UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

Intelligence-aided Channel Estimation and Detection for Multi-dimensional Index Modulation

by

Xinyu Feng
ORCiD: 0009-0006-8363-4771

A thesis for the degree of Doctor of Philosophy

Supervisors: Prof.Mohammed El-Hajjar
Dr.Chao Xu
Prof.Lajos Hanzo

May 2024

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

A progress report submitted for continuation towards a PhD

Intelligence-aided Channel Estimation and Detection for Multi-dimensional Index Modulation

by Xinyu Feng

Multiple-Input Multiple-Output (MIMO) technology is a cornerstone of the Fifth Generation Mobile Systems (5G) wireless communication systems. Although it offers significant gains in Energy Efficiency (EE), Spectral Efficiency (SE), performance and throughput, supporting the escalate increasing and dynamically changing demands of large wireless environments remains a significant challenge. In this case, a prominent technique in next-generation wireless communications known as Index Modulation (IM) building on the MIMO system, develops its unique advantages. The concept of index modulation (IM) has been actively researched due to its flexible performance, SE, EE and hardware trade-offs. In order to further improve the SE, Compressed Sensing (CS) and Multi-dimensional Index Modulation (MIM) have been investigated. However, the flexibility and performance gains achieved with IM come at the cost of increased detection complexity. Channel Estimation (CE) for IM systems presents a challenging problem due to the more complex transmission conditions and the pilot allocation requirements. Conventional detection such as Maximum Likelihood (ML) could achieve near-capacity performance, while the detection complexity of computation will exponentially increases with the increase of the degrees of freedom of the IM. Nowadays, machine learning technology has been widely used in wireless network solving massive data problem by its robust computing capability. With the success of many attempts to combine the complex MIMO system and machine learning to detect the signal, it have been proved that learning-aided detection could significantly reduce the detection complexity. Hence, this thesis focus on designing composite IM system, which may employ CS and learning based reduced complexity detector.

First, we propose Deep Learning (DL) based detection for CS-MIM, where both Hard-Decision (HD) as well as Soft-Decision (SD) detection combined with iterative decoding are conceived. We propose two novel neural network aided methods for Iterative Soft Detection (ISD), where iterations are carried out between the CS-MIM detector and a channel decoder. In contrast to the conventional detection of CS-MIM system, which critically relies on the knowledge of Channel State Information (CSI) at the receiver, the proposed learning-aided methods are capable of eliminating the overhead and complexity of CE, which results in an improved transmission rate. Our simulation results demonstrate that the proposed learning techniques conceived for SD CS-MIM combined with iterative detection are capable of achieving near-capacity performance at a reduced complexity compared to the conventional model-based SD relying on CSI acquisition.

However, having accurate CSI is essential for reliable MIM, which requires high pilot overhead. Hence, Joint Channel Estimation and data Detection (JCED) is harnessed to reduce the pilot overhead and improve the detection performance at a modestly increased estimation complexity. Then this thesis circumvents this by proposing Deep Learning (DL) based JCED for CS-MIM for significantly reducing the complexity, despite reducing the pilot overhead needed for Channel Estimation (CE). Our simulation results confirm a Deep Neural Network (DNN) is capable of near-capacity JCED of CS-MIM at a reduced pilot overhead and reduced complexity both for Hard-Decision (HD) and SD detection.

On the other hand, while MIM uncovers the possibility of combination of IM in different dimensions, roughly assemble of different dimension IM does not fully exploit the flexibility of MIM. In this thesis, we propose Joint Multi-dimensional Index Modulation (JMIM) that can utilize the time-, space- and frequency-dimensions in order to increase the IM mapping design flexibility. Explicitly, this thesis develops a jointly designed MIM architecture combined with CS. Three different JMIM mapping methods are proposed for a space- and frequency-domain aided JMIM system, which can attain different throughput and diversity gains. Then, this thesis extend the proposed JMIM design to three dimensions by combining it with the time domain. To circumvent the high detection complexity of the proposed CS-aided JMIM design, a large fraction of the overall computational complexity is relegated to the offline model training phase. Furthermore, we integrate the different communication tasks, such as CE, decoding and detection into a simple neural network model. Additionally, this thesis investigates the adaptive design of the proposed CS-aided JMIM system, where a learning-based adaptive modulation configuration method is applied. Our simulation results demonstrate that the proposed CS-aided JMIM (CS-JMIM) is capable of outperforming its CS-aided separate-domain MIM counterpart. Furthermore, the learning-aided adaptive scheme is capable of increasing the throughput while maintaining the required error probability target.

Contents

Acknowledgements **xi**

Declaration of Authorship **xv**

List of Publications **xvii**

1 Introduction **1**
1.1 Thesis Structure 2
1.2 Novel Contribution of The Thesis 4

2 Literature Review **7**
2.1 Index Modulation 7
2.1.1 Spatial Modulation 7
2.1.2 Frequency Domain Index Modulation 8
2.1.3 Multi-dimensional Index Modulation 9
2.2 Compressed Sensing With Index Modulation 11
2.3 Detection 12
2.3.1 Channel Estimation 12
2.3.1.1 Estimation method 13
2.3.1.2 Pilot Allocation 13
2.3.1.3 Interpolation 14
2.3.1.4 CE methods for IM 14
2.3.2 Hard Decision Detection 15
2.3.3 Soft Decision Detection 16
2.3.4 Joint Channel Estimation and Detection for IM 16
2.3.5 Machine Learning Based Detection for IM 16
2.3.5.1 Deep Learning 17
2.3.5.2 Learning aided Channel Estimation and Detection . . . 18

3 Deep Learning-Based Soft Iterative-Detection of Channel-Coded Compressed Sensing-Aided Multi-Dimensional Index Modulation **21**
3.1 Introduction 21
3.2 System Model 24
3.2.1 Transmitter 25
3.2.1.1 SI Selection 25
3.2.1.2 STSK Encoder 26
3.2.1.3 ST Transmitter Model 26
3.2.1.4 Antenna Selection 27

3.2.2 Received Signal Model . . . 28
3.3 Detection . . . 29
3.3.1 Hard-Decision Detection . . . 29
3.3.1.1 ML Detection . . . 30
3.3.1.2 CE Methods . . . 30
3.3.1.3 Reduced Complexity S-MP Detection . . . 31
3.3.1.4 AMP-Aided Detection . . . 31
3.3.1.5 Proposed Learning-Aided HD Detection . . . 32
3.3.2 Soft-Decision Detection . . . 35
3.3.2.1 Conventional SD Detection . . . 35
3.3.2.2 Single-Stage DNN-Aided SD Detector . . . 36
3.3.2.3 Iterative DNN-Aided MAP SD Detection . . . 38
3.4 Performance Analysis . . . 40
3.5 Summary . . . 50

4 Reduced Complexity Learning-Based Joint Channel Estimation and Detection of Compressed Sensing-aided Multi-Dimensional Index Modulation 51
4.1 Introduction . . . 51
4.2 System Model . . . 54
4.3 CE and Detection for CS-MIM . . . 55
4.3.1 Conventional CE and HD Detection . . . 56
4.3.1.1 CE Methods . . . 57
4.3.1.2 ML Detection . . . 58
4.3.1.3 JCED . . . 58
4.3.2 SD Detection . . . 60
4.3.2.1 Conventional SD Detection . . . 61
4.3.2.2 SD-JCED . . . 62
4.3.3 Proposed Learning based Channel Estimation and Detection . . . 63
4.3.3.1 Conventional DNN-aided CE and HD detection . . . 63
4.3.3.2 Separate DNN-aided CE and detection . . . 65
4.3.3.3 Proposed DNN-aided JCED . . . 67
4.4 Performance Analysis . . . 71
4.5 Summary . . . 80

5 Near-Instantaneously Adaptive Learning-Assisted and Compressed Sensing-Aided Joint Multi-Dimensional Index Modulation 81
5.1 Introduction . . . 81
5.2 System Model . . . 84
5.2.1 Transmitter . . . 85
5.2.1.1 JIM . . . 85
General JIM . . . 86
Grouped JIM . . . 86
Coded JIM . . . 88
3-Dimensional JIM-DM Design . . . 89
5.2.1.2 CS and Block Assembly . . . 92
5.2.2 Receiver Processing . . . 94
5.3 CS-JMIM Signal Detection . . . 95

5.3.1 HD Detection . . . 95
5.3.1.1 ML Detection . . . 96
5.3.1.2 DNN-based Detection . . . 96
5.3.2 SD Detection . . . 98
5.3.2.1 Conventional SD Detection . . . 98
5.3.2.2 DNN-based SD Detection . . . 100
5.4 Adaptive design . . . 101
5.4.0.1 Conventional Threshold-based Adaptive Design . . . 101
5.4.0.2 Learning aided adaptive modulation . . . 103
KNN-based Adaptive Design . . . 103
DNN-aided Adaptive Design . . . 104
5.5 Performance Analysis . . . 105
5.6 Summary . . . 117

6 Conclusions and Future work **119**
6.1 Conclusion . . . 119
6.2 Future work . . . 120
6.2.1 Graph Neural Network (GNN)-Based Detection of Multi-User (MU)- IM . . . 120
6.2.2 Downlink Communication in MU- IM Systems . . . 121
6.2.3 Learning-Based Detection in Channel Coded IM Systems . . . 121
6.2.4 Multi-cell and Stacked Intelligent Meta-surfaces . . . 121

References **123**

Acknowledgements

I would like to sincerely thank my supervisor Professor Mohammed El-Hajjar for his Patient and professional guidance, I could not finish this thesis without his generous support.

I would also thank my supervisor Dr.Chao Xu for his outstanding help and support, whose passion for research has always inspired me.

I would also like to thank my supervisor Professor Lajos Hanzo for his timely concern and help. He always gave me many wise advice, guide me both in research and life.

Finally, I would like to express my gratitude to my parents Xueqiang Feng and Libo Sun, for their encouragement and unconditional support. Furthermore, I would like to thank my Sister Luxia Feng and all my friends for their insightful advice and concern for me.

To my beloved family and friends...

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published in the provided list of publications.

Signed:..... Date:.....

List of Publications

[1] X. **Feng**, M. EL-Hajjar, C. Xu, and L. Hanzo, "Deep learning-based soft iterative-detection of channel-coded compressed sensing-aided multi-dimensional index modulation," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 7530–7544, 2023.

[2] X. **Feng**, M. EL-Hajjar, C. Xu, and L. Hanzo, "Reduced complexity learning-assisted joint channel estimation and detection of compressed sensing-aided multi-dimensional index modulation," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 78–94, 2024.

[3] X. **Feng**, M. EL-Hajjar, C. Xu, and L. Hanzo, "Near-instantaneously adaptive learning-assisted and compressed sensing-aided joint multi-dimensional index modulation," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 893–912, 2023.

[4] X. **Feng**, M. El-Hajjar, C. Xu, and L. Hanzo, "Graph nerual network based low complexity detection of multi-user compressed sensing-aided space-frequency index modulation," in preparation.

List of Symbols

The next list describes several symbols that will be later used within the body of the document

General Notation

N_r	number of RAs
N_t	number of TAs
N_c	number of subcarrier for each frame
G	number of group
N_f	number of subcarrier per group in FD
N_v	number of subcarriers per group in VD
$b_{g,1}$	bits for SM at g-th group subcarriers of CS-MIM
$b_{g,2}$	bits for frequency index modulation at g-th group subcarriers of CS-MIM
$b_{g,3}$	bits for STSK mapping at g-th group subcarriers of CS-MIM
b	total bits number of each frame
M	number of TAs for STSK dispersion matrices
N	number of RAs for STSK dispersion matrices
T	number of time slots for STSK dispersion matrices
b_g	bits number per group
K	number of activated index in each subcarrier group in VD
K	number of activated index in JMIM matrix
Q	number of dispersion matrices for STSK mapping
L	number of constellation realizations for STSK mapping

$\boldsymbol{S}$ Space-time symbol block of each subcarrier group in VD

$\boldsymbol{s}$ Space-time symbol sequence of each subcarrier group in VD

T number of TS for transmitting in TD

$\hat{\boldsymbol{b}}_r$ detected data bits

$\hat{\boldsymbol{C}}$ detected coded data bits

$\boldsymbol{A}$ CS measurement matrix

$\boldsymbol{h}_\alpha$ channel model in TD

$\boldsymbol{H}_\alpha$ channel model in FD

$\boldsymbol{Y}$ Received signal

$\boldsymbol{W}$ Additive White Gaussian noise

N_0 noise power of AWGN

$\boldsymbol{W_n}$ weights of n-th neuron

$\boldsymbol{\theta_n}$ bias of n-th neuron

B sample size of current iteration of DNN training phase

L_a de-interleaved LLR sequence of L_e

L_e output extrinsic LLR after soft demodulation

$\mathbf{c}$ coded bit sequence of information bits

$\mathbf{u}$ information bit sequence stream

$\mathcal{X}_1^l$ subset of the legitimate equivalent signal $\mathcal{X}$ corresponding to bit $u_l = 1$

$\mathcal{X}_0^l$ subset of the legitimate equivalent signal $\mathcal{X}$ corresponding to bit $u_l = 0$

