ as an activation function to calculate the dot-product similarity, which can be formulated as

$$f(x_i, x_j) = e^{{x_i}^T x_j}. \tag{5.7}$$

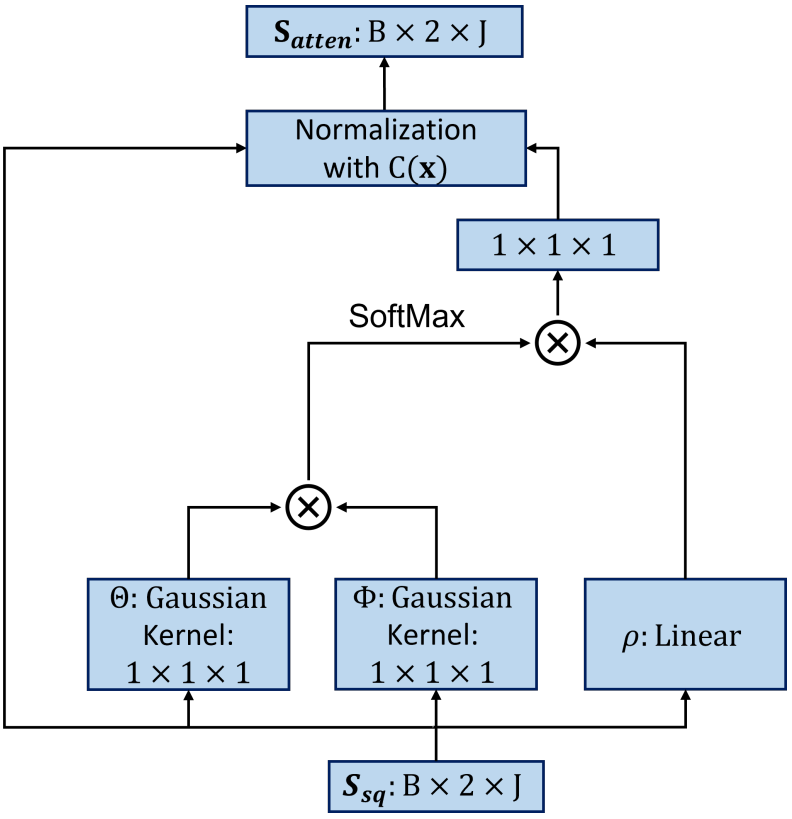



FIGURE 5.2: Attention block diagram.

With the extension of the Gaussian activation function, we introduce the embedded Gaussian function to map the subcarrier information to embedding space, as given by

$$\Theta(x_i) = W_\theta x_i, \quad \Phi(x_j) = W_\phi x_j, \tag{5.8}$$

where $W_\theta$ and $W_\phi$ denote the weights for output subcarriers and all possible deep fading subcarriers, respectively. As shown in Fig. 5.2, the kernel size is $1 \times 1 \times 1$, which can highly reduce the computational complexities. Thus, the weighted embedded Gaussian function can be defined as

$$f(x_i, x_j) = e^{\Theta(x_i)^T \Phi(x_j)}. \tag{5.9}$$

To normalize the weighted embedded Gaussian function, we denote the function $C(\mathbf{x})$ as

$$C(\mathbf{x}) = \sum_{\forall j} f(x_i, x_j). \tag{5.10}$$

This can be seen from the fact that for a given $i$, $\frac{1}{C(\mathbf{x})} f(x_i, x_j)$ becomes the Soft-Max computation along the dimension $j$. Following that, we introduce the linear embedding for embedding space $\rho$, which can be formulated as

$$\rho(x_i) = W_\rho x_i. \tag{5.11}$$

The Relu activation function can be defined as

$$Relu(x) = \begin{cases} 0, & x \leq 0, \\ x, & x > 0. \end{cases} \tag{5.12}$$

Thus, the whole attention block can be formulated as

$$y_{(atten,i)} = Relu(\frac{1}{\sum_{\forall j} f(x_i, x_j)} \sum_{\forall j} e^{\Theta(x_i)^T \Phi(x_j)} W_\rho x_j). \tag{5.13}$$

### 5.3.2 Comm-Transformer Network

In this section, the proposed Comm-Trans Net will be discussed in detail, which consists of several parts including an attention block, feature extractor, and gated recurrent unit (GRU) as shown in Fig. 5.3. The proposed Comm-Trans Net serves

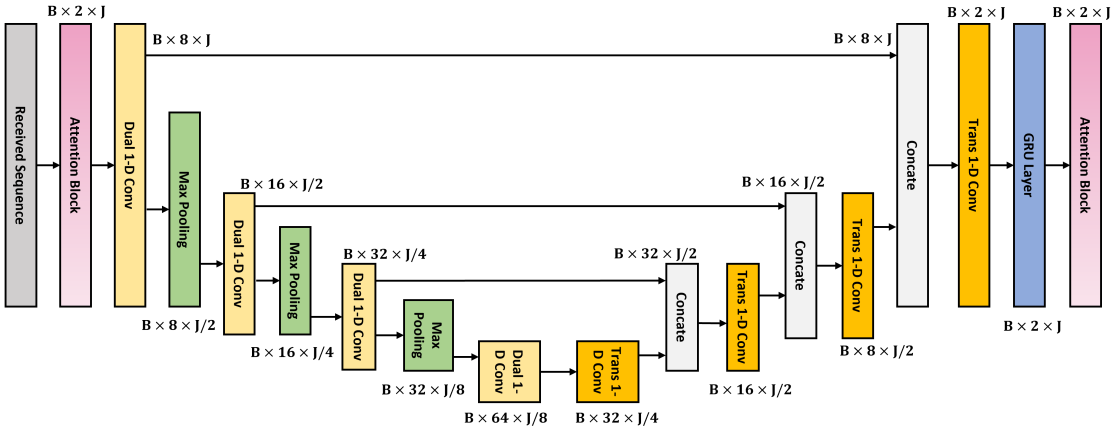



FIGURE 5.3: Comm-Trans Net structure.

as an end-to-end model, which is under the two-phase training procedure for channel estimation and bit recovery.

Overall, the proposed Comm-Trans Net consists of three different components. The received signal $y_n$ is reshaped into $\mathbf{S}_{sq}$ as the input of the Comm-Trans Net, with batch size $M$ and block length $J = N\zeta$, as given by

$$\mathbf{S}_{sq} = \begin{bmatrix} y^{Quad}_{(1,1)} & y^{Quad}_{(1,2)} & \cdots & y^{Quad}_{(1,J)} \\ y^{In}_{(1,1)} & y^{In}_{(1,2)} & \cdots & y^{In}_{(1,J)} \\ y^{In}_{(2,1)} & y^{In}_{(2,2)} & & y^{In}_{(2,J)} \\ \vdots & & \ddots & \\ y^{In}_{(M,1)} & & & y^{In}_{(M,J)} \end{bmatrix}^T, \tag{5.14}$$

where $y^{In}_{(i,j)}$ and $y^{Quad}_{(i,j)}$ denote the in-phase and quadrature components of the reshaped signal $y_n$, respectively.

We implement the attention block to the reshaped signal $y_n$ according to Eq. (5.13), and for each batch, the channel positional encoding will be applied as follows,

$$\mathbf{S}_{atten} = \begin{bmatrix} y^{Quad}_{atten(1,1)} & y^{Quad}_{atten(1,2)} & \cdots & y^{Quad}_{atten(1,J)} \\ y^{In}_{atten(1,1)} & y^{In}_{atten(1,2)} & \cdots & y^{In}_{atten(1,J)} \\ y^{In}_{atten(2,1)} & y^{In}_{atten(2,2)} & & y^{In}_{atten(2,J)} \\ \vdots & & \ddots & \\ y^{In}_{atten(M,1)} & & & y^{In}_{atten(M,J)} \end{bmatrix}^T. \tag{5.15}$$

After the attention block, we apply the dual-1D convolution block for feature extraction. To extract the different levels of the feature, multiple 1D convolution kernels are used. Each dual-1D convolution block follows the sequence of 1D-convolution layer $\rightarrow$ batch normalization layer $\rightarrow$ 1D-convolution layer. We denote $\mathbf{K} = [w_1 \ w_2 \ ... \ w_I]^T$ as the convolutional kernel, where each $w_i$ in the kernel denotes the weight (learnable parameters) learned during the training procedure,

and $I$ denotes the length of the kernel. Each 1D-convolution layer between the input sequence and the kernel can be expressed as

$$S^l_{conv\ (m,j)} = \sum_{i=0}^{I-1} w^l_{I-i} S_{atten\ (m,j+i)} + b^l{}_{(m,j)}, \tag{5.16}$$

where $l$ denotes the $l^{\text{th}}$ kernel, and $\mathbf{S}_{con} = S^l_{con(m,j)}$ is the output of the 1D convolutional layer. The bias $b^l{}_{(m,j)}$ is considered to help the model fit more complex and non-linear functions with zero padding added to maintain the feature size. After each 1D-convolution layer, the batch normalization layer [55] is applied to improve the training procedure, where the mean $\mu_B$ and variance $\sigma^2_B$ are calculated as

$$\mu_B = \frac{1}{JLM} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{m=1}^{M} S^l_{conv\ (m,j)}, \tag{5.17}$$

$$\sigma^2_B = \frac{1}{JLM} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{m=1}^{M} (S^l_{conv\ (m,j)} - \mu_B)^2. \tag{5.18}$$

Note that the sizes of the tensor used to compute the mean and variance are denoted by $J, L$, and $M$. Then we use the parameters $\mu_B$ and $\sigma^2_B$ to normalize the input data, which can be formulated as

$$\widehat{S}^l_{conv\ (m,j)} = \frac{S^l_{conv\ (m,j)} - \mu_B}{\sqrt{{\sigma_B}^2 + \varepsilon}}, \tag{5.19}$$

$$S^l_{oc\ (m,j)} = \gamma \widehat{S}^l_{conv\ (m,j)} + \beta. \tag{5.20}$$

To enhance numerical stability, a minuscule constant $\varepsilon$ is introduced into the denominator. The learnable parameters $\gamma$ and $\beta$ are updated during the training process. The batch normalization output is represented as $\mathbf{S}_{oc}$.

To compress the extracted feature, the max-pooling operation is used. This operation gradually concentrates the extracted information from low level to high level

(e.g., local feature and global feature), resulting in a formulation as:

$$S^{l}_{sp\ (m,n)} = \max\left(S^{l}_{con\ (m,2n-1)}, S^{l}_{con\ (m,2n)}\right). \tag{5.21}$$

The transpose convolution operation is used as an up-sampling process [88]. We denote the output of the transpose convolution operation as $\mathbf{S}_{ot}$.

$$Cat(\mathbf{S}_{oc}, \mathbf{S}_{ot}) = \begin{bmatrix} S^{L} \cdots S^{L} & S^{L} \cdots S^{L} \\ S^{2} \cdots S^{2} & S^{2} \cdots S^{2} \\ S^{1}_{oc\ (m,1)} \cdots S^{1}_{oc\ (m,J)} & S^{1}_{ot\ (m,1)} \cdots S^{1}_{ot\ (m,J)} \end{bmatrix}^{T}_{l}. \tag{5.22}$$

The process involves performing multiple 1D-convolution and max-pooling operations, which extract different levels of features containing both local and global information. These features are then flattened before being sent to the GRU layers, which are chosen for their ability to process time series sequences [66]. The extracted feature map is processed by GRU ergodically, due to the uncertainty for the range of influenced subcarriers caused by the deep fading effect. The output of the GRU layer is denoted as $g_k$, and details of the GRU are formulated as

$$z_k = \sigma_{sig}(w_z S_{cat\ k} + U_z g_{k-1} + b_z), \tag{5.23}$$

$$r_k = \sigma_{sig}(w_r S_{cat\ k} + U_r g_{k-1} + b_r), \tag{5.24}$$

$$\hat{g}_k = \phi_{tanh}(w_g S_{cat\ k} + U_h(r_t \odot g_{k-1} + b_g), \tag{5.25}$$

$$g_k = z_k \odot \hat{g}_k + (1 - z_k) \odot g_{k-1}, \tag{5.26}$$

where $z_k$ and $r_k$ denote the update gate and reset gate, respectively. $\hat{g}_k$ is the candidate activation vector, which holds the possible values for the final output $g_k$. Note that $\odot$ represents the Hadamard product in the gate calculation functions.

### 5.3.3 Training and Testing

#### 5.3.3.1 Problem Formulation

To pre-process the distorted received signal $r_n$ for the Comm-Trans Net receiver, we first eliminate the cyclic prefix and then use fast Fourier transform (FFT) to generate the pre-processed signal $\tilde{y_n}$. As for the task of integrating the demodulation for modulated symbols $a_k$, and detecting the transmitted bits $b_k$, we can frame the problem as a classification task. The input for the network is the pre-processed signal $\mathbf{y} = \{\tilde{y_n},\ n = 1, ..., N\zeta\}$ in frequency domain. The transmitted bits $b_k$ and the estimated bits $\hat{b}_k$ serve as the ground truth label and target, respectively. Both $b_k$ and $\hat{b}_k$ are in frequency domain, and the value of $k$ ranges from 1 to $N$. To solve this problem, we aim to maximize the posterior probability of the estimated bits $\hat{b}_k$, which is given by

$$\hat{b}_k = \arg\max_{b_k} P(b_k|\mathbf{y}). \tag{5.27}$$

#### 5.3.3.2 Loss Design

The loss design includes two parts. First, the *Sigmoid* function is applied to binary classification problem and calculates the probability of estimated bits $\hat{b}_k$, which is given by

$$P(\hat{b}_k) = \frac{1}{1 + \exp(-g_k)}, \tag{5.28}$$

where $P(\hat{b}_k)$ is the probability of estimated bits. Thus, the binary cross-entropy loss function can be formulated as

$$\mathcal{L}_{CE} = -\sum_{k=0}^{N-1} P(b_k) \log P(\hat{b}_k) + (1 - P(b_k)) \log(1 - P(\hat{b}_k)), \tag{5.29}$$

where $P(b_k)$ denotes the *a priori* probability of the transmitted bits.

Second, the channel estimation loss is the mean square error for the regression problem, where frequency domain $\hat{H}_i$ denotes the estimated channel coefficient, and frequency domain $H_i$ denotes the ground truth channel coefficient, as given by

$$\mathcal{L}_{MSE} = -\sum_{i=0}^{N-1} (H_i - \hat{H}_i)^2. \tag{5.30}$$

Thus, the overall loss function can be formulated as

$$\mathcal{L}_{loss} = \lambda \mathcal{L}_{MSE} + (1 - \lambda)\mathcal{L}_{CE}, \tag{5.31}$$

where $\lambda$ is the learning parameter trained by the network. The two-phase training procedure is illustrated in Algorithm 1.

## 5.4 Computational Complexity Analysis

The computational complexity of the Comm-Trans Net during online testing stage is analyzed in this section. We use the number of multiplications as a measure of computational complexity. The complexity of the proposed Comm-Tran Net is composed of three main components: the attention block, the 1D U-shape convolutional operation, and the GRU recurrent layer. The number of multiplications for each attention block can be expressed as $\mathcal{O}(J^2 d_a + d_a^2 J)$, where $J$ represents the number of subcarriers, and $d_a$ represents the number of channels for each attention block. In terms of the 1D U-shape convolutional operation, the number of multiplications for each dual 1D-convolutional layer can be computed as

**Algorithm 1** Comm-Trans Net based Receiver Training Algorithm.

**Initialization:** The transmitted sequences, $X_{train}$, the labels for training $Y_{train}$ and the DL network setting *options*.

**Phase:** 1. Training of channel estimation for TDL

**Input:** The received signals through different sub-type of TDL channels $\mathbf{S}_{sq}$, and ground truth channel state information $H_i$.
**while** iteration $\in$ setting epoch, and stop criterion is not met **do**
    Shuffle the training dataset, initialize the parameters $\theta$,
    and split with min-batch size $M$.
    **for** $i \in$ min-batch $M$ **do**
        Calculate the MSE loss $\mathcal{L}_{MSE}$ by Eq. (30).
        Calculate the gradient for all the weights and loss:
        $g_i = \frac{1}{M}\nabla_{\theta_i}\mathcal{L}_{MSE}$.
        Update the weight and loss: $\theta_{i+1} = \theta_i - \alpha \cdot \frac{\hat{m}_i}{\sqrt{\hat{v}_i}+\epsilon}$,
        based on ADAM optimizer [89].
    **end for**
**end while**
Save the weights for Comm-Trans Net as the pre-trained model and Estimated CSI $\hat{H}_i$.