$\boldsymbol{Y}_p$ Received signal of transmitted pilot symbol of each group

Special Symbols

Chapter 3

$\boldsymbol{X}$ possible realization of STSK codewords

$\boldsymbol{X}_t^m$ STSK slice at t-th TS and m-th activated antennas

$\boldsymbol{S}$ ST matrix

$\boldsymbol{s_a}$ ST vector in one TS for single TS

$\boldsymbol{s_m^{FD}}$ compressed FD ST vector of single TS

$\boldsymbol{S^{FD}}$ compressed FD symbol

$\boldsymbol{I_{AC}}$ sub-matrix of antenna index selection pattern

$\boldsymbol{I_{SI}}$ sub-matrix of suncarrier index selection pattern

$\bar{\mathcal{I}}_{AC}$ possible realizations of the active TA indices

$\bar{\mathcal{I}}_{SI}$ possible realizations of the active subcarrier indices

$\mathcal{X}_{q,l}$ possible realizations of STSK codes blocks over q DMs and l constellation

$\boldsymbol{S}_p$ compressed FD pilot symbol of each group

$\boldsymbol{R}_H$ channel correlation matrix

$\boldsymbol{\Phi}$ possible joint identifications of activated TAs and subcarriers

$\boldsymbol{Y}^{FD}$ Equivalent received signal without the identification of activated antenna index

$\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)}$ the updated FD symbol in AMP detection

$\boldsymbol{z}^{[t]_{a,\tau}}$ residual of current iteration

C_k cell state in LSTM unit

x_k input of LSTM

z_k output of LSTM

ϕ_k parameter of LSTM

$\boldsymbol{W}_n$ weight of each neuron

$\boldsymbol{\theta}_n$ bias of each neuron

$\mathcal{X}_{\gamma,\beta,\varphi}$ realization of STSK codewords

ϵ training sets of iterative DNN-MAP-based model

$\lambda_{\gamma,\beta,\varphi}$ intrinsic metrix of STSK realization

Chapter 5

N_{vt} number of TAs in VD

Q number of total realizations of JMIM mapping DMs

$\boldsymbol{S}$ ST block

$\boldsymbol{D}_q$ The q-th DM of Q total JIM DMs

$\boldsymbol{D}_i$	Selected DM of Q total JIM DMs
χ	M-ary PSK or QAM constellation
x	Selected constellation symbol from χ
$\boldsymbol{S}_g$	combined signal after JMIM and PSK/QAM
gs	number of sub-groups in grouped JIM methods of JMIM
gsx	number of split groups in virtual FD
gsy	number of split groups in virtual spatial domain
N_q	column length of coded JMIM DM
n_q	number of active subcarriers in each column of coded JMIM
T_v	number of TS for transmitting in VD
gsz	number of split groups in virtual TD
$\boldsymbol{s}$	JMIM vector after block creator in VD
$\boldsymbol{s}^{(RD)}$	JMIM vector after CS
$\bar{\mathcal{D}}(\beta)$	all the possible realizations of the JMIM DM
$\bar{\mathcal{X}}$	all the possible realizations of the selected PSK/QAM symbol
$\mathcal{X}_{\gamma,\beta}$	the PSK/QAM symbol at the β-th CS-JMIM DM
ξ	the SNR value of a symbol with a BER lower than the target BER value for adaptive modulation
$\mathcal{T}$	training sets of SNR with different the adaptive mode
N_p	total number of instantaneous SNR values with BER under the target

List of Acronyms

5G Fifth Generation Mobile Systems.

ADC Analog-to-Digital Converter.

AMP Approximate Message Passing.

Approx-Log-MAP Approximate Log Maximum A Posteriori.

AWGN Additive white Gaussian Noise.

BER Bit Error Ratio.

CE Channel Estimation.

CNN Convolutional Neural Networks.

CoE Cost Efficiency.

CP Cyclic Prefix.

CS Compressed Sensing.

CS-JMIM Compressed Sensing-aided Joint Multi-dimensional Index Modulation.

CS-MIM Compressed Sensing-aided Multi-dimensional Index Modulation.

CS-STFIM Compressed Sensing-aided Space-Time-Frequency Index Modulation.

CSI Channel State Information.

D2D Device-to-Device.

DCMC Discrete-Input Continuous-Output Memoryless Channels.

DDCE Decision Directed Channel Estimation.

DL Deep Learning.

DM Dispersion Matrix.

DNN Deep Neural Network.

EE Energy Efficiency.

EM Expectation-Maximization.

ESM Enhanced Spatial Modulation.

EXIT Extrinsic Information Transfer Chart.

FC Fully-Connected.

FD Frequency Domain.

FFT Fast Fourier Transform.

FSK Frequency Shift Keying.

GNN Graph Neural Network.

GSIM-OFDM Generalized Subcarrier Index Modulation.

GSM Generalized Spatial Modulation.

GSTSK Generalized Space-Time Shift Keying.

HD Hard-Decision.

ICI Inter-Carrier Interference.

IFFT Inverse Fast Fourier Transform.

IM Index Modulation.

ISD Iterative Soft Detection.

JCED Joint Channel Estimation and data Detection.

JMIM Joint Multi-dimensional Index Modulation.

KNN *K*-nearest Neighbour.

LDPC Low-Density Parity-Check.

LLR Log Likelihood Ratio.

LSCE Least Squared Channel Estimation.

LSTM Long Short-Term Memory.

MAP Maximum A Posteriori.

Max-Log-MAP Maximum Log-MAP.

MF Matched Filter.

MIM Multi-dimensional Index Modulation.

MIMO Multiple-Input Multiple-Output.

ML Maximum Likelihood.

MMSE Minimum Mean Square Error.

MMSE-CE MMSE based Channel Estimation.

mmWave Millimeter Wave.

MS-STSK Multi-Set Space-Time Shift Keying.

MSE Mean Square Error.

MU Multi-User.

NN Neural Network.

NOMA Non-Orthogonal Multiple Access.

OFDM Orthogonal Frequency Division Multiplexing.

OFDM-IM IM-aided OFDM.

PAPR Peak-to-Average Power Ratio.

QAM Quadrature Amplitude Modulation.

QSM Quadrature Spatial Modulation.

RA Receive Antenna.

RC Reduced-Complexity.

Relu Rectified linear unit.

RF Radio Frequency.

RLS Recursive Least Square.

RMSE Root Mean Square Error.

RSC Recursive Systematic Convolutional.

S-MP Simultaneous Matching Pursuit.

SD Soft-Decision.

SE Spectral Efficiency.

SI Subcarrier Index.

SIM Subcarrier-Index Modulation.

SM Spatial Modulation.

SNR Signal-to-Noise Ratio.

SpD Spatial Domain.

SSK Space Shift Keying.

ST Space-Time.

ST-IM Space-Time Index Modulation.

STBC Space-Time Block Codes.

STFSK Space-Time-Frequency Shift Keying.

STSK Space-Time Shift Keying.

TA Transmit Antenna.

TD Time Domain.

TD-IM Time Domain Index Modulation.

TS Time Slot.

VD Virtual Domain.

ZF Zero Forcing.

Chapter 1

Introduction

With the advent of the Fifth Generation Mobile Systems (5G), wireless communication systems are expected to deliver high-quality communication at high spectrum efficiency to a wide range of smart devices [1]. To achieve this goal, several novel physical layer concepts have been proposed, such as Multiple-Input Multiple-Output (MIMO) and Non-Orthogonal Multiple Access (NOMA) scheme [2]. Recently, the concept of Index Modulation (IM) attracted significant research interest, which offers high Energy Efficiency (EE), providing an alternative to traditional digital modulation [3] [4] [5]. Briefly, the IM technique relies on innovative methods of conveying additional information bits [6]. This approach can activate/deactivate various transmission entities, including transmit antennas [7] [8], subcarriers [9] [10], radio frequencies, LEDs, relays, modulation types [11], time slots [12], pre-coder matrices, dispersion matrices [13], spreading codes [14], and signal powers. It introduces a new dimension for data transmission and can be integrated with conventional techniques [15]. Specifically, the concept of Spatial Modulation (SM) can be regarded as a competitive candidate for next generation wireless networks due to the significant advantages of its high Spectral Efficiency (SE) and EE as well as simplicity in hardware [4]. Furthermore, the concept of IM in the frequency domain has been explored, relying on the Subcarrier-Index Modulation (SIM) idea [16].

SM requires a specifically controlled set of antennas to transmit extra information, where some of the antennas remain inactive [15], while SIM has subcarriers unused and embeds extra information in to their indices, however, deactivating some of the antennas of SM or the subcarriers of SIM potentially reduces the throughput compared to 'fully-loaded' Orthogonal Frequency Division Multiplexing (OFDM) schemes [10]. In this situation, the Compressed Sensing (CS) techniques can play a significant role in reversing the throughput reduction, while increaseing the detection performance [17]. To further improve the performance and provide a flexible trade-off between diversity, SE, EE and error performance, Multi-dimensional Index Modulation (MIM) systems have been developed [18] [19] [20] [21].

Detection in single-dimension IM systems is relatively straightforward, owing to the relative simplicity inherent in these techniques [22] [23]. However, the complexity escalates significantly when IM is developed for the CS-aided IM schemes [17]. Hence, several methods have been proposed for mitigating the detection complexity, such as Approximate Message Passing (AMP) [24] [25]. The computational complexity is further increased by Channel Estimation (CE) employed for acquiring the Channel State Information (CSI) [26] [27]. Specifically, the design of pilots for IM imposes high complexity, which also depends on the channel conditions. Moreover, pilot-based CE introduces additional overhead, consequently diminishing the EE gains of IM [16]. In the realm of Compressed Sensing-aided Multi-dimensional Index Modulation (CS-MIM), the modulation of indices across multiple dimensions exponentially increases the search-space, necessitating advanced algorithms for accurate signal detection. CS further complicates this scenario by requiring the recovery of these high-dimensional IM signals from compressed datasets.

When aiming for solving similar large-scale search problems, learning techniques have shown substantial promise in different areas of wireless communications system [28] [29] [30]. Inspired by its success in pattern recognition, finance, marketing, and healthcare, machine learning is increasingly recognized as a valuable tool in artificial intelligence [31]. To elaborates machine learning techniques have proven effective in addressing various technical challenges in next-generation systems, including massive MIMO [32] [33], Device-to-Device (D2D) networks [34] and detection [35]. These learning-aided methods can significantly reduce the complexity of detection in advanced systems, offering improved adaptability to next-generation systems.

Deep Learning (DL) algorithms, known for their proficiency in handling high dimensional data and extracting intricate patterns, are ideally suited for CS-MIM schemes. They can be trained to interpret the sparse representations typical in CS, thereby facilitating more accurate and efficient signal reconstruction. This is particularly beneficial in MIM systems, where the modulated signal is complex due to having activated indices across multiple dimensions such as frequency, time, and space. Moreover, DL is capable of optimizing the signal recovery process. By the channel conditions, Neural Network (NN) can adaptively refine the reconstruction algorithms, leading to improved accuracy at a reduced complexity. This adaptability is crucial in dynamic wireless environments, where the channel conditions and system parameters can vary rapidly.

1.1 Thesis Structure

Fig.1.1 shows the structure of this thesis.

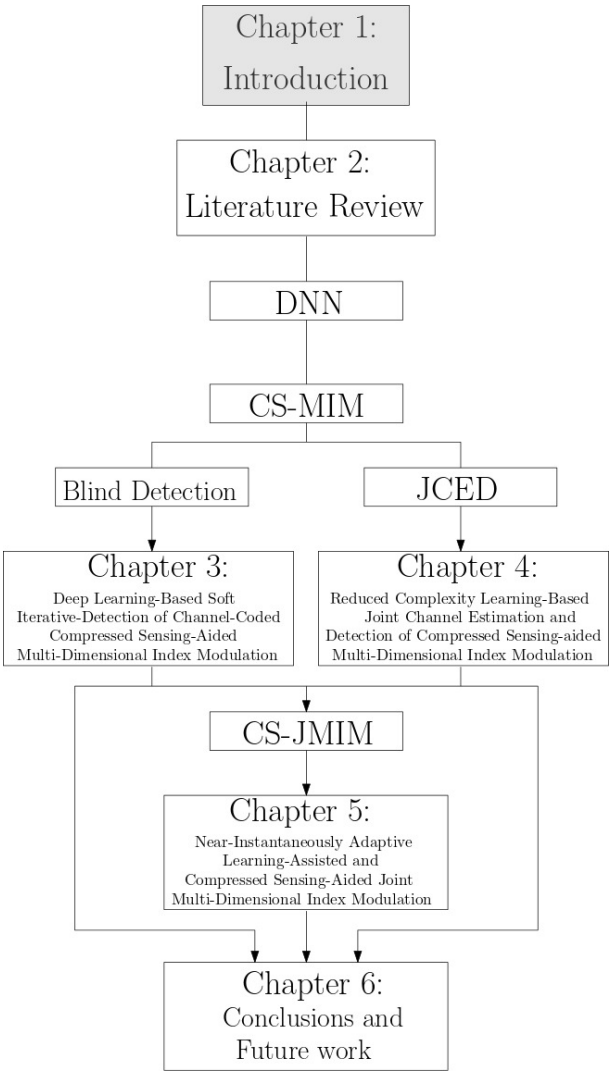

FIGURE 1.1: The organization of this thesis

In Chapter 2 we present the background knowledge of concerning the main techniques employed in the thesis.