**Phase:** 2. Training of Demodulation and Bits Recovery for Comm-Trans Net Receiver

**Input** The received signals through different sub-type of TDL channels $\mathbf{S}_{sq}$, Estimated CSI $\hat{H}_i$, and ground truth bit sequence $b_k$.
Concatenate Estimated CSI $\hat{H}_i$ to received signal $\mathbf{S}_{sq}$.
Shuffle the training dataset, and split with min-batch size $M$.
Load the pre-trained Comm-Trans Net from Phase 1.
Train the Comm-Trans Net with $\mathcal{L}_{loss}$ by Eq. (31) with ADAM optimizer to get the estimated bits;
Test with the trained network to get BER versus SNR curves;
**return** BER Score;

$\mathcal{O}(IJC_{out}(C_{in} + C_{out}))$, where $J$ and $I$ represent the length of input feature and each kernel, respectively. We denote $C_{in}$ and $C_{out}$ as the number of input and output kernels, respectively. The computational complexity of the 1D U-shape convolutional operation can be computed as $\mathcal{O}(\sum_{k=1}^{K} I_k J_k C_{out}^k (C_{in}^k + C_{out}^k))$, where $k$ denotes the $k^{\text{th}}$ convolutional layer. Finally, for the GRU layers, the complexity can be expressed as $\sum_{s=1}^{S}(h_g^s(h_g^s + J_s) + h_g^s)$, where $h_g$ represents the number of hidden neurons in each GRU layer, and $s$ denotes the $s^{\text{th}}$ GRU layer. Hence, the overall computational complexity for online testing of the Comm-Trans Net can be stated as the combination of the sum of the attention block, the 1D U-shape

convolutional operation, and the GRU layer, which can be represented as

$$\mathcal{O} \sim (\sum_{a=1}^{A}(J^2 d_a + d_a^2 J) + \sum_{k=1}^{K} I_k J_k C_{out}^k (C_{in}^k + C_{out}^k) + \sum_{s=1}^{S}(h_g^s(h_g^s + J_s) + h_g^s)). \quad (5.32)$$

In summary, the computational complexities of different conventional and DL-based receivers are compared in Table 5.1.

TABLE 5.1: Computational Complexity Comparison

| **Algorithm** | **Computational Complexity** |
|---|---|
| LS-GAMP 64 [51] | $\mathcal{O} \sim (2t_{max} J \log_2(\mathrm{J}) + 36 t_x J N_x^2 + L(2M+1))$ |
| MMSE-GAMP 64 [51] | $\mathcal{O} \sim (2t_{max} J \log_2(J) + 36 t_x J N_x^2 + L^{2.37}(2M+1))$ |
| LSTM Receiver [49] | $\mathcal{O} \sim \sum_{s=1}^{S} 4(h_{lstm}^s(h_{lstm}^s + J_s) + 3h_{lstm}^s))$ |
| DNN Receiver [5] | $\mathcal{O} \sim \sum_{d=1}^{D} J_d I_d$ |
| DeepRx Receiver [50] | $\mathcal{O} \sim \sum_{bk=1}^{BK} N_{bk} \sum_{k=1}^{K} I_k J_k (C_{in}^k + C_{out}^k)^2$ |
| Proposed Comm-Trans Net | $\mathcal{O} \sim (\sum_{a=1}^{A}(J^2 d_a + d_a^2 J) + \sum_{k=1}^{K} I_k J_k C_{out}^k (C_{in}^k + C_{out}^k) + sum_{s=1}^{S}(h_g^s(h_g^s + J_s) + h_g^s))$ |

[a] $t_{max}$ denotes the maximum iteration of the GAMP algorithm, and $N_x$ denotes the discrete samples within the range in each direction are taken for summation. $M$ denotes the modulation order.

[b] $h_{lstm}$ denotes the number of hidden neurons in each LSTM layer, and $s$ denotes the $s^{\text{th}}$ LSTM layer.

[c] $J_d$ and $I_d$ denotes the input and output dimension of the $d^{th}$ fully connected layer of DNN receiver, respectively.

[d] $N_{bk}$ denotes the number of ResNet blocks.

## 5.5 Simulation Results and Discussions

The proposed Comm-Trans Net-based receiver is assessed for the OFDM system under different TDL channels. Our simulation results consist of several parts. The first part is the channel estimation results for different sub-types of TDL channels according to various communication scenarios. Second, based on the channel estimation results, our proposed Comm-Trans Net serves as an end-to-end receiver to recover the data, and BER performance is illustrated. Moreover, the computational

complexities of our proposed Comm-Trans Net, state-of-the-art deep learning-based receiver, and conventional methods are studied. Based on the system model presented in Section II, the OFDM system uses 64 subcarriers, with QPSK modulation applied to each subcarrier. The bit sequences are binomially distributed for data balance. For the condition of insufficient pilots, different numbers of orthogonal pilots are sent for each OFDM frame evenly distributed as shown in Table 5.2. For the 5G channel model, we assume that the TDL-A, TDL-B, TDL-C, TDL-D, and TDL-E CSI are for both offline and online channel realizations. All the different sub-types of TDL channels are generated by the 5G Toolbox supported by MATLAB 2021a. In order to avoid the data imbalance, we equally generate the dataset for each sub-type TDL channel with a size of 100,000. The dataset is randomly split for training, testing, and validation with the ratio of 80%, 10%, and 10%.

The channel state information and response can be generated using the parameters specified in Tables 5.2 and 5.3, and their magnitudes for different paths are visualized in Fig. 5.4.

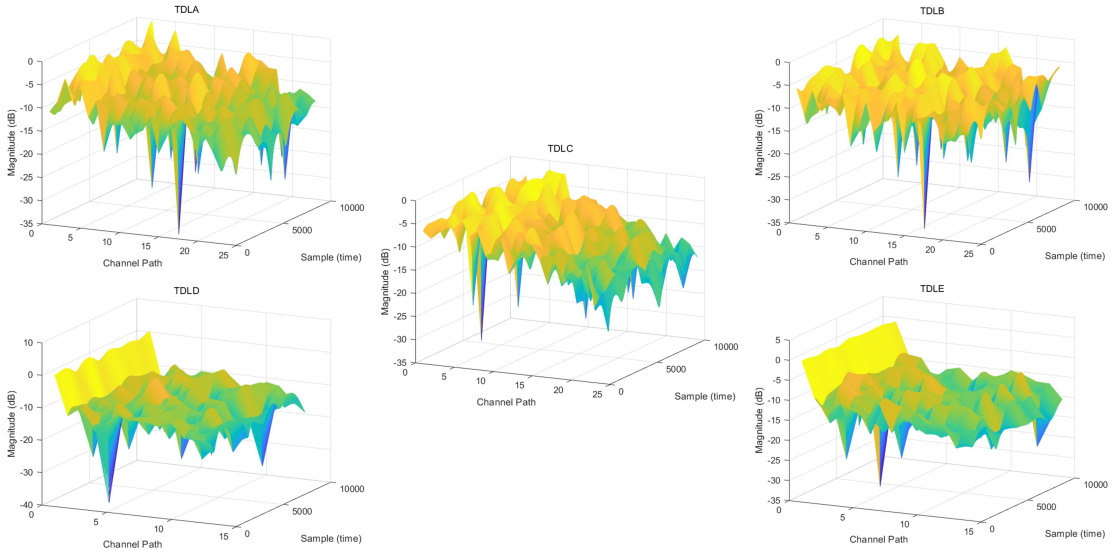



FIGURE 5.4: Different TDL sub-type channels.

In the training procedure, we involve setting the batch size $M = 256$ for 1000 epochs. The learning rate is set to 0.001 with the linear decay schedule for 10 % for every 100 epochs. For dual 1-D convolutional operations, the kernel size is $1 \times 3$.

The number of kernels available for dual 1-D convolutional operations starts at 8 and increases in steps of 2, resulting in a set of available options that include 8, 16, 32, and 64 kernels. The well-known ADAM optimizer is used during the training procedure. We train the network with different SNR strategies, one is assuming that the network has prior knowledge of the SNR information, and the other is training the network following the same SNR strategy as in [5]. The Python 3.8 environment is utilized for training purposes, employing the Pytorch deep learning framework and powered by the NVIDIA RTX 3080 GPU.

To generate the bit sequence, we use the binomial distribution. The bits are then mapped to symbols using QPSK modulation, following the system model in each simulation to generate the transmission sequence before passing through the OFDM system. Subsequently, we transmit the signal randomly over the TDL channel with AWGN noise under different SNR conditions, treating the channel coefficient for each OFDM frame as approximately constant according to the maximum Doppler shift [5]. During the offline training stage, each subcarrier is trained separately under different SNR conditions, and the ground truth is the same as the transmitted bits. In the online testing stage, the trained network does not require channel state information.

TABLE 5.2: System Model Parameters

| **Parameters** | **Value** |
|---|---|
| Carrier Frequency | 4 GHz |
| Maximum Doppler Shift | 111.18Hz |
| Mobile Speed | 8.32 m/s |
| Oversampling Factor | 4 |
| Length of Cyclic Prefix | 16 |
| Number of Subcarriers | 64 |
| Number of Pilots | 64,16,8 |
| Number of Subframe | 2 |
| Block Length | 128 |

TABLE 5.3: Channel Model Parameters

| **Delay Profile** | **TDL-A** | **TDL-B** | **TDL-C** |
|---|---|---|---|
| Number of Paths | 23 | 23 | 24 |
| LOS Path | N/A | N/A | N/A |
| Fading Distribution for Each Path | Rayleigh | Rayleigh | Rayleigh |
| Ricean K Factor | N/A | N/A | N/A |

| **Delay Profile** | **TDL-D** | **TDL-E** |
|---|---|---|
| Number of Paths | 13 | 14 |
| LOS Path | $1^{st}$ path | $1^{st}$ path |
| Fading Distribution for Each Path | Ricean and Rayleigh | Ricean and Rayleigh |
| Ricean K Factor | 13.3 dB | 22 dB |

### 5.5.1 Channel estimation results

Simulation results of different sub-types of TDL channels are presented as examples in Figs. 5.5 and 5.9, where they represent the frequency channel estimation results for TDL-A and TDL-E channels. For a fair comparison, we use the same pilot structure (16 pilots uniformly distributed into 64 subcarriers) for each sub-type TDL channel. From the simulation results, it is observed that the TDL-A channel has the strongest frequency selectivity. The magnitude range for channel frequency response is from -30 dB to 5 dB. Thus, TDL-A is the most challenging channel for estimation. Due to the difference between NLoS and LoS paths for different sub-type of TDL channels, TDL-E will generally be less challenging for channel estimation. The normalized mean square error (NMSE) score is shown in the caption for each figure, which is formulated as follows,

$$\text{NMSE} = \mathbb{E}\left[\frac{||\mathbf{H} - \hat{\mathbf{H}}||_2^2}{||\mathbf{H}||_2^2}\right]. \tag{5.33}$$

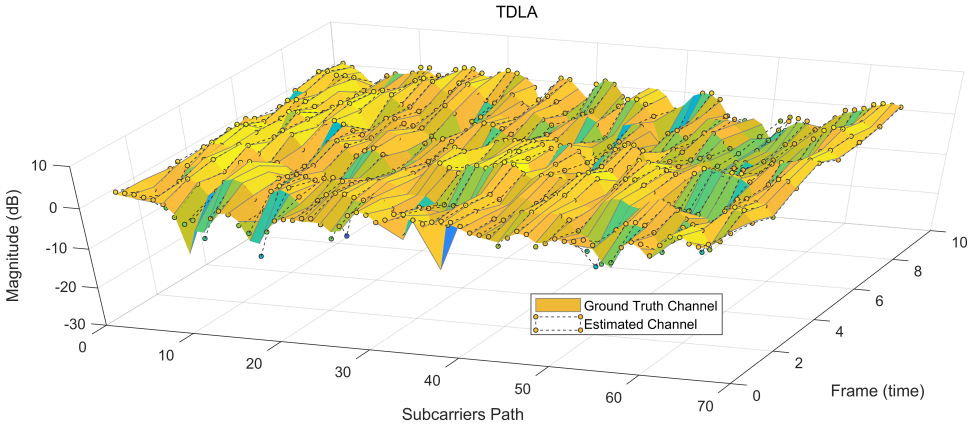



FIGURE 5.5: TDL-A channel estimation result under SNR = 20$dB$, NMSE = -15.7 dB.

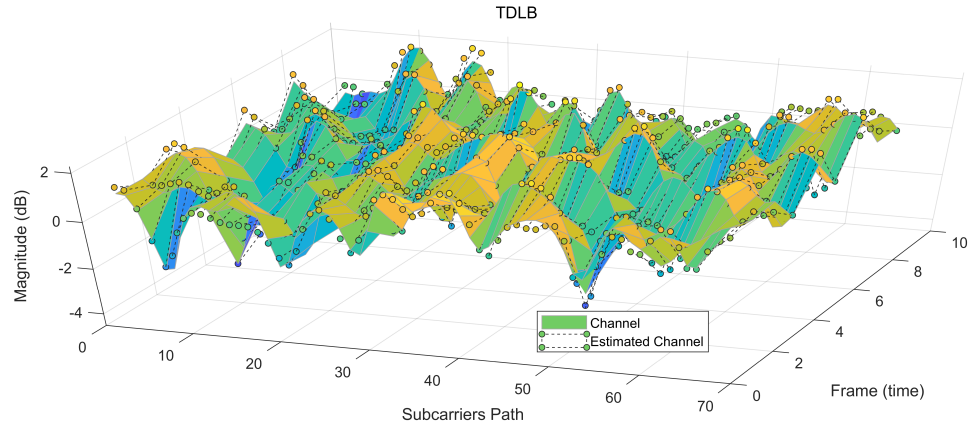



FIGURE 5.6: TDL-B channel estimation result under SNR= 20$dB$, NMSE = -16.2 dB.

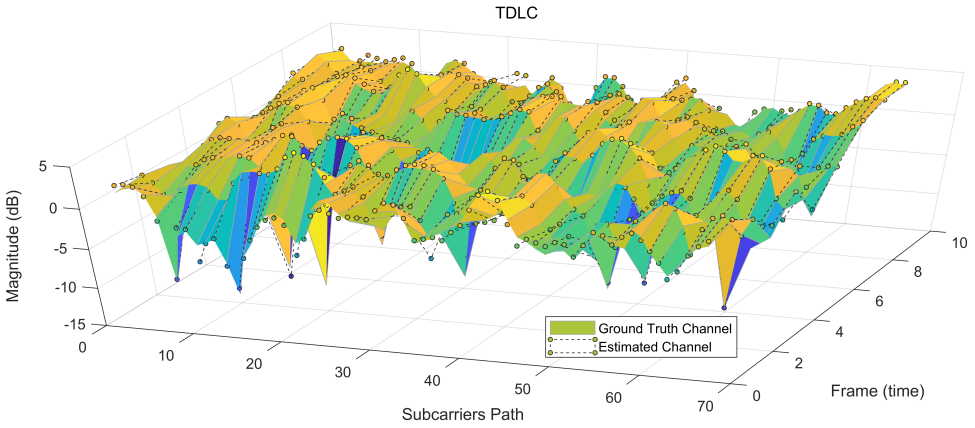



FIGURE 5.7: TDL-C channel estimation result under SNR= 20$dB$, NMSE = -17.5 dB.

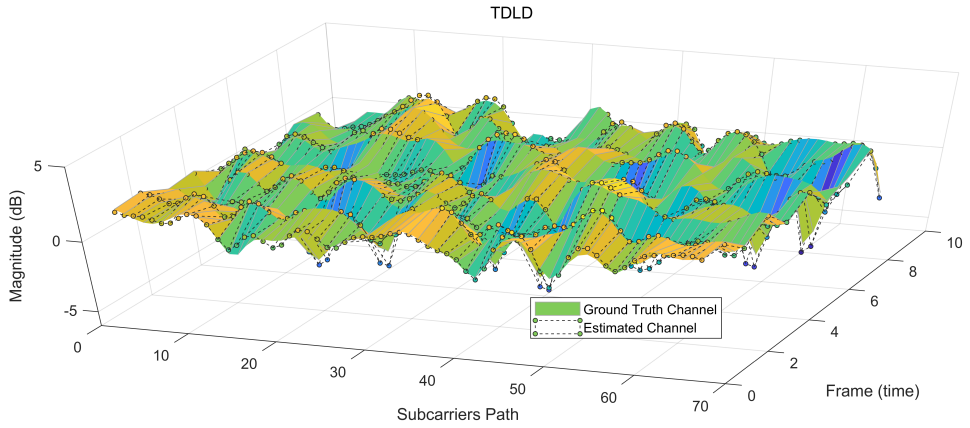



FIGURE 5.8: TDL-D channel estimation result under SNR= 20$dB$, NMSE = -19.6 dB.

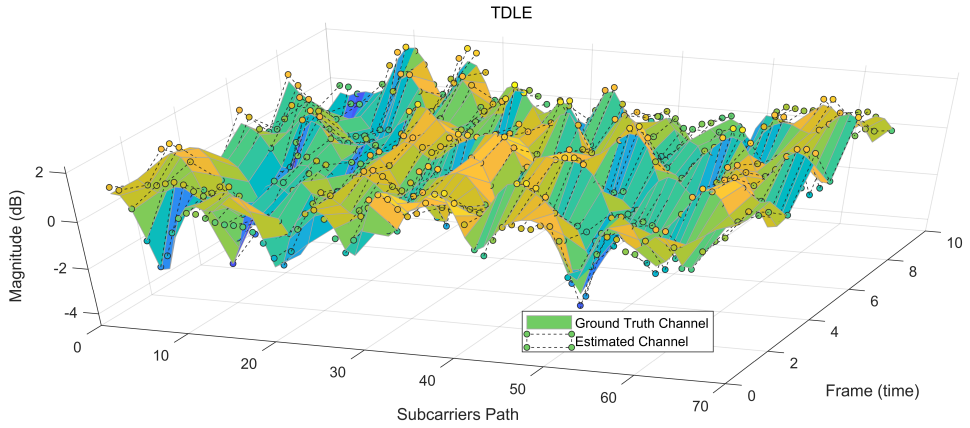



FIGURE 5.9: TDL-E channel estimation result under SNR = 20$dB$, NMSE = -19.0 dB.

The NMSE results in Figs. 5.10 and 5.11 show that our proposed Comm-Trans Net receiver achieves around 3.3 dB for channel estimation from the state-of-the-art methods. Moreover, our proposed Comm-Trans Net also shows the robustness of different sub-types of the TDL channel. When compared to the conventional LS and MMSE channel estimation methods, our proposed scheme achieves a significant NMSE gain across the entire range of SNR. The LS and MMSE methods struggle to handle ISI and ICI under conditions of limited pilot information. Also, our proposed Comm-Trans Net outperforms DNN (Fully connected neural network receiver) and LSTM receives, since the simple structure DL-based receivers

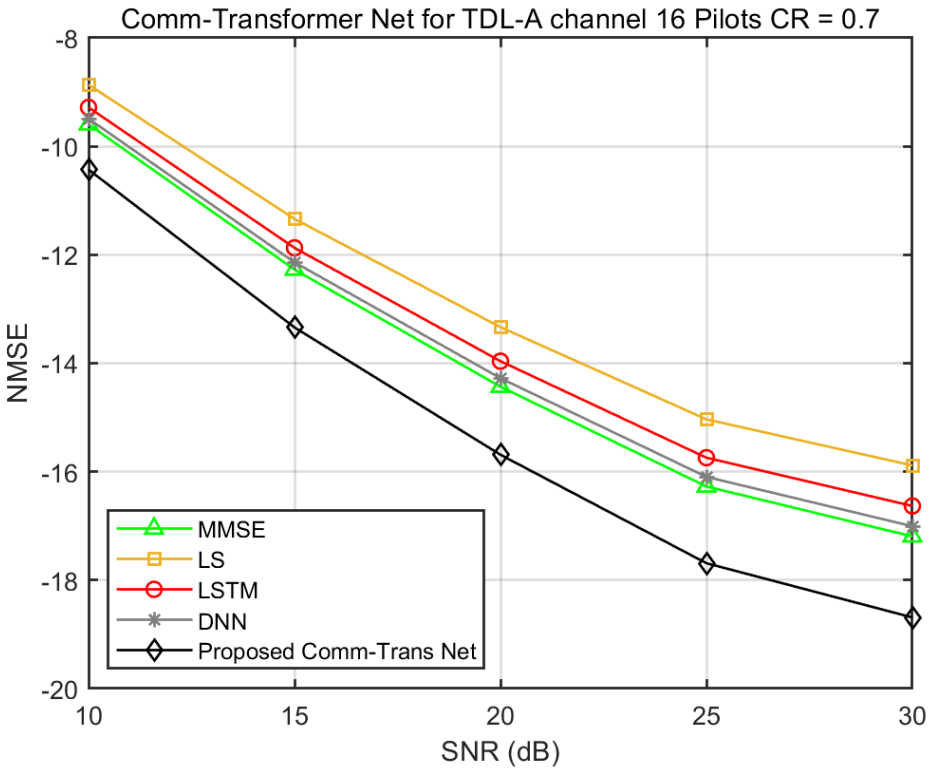



FIGURE 5.10: NMSE simulation results for TDL-A channel.

will lack of ability to extract sufficient features to recover the channel state information. From visualization channel estimation, we observe that the estimated channel state information (CSI) is very close to the ground truth CSI. Compared to conventional channel estimation methods such as LS and MMSE, our proposed Comm-Trans Net could take advantage of channel positional encoding to analyze the deep fading effect and fading spreading to the surrounding subcarriers. The LS or MMSE methods highly rely on the pilots. Under insufficient pilot conditions, the conventional estimation results will be severely affected.

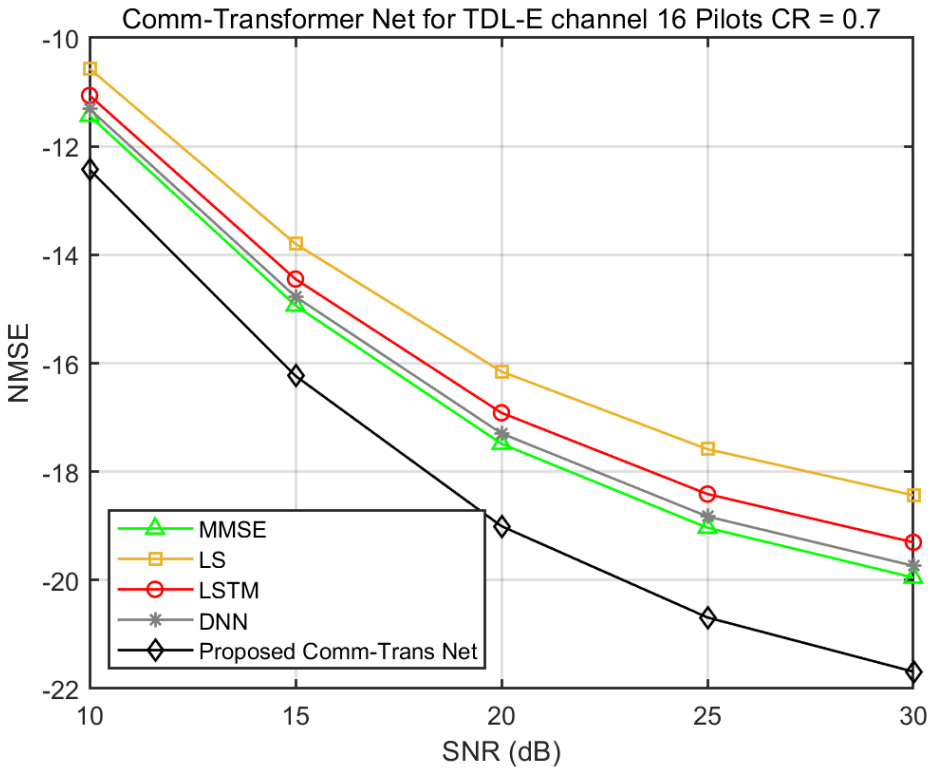



FIGURE 5.11: NMSE simulation results for TDL-E channel.

### 5.5.2 Robustness of the proposed Comm-Trans Net under different TDL sub-types

To study the robustness of the proposed algorithm and existing methods, we test the proposed Comm-Trans Net and existing methods under different sub-types of TDL channels to imitate the practical 5G communication scenarios, compared with DNN [5], LSTM [49], DeepRx [50] and GAMP [51] receivers. The simulation results are presented in Figs. 5.12 to 5.17. The TDL-A channel will be suitable for the urban area with more NLoS, reflection, and refraction. The TDL-E channel will be suitable for the suburban area with more LoS and less frequency selectivity. We test all the algorithms for individual sub-type of TDL channels and mixed TDL channels. Note that we create the dataset with randomly selected TDL sub-types. To avoid data imbalance, the size for each sub-types is 20,000, and the total data size will be 100,000. From the BER results, we can observe that our proposed

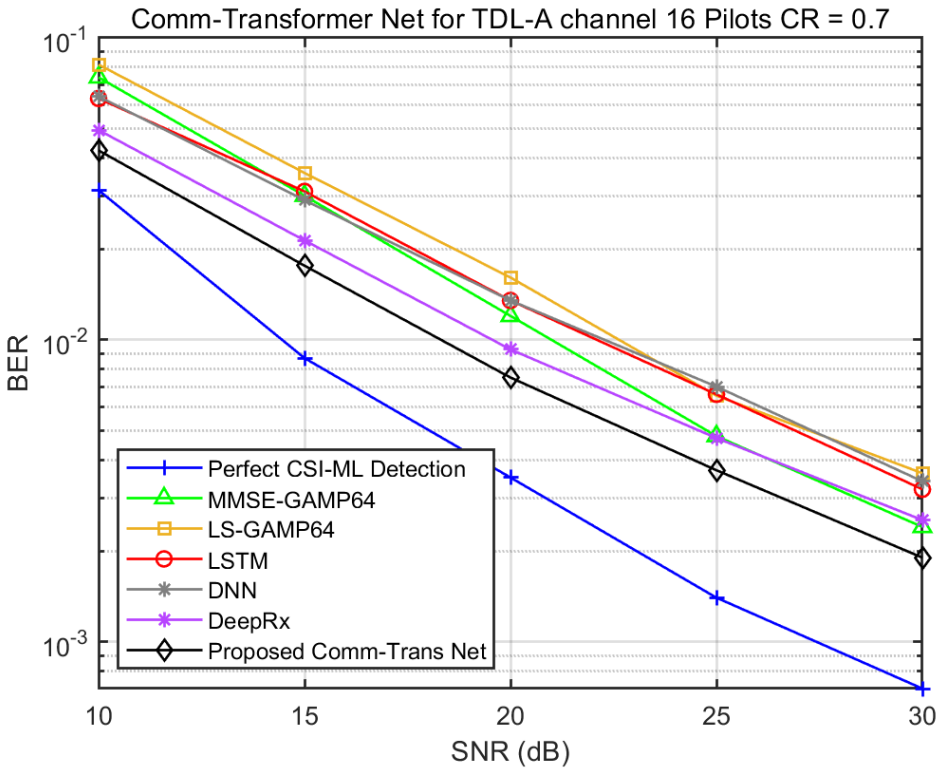



FIGURE 5.12: BER simulation results for TDL-A channel.

Comm-Trans Net outperforms the existing methods, and it can achieve around 3 dB gain at SNR = 20 dB.

We have studied the robustness of prior SNR knowledge by following the SNR training strategy as [5], chosen high SNR for training, and lower SNR for testing, which are shown in Fig. 5.13 and 5.15. We can observe that, the proposed Comm-Trans Net receiver is robust to AWGN noise since the Comm-Trans Net receiver can achieve BER performance close to around 0.6 dB gap without specific prior SNR knowledge. Thus, the proposed Comm-Trans Net is not overfitted to a specific SNR, demonstrating its robustness and adaptability across various SNR ranges.

Moreover, we have studied the adaptability of our proposed Comm-Trans Net receiver for cross-domain subtype of TDL channels, which is shown in Fig. 5.17. Specifically, we train the Comm-Trans Net receiver by TDL-A, TDL-B, TDL-C and TDL-D channels, and test the Comm-Trans Net receiver with the TDL-E channel. The cross-domain result is labeled as "Proposed Comm-Trans Net ex" in Fig.

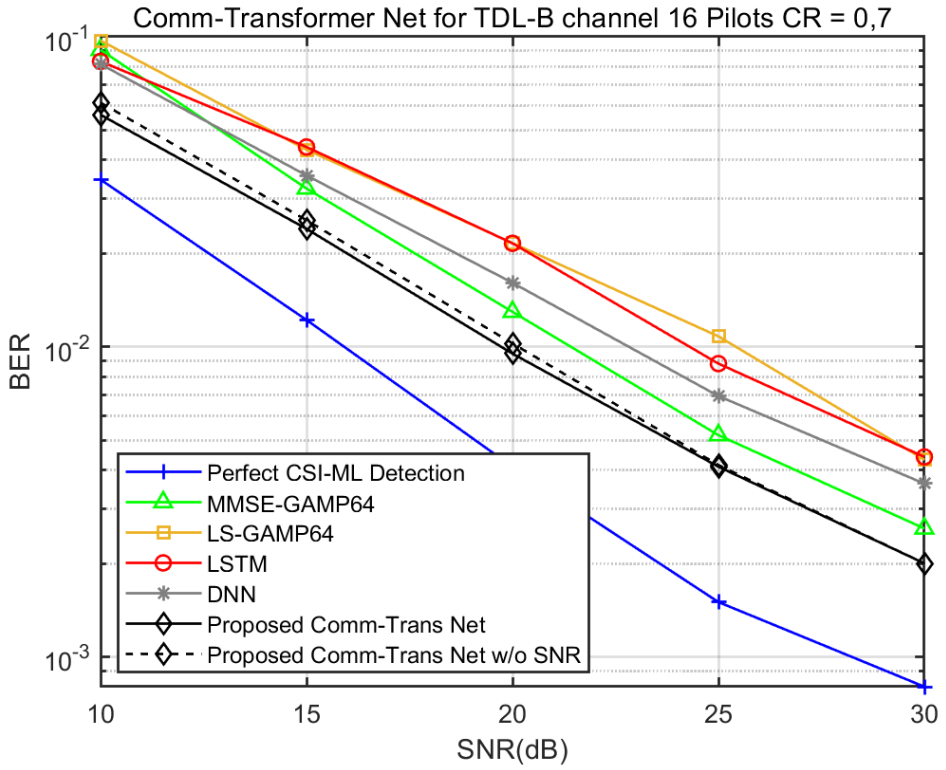



FIGURE 5.13: BER simulation results for TDL-B channel.

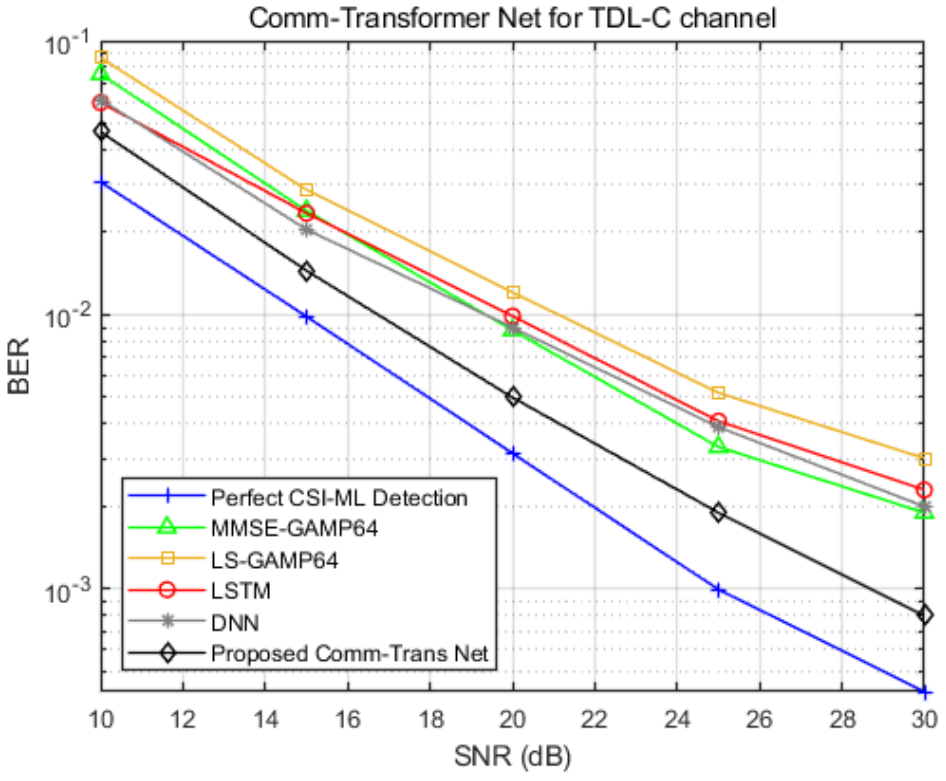



FIGURE 5.14: BER simulation results for TDL-C channel.

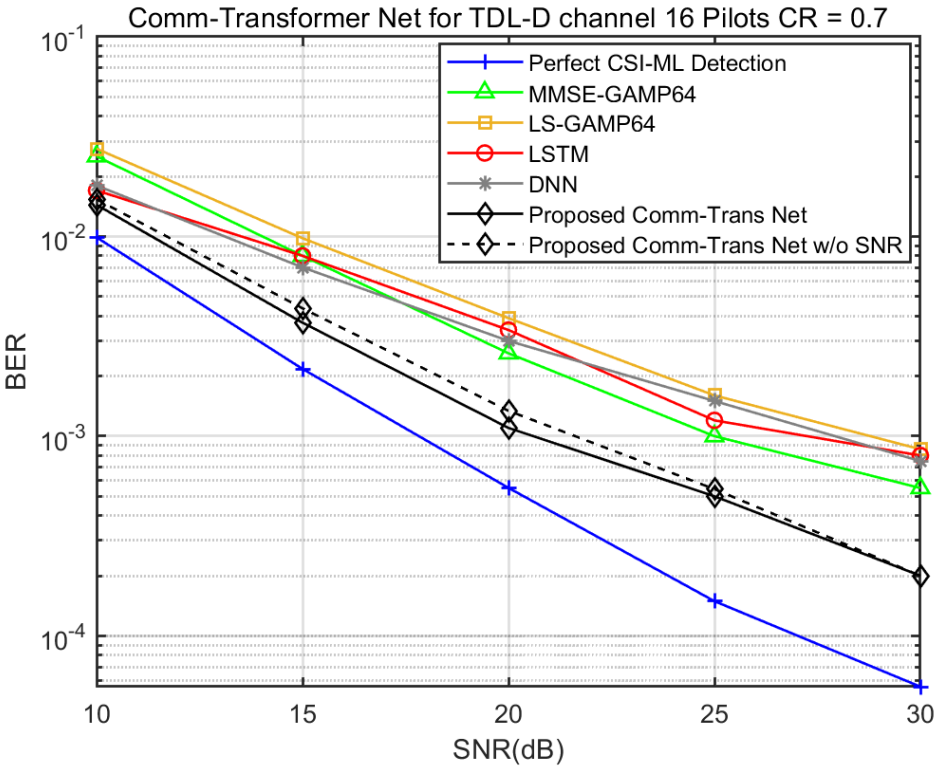



FIGURE 5.15: BER simulation results for TDL-D channel.