Then, in Chapter 3, a Deep Learning (DL) based detection techniques in conceived for Compressed Sensing-aided Multi-dimensional Index Modulation (CS-MIM), where both Hard-Decision (HD) as well as Soft-Decision (SD) detection are considered. More explicitly, firstly, we propose learning aided hard and soft detection for CS-MIM. Secondly, a pair of novel neural network aided methods is proposed for Iterative Soft Detection (ISD), where iterations are carried out between the CS-MIM detector and a channel decoder. In contrast to the conventional CS-MIM system, which critically relies on the knowledge of Channel State Information (CSI) at the receiver, the proposed learning-aided methods are capable of mitigating the overhead and complexity of CE.

Chapter 4 employs the pilot-based CE and detection to have accurate Channel State Information (CSI) for reliable Multi-dimensional Index Modulation (MIM). Then, Joint

Channel Estimation and data Detection (JCED) is harnessed to reduce the pilot overhead and improve the detection performance at a modestly increased estimation complexity, which we then circumvent this by proposing DL based JCED for CS-MIM, which significantly reducing the complexity, despite reducing the pilot overhead needed for CE. Furthermore, we conceive training-aided SD detection.

In Chapter 5, we propose Joint Multi-dimensional Index Modulation (JMIM) relying on the time-, space- and frequency-dimensions in order to increase the IM design flexibility. Three different JMIM mapping methods are proposed for a space- and frequency-domain aided JMIM system, which can attain different throughputs and diversity gains. Then, we extend the proposed JMIM design to three dimensions by combining it with the time domain. Additionally, to circumvent the high detection complexity of the proposed CS-JMIM design, we propose DL based detection. Again, both HD as well as SD detection are conceived. Additionally, we the design an adaptive CS-JMIM system, where a learning-based adaptive modulation configuration is applied.

Finally, Chapter 6 summarize the main research and provides the conclusion of this thesis, as well as further outlook for possible future works.

1.2 Novel Contribution of The Thesis

Chapter 3: Learning-aided blind detection is proposed for CS-MIM. The novel contribution of Chapter 3 are listed below:

1. In this chapter, We propose the first learning-based blind Hard-Decision (HD) CS-MIM systems by harnessing NNs. We demonstrate that the proposed DL architecture is capable of attaining near- ML performance at a significantly reduced number of cost-function evaluations.
2. Furthermore, we propose the first learning-based blind SD CS-MIM system, which is capable of achieving near-capacity performance with the aid of ISD.
3. Our capacity analysis and simulation results demonstrate that the proposed learning-based blind detector is capable of outperforming the conventional coherent detection techniques in the presence of realistic CSI estimation errors. This is achieved at a low learning-based detection complexity, which is several orders of magnitude lower than that of its ML counterpart.

Chapter 4: This chapter employs JCED techniques for improving the detection performance of both HD and SD-aided CS-MIM systems. The contributions of Chapter 4 are summarized as follows:

1. In this chapter, we propose a reduced-complexity JCED for HD CS-MIM, employing a data driven DNN. The proposed learning aided JCED method is capable of attaining near- ML performance at a low pilot overhead and complexity.

2. We then further extend this DNN-aided JCED CS-MIM scheme for producing soft information, where we combine our system with channel coding in order to attain an improved BER performance.

3. Our simulation results demonstrate that the proposed MMSE-based JCED DNN model is capable of outperforming the conventional MMSE-CE and detection scheme in different channel environments at a reduced pilot overhead, while also approaching the performance of conventional JCED, despite its reduced complexity and pilot overhead.

Chapter 5: The CS-MIM system discussed in Chapter 3 and Chapter 4 only considered single-dimensional IM. By contrast, this chapter proposes a novel CS-JMIM system. Then, learning-aided detection is proposed for both HD and SD to reduce the complexity without degrading the performance. Adaptive modulations is introduced for improving the flexibility of JMIM in different wireless channel environment. The contributions of Chapter 5 are listed as follow:

1. we propose the CS-JMIM system concept and present several JMIM mapping matrix designs. Then, we demonstrate that the proposed JMIM mapping is capable of striking an attractive trade-off between the diversity order and throughput.

2. We propose a DL-based HD detection aided CS-JMIM system that achieves near-ML performance, despite imposing significantly reduced complexity. Furthermore, we propose a DNN-aided SD detector for the proposed CS-JMIM that is capable of achieving near-capacity performance.

3. A K-nearest Neighbour (KNN) algorithm based and a DL-assisted adaptive modulation scheme is proposed for CS-JMIM. We demonstrate that the learning-assisted adaptive CS-JMIM scheme is capable of selecting more appropriate CS-JMIM mapping schemes than its conventional threshold-based adaptive counterparts. Hence it can obtain a significant throughput gain over the conventional threshold-based adaptive method.

4. Our simulation results demonstrate that the proposed learning-based detector is capable of approaching the performance of the conventional coherent detection techniques at a reduced detection complexity. We also provide the associated capacity and throughput analysis, for characterising the trade-off between each mapping scheme and the benefits of the learning-assisted adaptive method.

Chapter 2

Literature Review

In this chapter, we present the background literature review for the research detailed in later chapters in this thesis, as shown in Fig. 1.1. Section 2.1 briefly introduces the history of the IM concept and how IM techniques operate in different domains as well as highlights Multi-dimensional Index Modulation (MIM). Section 2.1 introduces the main IM techniques and MIM. The CS techniques used for IM are introduced in Section 2.2 and the detection of IM is discussed in Section 2.3. Finally, Section 2.3.5 illustrates how DL can be harnessed for IM detection.

2.1 Index Modulation

At the transmitter side of an IM-aided system, bits are mapped to the modulated signal that can be transmitted over the wireless channel in the frequency, space, time and code domains [36].

2.1.1 Spatial Modulation

The concept of SM was first proposed in 2001 [37], which was referred to as Space Shift Keying (SSK) [38]. One of the TAs was activated to transmit the signal in single Time Slot (TS) and information was embedded into the antenna index. In 2009, SM was explicitly distinguished form SSK modulation, where SM transmits a QAM symbol by the activated TA index. As shown in Fig. 2.1, it will select one of several antennas to transmit the conventional modulated signal.

An optimal Maximum Likelihood (ML) detector was proposed in [22] for SM techniques. The system structure of SM is shown in Fig. 2.2 observe that the information bits sequence b is forwarded to the 'P/S' block, which is a serial to parallel module to separate it into two parts: b_1 for QAM constellation selection and b_2 for the SM index

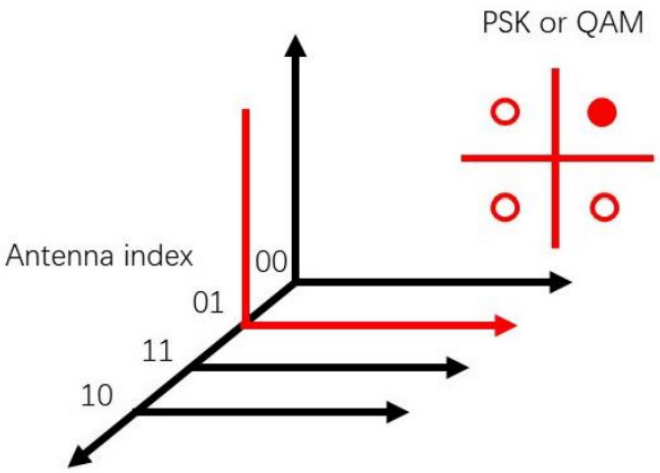

FIGURE 2.1: Constellation of SM.

selection. Then, only a single antenna is activated to transmit the signal based on the SM index selection bits. After corruption by the wireless channel and noise, the signal y is received by multiple Receive Antenna (RA)s and detect by the detection module. The research of SM has developed rapidly [39] [40]. Afterwards, SM was extended to generalized SM [38], which activates multiple TAs for the transmission of more bits.

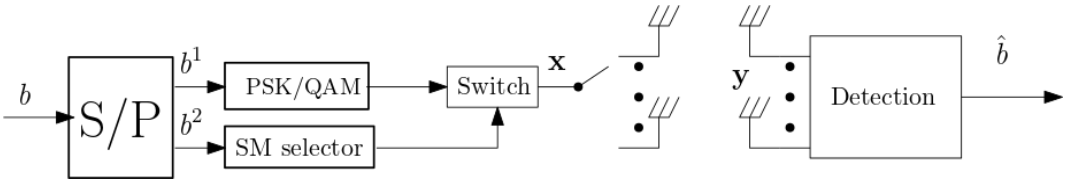

FIGURE 2.2: SM transceiver structure.

To further improve the data rate, Quadrature Spatial Modulation (QSM) [41] was proposed to fully exploit both the in-phase and quadrature dimensions, where the information bits are mapped to the in-phase and quadrature-phase separately and transmitted by two different activated antennas.

An Enhanced Spatial Modulation (ESM) scheme was then proposed in 2015 [42] to mitigate the inherent throughput loss of SM schemes imposed by the fact that only a fraction of the TAs are activated. For each TS, the number of activated antennas is variable and the modulation scheme selected for the different antennas can be different. This may lead to an enhanced throughput.

2.1.2 Frequency Domain Index Modulation

The SM concept was then also transferred to the frequency domain of OFDM, which was later termed as SIM [43] or OFDM-IM [16]. Afterwards, the concepts of SM and

SIM have been unified as the family of IM, which was then extended to both the Time Domain (TD) [12] and to the code domain. SIM can be considered as an improved version of OFDM, which maintains low Peak-to-Average Power Ratio (PAPR). As shown in Fig. 2.3, the transmitter of a SIM scheme has a similar structure to that of SM seen in Fig. 2.2 with the difference that the index selection is applied to the subcarriers of the FD. With the assumption that the length of subcarrer block is $M = 4$ and the number of activated subcarrier is $K = 1$, we can have use 2 bits for arranging 4 possible realizations of the subcarrier index selection. Then the selected QAM symbol is mapped to the activated subcarrier and amplified in the power allocation module by a factor of $\alpha = \sqrt{M/k}$. Finally, the FD symbol extended by the Cyclic Prefix (CP) will be converted into the TD signal by the Inverse Fast Fourier Transform (IFFT) module.

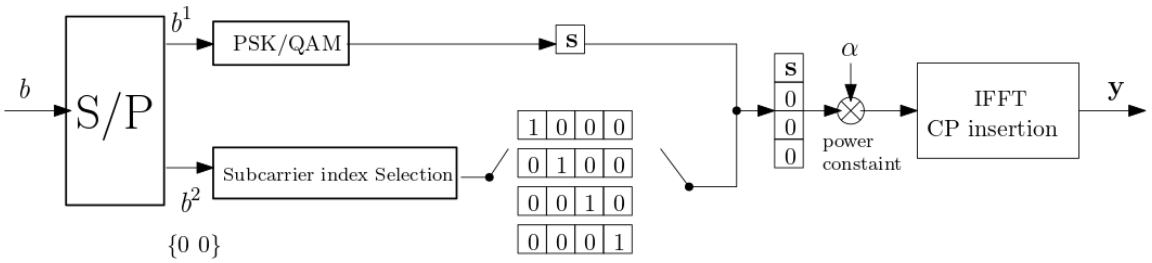

FIGURE 2.3: SIM transmitter structure associated with $M = 4$ and $K = 1$.

In 2013, Generalized Subcarrier Index Modulation (GSIM-OFDM) was designed [9] to split the OFDM subcarriers into blocks, where multiple subcarriers may be activated for striking a flexible throughput *vs.* BER trade-off. To strike flexible trade-offs, later MIMO-OFDM-IM [44] [45] was developed, which used coordinate interleaving [46] to enhance the system performance and and compressive sensing to increase both the energy and spectral efficiency [17].

2.1.3 Multi-dimensional Index Modulation

Recall that SM transmits information bits by activating TA indices, while SIM activates the OFDM subcarriers. This concept was then extended to different dimensions,leading to the concept of Multi-dimensional Index Modulation (MIM) relying on various combinations of separate IM dimension to provide higher SE and enhanced BER performance at the increased complexity [47]. In [48], the concept of Space-Time Index Modulation (ST-IM) was conceived, where each frame was split into several TSs and both SM and Time Domain Index Modulation (TD-IM) were applied to convey extra information bits by the TS index and activated TA index [49]. Furthermore, TD-IM was also combined with the Space-Time Block Codes (STBC) in [13], where only a subset of the Dispersion Matrices (DM) was activated for data transmission, where their indices carry extra information bits [18].

Space-Time Shift Keying (STSK) on the other hand combined IM in the TD and SpD, where Q DMs spread over the TD and SpD and only one of Q DMs was activated for transmission. The index of the DM carries additional information. The STSK can strike a complexity of diversity *vs.* multiplexing gain trade-off and totally avoids the ICI in STSK based systems [13].