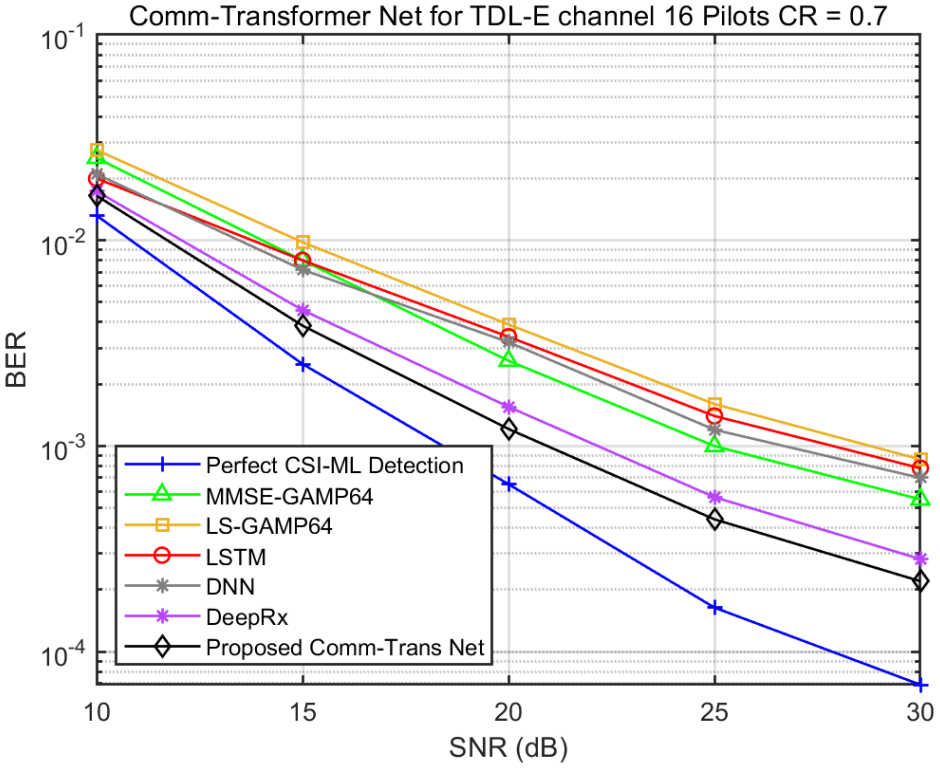



FIGURE 5.16: BER simulation results for TDL-E channel.

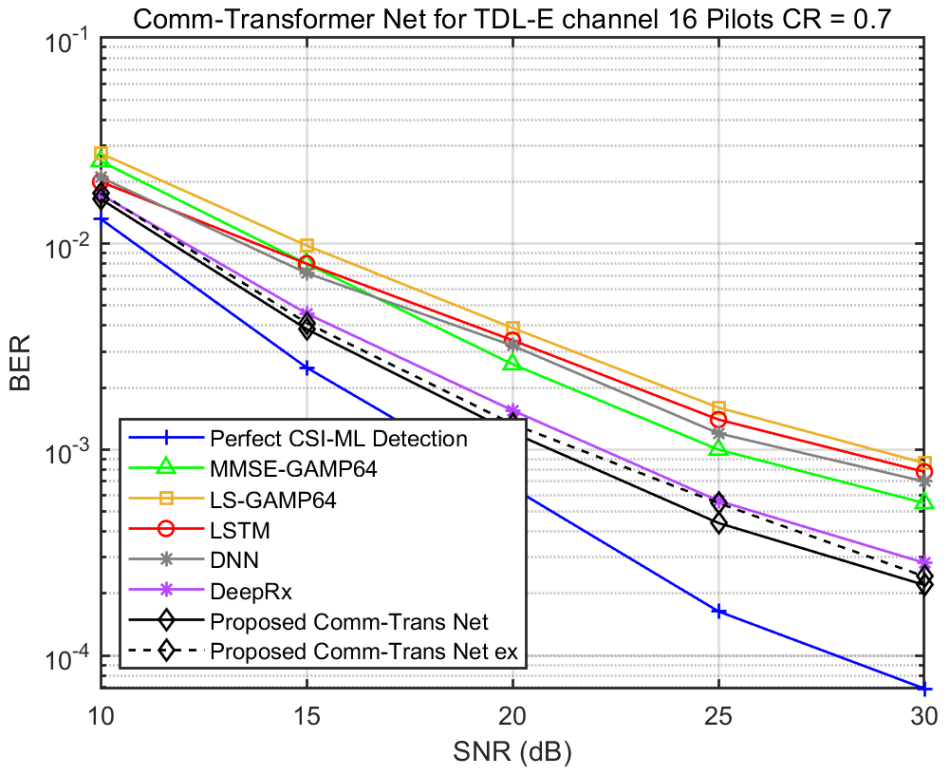



FIGURE 5.17: BER simulation results for cross domain validation of TDL-E channel.

5.17. Compared with the Proposed Comm-Trans Net curve which is trained with TDL-E domain knowledge, we observe a relatively close BER performance with a gap of 1.3 dB. This indicates that our proposed Comm-Trans Net demonstrates adaptability to cross-domain knowledge for different subtypes of TDL channels.

The proposed Comm-Trans Net receiver compensates for the deep feeding effect by relying on the attention mechanism and it handles the inter-carrier interference by the 1D U-shape feature extractor. Moreover, for the decision and classification of the bits, GRUs provide a vital role in handling the sequence relationship for the received signal.

### 5.5.3 Impact of different pilot conditions

Moreover, we also conduct the BER simulations for our proposed Comm-Trans Net, conventional methods, and state-of-the-art deep learning methods such as DNN and LSTM receivers under different pilot conditions as shown in Table 5.2. The simulation results in Fig. 5.18 and 5.19 show that our proposed Comm-Trans Net receiver can outperform conventional methods and DL-based methods under different numbers of pilots.

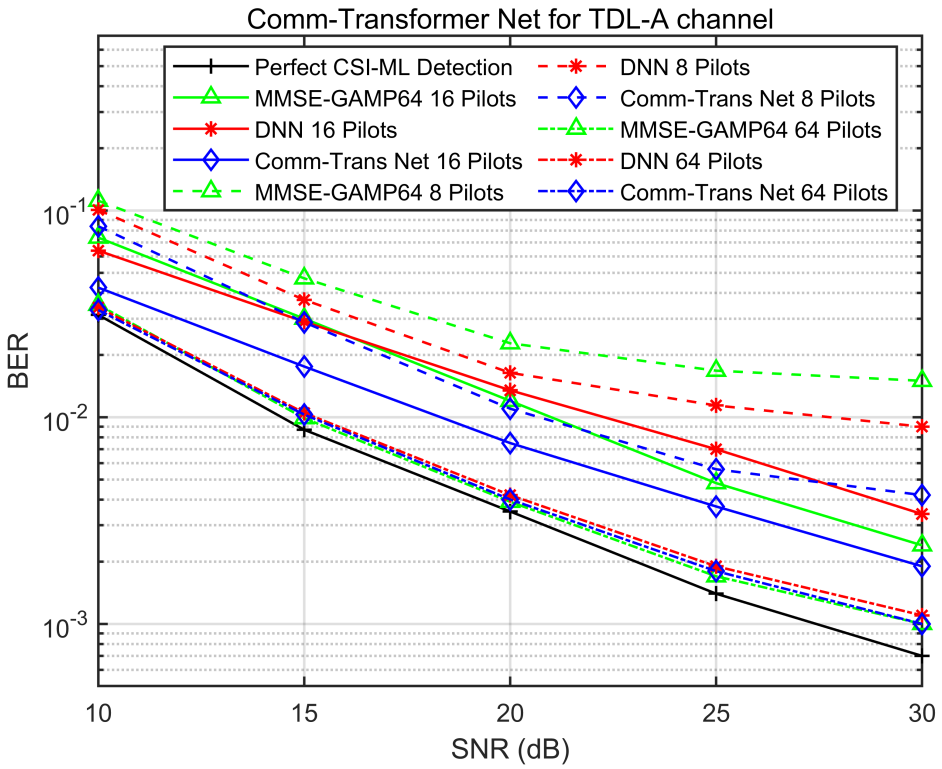



FIGURE 5.18: BER simulation results for TDL-A channel with different number of pilots.

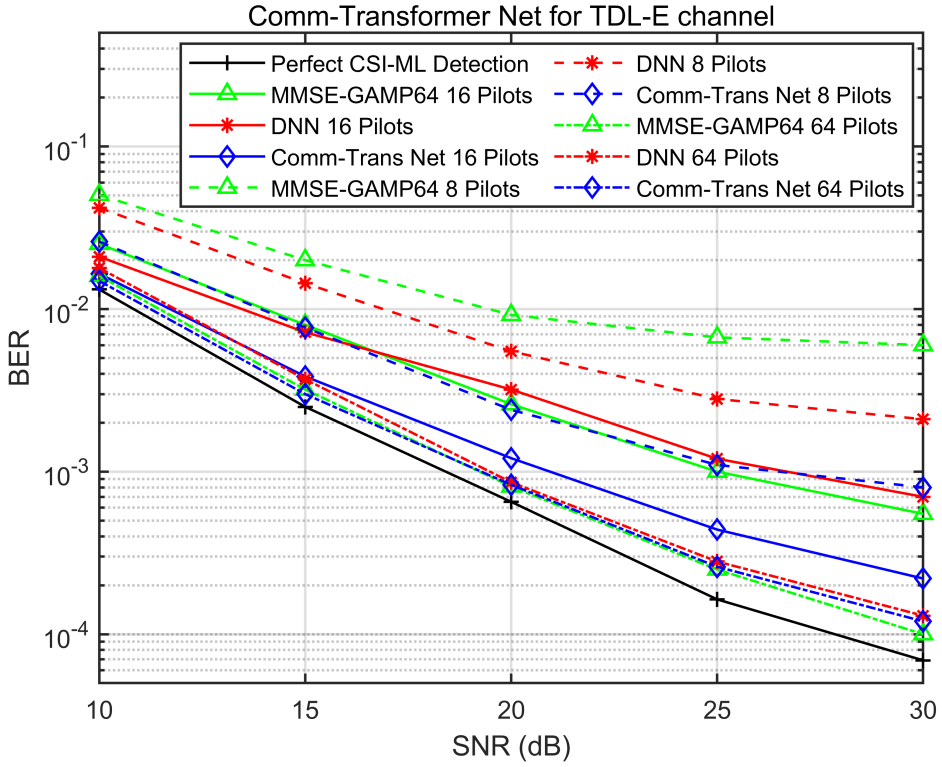



FIGURE 5.19: BER simulation results for TDL-E channel with different number of pilots.

### 5.5.4 Ablation Experiment

In this section, we present the results of two ablation experiments aimed at understanding the impact of different network structures on the BER performance of the Comm-Trans Net model in practical 5G TDL communication scenarios. The experiments focus on fine-tuning specific components of the network architecture and evaluating their influence on the BER at different SNR levels.

In the first ablation experiment, we investigate the effect of different kernel sizes on the BER performance of the Comm-Trans Net model. Four configurations are compared, which include using 1x1 1x3, 1x5, and 1x7 kernels independently. As depicted in Fig. 5.20, the choice of kernel size significantly influences the BER performance. Smaller kernel sizes, such as 1x3, excel at capturing local features and exhibit lower BER values. Conversely, larger kernel sizes such as 1x5 and 1x7 capture broader spatial information, resulting in slightly higher BER values.

The second ablation experiment focuses on the impact of varying the number of convolutional blocks on the BER performance of the Comm-Trans Net model. Four configurations are compared, with the number of convolutional blocks set to 2, 3, 4, and 5, while keeping the kernel size fixed at 1x3. As illustrated in Fig. 5.21, increasing the number of convolutional blocks generally lead to improved BER performance, suggesting enhanced channel state information estimation. However, beyond a certain number of blocks, the performance gains become less significant, indicating a trade-off between model complexity and BER improvement.

The ablation experiments provide valuable insights into the influence of different network components on the BER performance of the Comm-Trans Net model. Smaller kernel sizes favor capturing local details, while larger kernel sizes facilitate the extraction of global spatial information. The combined use of multiple kernel sizes achieves a balance between the two. Moreover, increasing the number of convolutional blocks enhances BER performance, but this improvement reaches a saturation point with excessive blocks.

These findings offer essential guidelines for fine-tuning the Comm-Trans Net model and customizing it for various communication scenarios. By making informed decisions about network architecture and hyperparameters, we can optimize the model's BER performance and achieve accurate channel state information estimation in practical 5G TDL communication systems.

By performing this ablation experiment, we aim to gain insights into the contribution and importance of different structural components in the network. This information can help us understand which parts of the architecture are crucial for achieving optimal performance and guide further fine-tuning or architectural improvements. To ensure the validity of the experiment, it is important to carefully define evaluation metrics and perform multiple runs to account for variability. Additionally, appropriate control experiments should be conducted to compare against randomly modified or shuffled networks to establish the significance of the observed performance changes.

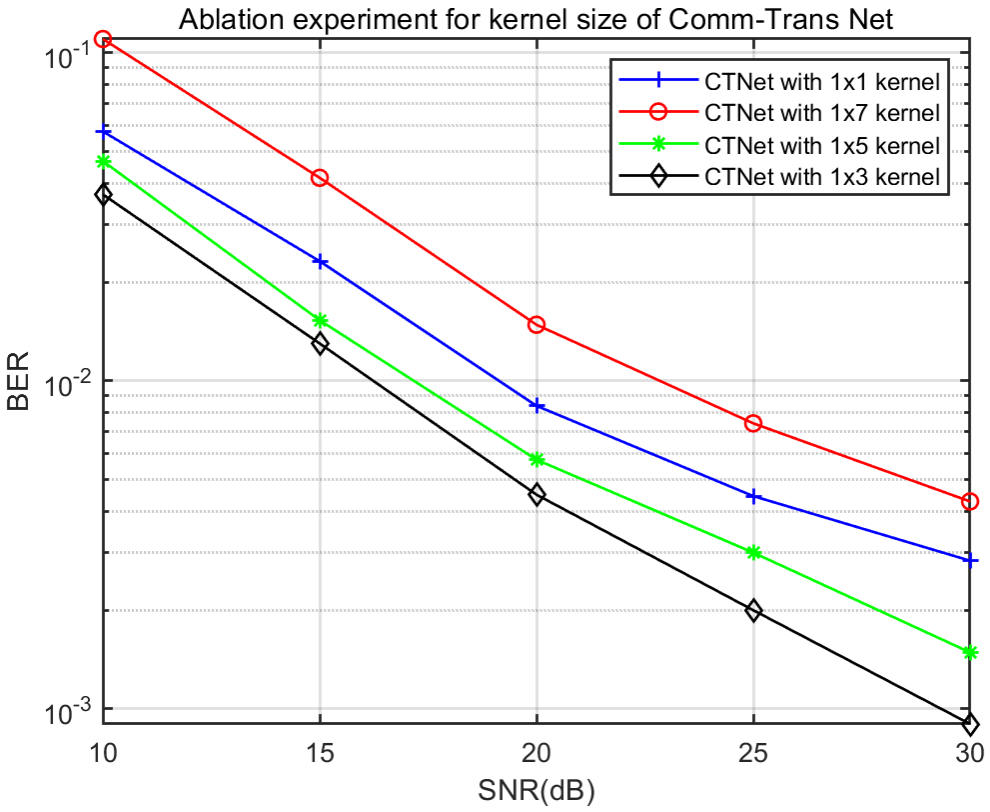



FIGURE 5.20: Ablation experiment of kernel size.

The results of the ablation experiment can provide valuable insights into the network structure, identifying critical components and potentially highlighting areas where further optimization or modification may be beneficial.

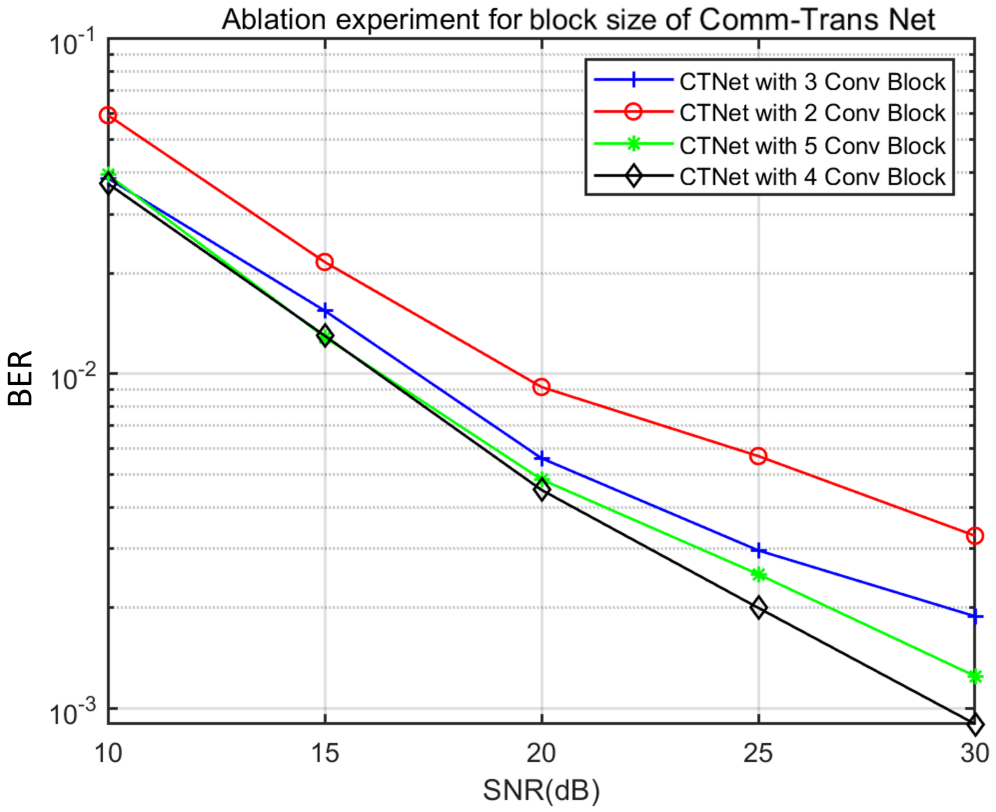



FIGURE 5.21: Ablation experiment of block size.

### 5.5.5 Comparisons of Online Testing Computational Complexity

In this section, we compare the computational complexities of LS-GAMP and MMSE-GAMP algorithms with the step size $N_x = 64$, DL-based receivers such as LSTM, DNN algorithms, and the proposed Comm-Trans Net receiver. For the LS-GAMP and MMSE-GAMP methods, the typical number of iterations required is set to $t_{max} = 5$ [31]. We calculate the online testing computational complexity for each batch. Table 5.4 displays the required number of multiplications for different algorithms. Our proposed Comm-Trans Net receiver exhibits lower computational complexity compared to the conventional LS-GAMP and MMSE-GAMP algorithms with the step size $N_x = 64$ [51]. Furthermore, our proposed Comm-Trans Net receiver can achieve a BER performance gain of approximately 3 dB when compared with the conventional MMSE-GAMP algorithm for $N_x = 64$.

TABLE 5.4: Comparisons of Online Testing Computation Complexities

| **Algorithm** | **Number of multiplications** | **Memory** |
|---|---|---|
| LS-GAMP 64 [51] | $3.78 \times 10^8$ | – |
| MMSE-GAMP 64 [51] | $4.27 \times 10^8$ | – |
| LSTM Receiver [49] | $2.7 \times 10^6$ | 0.9 MBytes |
| DNN Receiver [5] | $1.13 \times 10^6$ | 0.6 MBytes |
| DeepRx Receiver [50] | $3.85 \times 10^7$ | 4.6 MBytes |
| Proposed Comm-Trans Net | $8.64 \times 10^6$ | 3.6 MBytes |

## 5.6 Conclusion

In this chapter, a robust Comm-Trans Net-based receiver has been proposed to handle the 5G doubly selective fading channel with insufficient pilots. Our proposed Comm-Trans Net receiver combines the attention mechanism for channel positional encoding, U-shape CNN feature extractor, and GRU for sequence relationships between received signals. Simulation results have shown that our proposed Comm-Trans Net outperforms conventional methods such as LS-GMAP, MMSE-GAMP, and state-of-the-art deep learning receivers such as DNN and LSTM. Furthermore, we have conducted computational complexity analysis, and it has been shown that Comm-Trans Net requires less computational complexity than the MMSE-GAMP methods. Based on the computational complexity analysis, the attention mechanism for channel positional encoding is a cost-efficient way to improve the QoS. Also, the robustness of our proposed Comm-Trans Net is validated for different sub-type of TDL channels for different 5G communication scenarios.

In the next chapter, we delve into the exciting realm of "Transmission Beyond Bits" by exploring a novel approach based on the Diffusion Model. Traditional communication systems primarily focus on the transmission and reception of binary data (bits), limiting the achievable transmission rates and spectral efficiency. However, with the ever-growing demand for higher data rates, low latency, and massive connectivity in modern communication networks, there is a need to explore innovative methods that transcend the binary data paradigm. We investigate the theory and practical implementation of the diffusion model for wireless communication systems. We analyze how continuous data can be modulated, transmitted, and received in a wireless channel, taking advantage of the unique properties of diffusion to achieve superior performance.

# Chapter 6

# Generative Semantic Communication System: Transmission Beyond Bits with Diffusion Model

## 6.1 Introduction

In this chapter, we build upon previous research by studying the application of deep learning-based semantic communication in the context of source coding for 5G communication and beyond. The motivation for this research stems from the need to address the challenges of efficient data compression and transmission in communication systems.

Previous works have laid the foundation for DL-based approaches in various aspects of communication systems, including channel estimation, equalization, and joint detection. Drawing upon this lineage, we now delve into the specific area of source coding. By leveraging the power of DL algorithms and the insights gained from

previous DL-based techniques, we aim to enhance the source coding process in terms of higher transmission rates and improved privacy.

The focus of this chapter is on developing and exploring DL-based techniques that can effectively encode and decode data while preserving its semantic information. By incorporating semantic meaning or context into the encoding process, we aim to achieve a more efficient and effective representation of the data. This, in turn, can lead to higher transmission rates and improved privacy in future communication systems.

By building upon the lineage of DL-based research and focusing on the application of semantic encoding communication, this chapter contributes to the broader understanding and advancement of DL techniques in the physical layer of wireless communication systems. It explores new possibilities for source coding and emphasizes the potential benefits that DL-based semantic encoding communication can bring to future communication systems.

Semantic communication is a novel paradigm that aims to transmit information beyond the symbols or physical signals themselves and instead focuses on the meaning or semantics of the information being transmitted. In traditional communication systems, the emphasis is on encoding and transmitting symbols efficiently [43]. However, as communication systems become increasingly complex and diverse, there is a growing need for more intelligent and efficient communication techniques to handle the vast amount of data and diverse applications used for today and future. Semantic communication is one of the techniques that promises to address this need [90].

Semantic communication involves encoding and transmitting only the semantic information required to perform a specific task or achieve a specific goal rather than transmitting all available data [91]. This approach can significantly reduce data traffic and increase communication efficiency, eliminating the need to transmit large amounts of unnecessary data. Semantic communication systems are designed

to facilitate communication between individuals or machines by leveraging shared semantic models or ontologies [92], [93].

Recently, with the development of deep learning techniques and computational capacity, DL-aided physical layer communication systems have achieved creditable performance [81], [94]. DL techniques also have drawn tremendous attention to semantic communication systems. In [95], [96], the proposed semantic communication system considers various factors that influence the spread of information, such as oriented tasks, features of data, and the content of the information being transmitted. To further accelerate the progress of artificial intelligence beyond fifth generation (AI-B5G) evolution, the generative semantic communication (GSC) system is one of the possible solutions.

GSC system combines the AI generative content (AIGC) and semantic communication system, which could further save the spectrum resource and provide a cost-efficient solution under specific communication scenarios [97], [98], e.g., unmanned aerial vehicle (UAV) communication, device-to-device vehicular networks, which have limited spectrum resource. To utilize the GSC system, we propose the diffusion model-based GSC (DM-GSC) system, which relies on mathematical models of the forward diffusion process and the reverse denoise process to simulate and analyze the spread of semantic information in communication systems. The proposed DM-GSC system aims to accurately model and understand the flow of semantic information through the transmission of the physical communication channel and generate the semantic content of the semantic information being transmitted.

The proposed DM-GSC system takes advantage of the diffusion model, originally designed for AIGC problems for high-quality synthesis results [99, 100]. The source information is considered as latent variables by using parameterized Markov chain process in our proposed semantic encoder and decoder. The diffusion process semantic encoder is a Markov chain process that gradually adds noise to the source information, and the reverse denoise semantic decoder is the process of gradually removing the predicted noise. The training process of the DM-GSC system involves

optimizing the parameters of the diffusion model so that it accurately predicts the generative content of the semantic information in the communication system.

The contribution of this chapter can be summarized as follows. We investigate the generative semantic communication system by applying the diffusion mechanism to physical layer communications. The diffusion model is adapted to the GSC system as the semantic encoder and decoder using the forward diffusion process and reverse denoise process under conditional tokenized semantic features. Based on the tokenized semantic features and the trained diffusion model, the proposed DM-GSC system achieves relatively high performance for predicting the generative information. Furthermore, the robustness of the proposed DM-GSC system to different physical channel conditions is studied.

## 6.2 System Model

The communication system depicted in Fig. 6.1 involves the transmission of source information $\mathbf{s}$ (e.g., images), which is first processed by the semantic encoder $E_{se}(\cdot)$ to extract the semantic information, which is considered as the tokenized tag. After this, the tokenized tag and diffusion noise distribution parameters are processed by the channel encoder $C_h(\cdot)$ which consists of the analog-to-digital converter (ADC) and modulation scheme. Since the source information will be sent through the physical layer communication system, the quantization layer achieved by ADC and quadrature phase-shift keying (QPSK) are applied to our system model, which can be formulated as

$$\mathbf{t} = C_h(E_{se}(\mathbf{s})). \tag{6.1}$$

The transmitter comprises a semantic encoder and a channel encoder, as shown in Fig. 6.1, to extract the semantic information and ensure successful transmission over the physical channel. The mapped symbol stream $\mathbf{s}$ is transmitted through the air with a multi-path fading channel and additive white Gaussian noise (AWGN), which can be presented as

$$\mathbf{r} = \mathbf{t} * \mathbf{h} + \alpha, \tag{6.2}$$

where $*$ denotes the convolution operation, $\mathbf{h}$ is the multi-path channel response, and $\alpha$ denotes the AWGN with power $\sigma_\alpha^2$.

At the receiver, the received signal, $\mathbf{r}$, is processed by the channel decoder and semantic decoder to predict the generative information $G(\mathbf{s})$. The semantic receiver consists of two main parts, channel decoder ${C_h}^{-1}(\cdot)$ and semantic decoder ${E_{se}}^{-1}(\cdot)$. To handle the physical channel interference, the channel decoder integrates channel estimation, equalization, demodulation, and dequantization. The semantic decoder is based on a reverse denoise process and transmitted tokenized tag (known as the conditional token) to generate the content as $G(\mathbf{s})$. Thus, the whole receiver can be presented as

$$G(\mathbf{s}) = {E_{se}}^{-1}({C_h}^{-1}(\mathbf{r})). \tag{6.3}$$

## 6.3 Proposed Diffusion Model-Based Receiver

### 6.3.1 The Proposed DM-GSC Structure

Overall, the proposed DM-GSC communication system is presented in Fig. 6.1, which consists of two major components, diffusion process semantic encoder and reverse denoise semantic decoder. In general, the source information $\mathbf{s}$ is considered to add a certain noise distribution in the diffusion process, and a trained DM neural network is used for denoising during the reverse process, which is shown in Fig. 6.2.

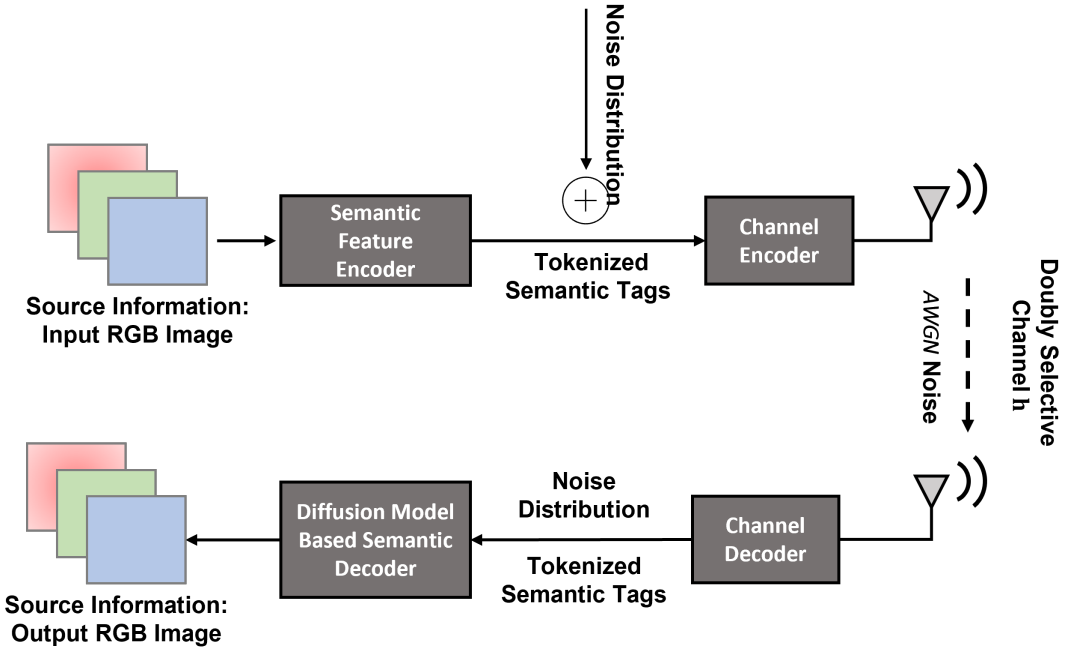



FIGURE 6.1: The structure of the proposed DM-GSC system.

### 6.3.2 Diffusion Process Semantic Encoder

First, the source information $\mathbf{s}$ is treated as a latent variable model of the form $p_\theta(\mathbf{s}_0) = \int p_\theta(\mathbf{s}_{0:T})d\mathbf{s}_{1:T}$. The latent variables $\mathbf{s}_{1:T}$, which consist of $\mathbf{s}_1, \cdots, \mathbf{s}_T$, have the same dimension as that of the source information. Additionally, the Markov chain process is applied to the forward diffusion process, which considers

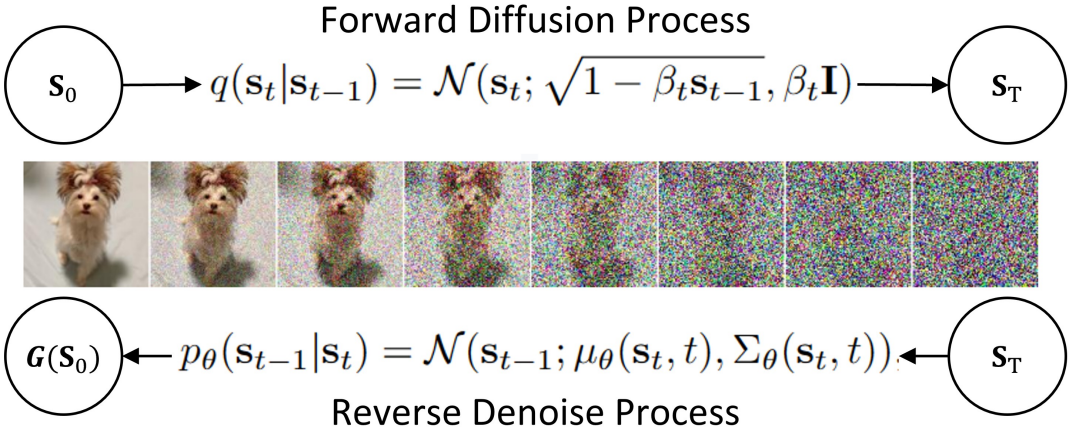



FIGURE 6.2: The example for forward diffusion process and reverse denoise process.

the approximate posterior $q(\mathbf{s}_{1:T}|\mathbf{s}_0)$. This process gradually adds Gaussian noise to the data based on a variance schedule $\beta_t$. This process can be formulated as follows:

$$q(\mathbf{s}_{1:T}|x_0) = \prod_{t=1}^{T} q(\mathbf{s}_t|\mathbf{s}_{t-1}), \tag{6.4}$$

$$q(\mathbf{s}_t|\mathbf{s}_{t-1}) = \mathcal{N}(\mathbf{s}_t; \sqrt{1-\beta_t}\mathbf{s}_{t-1}, \beta_t\mathbf{I}), \tag{6.5}$$

where $\mathbf{I}$ denotes the identity matrix, and $\mathcal{N}(\cdot)$ denotes the Gaussian distribution.

Second, we define the joint distribution $p_\theta(\mathbf{s}_{0:T})$ as a Markov chain with learned Gaussian process starting at $p(\mathbf{s}_T) = \mathcal{N}(\mathbf{s}_T; \mathbf{0}, \mathbf{I})$, which can be formulated as

$$p_\theta(\mathbf{s}_{0:T}) = p(\mathbf{s}_T) \prod_{t=1}^{T} p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t), \tag{6.6}$$

$$p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t) = \mathcal{N}(\mathbf{s}_{t-1}; \mu_\theta(\mathbf{s}_t, t), \Sigma_\theta(\mathbf{s}_t, t)), \tag{6.7}$$

where $p_\theta(\mathbf{s}_{0:T})$ represents the joint probability distribution over a sequence of random variables $\mathbf{s}_{0:T}$. Note that $p(\mathbf{s}_T)$ represents the probability of the final random variable $\mathbf{s}_T$. (6.6) is a product over time steps $t$ of conditional probabilities $p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t)$, which represent the probability of the random variable $\mathbf{s}_{t-1}$ given the

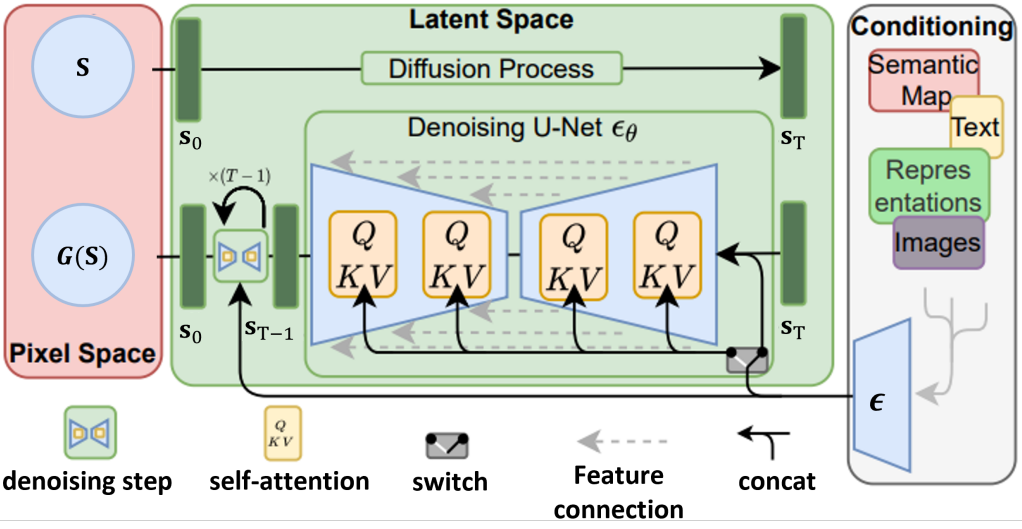



FIGURE 6.3: The network structure of the proposed DM-GSC receiver.

value of $\mathbf{s}_t$. These conditional probabilities are assumed to be Gaussian, with mean $\mu_\theta(\mathbf{s}_t, t)$ and covariance matrix $\Sigma_\theta(\mathbf{s}_t, t)$, both of which are functions of the input $\mathbf{s}_t$ and the time step $t$.