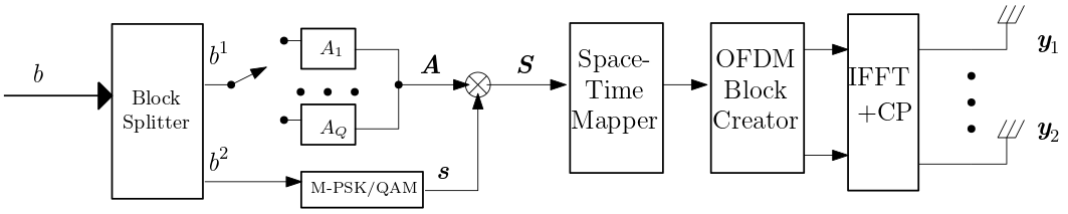

FIGURE 2.4: System structure of OFDM-STSK transmitter.

To further enhance the achievable system throughput, Generalized Space-Time Shift Keying (GSTSK) [50] [51] was developed by activating more than one DMs simultaneously, which is capable of improving the system's performance at the cost of increasing the detection complexity. The structure of GSTSK is highly flexible and it subsumes all the existing MIMO structures, such as STBC [52], BLAST, SSK and SM, as its special cases. Then a Space-Time-Frequency Shift Keying (STFSK) scheme was proposed later by combining the IM in the FD with STSK [53]. The information bits are modulated using Frequency Shift Keying (FSK) in the FD along with employing the DMs of STSK, which can achieve significant performance improvements in frequency-selective MIMO channels. Additionally, another extension to the STSK concept is differential STSK [13], where the signals are differentially encoded at the transmitter. Hence no CSI is required at the receiver [54].

Multi-Set Space-Time Shift Keying (MS-STSK) was proposed in [55], which may also be combined with SIM-OFDM techniques. In the resultant OFDM-STSK some bits select the corresponding DM, while the rest of them are mapped to the index of activated subcarriers [56]. Fig. 2.4 shows the process of the OFDM-STSK, where b bits are separated into two parts: b_1 for STSK DM selection and b_2 for QAM symbol selection. Afterwards, the ST mapper will combine conventional modulation with the DM selection into STSK symbols and embed them into the FD. Then, the IFFT converts the FD-ST symbols into TD symbol at the transmitter.

The implementation of IM in MIM systems introduces a significant level of complexity. This complexity arises from the intricate interplay between the high-dimensional signal spaces inherent in MIM and the sparse signal reconstruction demands of CS. In MIM, the modulation of indices across multiple dimensions exponentially increases the number of potential signal configurations, requiring advanced algorithms for accurate signal detection and processing. CS further complicates this scenario by requiring the recovery of these high-dimensional signals from under-sampled or incomplete data sets. The challenge is compounded by the need to maintain the fidelity of signal

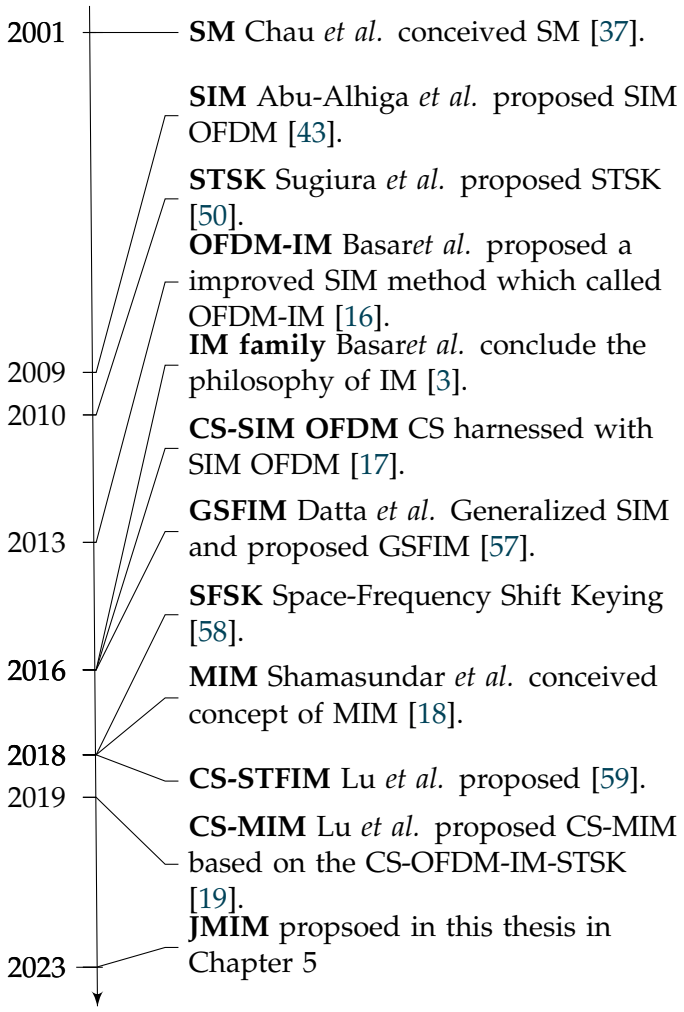

FIGURE 2.5: Milestones of the index modulation family evolving from single-dimensional index modulation to MIM.

reconstruction, while operating under the constraints of CS, such as limited number of measurements and noise sensitivity. This combination demands sophisticated optimization and signal processing techniques, often leading to increased computational requirements and algorithmic complexity. As a result, the integration of CS-MIM systems, while offering benefits in terms of spectral efficiency and data throughput, poses significant challenges in terms of computational load and algorithm design. Fig. 2.5 shows the milestone of the IM family development.

2.2 Compressed Sensing With Index Modulation

CS offers an alternative to traditional Shannon/Nyquist sampling, enabling the reconstruction of sparse or compressible signals with the aid of a reduced number of measurements [60], mitigating the transmission and storage overload [61]. CS capitalizes on the fact that many natural signals can be succinctly represented in certain bases

with a limited number of non-zero coefficients. Hence, the CS technique may be beneficially harnessed by the IM scheme, which exhibits sparsity in the index dimension. OFDM-IM was first combined with CS in [17]. Furthermore, CS techniques have also been applied for OFDM-IM relying on a MIMO scheme to improve the SE [62].

For the sake of convenience, we assume that $\boldsymbol{s} \in \mathbb{R}^N$is a K-sparse random IM signal having K non-zero entries in some transform domain. The mathematical expression of the associated received signal is formulated as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n} \tag{2.1}$$

where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a known sensing measurement matrix associated with $M < N, \boldsymbol{n} \in \mathbb{R}^{M \times 1}$ denotes AWGN noise with 0 mean and variance $\boldsymbol{N_0}$, while $\boldsymbol{y} \in \mathbb{R}^{M \times 1}$ represents linear measurements, or the received signal of a wireless system.

2.3 Detection

Detection is the process of recovering information from the transmitted signal in the face of channel corruption. In wireless systems, the radio channel is highly environment-dependent and many detrimental effects will corrupt the transmitted signal such as multi-path prorogation, at the mobility of the transmitter as well as receiver, which limits the performance of the system [36]. In this case, channel estimation plays a significant role in determining the performance of the system. Conventional channel estimation methods typically rely on pilot based estimators such as least square and minimum mean square error estimators [63]. With appropriate channel estimation, the influence of dispersive channels will be reduced and conventional full search detection methods, such as Maximum Likelihood (ML) receivers may be used to recover the bits from the received symbol.

In the following section, we will discuss channel estimation and detection conceived for IM system, including joint channel estimation and detection.

2.3.1 Channel Estimation

Transmitted signal is affected by numerous factors in the wireless channel, including the receiver's noise, which is assumed to follow the Gaussian distribution. In addition to the Additive white Gaussian Noise (AWGN), there are other factors such as path loss, shadowing and multi-path propagation [36].

Many channel estimator developed for single carrier systems may also be applied to multi-carrier systems, there is also a family of methods developed specifically for multi-carrier systems. Popular estimators used in wireless system are constituted by Least Squared Channel Estimation (LSCE) or Minimum Mean Square Error (MMSE). The LSCE estimation method may be utilized to get initial channel estimates at the pilot subcarriers, which offers acceptable performance at a reasonable complexity [64]. MMSE is optimum in terms of minimizing the Mean Square Error (MSE) of the channel estimates which is widely used in OFDM [65].

2.3.1.1 Estimation method

Typically, CE methods can be grouped into blind [36] and pilot-based [66] categories. Blind channel estimators estimate the channel based on the statistical form of the received signals and require a large amount of data, which cis hard to acquire in rapidly fast fading channels [26]. Pilot-based CE can also be classified into two main groups: data-aided and decision direct channel estimation.

In pilot symbols aided channel estimation (PSA-CE) [36], pilot symbols known by the receiver are transmitted to estimate the channel. Pilots can be either inserted into all subcarriers of a specific OFDM symbol periodically under the assumption of slow fading channels or inserted at regularly spaced subcarriers [67] [68].

Decision Directed Channel Estimation (DDCE) estimates the channel of the current symbol, while also exploiting the estimated channel information of the previous symbol [69]. These estimates are less reliable as the channel can vary severely from symbol to symbol. In this case, additional information is usually incorporated in DDCE such as periodically inserted training symbols [70].

2.3.1.2 Pilot Allocation

The specific choice of the pilot allocation method can significantly affect the estimation performance and further influence the BER performance [63]. Specifically, pilots, which are symbols known by the receiver, can be transmitted over one or more OFDM symbols. For instance, in the pilots aided channel estimation of OFDM system, the pilots are inserted systematically in the FD and multiplexed with the data [71]. For example, using constant modulus pilots can simplify the channel estimation algorithms, since the associated matrix operations become less complex [72] [73].

The design of the pilot pattern can also affect the MSE performance [74]. In [75], the CE accuracy was improved by allocating more power to the pilots than to data symbols, while the power of pilot and data symbols tends to be identical in conventional scenarios. Since there is a trade-off between the extra overhead and the accuracy of estimation,

adaptive allocation of pilots based on the channel length estimation is proposed in [75] to offer a better trade-off [66]. Generally, when using pilot symbol assisted modulation, we assume that the channel is slowly fading and then the channel estimated at the pilot symbol is used for subsequent symbols until the following CE. However, this method degrades the CE performance as the channel fading become faster [27] [26]. Increasing the rate at which the pilot symbols are sent could achieve more accurate estimated channel, while increasing the the overhead or reducing the SE [26]. Given the channel estimates at the pilot symbols, interpolation must be used between them.

2.3.1.3 Interpolation

Two of the simplest ways of interpolation are the use of piecewise constant [76] and linear interpolation approximations [77]. In piecewise linear interpolation, the channel of data symbol subcarriers is estimated using a straight line between two adjacent pilot subcarriers [64] [78]. Upon using extra information about the frequency- or time-selectivity of the channel, higher order interpolation can be used for improving the performance. The most common higher-order interpolation is spline interpolation [70], Gaussian interpolation [79] and polynomial fitting [74]. Higher order polynomials can approximate the wireless channels more accurately [80]. For highly time- and frequency- selective channels, higher polynomial order can achieve more accurate estimation of channel with higher computational complexity. Adaptive polynomial fitting performs better than static polynomial fitting, when the channels become more selective [27]. However, in slow-fading channel condition, the use of very high order polynomials may in fact degrade the performance, as the modelling misinterpret noise as the channel envelope [80].

As a further advance, researcher proposed 2-D regression models for the pilot subcarriers scattered in FD and TD [81]. In case of regression models, a 2-D polynomial with coefficients obtained using the channel correlation and the initial LSCE at the pilot subcarriers was developed in [82].

2.3.1.4 CE methods for IM

Typical IM scheme such as SM significantly reduce the ICI between the TAs by activating either only one or a small fraction of TAs at each time slot. SM has a very flexible mechanism that provides high spectral efficiency at a low complexity. To achieve these benefits of the SM, accurate CSI must be available at the receiver [83].

However, in practical SM schemes, the active antenna is determined by the spatial modulated information bits. Hence, conventional pilot-aided CE can not be direct by harnessed in this case due to having randomly activated transmit antennas. In [84],

training based CSI estimation was used, where the pilots are transmitted in advance of the frame of data symbols and linear MMSE- CE is used. In conventional CE methods used for SM, all transmit antennas have to be sequentially activated for sending pilots. At the receiver, the channel of the current activated antenna is estimated by estimation methods such as the LSCE, MMSE, Recursive Least Square (RLS) methods [85]. In [86], the authors used the difference between the channel fading coefficients of two transmit antennas to obtain the full CSI of all channels. In this case, only one transmit antenna is required to send the pilot signal to estimate the entire MIMO channel, which reduces the CE duration in comparison to conventional CE.

For OFDM-IM, the pilots are inserted systematically in the subcarriers to track the frequency selective channel [74]. Piecewise linear interpolation can also be applied to increase the performance of channel estimation [87]. This method can also be extended to combined SM and OFDM-IM systems [71].