One way to study the forward process variances $\beta_t$ is through reparameterization [101], or they can be set as constant hyperparameters. The choice of Gaussian conditionals in $p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t)$ ensures the expressiveness of the reverse process with the same functional form. The forward process has a notable property of allowing closed-form sampling of $\mathbf{s}_t$ at any time step $t$, which can be achieved by defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. Thus, we can formulate the relationship between $\mathbf{s}_t$ and $\mathbf{s}_0$ as

$$q(\mathbf{s}_t|\mathbf{s}_0) = \mathcal{N}\left(\mathbf{s}_t; \sqrt{\bar{\alpha}_t}\mathbf{s}_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \tag{6.8}$$

where $\beta_t$ is the constant of the unnormalized distribution at time $t$, and $\bar{\alpha}_t$ is the normalized constant that ensures that $q_\theta(\mathbf{s}_t|\mathbf{s}_{1:t})$ follows the same probability distribution. Thus, the ultimate goal for training is performed by optimizing the variational bound on negative log-likelihood based on [100] as follows

$$\mathcal{L} = \mathbb{E}_q \left[ -\log p(\mathbf{s}_T) - \sum_{t=1}^{T-1} \log \frac{p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t)}{q(\mathbf{s}_t|\mathbf{s}_{t-1})} \right], \tag{6.9}$$

which can be further represented in Kullback–Leibler divergence form follow the instruction [100] as

$$\begin{aligned} &\mathcal{L} = \mathbb{D}_{KL}(q(\mathbf{s}_T|\mathbf{s}_0)||p(\mathbf{s}_T)) + \\ &\sum_{t=1}^{T-1} \mathbb{D}_{KL}(q(\mathbf{s}_{t-1}|\mathbf{s}_t, \mathbf{s}_0)||p_\theta(\mathbf{s}_{t-1}|\mathbf{s}_t)) - \log p_\theta(\mathbf{s}_0|\mathbf{s}_1), \end{aligned} \tag{6.10}$$

where the first term can be approximately considered as zero since $q(\mathbf{s}_T|\mathbf{s}_0)$ and $p(\mathbf{s}_T)$ are close to $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

### 6.3.3 Reverse Denoise Semantic Decoder

In this section, the reverse denoise semantic decoder is discussed in detail. Equation (6.9) can be reparameterized and optimized for the training procedure during the reverse denoise process. The term $p_\theta(\mathbf{s}_{t-1} \mid \mathbf{s}_t) = \mathcal{N}(\mathbf{s}_{t-1}; \mu_\theta(\mathbf{s}_t, t), \Sigma_\theta(\mathbf{s}_t, t))$ for $1 < t \leq T$ is first studied. We define the untrained time-dependent constants by setting $\Sigma_\theta(\mathbf{s}_t, t) = \sigma_t^2 \mathbf{I}$. In our experiments, both choices of $\sigma_t^2$, namely $\beta_t$ and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, produce similar results. The former is optimal when $\mathbf{s}_0$ follows a normal distribution with zero mean and unit variance, while the latter is optimal when $\mathbf{s}_0$ is set deterministically to a single point. These two extreme options represent upper and lower bounds on the entropy of the reverse process for data with coordinate-wise unit variance.

Furthermore, we propose a specific parameterization to represent the mean $\mu_\theta(\mathbf{s}_t, t)$ based on our analysis of $\mathcal{L}$. With $p_\theta(\mathbf{s}_{t-1} \mid \mathbf{s}_t) = \mathcal{N}(\mathbf{s}_{t-1}; \mu_\theta(\mathbf{s}_t, t), \sigma_t^2 \mathbf{I})$, we can reformulate $\mathcal{L}$ as

$$\begin{aligned} \mathcal{L} &= \mathbb{D}_{KL}(\mathcal{N}(\mathbf{s}_{t-1}; \tilde{\mu}_t(\mathbf{s}_t, \mathbf{s}_0), \sigma_t^2 \mathbf{I}) || \mathcal{N}(\mathbf{s}_{t-1}; \mu_\theta(\mathbf{s}_t, t), \sigma_t^2 \mathbf{I})) \\ &= \mathbb{E}_q \left[ \frac{1}{2\sigma_t^2} \| \tilde{\mu}_t(\mathbf{s}_t, \mathbf{s}_0) - \mu_\theta(\mathbf{s}_t, t) \|^2 \right] + C, \end{aligned} \tag{6.11}$$

where $\| \cdot \|^2$ denotes the Euclidean norm, the learnable parameter $\tilde{\mu}_t$ is the predicted mean at time $t$ given the current observation $\mathbf{s}_t$ and the initial observation $\mathbf{s}_0$. Furthermore, $\tilde{\mu}_t(\mathbf{s}_t, \mathbf{s}_0)$ can be defined as $\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{s_0} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}}\mathbf{s_t}$. Note that $C$ is defined as a constant.

We reparameterized $\mathbf{s}_t(\mathbf{s}_0, \epsilon) = \sqrt{\bar{\alpha}_t}\mathbf{s}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ for $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\bar{\alpha}_t$ is defined in (6.8). This can be derived by applying the variational bound on negative log-likelihood as follows

$$\begin{aligned} &\mathcal{L} - C = \mathbb{E}_{\mathbf{s}_0, \epsilon} \\ &\left[ \frac{1}{2\sigma_t^2} \| \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{s}_t(\mathbf{s}_0, \epsilon) - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon \right) - \mu_\theta \left( \mathbf{s}_t(\mathbf{s}_0, \epsilon), t \right) \|^2 \right]. \end{aligned} \tag{6.12}$$

By further replacing the learnable parameters $\mu_\theta$ in (6.12), we consider to parameterize $\mu_\theta(\mathbf{s}_t, t)$, which can be presented as

$$\begin{aligned}\mu_\theta(\mathbf{s}_t, t) &= \tilde{\mu}_t\left(\mathbf{s}_t, \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{s}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{s}_t))\right) \\ &= \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{s}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{s}_t, t)\right).\end{aligned} \tag{6.13}$$

Thus, the variational bound on the negative log-likelihood function can be expressed as

$$\mathbb{E}_{\mathbf{s}_0,\epsilon}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\|\epsilon - \epsilon_\theta\left(\sqrt{\alpha_t}\mathbf{s}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\|^2\right], \tag{6.14}$$

where $\epsilon_\theta$ is the learned noise scaled by the diffusion model, which achieves the revere denoise process.

### 6.3.4 Problem Formulation

The generation of semantic information $G(\mathbf{s})$ from source information $\mathbf{s_0}$ can be converted to minimize the variational bound from (6.14) by estimation of the noise $\epsilon_\theta$. Thus, the loss function for the DM-GSC system can be formulated as

$$\arg\min_{\epsilon_\theta} L(\epsilon_\theta) = \mathbb{E}_{t,\mathbf{s}_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\mathbf{s}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2\right], \tag{6.15}$$

where $\mathbb{E}_{t,\mathbf{s}_0,\epsilon}$ denotes the expectation over $t$, $\mathbf{s}_0$, and $\epsilon$.

## 6.4 Training Procedure and Implementation Setting

### 6.4.1 Forward diffusion semantic encoder

To ensure that the forward diffusion semantic encoder has a similar functional form and minimizes $\mathbb{D}_{KL}(q(\mathbf{s}_T|\mathbf{s}_0)||p(\mathbf{s}_T))$, we set the variance schedule $\beta_t$ as constants that increase linearly from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$. These constants were specifically selected to be small relative to the data scaled to the interval $[-1, 1]$.

### 6.4.2 Reverse denoise semantic decoder

For training the reverse denoise semantic decoder, we choose the U-Net backbone with the self-attention mechanism, as shown in Fig. 6.3. Network parameters are shared with time step $t$ during the Markov chain, and sinusoidal position embedding is applied to the self-attention with feature map resolution of $16 \times 16$.

To measure the quality of the generative information, the structural similarity index (SSIM) is used for measuring the similarity between two images [102]. SSIM index considers both the images' pixel values and structural information. The SSIM index assumes that the human visual system is highly sensitive to changes in structural information and less sensitive to changes in pixel values. It can be formulated as

$$\mathrm{SSIM}(\mathbf{s}, G(\mathbf{s})) = \frac{(2\mu_{\mathbf{s}}\mu_{G(\mathbf{s})} + c_1)(2\sigma_{\mathbf{s}G(\mathbf{s})} + c_2)}{(\mu_{\mathbf{s}}^2 + \mu_{G(\mathbf{s})}^2 + c_1)(\sigma_{\mathbf{s}}^2 + \sigma_{G(\mathbf{s})}^2 + c_2)}, \tag{6.16}$$

where $\mathbf{s}$ and $G(\mathbf{s})$ are source and generated images being compared, $\mu_{\mathbf{s}}$ and $\mu_{G(\mathbf{s})}$ are the means of the two images, $\sigma_{\mathbf{s}}$ and $\sigma_{G(\mathbf{s})}$ are the standard deviations of the two images, $\sigma_{\mathbf{s}G(\mathbf{s})}$ is the covariance between the two images, and $c_1$ and $c_2$ are constants to prevent division by zero.

The physical channel conditions are presented in Table 6.1. The multi-path Rayleigh time-varying fading and AWGN channels are considered for testing the robustness of the proposed DM-GSC system. Moreover, different levels of quantization are also considered to emulate the practical situation.

TABLE 6.1: Parameters for generating the channel conditions

| **Parameters** | **Value** |
|---|---|
| Number of multi-paths | 4 |
| Normalized path delays | [0, 0.001, 0.002, 0.003] |
| Average path gain (dB) | [0, -3, -6, -9] |
| Fading model | Rayleigh fading |
| Quantization depths | 8, 16 bits |

## 6.5 Numerical Results and Discussions

In this section, we present the simulation details for the proposed DM-GSC systems. We test our trained diffusion model for different semantic information categories and investigate robustness for different physical channel conditions. For each category, we calculate the SSIM metric to measure the quality of the generative information.

For implementation details, we train the semantic system model based on the pre-trained stable diffusion model version 1.5 [103], which saves tremendous computational resources. We used the LAION dataset [104], for training and testing the proposed model. For channel encoding and decoding, we choose quadrature phase-shift keying (QPSK) modulation to map the bit sequence and use different levels of bits for the quantization layer to convert the feature map and tokenized

tag for semantic information during transmission. The simulation for the DM-GSC system is based on the Python programming environment with PyTorch deep learning framework deploying with NVIDIA RTX 3090 GPUs support.

### 6.5.1 Simulation Results for Tokenzied Feature Map Through Different Channels

The simulation results compare the performance of different encoding schemes and channels in the presence of noise. Specifically, the DM-GSC simulations are run on the trained forward diffusion semantic encoder and reverse denoise decoder under different channels, such as Rayleigh fading and AWGN, with different quantization depths (8 and 16 bits). The results are presented in Fig. 6.4, which consider the mean squared error (MSE) metric for measuring the transmission loss between the transmitted and received tokenized feature map (also known as the semantic information).

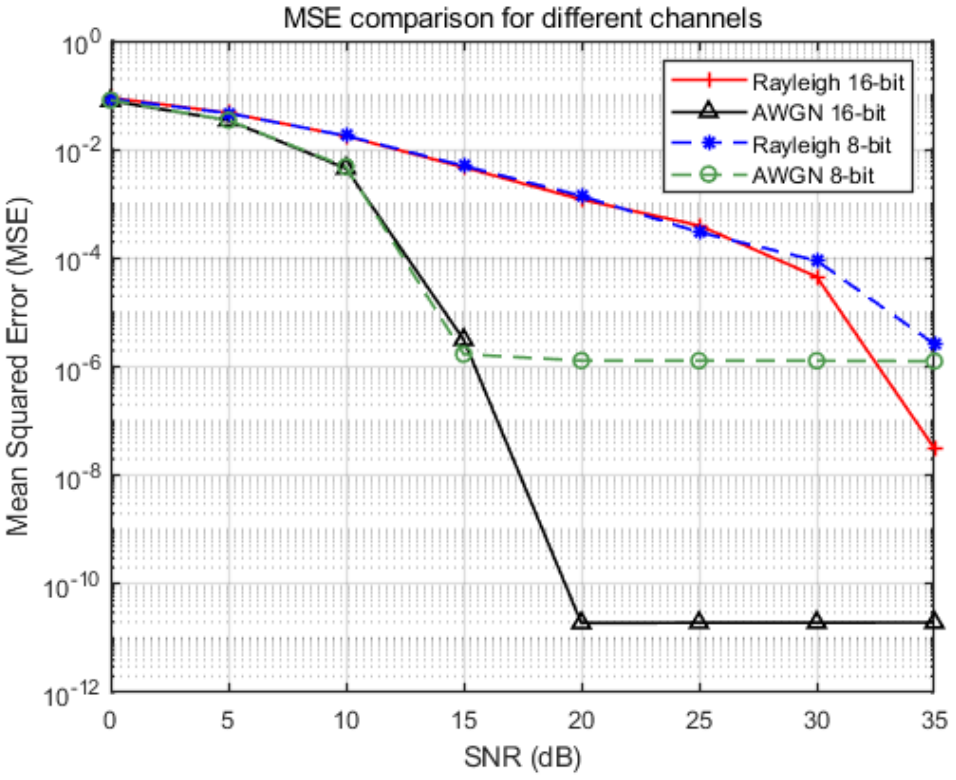



FIGURE 6.4: MSE results for the transmission loss under different channels.

The results show that increasing the SNR generally leads to better performance in both MSE for all the channel encoding schemes and conditions. Furthermore, the results indicate that increasing the quantization depth from 8 to 16 bits improves performance in terms of MSE since the error flow will significantly reduce.

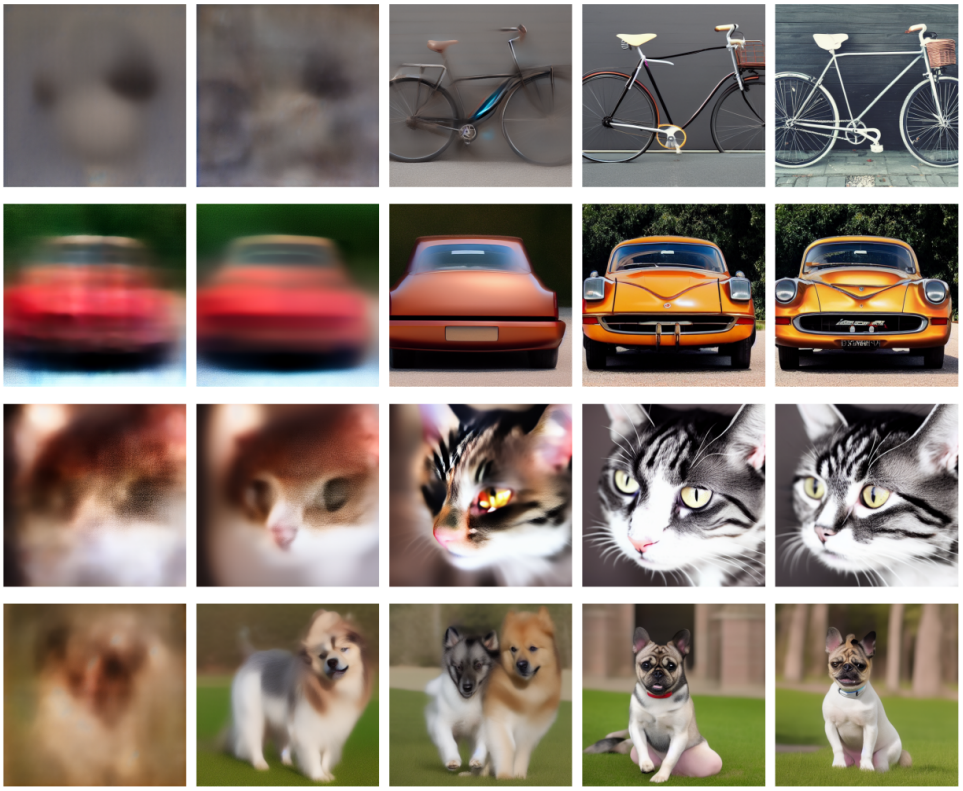


FIGURE 6.5: DM-GSC generative examples for the semantic receiver.

### 6.5.2 Generative Content Results for Diffusion Model-Based Semantic Communication

The forward diffusion process gradually adds noise to the source information to target the criterion noise level. The reverse denoise semantic decoder removes noise from the image by gradually predicting the noise and revealing the underlying structure of the image, which is shown in Fig. 6.5. The diffusion model that reverses the denoising process is a Markov chain process. It involves gradually removing noise from an image and then adding it back in to restore the original image according to the tokenized semantic conditions.

Considering the distortion of different physical channels, the system-level simulations are studied to investigate the SSIM performance and robustness of channel conditions. The results in Fig. 6.6 show that the SSIM values increase as the SNR level increases. This is because the transmission loss decreases as the SNR increases. Comparing different channels, we can observe that the Rayleigh fading channel gives lower SSIM values than the AWGN channel, indicating that the Rayleigh fading channel will bring more distortion to the semantic encoder and decoder. Moreover, the 16-bit quantization channels outperform the 8-bit quantization channels by having higher SSIM values since the 16-bit quantization causes less MSE loss of the physical layer transmission.

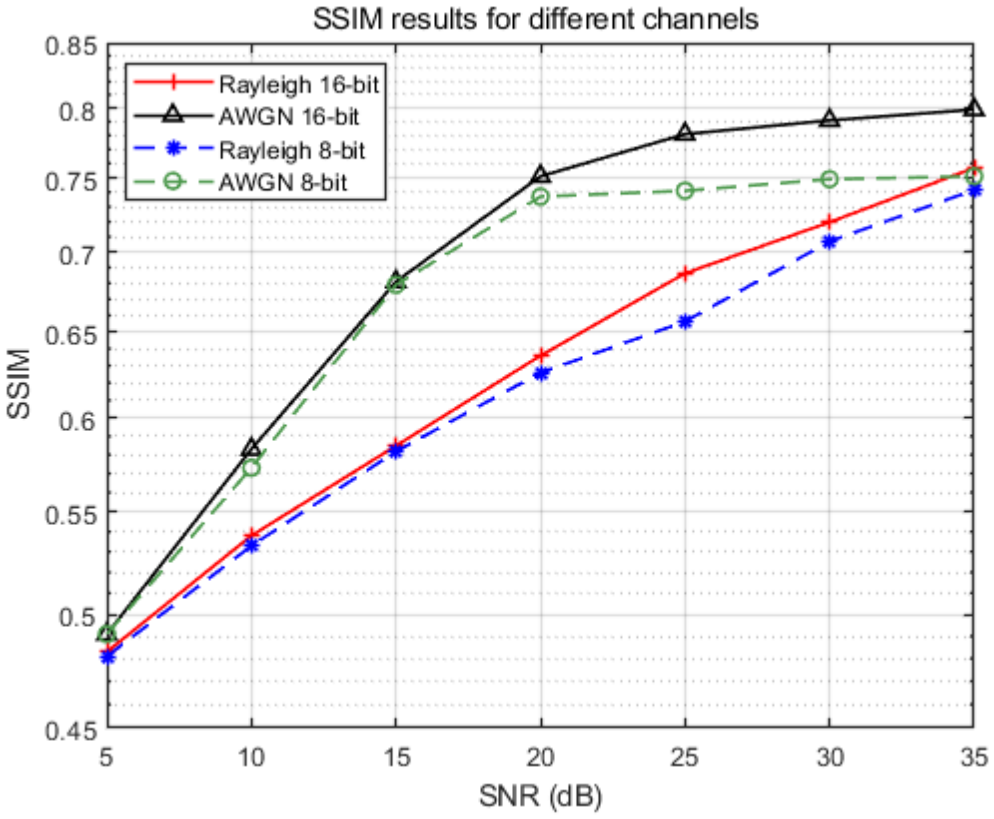



FIGURE 6.6: SSIM results for DM-GSC system under different channels.

## 6.6 Conclusion

In this chapter, we have proposed a novel semantic communication system based on the diffusion model. The proposed diffusion model-based generative semantic system aims to enhance the efficiency of artificial intelligence based beyond 5G wireless communication by leveraging the principles of the diffusion model and the semantic representation of source information. Forward diffusion semantic encoder and reverse denoise semantic decoder have been designed for the information through the Markov chain process. The experimental results demonstrate the performance of the proposed system in terms of SSIM for predicting the generative information. Overall, the proposed diffusion model-based semantic communication system has the potential for future B5G wireless communication scenarios, e.g., unmanned aerial vehicle communication, and device-to-device vehicular networks.

# Chapter 7

# Conclusion and Future Works

## 7.1 Conclusion

We focus on machine-learning solutions to problems in communication networks, across various layers and within a broad range of applications. The topics of interest include but are not limited to, machine learning, especially deep learning, for signal detection, 5G channel modeling, deep learning based joint detection, compressed CSI feedback, non-linear distortion recovery, deep learning based channel estimation, and transmission beyond bits. The objective of my Ph.D. is to bring together state-of-the-art research results and industrial applications of machine learning for wireless communication systems. Based on our experiments and research studies, DL-based algorithms can achieve outstanding performance compared with state-of-the-art conventional methods in some particular communication scenarios.

Firstly, in the NOMA system with imperfect SIC conditions, our proposed DL-based algorithm significantly reduces the error floor under insufficient pilots and achieves around 4 dB gain in SER performance compared with the conventional MMSE-SIC method. Also, our proposed DL-based receiver is an end-to-end model,

which integrates the function of channel estimation, equalization, and demodulation. The proposed DL-based algorithm has less computational complexity as compared with the MMSE-SIC method since the matrix inversion operation is required for the MMSE-SIC method. Moreover, as compared with the existing methods, the proposed DL-based method is more robust against a smaller number of pilots.

Secondly, we investigate the autoencoder-decoder (AED)-based CSI feedback for massive MIMO communication systems. A novel CSI sensing and recovery mechanism has been proposed. AED performs well at low compression ratios and reduced time complexity. As tested under practical 5G channel response, the significance of DL-based application has more reliability. We believe that its reconstruction quality can be further improved by applying advanced DL technology.

Thirdly, we propose 1D-TRNet for the non-linearly distorted OFDM system. The clipped OFDM system with an out-of-band (OOB) filter is a critical problem because both clipping and OOB filters will bring in inter-carrier interference. Conventional GAMP algorithm has the problem of large computational complexity to reach the optimal performance, however, the performance of F-GAMP is sacrificed to reduce the computational complexity. Our proposed 1D-TRNet gives a new direction to use a DL-based receiver to solve this problem. Based on our experiments, our proposed 1D-TRNet-based receiver outperforms the conventional methods such as GAMP16 and F-GAMP, as well as the DNN receiver.

Fourthly, we propose Comm-Trans Net for severe doubly selective fading channels under 5G communication scenarios. The Comm-Trans Net backbone conducts with attention mechanism and U-Net structure for channel positional encoding and feature extraction. The proposed Comm-Trans Net will focus on the deep-fading subcarriers caused by doubly selective fading channels, e.g., TDL channels. Moreover, the robustness of the proposed Comm-Trans Net is validated for different communication scenarios, from TDL-A to TDL-E. Our proposed Comm-Trans Net achieves around 3 dB BER performance gain compared to conventional methods and state-of-the-art DL methods.

Lastly, we investigate the diffusion model based semantic communication system. The proposed diffusion model-based generative semantic system aims to enhance the efficiency of artificial intelligence based beyond 5G wireless communication by leveraging the principles of the diffusion model and the semantic representation of source information. Forward diffusion semantic encoder and reverse denoise semantic decoder have been designed for retrieving the information through the Markov chain process. Thus, noise schedule, trained DM score, and conditional embedding are required for semantic communication, which is more efficient compared to transmitting the full source information.

In conclusion, from different applications in physical layer communication problems, DL has a larger potential to solve critical problems. To realize this vision of intelligent processing and operation, there is a need to integrate deep learning, known to be the vessel that carries artificial intelligence, into the design, planning, and optimization of future communication networks. Particularly, the emerging framework of deep learning can be a key enabler for intelligent processing in a broad range of scenarios. Modern machine learning techniques provide ample opportunities to enable intelligent communication designs while addressing various problems ranging from end-to-end communication design, channel estimation, CSI feedback, channel equalization, and semantic communication systems, etc.

## 7.2 Future Work

Based on the research works and results, we have several visions for future works.

Firstly, for multi-user communication systems, several constraints are not considered yet. As an example, dynamic power allocation is still a critical problem for the multi-user system. Recent research works show that deep reinforcement learning (DRL) could provide optimal performance in the ideal condition [105], thus it is possible to combine the DRL strategy and DL-based receiver to further improve

the overall performance of the multi-user system. In this case, we aim to design a model to achieve efficient resource allocation in the multi-user system.

Secondly, non-linear distortion in the wireless communication system is also a critical problem. For example, the impulse noise and analog-to-digital converter will bring in non-linear distortion. To solve these problems, the DL-based method can serve an important role. Based on our experience with the clipping effect, the design of the feature extractor is a crucial step. Moreover, the emergence of a transformer network structure also brings in the opportunity to enhance the network's ability. The attention mechanism breaks the limitation of recurrent models. Aligning the positions to steps in computation time, conventional recurrent models generate a sequence of hidden states $h_t$, as a function of the previous hidden state $h_{t-1}$ and the input for position $t$. This inherently sequential nature precludes parallelization within training examples, which becomes critical at longer sequence lengths, as memory constraints limit batching across examples.

Thirdly, solving the ISI caused by the high Doppler shift effect is also a potential research direction. In the future 5G communication system, the mobile speed will become much faster, which brings a high Doppler shift. Solving this problem has significance in the practical 5G communication scenarios. To handle severe ISI, the feature extractor becomes a crucial point. Based on 1D-TRNet in Chapter 4, low-level to high-level features carry different information between a different range of the received signals. This idea has the potential to solve the ISI problem since the Doppler shift also will interfere uncertain range of the transmitted signal.

Moreover, deep learning techniques also show potential for multi-cell wireless communication problems, such as interference management, resource allocation, and handover management [106]. CNNs and recurrent neural networks (RNNs) help predict and mitigate interference by learning complex patterns between cells. Reinforcement learning algorithms, like deep Q-networks (DQNs), optimize resource allocation based on current network states, enabling efficient resource distribution.

RNNs and LSTM networks predict user mobility patterns to enhance handover management, reducing call drops and improving network performance.

To further improve the performance of the DL-based receiver, a DL-based error correcting autoencoder-decoder can be cascaded. Additionally, the practical implementation of the DL-based receiver can be explored through ensemble learning [107], [108] and low-rank adaptation [109] techniques to further enhance its robustness in complicated wireless communication systems. These are some potential future works to be investigated.

# List of Author's Awards, Patents, and Publications[1]

## Journal Articles

- **Yihang Xie**, Kah Chan Teh, Alex Chichung Kot, "Deep learning-based joint detection for OFDM-NOMA scheme," IEEE Communications Letters, 2021, 25(8): 2609-2613.
- **Yihang Xie**, Xiaobei Liu, Kah Chan Teh, Yong Liang Guan, "Robust Deep learning based end-to-end receiver for OFDM system with non-linear distortion," IEEE Communications Letters, 2022, 26(2): 340-344.
- **Yihang Xie**, Kah Chan Teh, Alex Chichung Kot, "Comm-Transformer: A Robust Deep Learning Based Receiver for OFDM System under TDL Channel," Accepted in IEEE Transactions on Communications, 2023, DOI: 10.1109/TCOMM.2023.3343787.
- **Yihang Xie**, Kah Chan Teh, Alex Chichung Kot, "Diffusion model based generative semantic communication system," Submitted for review IEEE Communications Letters, 2023.
- Zhengyang Su, Kah Chan Teh, **Yihang Xie**, Sirajudeen Gulam Razul, Alex Chichung Kot, "Signal enhancement aided end-to-end deep learning approach for joint denoising and spectrum sensing," Accepted in IEEE Transactions on Vehicular Technology, DOI: 10.1109/TVT.2023.3324826.

[1]The superscript * indicates joint first authors

- Xiaobei Liu, Yong Liang Guan, **Yihang Xie**, Qin Yan, Kah Chan Teh, "Two-stage BEM-ANN receiver with linear complexity for severe doubly-selective fading channels," Submitted for review IEEE Transactions on Vehicular Technology, 2023.

# Academic Service

- Reviewer for IEEE Journal on Selected Areas in Communications.
- Reviewer for IEEE Transactions on Wireless Communications.
- Reviewer for IEEE Transactions on Communications.
- Reviewer for IEEE Transactions on Vehicular Technology.
- Reviewer for IEEE Transactions on Cognitive Communications and Networking.
- Reviewer for IEEE Wireless Communications Letters.
- Reviewer for IEEE Communications Letters.

# Bibliography


[1] Zhiguo Ding, Xianfu Lei, George K Karagiannidis, Robert Schober, Jinhong Yuan, and Vijay K Bhargava. A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends. *IEEE J. Sel. Areas Commun.*, 35(10):2181–2195, 2017. 5

[2] SM Riazul Islam, Nurilla Avazov, Octavia A Dobre, and Kyung-Sup Kwak. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surveys Tuts.*, 19(2):721–742, 2016. 5

[3] Alessio Zappone, Marco Di Renzo, and Mérouane Debbah. Wireless networks design in the era of deep learning: Model-based, AI-based, or both? *IEEE Trans. on Commun.*, 67(10):7331–7376, 2019. 5, 72

[4] Alessio Zappone, Marco Di Renzo, Merouane Debbah, Thanh Tu Lam, and Xuewen Qian. Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization. *IEEE Veh. Tech. Mag.*, 14(3):60–69, 2019. 5

[5] Hao Ye, Geoffrey Ye Li, and Biing-Hwang Juang. Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wirel. Commun. Lett.*, 7(1):114–117, 2017. 5, 6, 8, 9, 10, 14, 18, 26, 27, 28, 31, 32, 52, 59, 61, 64, 73, 74, 76, 88, 90, 95, 96, 105

[6] Mehran Soltani, Vahid Pourahmadi, Ali Mirzaei, and Hamid Sheikhzadeh. Deep learning-based channel estimation. *IEEE Commun. Lett.*, 23(4):652–655, 2019. 5, 10, 73

[7] Qinbo Bai, Jintao Wang, Yue Zhang, and Jian Song. Deep learning-based channel estimation algorithm over time selective fading channels. *IEEE Trans. Cogn. Comm. and Netw*, 6(1):125–134, 2019. 5, 21, 74

[8] Chuan Lin, Qing Chang, and Xianxu Li. A deep learning approach for MIMO-NOMA downlink signal detection. *Sensors*, 19(11):2526, 2019. 5, 18, 21, 27

[9] Hengtao He, Shi Jin, Chao-Kai Wen, Feifei Gao, Geoffrey Ye Li, and Zongben Xu. Model-driven deep learning for physical layer communications. *IEEE Wireless Commun. Mag.*, 26(5):77–83, 2019. 6, 18, 72

[10] Yanyu Cheng, Kwok Hung Li, Kah Chan Teh, Sheng Luo, and Wei Wang. Two-step user pairing for OFDM-based cooperative NOMA systems. *IEEE Commun. Lett.*, 24(4):903–906, 2020. 6, 18

[11] David J Love, Robert W Heath, Vincent KN Lau, David Gesbert, Bhaskar D Rao, and Matthew Andrews. An overview of limited feedback in wireless communication systems. *IEEE J.Sel. Areas Commum.*, 26(8):1341–1365, 2008. 6

[12] Ping-Heng Kuo, HT Kung, and Pang-An Ting. Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 492–497. IEEE, 2012. 6, 7

[13] Xiongbin Rao and Vincent KN Lau. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. Signal Process.*, 62(12):3261–3271, 2014. 6

[14] Jiajia Guo, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Convolutional neural network-based multiple-rate compressive sensing for massive mimo csi feedback: Design, simulation, and analysis. *IEEE Transactions on Wireless Communications*, 19(4):2827–2840, 2020. 7

[15] Yangyang Zhang, Xichang Zhang, and Yi Liu. Deep learning based csi compression and quantization with high compression ratios in fdd massive mimo systems. *IEEE Wireless Communications Letters*, 10(10):2101–2105, 2021. 7

[16] Yuting Wang, Jinlong Sun, Jie Wang, Jie Yang, Tomoaki Ohtsuki, Bamidele Adebisi, and Haris Gacanin. Multi-rate compression for downlink csi based on transfer learning in fdd massive mimo systems. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pages 1–5. IEEE, 2021. 7

[17] Zhilin Lu, Jintao Wang, and Jian Song. Multi-resolution CSI feedback with deep learning in massive MIMO system. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2020. 7, 47, 48

[18] Zhilin Lu, Jintao Wang, and Jian Song. Binary neural network aided csi feedback in massive mimo system. *IEEE Wireless Communications Letters*, 10(6):1305–1308, 2021. 7

[19] Yuwen Yang, Feifei Gao, Xiaoli Ma, and Shun Zhang. Deep learning-based channel estimation for doubly selective fading channels. *IEEE Access*, 7: 36579–36589, 2019. 7, 10, 73

[20] Guanghui Fan, Jinlong Sun, Guan Gui, Haris Gacanin, Bamidele Adebisi, and Tomoaki Ohtsuki. Fully convolutional neural network-based csi limited feedback for fdd massive mimo systems. *IEEE Transactions on Cognitive Communications and Networking*, 8(2):672–682, 2021. 7

[21] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, and Shi Jin. Deep learning for wireless physical layer: Opportunities and challenges. *China Communications*, 14(11):92–111, 2017. 8

[22] Hengtao He, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Deep learning-based channel estimation for beamspace mmwave massive MIMO systems. *IEEE Wireless Commun. Lett.*, 7(5):852–855, 2018. 8, 10, 73

[23] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. Deep learning for massive MIMO CSI feedback. *IEEE Wireless Commun. Lett.*, 7(5):748–751, 2018. 8, 40, 46, 48, 72