As pilot is the only information available to estimate the channel, the performance of CE for Index Modulation (IM) will highly depend on the spacing of the inserted pilots [26]. In the next section, the conventional detection of IM systems is elaborated. Furthermore, thr concept of Joint Channel Estimation and data Detection (JCED) is introduced.

2.3.2 Hard Decision Detection

Hard-Decision (HD) detection also known as hard-in-hard-out detection detects the bits by considering the signal power over a specific threshold value as logic '1', otherwise, a logic '0'. Researchers developed several conventional detection methods for IM system under the assumption that perfect CSI is known by the receiver for MIMO systems [88]. However, it exhibits an exponential complexity escalation, when the number of dimensions increases, especially in multi-user scenarios [89].

In MIM scenarios, it is reasonable to investigate sub-optimal detector to provide a trade-off between the performance and detection complexity [18]. It was shown in [90] that some sub-optimal detectors such as Matched Filter (MF), Zero Forcing (ZF) detector and MMSE detector perform well.

Recently, an iterative greedy detection algorithm family was developed for reducing the complexity, while attaining a good performance [91]. The orthogonal matching pursuit (OMP) [92] algorithms was proposed for signal recovery from random measurements and then it was further developed to a generalized form (GOMP) [93], to Compressive Sampling-based Matching Pursuit (CoSaMP) [94], subspace-based pursuit [95] and multi-path matching pursuit [96].

2.3.3 Soft Decision Detection

Soft-decision detection utilises the symbol probabilities estimated at the output of the demodulater to produce Log Likelihood Ratio (LLR)s, which can be used to input to channel decoders for the sake of attaining near-capacity performance [97]. However, the exponentially escalating complexity of the optimal maximum a posterior probability detector rapidly becomes prohibitive [98]. This motivates the design of sub-optimal soft detectors with acceptable complexity. Iterative decoding based turbo coding was proposed in [99] [100], which was shown to achieve near-capacity performance [101] [102].

Another popular channel coding technique is constituted by the family of Low-Density Parity-Check (LDPC) codes, which were conceived by Gallager in this PhD thesis in the 1960s [103]. They were then further analysed in [104]. LDPC codes have been combined with IM to attain near-capacity performance, while utilising low complexity SD decoding [105]. Additionally, an ISD-aided CS-MIM scheme was proposed in [19].

2.3.4 Joint Channel Estimation and Detection for IM

As discussed in Section 2.3.1, when transmitting more pilots, the CE accuracy tends to improve [63]. Hence iterative methods such as Expectation-Maximization (EM) algorithm, which is a popular technique of computing Maximum Likelihood (ML) estimates can be utilised to improve the channel estimation accuracy [106]. An iterative Space-Time (ST) receiver based on the EM algorithm was proposed in [107] for transmission over Rayleigh fading channels.

In [108], an EM algorithm based iterative Joint Channel Estimation and data Detection (JCED) system was used for high speed train communication, here, the reliably detected symbols were also exploited to update the estimated channel and hence to improve the data detection performance. The JCED philosophy has been shown to work efficiently both in a ST system [109] and in a SM system [84].

2.3.5 Machine Learning Based Detection for IM

The term 'machine learning' was first coined by Arthur Samuel in 1959 and developed as a tool in artificial intelligence [110]. Machine learning has been considered in many research fields as a powerful tool, including wireless communications [34].

For example, it has been employed for detection in wireless communications in order to reduce the detection complexity [35]. In contrast to conventional model-based solutions, machine learning collects data samples for the selected algorithm without any field knowledge [111]. Given enough desired samples, a machine can be trained by the

algorithm to solve complex problems at an acceptable performance even in the face of uncertainty [112].

A typical machine learning system has three parts, which include the input, the trained model and the output. In wireless communication systems, the input data can be the SNR and other performance measures, such as the CSI [112]. Then, for this input data to be recognized by a machine learning system, it must be transformed into data exhibiting unique features processable by machine learning. Second part is constituted by the feature extraction and processing layer, where single or multiple layers can be used to process the data to achieve the desired purpose. The final part is the output layer, which is the decision layer of the whole system to be used for classification, regression, clustering, density estimation, or dimensionality reduction based on the machine learning training model target [35]. As for detection, especially IM based detection, Deep Learning (DL) may be harnessed for low complexity detection and yet accurate detection.

2.3.5.1 Deep Learning

DL is a special type of machine learning, which uses an abstract concept by considering the world as a nested hierarchy that can achieve magnificent power and flexibility [31]. The original data entered into the machine learning model could be diverse, such as images, audio or signals in communication, which require pre-processing to create data processable by the machine learning training model [110]. It has to transform the information into valid data, where the performance of the trained model is profoundly influenced by the characteristics of the processed data.

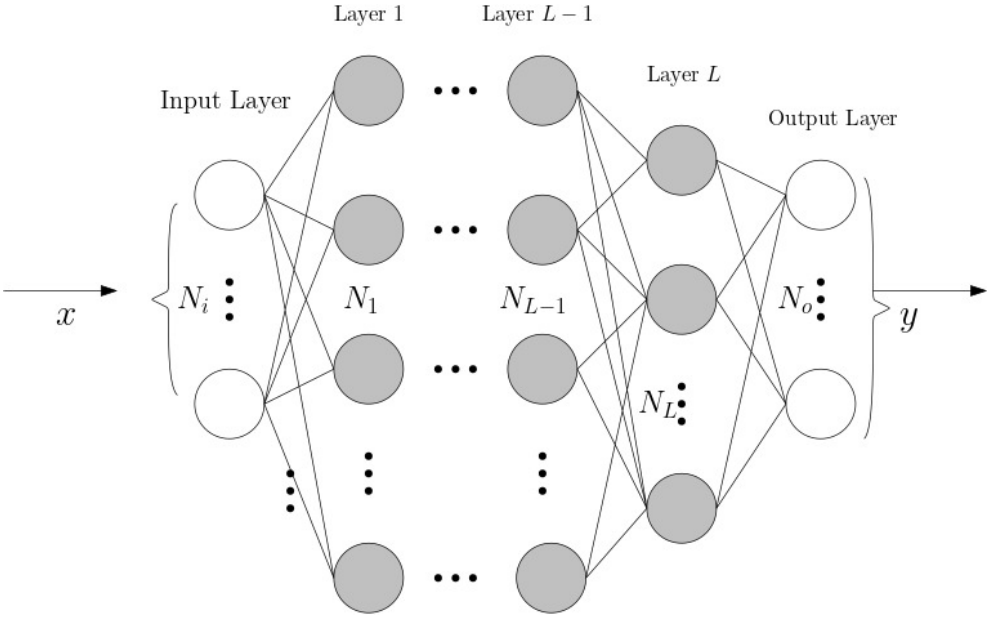

FIGURE 2.6: A fully connected Deep Neural Network (DNN) structure.

Fig. 2.6 shows an a example of the typical Fully-Connected (FC) neural network model having L layers with N_i inputs and N_o outputs. This Deep Neural Network (DNN) is characterized by the weight $\boldsymbol{W_n}$ and the bias $\boldsymbol{\theta_n}$, using a function expressed as

$$y = f_o(\boldsymbol{W_n}...f_{Relu_3}\{\boldsymbol{W_2}(f_{Relu_2}[\boldsymbol{W_1}f_{Relu1}(\boldsymbol{x}) + \boldsymbol{\theta_1}]) + \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n), \tag{2.2}$$

where f_{Relu_l} represents a Rectified linear unit (Relu) activation function for the l^{th} layer. DL can automatically extract appropriate features of raw data by directly using these data at the input layer [29]. All these processes will operate in the middle part of Fig.2.6in the so-called hidden layer. The number of hidden layers may be adjusted by the feedback of a trained model. When processing natural data, the layer will extract the dominant features from the input data and emphasizes them in support of the decision making process. When the natural data is input to the DL model, each layer extracts different features from the raw data, where the relevant features will be emphasized and irrelevant features will be gradually eliminated. Each units in each layer will connect to other units in different layer with the aid of different weights. By appropriately adjusting the number of layers, deep learning can be used to solve high-dimensional problem [113].

2.3.5.2 Learning aided Channel Estimation and Detection

DL has been applied in many areas of wireless communications such as CE [114], signal detection [29], channel decoding [115] or positioning [116].

Learning-based detection can significantly reduce the complexity of signal recovery [117] and CSI estimation [118]. In [28], the authors proposed a Deep Neural Network (DNN) for MIMO channel estimation. Indeed, several different MIMO detection network models have been developed. For instance, Samuel *et al.* [117] proposed a detection network (DetNet) which is based on deep neural network to detect MIMO signals in time varying channel. The performance of the learning-aided method of [119] is strongly related to the training data and target data, as well as to the trained model structure.

Then a parallel deep neural network was developed in [30], which improved the performance further by separating the detection problem parallel several small tasks, where the intelligence-based detection method has reduced complexity while attaining acceptable performance for an IM based system [120]. In [121], a DNN based model was proposed to detect the OFDM-IM signal. Both [122] and [123] harnessed Convolutional Neural Networks (CNN) for IM. As a further advance, a blind learning based detection was proposed in [124] for mmWave IM systems. Furthermore, a learning aided blind SD detection-aided MS-STSK mmWave system was proposed in [125].

Additionally, [19] and [54] discuss reduced complexity SD detection, while [125] investigated the combination of basic SD detection and DL methods.

Naturally, rapidly time varying channels limit the performance of learning aided blind detection unless frequent retraining is used [126]. In [114], DL was proposed for CE and signal detection in OFDM systems. Then a CNN based CE was developed for CE in MIMO systems using low-resolution Analog-to-Digital Converter (ADC) [127] and for mmWave systems [128] [129]. In [130], the authors proposed a novel NN model for iteratively updating the CSI, which outperformed the conventional NN model operating without iterative updates.

Against the above background, in this thesis, we investigate the application of DL in the detection of MIM, where we show how DL can be applied for ISD. Then we also explore the application of NN-aided CE and combine it with NN aided detection. Furthermore, DL can also deal with complex IM models, such as our proposed Joint Multi-dimensional Index Modulation (JMIM) system both in HD and SD.

Chapter 3

Deep Learning-Based Soft Iterative-Detection of Channel-Coded Compressed Sensing-Aided Multi-Dimensional Index Modulation

3.1 Introduction

As discussed in Chapter 2, the evolution of IM techniques in wireless communication systems aim to enhance system performance without incurring additional hardware cost. However, advanced schemes proposed by Shamasundar *et al.* [18] and Lu *et al.* [59], which integrate CS with STSK and OFDM-IM, while beneficial in terms of performance, introduce significant challenges in signal detection due to increased complexity, especially when considering CE. In response to these challenges, this chapter explores the potential of machine learning, particularly DL, which harnessed with IM schemes in the context of blind learning-based detection both in HD and SD. Furthermore, learning-aided CE and detection are investigated in Chapter 4 as shown in Fig. 3.1.

The classic ML detection has been widely employed for IM systems, which however suffers from an escalating complexity upon increasing the degrees of freedom or dimensions [6]. Moreover, coherent detection requires the knowledge of CSI at the receiver side, which has to be estimated by using pilot symbols, leading to a loss in effective throughput [63]. Moreover, both the pilot overhead as well as the associated CE complexity increases with the number of antennas [131] [2]. As a result, the excessive

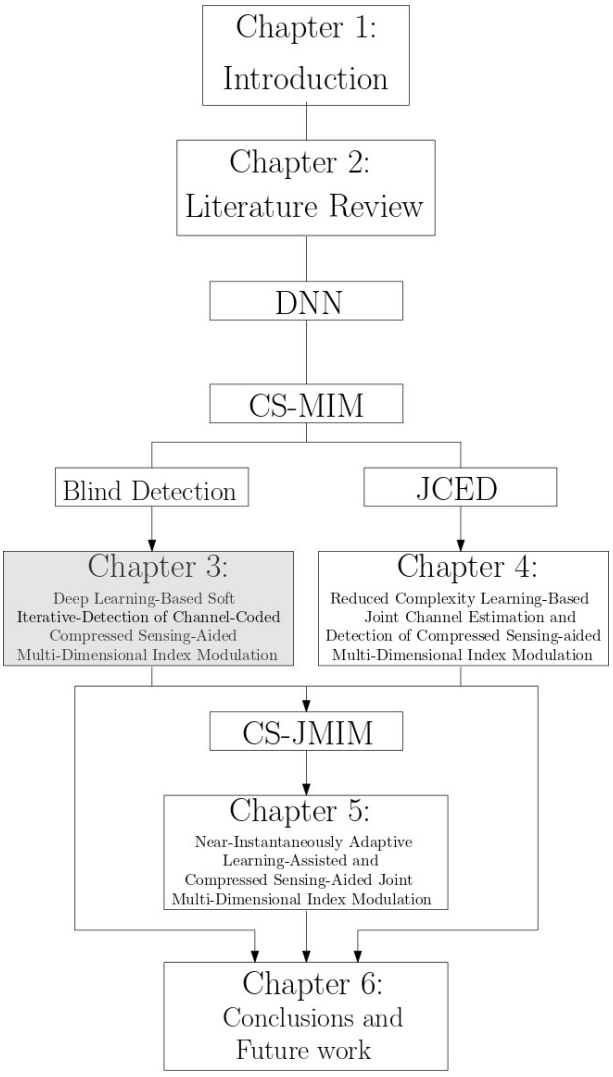

FIGURE 3.1: The organization of this thesis-Chapter3.

complexity of CE-aided ML detection of MIM schemes becomes unrealistic in practice. Furthermore, the Shannon capacity can only be approached, when SD detection is used together with the powerful tool of channel coding using ISD [100]. With the guidelines of ISD design in [132], Xu *et al.* discuss the trade-off between complexity and performance in the context of the model-based MIM solutions of [54] and [19]. In [59], CS-aided space-time-frequency IM was presented, where multiple detection stages were required for recovering the data from the constituent CS, STSK, OFDM-IM and SM schemes. As a result, when ISD is invoked, the complexity of MIM escalates both with the number of IM dimensions and with that of the ISD iterations.

Hence we set out to tackle the challenge (I) of optimizing the multi-dimensional performance, (II) eliminating both the pilot overhead and the CE complexity, and (III) reducing the ISD complexity, which cannot be solved by conventional model-based techniques.

TABLE 3.1: Boldly and explicitly contrasting our contributions to the literature

Contribution	proposed*	[17]	[105]	[19]	[120]	[138]	[124]	[125]
Index modulation	✓	✓	✓	✓			✓	✓
Multi-dimensional index modulation	✓		✓	✓			✓	✓
CS at the transmitter	✓	✓		✓				
Imperfect CSI at the receiver	✓						✓	✓
Learning aided hard detector	✓				✓		✓	✓
Soft detector	✓		✓	✓		✓		✓
Learning aided reduced complexity soft detector	✓					✓		✓
Learning-aided low-complexity iterative soft detector	✓							

Given its success in pattern recognition, in finance, marketing and health, the benefits of machine learning have also been considered in wireless scenarios as a powerful tool [112], namely in channel decoding [28], networking [133,134], and mmWave communications [135]. DL has been shown to be efficient in signal detection [136], especially in MIMO detection [35]. For instance, a DNN was used for detecting MIMO signals in [117], while Samuel and Diskin [119] proposed a DNN model for MIMO detection in a time-varying channel. Parallel DNNs were designed for separating the problem into several parallel tasks by Jin and Kim [30]. In [114], a DNN model was also harnessed for estimating an OFDM channel. Learning has also been used for channel decoding by Al-Baidhani and Fan [115]. Additionally, Xu *et al* [137] integrated the DNN model into a powerful channel coded OFDM system.

Learning-based detection has been conceived for reducing the complexity of signal recovery, while dispensing with the requirement of explicit CSI estimation. In [121], a DNN based model is proposed for detecting the OFDM-IM signal by Luong *et al.* The authors of [122] and [123] investigated CNN harnessed for IM, when the CSI is available at the input of the detector. By contrast, blind learning based detection was designed for mmWave IM by Katla *et al.* in [124] and for multi-set STSK in [125]. However, the authors of [125] only investigated the combination of basic SD detection and DL. *Against this backdrop, we propose the first blind-detection-aided ISD assisted CS-MIM schemes facilitated by learning-aided blind detection, which is capable of achieving a near-capacity performance without CSI.*

Table 3.1 boldly contrasts the novelty of this chapter to the literature. Explicitly, the contributions of this chapter are boldly and explicitly contrasted to the state-of-the-art as follows:

1. In this chapter, we propose the first learning-based blind HD CS-MIM systems by harnessing NNs. We demonstrate that the proposed DL architecture is capable of attaining near- ML performance at a significantly reduced number of cost-function evaluations.

2. Furthermore, we propose the first learning-based blind SD CS-MIM system, which is capable of achieving near-capacity performance with the aid of ISD.

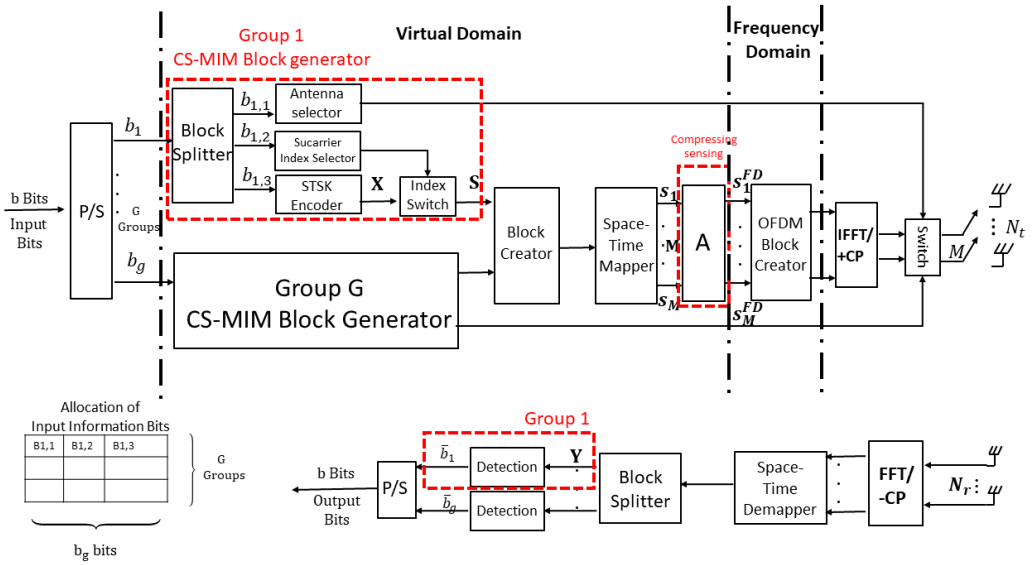

FIGURE 3.2: CS-MIM system block diagram.

3. Our capacity analysis and simulation results demonstrate that the proposed learning-based blind detector is capable of outperforming the conventional coherent detection techniques in the presence of realistic CSI estimation errors. This is achieved at a low learning-based detection complexity, which is several orders of magnitude lower than that of its ML counterpart.

The rest of this chapter is organized as follows. In Section 3.2, the system model of CS-MIM is presented. In Section 3.3, we characterize the conventional HD and SD detectors relying on the proposed learning-aided detector. Finally, our simulation results are provided in Section 3.4, while our conclusions are offered in Section 3.5.

3.2 System Model

Fig. 3.2 shows the block diagram of the CS-MIM transceiver employing N_t TAs and N_r RAs . As shown in Fig. 3.2, the input bit sequence is split into three parts: STSK in the VD[1], SIM in the FD[2] and SM in the spatial domain. We consider an OFDM symbol having N_c subcarriers, which are equally partitioned into G subcarriers groups, each containing $N_f = N_c/G$ subcarrier in the FD. In the transmitter of Fig. 3.2, the STSK symbol is allocated to the active subcarriers out of N_v subcarriers in the VD. Then, the FD signal is attained by compressing the VD signal using CS. For each group, N_f is

[1] VD is the actual domain for user data mapping/demapping, where SIM is applied before the CS process as shown in Fig. 3.2. This concept is firstly introduced in [17] to illustrate the CS techniques in IM system to improve the spectral efficiency.

[2] FD is the OFDM symbol domain after CS process that is transmitted in Fig. 3.2.

designed to be lower than N_v to increase the throughput. Following this, the OFDM-modulated data will be transmitted from the activated antennas, which is decided by the antenna selector of Fig. 3.2. At the receiver, the signal is transformed back to the subcarrier symbols and each IM group is detected separately. In the following sub-sections, we present the details of the processing stages both the transmitter and the receiver.

3.2.1 Transmitter

As shown in Fig. 3.2, the input bit sequence of b bits is split into G groups, where each group of bits b_g, $g \in [1, G]$, is input to the CS-MIM block generator. Then, each group of bits b_g is also split into three parts, $b_{g,1}$, $b_{g,2}$ and $b_{g,3}$, where $b_{g,1}$ is used for the antenna selector of Fig. 3.2. By contrast, $b_{g,2}$ is used for choosing the indices of the active subcarriers in each subcarrier group, which is called Subcarrier Index (SI) Selector. Finally, $b_{g,3}$ is used for choosing the indices of the selected DMs, which represents the STSK Encoder. The combination of the second and third parts $b_{g,2}$ and $b_{g,3}$ is identical to the CS-STFIM system of [59]. In the following we will detail the different blocks in the CS-MIM transmitter of Fig. 3.2.

3.2.1.1 SI Selection

N_c subcarriers of the OFDM symbol are divided into G groups, that corresponds to N_f subcarriers in the FD. The bit sequence $b_{g,2}$ is applied to select the active subcarrier in each group, as shown in Fig. 3.2. In each group, only K subcarriers are activated out of the N_v available subcarriers and the other subcarriers remain unused. In the following we consider an example to illustrate the subcarrier selection procedure. Explicitly, we consider the example of activating $K = 1$ subcarrier out of the $N_v = 4$ subcarriers in each group. Table 3.2 shows an example of subcarrier selection, where K_1 represents the active subcarriers and 0 represents an inactive subcarrier. Explicitly, when the input bits sequence is $b_{g,2} = [00]$, the first subcarrier will be activated, while when $b_{g,2} = [10]$ for example, the third subcarrier in the group is activated, as shown in Table 3.2.

TABLE 3.2: Example of subcarrier selection in our CS-MIM system for $K = 1, N_v = 4$

$b_{g,2}$	Indices	Allocation
[0 0]	(1)	[K_1 0 0 0]
[0 1]	(2)	[0 K_1 0 0]
[1 0]	(3)	[0 0 K_1 0]
[1 1]	(4)	[0 0 0 K_1]

3.2.1.2 STSK Encoder

The bit sequence $b_{g,3}$ of size $Klog_2(Q\mathcal{L})$ bits is fed into the STSK encoder of Fig. 3.2 to output K STSK codewords, which are mapped to the active subcarriers to $\{\boldsymbol{X}[1], \ldots, \boldsymbol{X}[i], \ldots, \boldsymbol{X}[K]\}$, where the DM spreads the information both over M TAs and over T TSs in each subcarrier. Each ST codeword $\boldsymbol{X}[i] \in \mathbb{C}^{M\times T}$ of Fig. 3.2 is generated by spreading a conventional $\mathcal{L}$-ary constellation symbol using a specific DM selected from Q available DMs. Then, the K STSK symbols generated are mapped to the K active subcarriers selected by the SI selector, while the inactive subcarriers are set to zero, which results in the ST Matrix $\boldsymbol{S}$ of Fig. 3.2.

3.2.1.3 ST Transmitter Model

After the CS-MIM block generator based processing of G groups, we obtain G ST matrices $\boldsymbol{S}$, which are then assembled by the block creator of Fig. 3.2 to form a long ST frame. Then the long ST frame is processed and output by the ST mapper for transmission from M TAs during T TSs. Fig. 3.3 shows an example ST matrix of a group, where we have $b_{g,2} = [01]$ and M1, M2 represent TA 1, 2 and T1, T2 represent TSs 1 and 2, respectively. Furthermore, N_v represents the subcarriers in the VD. Then, the ST matrix allocation becomes $\boldsymbol{S} = \{[\mathbf{0}\ \boldsymbol{X}[1]\ \mathbf{0}\ \mathbf{0}\}$ upon using $M = 2, T = 2, K = 1, N_v = 4, N_f = 2, N_t = 4, STSK(2,2,2,2,2)$. According to the example of Fig. 3.3, we can separate the ST matrix S into vector s_1 and s_2 over the antennas and TSs for first TS. Then we can obtain an STSK slice $\boldsymbol{X}_t^m[k]$ for each ST vector, where $t(t = 1, 2, \cdots, T)$ and $m(m = 1, 2, \cdots, M)$ represent the index of a TS and that of a TA, while $k(k = 1, 2, \cdots, K)$ is the index of the STSK codewords. As shown in Fig. 3.3, we can have $\boldsymbol{s}_1 = \{\mathbf{0}\ \boldsymbol{X}_1^2[1]\ \mathbf{0}\ \mathbf{0}\}$ and $\boldsymbol{s}_2 = \{\mathbf{0}\ \boldsymbol{X}_1^1[1]\ \mathbf{0}\ \mathbf{0}\}$ in a single TS.

The above processing is carried out in the VD, which is then compressed to the FD using CS, as shown in Fig. 3.3. The measurement matrix $\boldsymbol{A}$ compresses the symbols of dimension $N_v = 4$ in the VD into the symbols of dimension $N_f = 2$.The measurement matrix $\boldsymbol{A}$ of size $(N_f \times N_v)$ is applied to ST-mapped vectors for compressing the N_v-dimensional vectors $\boldsymbol{s}_m(m = 1, 2, \ldots, M)$ from the VD into the N_f-dimensional vectors $\boldsymbol{s}_m^{FD}(m = 1, 2, \ldots, M)$ in the FD, where M OFDM symbols are constructed from G groups of N_f-dimensional vectors, as shown in Fig. 3.3.