[24] Ali A Zaidi, Robert Baldemair, Vicent Molés-Cases, Ning He, Karl Werner, and Andreas Cedergren. OFDM numerology design for 5G new radio to support IoT, eMBB, and MBSFN. *IEEE Commun. Standards Mag.*, 2(2): 78–83, 2018. 8, 52

[25] Behrouz Farhang-Boroujeny and Hussein Moradi. OFDM inspired waveforms for 5G. *IEEE Commun. Surveys & Tuts.*, 18(4):2474–2492, 2016. 8, 52

[26] Jun-Young Woo, Hyun Seung Joo, Kee-Hoon Kim, Jong-Seon No, and Dong-Joon Shin. PAPR analysis of class-III SLM scheme based on variance of correlation of alternative OFDM signal sequences. *IEEE Commun. Lett.*, 19 (6):989–992, 2015. 8, 52

[27] Johannes Dommel, Gabriele Boccolini, Leszek Raschkowski, Stephan Jaeckel, Lars Thiele, Thomas Haustein, and Nuria González Prelcic. 5G in space: PHY-layer design for satellite communications using non-orthogonal multi-carrier transmission. In *IEEE ASMS/SPSC*, pages 190–196. IEEE, 2014. 8, 52

[28] Yan Zhao, Yue Xiao, Ping Yang, Shu Fang, and Wei Xiang. Iterative compensation for clipping noise in spatial modulation OFDM systems. *IEEE Trans. Veh. Technol.*, 68(12):12422–12426, 2019. doi: 10.1109/TVT.2019.2947193. 8, 52

[29] Shansuo Liang, Junjie Ma, and Li Ping. Clipping can improve the performance of spatially coupled sparse superposition codes. *IEEE Commun. Lett.*, 21(12):2578–2581, 2017. 8, 52

[30] Joao Guerreiro, Rui Dinis, and Paulo Montezuma. Optimum and sub-optimum receivers for OFDM signals with strong nonlinear distortion effects. *IEEE Trans. Commun.*, 61(9):3830–3840, 2013. 8, 52, 61

[31] Chaosan Yang, Xiaobei Liu, Yong Liang Guan, and Rongke Liu. Fast GAMP algorithm for nonlinearly distorted OFDM signals. Accepted for publications in the *IEEE Commun. Lett.*, 2021. 9, 14, 52, 54, 59, 61, 63, 64, 69, 104

[32] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *IEEE Int. Symp. Inf. Theory*, pages 2168–2172. IEEE, 2011. 9, 52

[33] Hongji Huang, Jie Yang, Hao Huang, Yiwei Song, and Guan Gui. Deep learning for super-resolution channel estimation and doa estimation based massive mimo system. *IEEE Transactions on Vehicular Technology*, 67(9): 8549–8560, 2018. 10

[34] Yuwen Yang, Feifei Gao, Geoffrey Ye Li, and Mengnan Jian. Deep learning-based downlink channel prediction for fdd massive mimo system. *IEEE Communications Letters*, 23(11):1994–1998, 2019. 10

[35] Sebastian Dörner, Sebastian Cammerer, Jakob Hoydis, and Stephan Ten Brink. Deep learning based communication over the air. *IEEE J. Sel. Topics Signal Process.*, 12(1):132–143, 2017. 10, 73

[36] Hao Ye, Geoffrey Ye Li, Biing-Hwang Fred Juang, and Kathiravetpillai Sivanesan. Channel agnostic end-to-end learning based communication systems with conditional GAN. In *2018 IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, UAE*, pages 1–5. IEEE, 2018. 10, 73

[37] Changyang She, Rui Dong, Zhouyou Gu, Zhanwei Hou, Yonghui Li, Wibowo Hardjawana, Chenyang Yang, Lingyang Song, and Branka Vucetic. Deep learning for ultra-reliable and low-latency communications in 6g networks. *IEEE network*, 34(5):219–225, 2020. 10

[38] Shen Gao, Peihao Dong, Zhiwen Pan, and Geoffrey Ye Li. Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs. *IEEE Communications Letters*, 23(11):1989–1993, 2019. 10

[39] Shiyu Liu, Ming Lei, and Min-Jian Zhao. Deep learning based channel estimation for intelligent reflecting surface aided MISO-OFDM systems. In *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pages 1–5. IEEE, 2020. 11

[40] Yindi Yang, Shun Zhang, Feifei Gao, Jianpeng Ma, and Octavia A Dobre. Graph neural network-based channel tracking for massive MIMO networks. *IEEE Communications Letters*, 24(8):1747–1751, 2020. 11

[41] Fayçal Ait Aoudia and Jakob Hoydis. End-to-end learning of communications systems without a channel model. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 298–303, 2018. doi: 10.1109/ACSSC.2018.8645416. 12

[42] Danlan Huang, Xiaoming Tao, Feifei Gao, and Jianhua Lu. Deep learning-based image semantic coding for semantic communications. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2021. doi: 10.1109/GLOBECOM46510.2021.9685667. 12

[43] Xuewen Luo, Hsiao-Hwa Chen, and Qing Guo. Semantic communications: Overview, open issues, and future research directions. *IEEE Wirel. Commun.*, 29(1):210–219, 2022. 12, 108

[44] Danlan Huang, Feifei Gao, Xiaoming Tao, Qiyuan Du, and Jianhua Lu. Toward semantic communications: Deep learning-based image semantic coding. *IEEE Journal on Selected Areas in Communications*, 41(1):55–71, 2023. doi: 10.1109/JSAC.2022.3221999. 12

[45] Guangming Shi, Yong Xiao, Yingyu Li, and Xuemei Xie. From semantic communication to semantic-aware networking: Model, architecture, and open problems. *IEEE Communications Magazine*, 59(8):44–50, 2021. 12

[46] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning enabled semantic communication systems. *IEEE Trans. Signal Process.*, 69:2663–2675, 2021. 12

[47] Zhenzi Weng and Zhijin Qin. Semantic communication systems for speech transmission. *IEEE Journal on Selected Areas in Communications*, 39(8): 2434–2444, 2021. 12

[48] Peihao Dong, Hua Zhang, Geoffrey Ye Li, Ivan Simoes Gaspar, and Navid NaderiAlizadeh. Deep CNN-based channel estimation for mmwave massive MIMO systems. *IEEE J. Sel. Top. Signal Process.*, 13(5):989–1000, 2019. 14, 74

[49] Abdul Karim Gizzini, Marwa Chafii, Shahab Ehsanfar, and Raed M Shubair. Temporal averaging LSTM-based channel estimation scheme for IEEE 802.11p standard. In *IEEE Global Communications Conference (GLOBECOM)*, pages 01–07. IEEE, 2021. 14, 88, 95, 105

[50] Mikko Honkala, Dani Korpi, and Janne MJ Huttunen. DeepRx: Fully convolutional deep learning receiver. *IEEE Trans. Wireless Commun.*, 20(6): 3925–3940, 2021. 14, 74, 88, 95, 105

[51] Sergey V Zhidkov and Rui Dinis. Belief propagation receivers for near-optimal detection of nonlinearly distorted OFDM signals. In *IEEE VTC 2019-Spring*, pages 1–6. IEEE, 2019. 14, 61, 63, 69, 88, 95, 104, 105

[52] Yunlong Cai, Zhijin Qin, Fangyu Cui, Geoffrey Ye Li, and Julie A McCann. Modulation and multiple access for 5G networks. *IEEE Commun. Surveys Tuts.*, 20(1):629–646, 2017. 21

[53] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz, TR 38.901. *3GPP Specification series*, Release 15, 2018. URL `http://www.3gpp.org/DynaReport/38901.htm`. 22, 77, 78

[54] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. IEEE, 2017. 23

[55] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 23, 83

[56] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015. 25

[57] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 25

[58] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 1–6, 1987. 32

[59] NAIC. Massive MIMO CSI feedback dataset. *National Artificial Intelligence Competition (NAIC)*, 2020. URL `https://naic.pcl.ac.cn/frame/1`. 39

[60] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015. 39

[61] Tianqi Wang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Deep learning-based CSI feedback approach for time-varying massive MIMO channels. *IEEE Wireless Commun. Lett.*, 8(2):416–419, 2018. 47, 48

[62] Yuyao Sun, Wei Xu, Lisheng Fan, Geoffrey Ye Li, and George K Karagiannidis. Ancinet: An efficient deep learning approach for feedback compression of estimated CSI in massive MIMO systems. *IEEE Wireless Commun. Lett.*, 9(12):2192–2196, 2020. 47, 48

[63] Jaakko Pihlajasalo, Dani Korpi, Mikko Honkala, Janne MJ Huttunen, Taneli Riihonen, Jukka Talvitie, Alberto Brihuega, Mikko A Uusitalo, and Mikko Valkama. Hybriddeeprx: Deep learning receiver for high-EVM signals. *arXiv preprint arXiv:2106.16079*, 2021. 52

[64] Alvis Logins, Jiale He, and Kirill Paramonov. Block-structured deep learning-based OFDM channel equalization. *IEEE Communications Letters*, 26(2): 321–324, 2021. 52

[65] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 58

[66] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014. 58, 84

[67] Shansuo Liang, Jun Tong, and Li Ping. On iterative compensation of clipping distortion in OFDM systems. *IEEE Wirel. Commun. Lett.*, 8(2):436–439, 2018. 61

[68] Lei Liu, Ying Li, Chongwen Huang, Chau Yuen, and Yong Liang Guan. A new insight into GAMP and AMP. *IEEE Trans. Veh. Technol.*, 68(8): 8264–8269, 2019. 62

[69] Zhijin Qin, Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. Deep learning in physical layer communications. *IEEE Wirel. Commun.*, 26(2): 93–99, 2019. 72

[70] Deniz Gündüz, Zhijin Qin, Inaki Estella Aguerri, Harpreet S. Dhillon, Zhaohui Yang, Aylin Yener, Kai Kit Wong, and Chan-Byoung Chae. Beyond transmitting bits: Context, semantics, and task-oriented communications. *IEEE J. Sel. Areas Commun.*, 41(1):5–41, 2023. doi: 10.1109/JSAC.2022.3223408. 72

[71] Jie Wang, Guan Gui, Tomoaki Ohtsuki, Bamidele Adebisi, Haris Gacanin, and Hikmet Sari. Compressive sampled CSI feedback method based on deep learning for FDD massive MIMO systems. *IEEE Trans. Commun.*, 69(9): 5873–5885, 2021. 72

[72] Zhengyang Su, Kah Chan Teh, Sirajudeen Gulam Razul, and Alex C. Kot. Deep Non-Cooperative Spectrum Sensing over Rayleigh Fading Channel. *IEEE Trans. Veh. Technol.*, 71(4):4460–4464, 2022. doi: 10.1109/TVT.2021.3138593. 72

[73] Wonjun Kim, Yongjun Ahn, and Byonghyo Shim. Deep neural network-based active user detection for grant-free NOMA systems. *IEEE Trans. Commun.*, 68(4):2143–2155, 2020. 72

[74] Wenchao Xia, Gan Zheng, Yongxu Zhu, Jun Zhang, Jiangzhou Wang, and Athina P Petropulu. A deep learning framework for optimization of MISO downlink beamforming. *IEEE Trans. Commun.*, 68(3):1866–1880, 2019. 72

[75] 3GPP. Study on enhancement for data collection for NR and ENDC, TR 37.817. *3GPP Specification Series*, Release 17, 2022. URL `https://www.3gpp.org/ftp/Specs/archive/37_series/37.817`. 73

[76] 3GPP. Study on artificial intelligence (AI)/machine learning (ML) for NR air interface. *3GPP Specification Series*, Release 18, 2022. URL `https://www.3gpp.org/ftp/Specs/archive/38_series/38.843`. 73

[77] Qiang Hu, Feifei Gao, Hao Zhang, Shi Jin, and Geoffrey Ye Li. Deep learning for channel estimation: Interpretation, performance, and comparison. *IEEE Trans. Wirel. Commun.*, 20(4):2398–2412, 2021. 73

[78] Eren Balevi, Akash Doshi, Ajil Jalal, Alexandros Dimakis, and Jeffrey G. Andrews. High dimensional channel estimation using deep generative networks. *IEEE Journal on Selected Areas in Communications*, 39(1):18–30, 2021. doi: 10.1109/JSAC.2020.3036947. 73

[79] Hao Ye, Le Liang, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels. *IEEE Trans. Wirel. Commun.*, 19(5):3133–3143, 2020. 73

[80] Tianyi Zeng, Yongyu Chang, Qi Zhang, Mengshi Hu, and Junyao Li. CNN-based LOS/NLOS identification in 3-D massive MIMO systems. *IEEE Communications Letters*, 22(12):2491–2494, 2018. doi: 10.1109/LCOMM.2018.2872522. 73

[81] Yihang Xie, Xiaobei Liu, Kah Chan Teh, and Yong Liang Guan. Robust deep learning-based end-to-end receiver for OFDM system with non-linear distortion. *IEEE Commun. Lett.*, 26(2):340–344, 2022. doi: 10.1109/LCOMM.2021.3132326. 73, 109

[82] Aidin Ferdowsi and Walid Saad. Deep learning for signal authentication and security in massive internet-of-things systems. *IEEE Trans. Commun.*, 67 (2):1371–1387, 2018. 73

[83] Peiwen Jiang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Dual CNN-based channel estimation for MIMO-OFDM systems. *IEEE Trans. Commun.*, 69 (9):5859–5872, 2021. 74

[84] Qiang Hu, Feifei Gao, Hao Zhang, Shi Jin, and Geoffrey Ye Li. Deep learning for channel estimation: Interpretation, performance, and comparison. *IEEE Trans. Wirel. Commun.*, 20(4):2398–2412, 2020. 74

[85] Qi Peng, Jinbao Li, and Hao Shi. Deep learning based channel estimation for OFDM systems with doubly selective channel. *IEEE Commun. Lett.*, 26 (9):2067–2071, 2022. 74

[86] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems (NIPS), Long Beach, USA*, 30, 2017. 75

[87] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah, USA*, pages 7794–7803, 2018. 75

[88] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *Stat*, 1050:11, 2018. 84

[89] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9,*, 2015. URL `http://arxiv.org/abs/1412.6980`. 87

[90] Deniz Gündüz, Zhijin Qin, Inaki Estella Aguerri, Harpreet S Dhillon, Zhaohui Yang, Aylin Yener, Kai Kit Wong, and Chan-Byoung Chae. Beyond transmitting bits: Context, semantics, and task-oriented communications. *IEEE J. Sel. Areas Commun.*, 41(1):5–41, 2022. 108

[91] Jincheng Dai, Ping Zhang, Kai Niu, Sixian Wang, Zhongwei Si, and Xiaoqi Qin. Communication beyond transmitting bits: Semantics-guided source and channel coding. *IEEE Wireless Commun.*, early access. doi=10.1109/MWC.017.2100705. 108

[92] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning based semantic communications: An initial investigation. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pages 1–6. IEEE, 2020. 109

[93] Shiva Raj Pokhrel and Jinho Choi. Understand-before-talk (UBT): A semantic communication approach to 6G networks. *IEEE Trans. Veh. Technol.*, 72 (3):3544–3556, 2023. 109

[94] Yihang Xie, Kah Chan Teh, and Alex C. Kot. Deep learning-based joint detection for OFDM-NOMA scheme. *IEEE Commun. Lett.*, 25(8):2609–2613, 2021. doi: 10.1109/LCOMM.2021.3077878. 109

[95] Huiqiang Xie, Zhijin Qin, and Geoffrey Ye Li. Task-oriented multi-user semantic communications for VQA. *IEEE Wirel. Commun. Letters*, 11(3): 553–557, 2021. 109

[96] Huiqiang Xie and Zhijin Qin. A lite distributed semantic communication system for internet of things. *IEEE J. Sel. Areas Commun*, 39(1):142–153, 2020. 109

[97] Minrui Xu, Hongyang Du, Dusit Niyato, Jiawen Kang, Zehui Xiong, Shiwen Mao, Zhu Han, Abbas Jamalipour, Dong In Kim, Victor Leung, et al. Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services. *arXiv preprint arXiv:2303.16129*, 2023. 109

[98] Jiawei Su, Zhixin Liu, Yuan-ai Xie, Kai Ma, Hongyang Du, Jiawen Kang, and Dusit Niyato. Semantic communication-based dynamic resource allocation in D2D vehicular networks. *IEEE Trans. Veh. Technol.*, early access. doi:10.1109/TVT.2023.3257770. 109

[99] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 109

[100] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations, Virtual Event, Austria*, 2021. 109, 114

[101] Diederik P. Kingma and Max Welling. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. 114

[102] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 117

[103] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 118

[104] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *arXiv preprint arXiv:2210.08402*, 2022. 118

[105] Chongwen Huang, Zhaohui Yang, George C Alexandropoulos, Kai Xiong, Li Wei, Chau Yuen, Zhaoyang Zhang, and Mérouane Debbah. Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design. *IEEE Journal on Selected Areas in Communications*, 39(6):1663–1677, 2021. 125

[106] Hoon Lee, Sang Hyun Lee, and Tony QS Quek. Deep learning for distributed optimization: Applications to wireless resource management. *IEEE Journal on Selected Areas in Communications*, 37(10):2251–2266, 2019. 126

[107] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain adaptive ensemble learning. *IEEE Transactions on Image Processing*, 30:8008–8018, 2021. 127

[108] Geoffrey I Webb and Zijian Zheng. Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques. *IEEE Transactions on Knowledge and Data Engineering*, 16(8):980–991, 2004. 127

[109] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=nZeVKeeFYf9`. 127