The FD symbol $\boldsymbol{s}_m^{FD}$ at the output of the CS block is evaluated as: $\boldsymbol{s}_m^{FD} = \boldsymbol{A}\boldsymbol{s}_m$. Afterwards, as in conventional OFDM, the FD symbol is transformed into the time domain symbols to be transmitted by the corresponding antennas and then a CP is added.

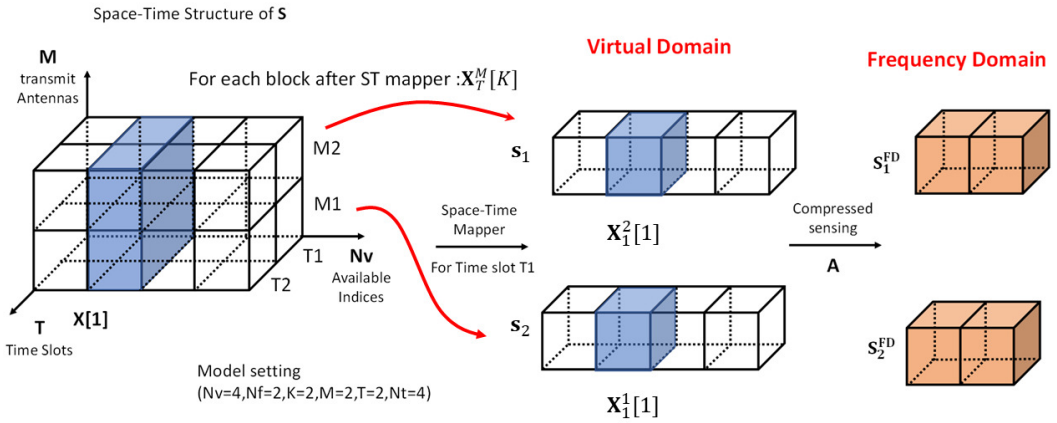

FIGURE 3.3: The process of single subcarrier group STSK symbol $\boldsymbol{S}$ with ST mapper and CS techniques employing $M = 2, T = 2, K = 1, N_v = 4, N_f = 2, N_t = 4$.

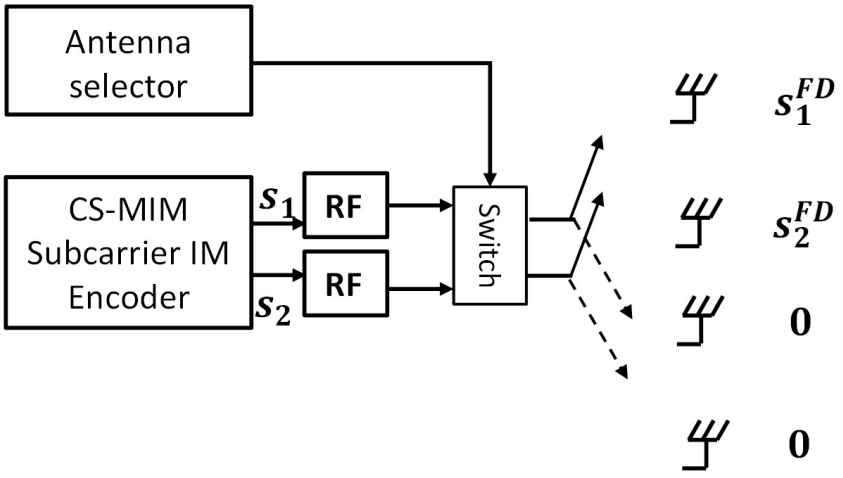

FIGURE 3.4: The process of selecting active TAs for $M = 2$ and $N_t = 4$.

3.2.1.4 Antenna Selection

After IFFT and CP, the time domain ST symbol is transmitted using M TAs during T TS using a limited number of RF chains. Similar to the concept of SM, not all antennas are activated, where the index of the active antennas is selected based on the antenna index bit sequence $b_{g,1}$, as shown in Fig. 3.2. Let us consider the FD signal, where the vectors $\boldsymbol{s}_m^{FD}(m = 1, 2, \ldots, M)$ per group at the M TAs are transmitted over the specific activated TA combinations out of N_{AC} available combinations obtained by the antenna selection. Then, the symbol vectors $\boldsymbol{s}_m^{FD}(m = 1, 2, 3, ..., M)$ in each subcarrier group are transmitted during a single TS over the data-specific activated TAs out of the N_t antennas available.

Let us consider a system using $M = 2, N_t = 4$ and $N_{AC} = 2$. As shown in Table 3.3, the antenna selection codebook will assign the RF chain to the activated TAs, while the others remain deactivated. When the input bit is $b_{g,1} = [0]$, both the first and second TAs are activated to transmit the modulated symbols in a specific subcarrier block, while the other two TAs remain inactive, as shown in Fig. 3.4. Similarly, if the incoming bit is $b_{g,1} = [1]$, then the third and fourth antennas are selected to transmit the symbols.

TABLE 3.3: A look-up table example for antenna selection in a CS-MIM system for $M = 2, N_t = 4$

$b_{g,1}$	Indices	Allocation
[0]	(1,2)	$[TA^1 TA^2\ 0\ 0]$
[1]	(3,4)	$[0\ 0\ TA^3 TA^4]$

3.2.2 Received Signal Model

The signal is assumed to be transmitted over an L_c-tap frequency selective channel, which is assumed to be known by the conventional model-based detection methods via CE [139]. Consider a $(N_r \times N_t)$- MIMO system, where N_r is the number of RAs. The signal received over the N_r RAs for the G groups over T TSs is then transformed to the FD using the FFT.

We use the channel model $\boldsymbol{h}_\alpha \in \mathbb{C}^{N_r \times N_t}$, which represents the FD channel between N_t TAs and N_r RAs. After the FFT, the FD channel matrix is $\boldsymbol{H}_\alpha \in \mathbb{C}^{N_r \times N_t}$ for $\alpha = 1, \ldots, N_f$. The ST demapper of Fig. 3.2 collects G groups of FD symbols from N_r RAs over T TSs, where the ST demapped symbols are split into G groups by the block splitter of Fig. 3.2. The group index g is omitted in the rest of the section for the sake of simplicity. Then, the symbols received by each subcarrier group are represented as $\boldsymbol{Y} = \{\boldsymbol{Y}[1]^T, \ldots, \boldsymbol{Y}[\alpha]^T, \ldots, \boldsymbol{Y}[N_f]^T\}^T$, where $\boldsymbol{Y} \in \mathbb{C}^{N_r N_f \times T}$ and $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T}$ denote the received ST signal per group and the ST symbol received at the α-th subcarrier of each subcarrier group, respectively.

Hence, the signal $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T} (\alpha = 1, \ldots, N_f)$ received during T TSs of each subcarrier group can be expressed as [19]

$$\boldsymbol{Y}[\alpha] = \boldsymbol{H}_\alpha \boldsymbol{I}_{AC} \boldsymbol{S}^{FD}[\alpha] + \boldsymbol{W}[\alpha], \tag{3.1}$$

where $\boldsymbol{S}^{FD}[\alpha] \in \mathbb{C}^{M \times T}$ denotes the ST symbols at α subcarriers transmitted from M TAs over T TSs and $\boldsymbol{W}[\alpha] \in \mathbb{C}^{N_r \times T}$ represents the AWGN obeying the distribution of $\mathcal{CN}(0, \sigma_n^2)$, and σ_n^2 is the noise variance. Finally, $\boldsymbol{I}_{AC} \in \mathbb{C}^{N_t \times M}$ denotes the $(N_t \times M)$-element sub-matrix, which describes the selection pattern of the active TAs for each subcarrier group at the transmitter. In practical model-based solutions, the channel

information is attained by employing CE techniques relying on known pilots, which is discussed in the next section.

3.3 Detection

As shown in Fig. 3.2, after the block splitter, the receiver recovers the information conveyed both by the STSK codeword, and the index of the subcarriers as well as the active TA index.

The received signal $\boldsymbol{Y}$ contains N_f ST symbols at N_f subcarriers in the FD of each subcarrier group. Given the received signal model $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T} (\alpha = 1, \ldots, N_f)$, we can rewrite $\boldsymbol{Y}$ of (3.1) in the following compact form [19]:

$$\boldsymbol{Y} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W}. \tag{3.2}$$

where $\bar{\boldsymbol{A}} \in \mathbb{C}^{MN_f \times MN_v}$ is the equivalent measurement matrix $\boldsymbol{A}$ shown in (3.1) used for compressing the M VD vectors. Similarly, $\bar{\boldsymbol{I}}_{\boldsymbol{AC}} \in \mathbb{C}^{N_r N_f \times N_i N_f}$ denotes the TA selection of N_f subcarriers in each subcarrier group. Observe from 3.1 and 3.2, that$\boldsymbol{S}^{FD}$ can be expanded as $\boldsymbol{S}^{FD} = \bar{\boldsymbol{A}}\boldsymbol{I}_{\boldsymbol{SI}}\boldsymbol{X}$, where $\boldsymbol{X} \in \mathbb{C}^{MK \times T}$ represents K STSK codewords and $\boldsymbol{I}_{\boldsymbol{SI}} \in \mathbb{C}^{MN_v \times MK}$ is the SI selection pattern.

The receiver may employ exhaustive search based ML detection, which may however lead to excessive complexity [3]. Furthermore, in case of SD, the received signal is converted into probability values, which are referred to as LLR [140], that can be forwarded to the channel decoder for attaining a near-capacity performance.

In the following section we present the conventional HD detector, where both the ML and a RC detector will be discussed. Then, we present our proposed NN aided HD detector, where the NN replaces the exhaustive search by a learning-based classification model in order to significantly reduce the complexity. Finally, upon considering SD reception, we discuss both non-iterative SD and ISD in Section3.3.2.1, where we first present the conventional SD detectors followed by our proposed non-iterative NN-aided SD and ISD receiver.

3.3.1 Hard-Decision Detection

In this section, we continue by presenting the conventional ML detector followed by RC Simultaneous Matching Pursuit (S-MP) based detector and AMP-aided detector. Then our learning-based blind detectors will be proposed, followed by the complexity study of these detectors.

3.3.1.1 ML Detection

The ML detector makes a joint decision concerning both the TA index, STSK codewords and the subcarrier activation. According to the receiver model of (3.2), $\bar{\boldsymbol{I}}_{AC}$ represents a specific realization of the active TA indices, $\boldsymbol{I}_{SI}$ is a specific realization of the active subcarrier indices in the VD of each subcarriers group and $\boldsymbol{X}$ represents K STSK codewords. To estimate the specific realization, we use $\bar{\mathcal{I}}_{AC}(\gamma)(\gamma = 1, 2, ..., N_{AC})$ to denote all the possible realizations of active TA indices. Furthermore, $\bar{\mathcal{I}}_{SI}(\beta)(\beta = 1, 2, ..., N_{SI})$ denotes all the possible realizations of the active TA indices and realizations of the active subcarrier indices. As there are $N_{q,l} = (QL)^K$ realizations of $\boldsymbol{X}$, we can use $\mathcal{X}_{q,l}(\varphi)(\varphi = 1, 2, ..., N_{q,l})$ to represent all realizations of the STSK blocks. Then, the ML detector can be modeled as [19]

$$\langle \hat{\gamma}, \hat{\beta}, \hat{\varphi} \rangle = \arg \min_{\gamma,\beta,\varphi} \|\boldsymbol{Y} - \boldsymbol{H}\bar{\mathcal{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\bar{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)\|^2, \tag{3.3}$$

where $\hat{\gamma}, \hat{\beta}$ and $\hat{\varphi}$ represent the estimates of the activated TA indices, the activated SI and the index of K STSK codewords in each subcarrier group, respectively [19]. Furthermore, $\boldsymbol{H}$ is the CSI, which is assumed to be perfectly known at the receiver. At the receiver, the ML detector carriers out a full search over all possible candidates, which has the complexity order of $\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$ per subcarrier group.

3.3.1.2 CE Methods

Conventional pilot based CE, which inserts pilots in each symbol can be ineffective due randomly activating both the subcarriers and TAs [141]. We circumvent this problem by constructing a pilot frame to estimate the CSI by the channel estimator for our CS-MIM receiver as shown in Fig. 3.5, in order to mitigate the randomness caused by the index of the antenna selection. The pilot frame has the same size as the information frame, while only one TA is activated for each subcarrier group. Here we assume that the number of subcarrier groups G and of the TAs N_t always meet the condition of $G > N_t$. Furthermore, each of the N_t TAs can be activated more than once in each frame. Then the CSI of every single TA and subcarrier group can be estimated by the channel estimator. Then, we can obtain the estimated CSI matrix $\hat{\boldsymbol{H}}$ of the equivalent subcarrier group by linear interpolation techniques [63].

Let us model the received ST pilot symbol based on (3.1) as

$$\boldsymbol{Y}_p = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\boldsymbol{S}_p + \boldsymbol{W}, \tag{3.4}$$

where the FD ST pilot symbol is $\boldsymbol{S}_p = diag\{\boldsymbol{S}_{p,1}, \boldsymbol{S}_{p,2}, \cdots, \boldsymbol{S}_{p,M}\}$, which has a fixed TA index known by the receiver of each group.

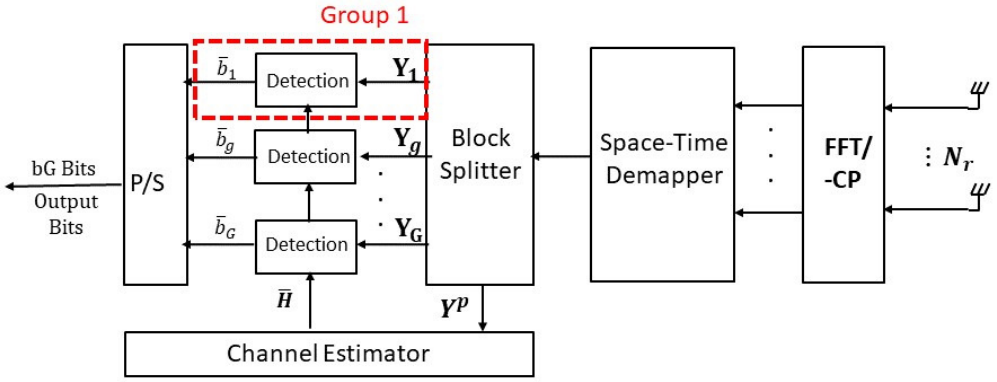

FIGURE 3.5: CS-MIM system receiver block diagram

Under the assumption that the fading envelope remains unchanged over a block of signal transmission, the LSCE and the MMSE-CE can be applied [142].

3.3.1.3 Reduced Complexity S-MP Detection

To reduce the search-space of the ML detector, a Reduced-Complexity (RC) detector was proposed in [19] based on the S-MP scheme. Here we rewrite (3.2) as:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W} = \boldsymbol{H}\bar{\boldsymbol{S}} + \boldsymbol{W} \\ &= \bar{\boldsymbol{H}}\boldsymbol{S}^{VD} + \boldsymbol{W} \\ &= \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W}, \end{aligned} \tag{3.5}$$

where $\bar{\boldsymbol{S}} \in \mathbb{C}^{N_t N_f \times T} = \bar{\boldsymbol{I}}_{AC}\boldsymbol{S}^{FD}$ denotes the equivalent data matrix that has $N_{AC}N_{SI}(Q\mathcal{L})^K$ possible realizations. Furthermore, $\bar{\boldsymbol{H}} \in \mathbb{C}^{N_r N_f \times MN_v} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}$ is used for representing the equivalent channel matrix that has N_{AC} possible realizations and $\boldsymbol{\Phi} \in \mathbb{C}^{N_r N_f \times MK}$ denotes the equivalent matrix, which represents $N_{AC}N_{SI}$ possible candidates for $\boldsymbol{I_{AC}}$ and $\boldsymbol{I_{SI}}$, including the active TAs and active subcarriers.

The S-MP algorithm can simultaneously detect the different components of $\boldsymbol{Y}$ for improving the performance. It uses the MMSE criterion to detect the TAs indices. Following this, the remaining joint signal is detected using the iterative S-MP algorithm, which has a similar structure to the conventional matching pursuit algorithm. Based on the S-MP associated with $n = N_{AC}$, the complexity order per subcarrier group becomes $\mathcal{O}(N_{AC}N_{SI} + Q\mathcal{L}K)$ [19].

3.3.1.4 AMP-Aided Detection

It has been shown in [143] that the AMP algorithm benefits from low complexity and rapid convergence. Inspired by the application of AMP in OFDM-IM [25] and GSM

[24], we apply the AMP detector for the CS-MIM system. We rewrite (3.2) as

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W} \\ &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\boldsymbol{S}^{\boldsymbol{FD}} + \boldsymbol{W}. \end{aligned} \tag{3.6}$$

To simplify the detection, we detect $\bar{\boldsymbol{I}}_{AC}$ and $\boldsymbol{S}^{\boldsymbol{FD}}$ separately. The index of antenna selection can be readily recovered by maximum signal energy estimation. Then the signal received based on the known spatial index can be formulated as

$$\boldsymbol{Y}^{FD} = \boldsymbol{H}^{FD}\boldsymbol{S}^{\boldsymbol{FD}} + \boldsymbol{W}^{FD}, \tag{3.7}$$

where $\boldsymbol{Y}^{FD} \in \mathbb{C}^{MN_f \times T}$ denotes the signal received by the activated antenna. Furthermore, $\boldsymbol{H}^{FD}$ and $\boldsymbol{W}^{FD}$ denote the activated paths' CSI and the Gaussian noise, respectively.

Firstly, let us proceed by setting the initial estimate of the FD symbol to $\boldsymbol{s}_{a,\tau}^{[0](FD)} = 0$ and the residual signal to $\boldsymbol{z}^{[0]} = \boldsymbol{y}_{a,\tau}^{FD}$. Again, since the detection procedure is the same for each subcarrier group, we also drop the group index g for the sake of simplicity. Then, the AMP employs iterative processing as follows

$$\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)} = \boldsymbol{s}_{a,\tau}^{[t](FD)} + \bar{\boldsymbol{H}}_{a,\tau}^{H(FD)}\boldsymbol{z}_{a,\tau}^{[t]}, \tag{3.8}$$

$$\boldsymbol{s}_{a,\tau}^{[t+1](FD)} = \eta(\check{\boldsymbol{s}}_{\boldsymbol{a,\tau}}^{\boldsymbol{[t](FD)}}; \boldsymbol{\sigma_{a,\tau,t}}), \tag{3.9}$$

$$\begin{aligned} \boldsymbol{z}_{a,\tau}^{[t+1]} &= \boldsymbol{y}_{a,\tau}^{FD} - \bar{\boldsymbol{H}}_{a,\tau}^{FD}\boldsymbol{s}_{a,\tau}^{[t+1](FD)} \\ &\quad + \frac{\boldsymbol{z}_{a,\tau}^{[t]}}{2N_f}\nabla\eta_{MMSE}(\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)}; \sigma_{a,\tau,t}), \end{aligned} \tag{3.10}$$

where $\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)}$ represents the updated FD modulated ST signal of antenna a and τ TS of the current iteration. Furthermore, $\eta(\boldsymbol{s}^{FD}; \sigma)$ represents the simplified MMSE denoiser discussed in [25], where σ characterizes the Root Mean Square Error (RMSE) of the FD symbol $\boldsymbol{s}^{[t](FD)}$ at iteration t. Upon using separate two-part detection, the complexity order of the AMP detection is $\mathcal{O}[N_{AC} + N_{SI}(Q\mathcal{L})^K]$.

3.3.1.5 Proposed Learning-Aided HD Detection

Although S-MP and AMP algorithm are capable of reducing the detection complexity, there exist a significant performance loss compared to the ML detector. Hence, to further reduce the detection complexity, learning based detection is considered in this section. As an additional benefit, in contrast to the conventional techniques described above, which require channel knowledge at the receiver, the proposed learning aided detector does not require any CE, hence it belongs to the family of blind algorithms.

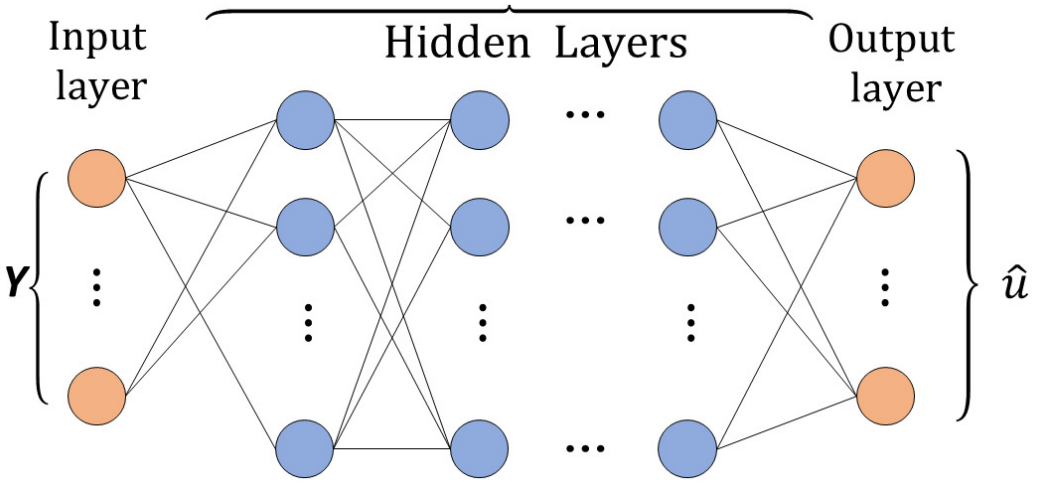

FIGURE 3.6: FC DNN model for CS-MIM detection system

First of all, HD detection may in fact be considered as a classification problem, where a single group of pre-processed symbols is input to a NN. Then the corresponding classification based candidates constitute the output. Here, to simplify the training phase, slowly varying fading channels are considered. Then, we can assume the CSI of each symbol to experience a block-fading channel.

The proposed NN model of HD detection is shown in Fig. 3.6. A single group of received symbols $\boldsymbol{Y}$ is considered as the input of this DNN model and the target output is the detected symbol $\hat{u}$. As the Long Short-Term Memory (LSTM) have strong capability in memory of long sequence data and deal with relevant information [31], to construct the associated model, we apply an LSTM in the first layer to capture the nonlinear relationship between the transmitted signal and its observations. This technique is eminently suitable for processing sequential data in DL and can be characterized by

$$\{C_k, z_k\} = LSTM(C_{k-1}, z_{k-1}, x_k; \phi_{k-1}), \tag{3.11}$$

where C_k is also often termed as the cell state [144] and represents the information flow versus time. Furthermore, x_k and z_k are the input and output at symbol instant k, while z_{k-1} is the input at k and ϕ_{k-1} is the parameter of LSTM, which is stored in the cell for the next state and shared with the next iteration.

Following this, the classic softmax activation is used for generating dependent probabilities at the output layer of our classification problem. Considering the weight $\boldsymbol{W_n}$ and the bias $\boldsymbol{\theta_n}$, the resultant function can be expressed as

$$\begin{aligned}\hat{u} =& f_{softmax}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{y}) + \boldsymbol{\theta_1}]) \\ &+ \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n).\end{aligned} \tag{3.12}$$

In order to increase the training accuracy, further input data pre-processing is required in the DNN which includes appropriately modifying the structure of $\boldsymbol{Y}$ before feeding it into the training model for locating the target activated subcarrier indices and STSK codewords.

The raw input data in complex-valued matrix form obtained from the received signal $\boldsymbol{Y}$ has to be vectorized first. We rearrange the complex values by separately extracting the real and imaginary parts and merging them into a real-valued vector. Then, we consider a slow time-varying channel so that the model can be trained with training samples and subsequently used for signal detection. In the training phase, we use randomly generated data, which are transmitted over the wireless channel using MIM, and then processed as the input data of the DNN. The number of training samples required is selected based on experimentation by gradually increasing the training size until acceptable MSE values are achieved. In each iteration, the classic back-propagation updates the weight values to compute the gradient of the loss function in the process of gradient descent. The cross-entropy-based cost function used for the training is

$$\mathcal{L}(\boldsymbol{u}_i, \hat{\boldsymbol{u}}_i) = -\sum_{i=1}^{B} \boldsymbol{u}_i \log(\hat{\boldsymbol{u}}_i), \tag{3.13}$$

where $\boldsymbol{u}$ represents the target class of corresponds bits, $\hat{\boldsymbol{u}}$ denotes the achieved class, which can be transformed into equivalent detected bits and B is the sample size of the current iteration. A stopping criterion can be defined either by the number of iterations or by an MSE threshold. Then, the parameter sets $\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\}$ can be updated in each training iteration based on the learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\} \leftarrow \{\boldsymbol{W}_n, \boldsymbol{\theta}_n\} - \alpha \nabla \boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\}),$$

where $\alpha > 0$ is the learning rate and $\nabla \boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\})$. In our proposed NN aided detection, we use $\alpha = 0.001$.

After the training phase, the DNN model learns the mapping from the received signal and stores both the weight as well as the bias information, which will be used for producing the desired outputs based on the real-life input data in the testing phase. The statistical properties of the input/output data have to remain the same as those used in training.

The detection complexity of the learning algorithm is dominated by the calculation of the layer weights and biases, which may be considered to be $\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$ [125], with n representing the number of neurons in each layer. Their complexity order is significantly lower than that of the ML or the S-MP detection schemes.

$$\begin{aligned}\boldsymbol{L}_e(u_l) &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) exp[\sum_{j\neq l} u_j L_a(u_j)]}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_0^l}} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) exp[\sum_{j\neq l} b_j L_a(b_j)]} \\ &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}} exp[-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2/N_0 + \sum_{j\neq l} u_j L_a(u_j)]}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_0^l}} exp[-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2/N_0 + \sum_{j\neq l} u_j L_a(u_j)]}.\end{aligned} \tag{3.15}$$

3.3.2 Soft-Decision Detection

SD detection is employed for achieving near-capacity performance when combined with channel coding, with or without ISD between the MIMO demapper and the channel decoder. Turbo coding was proposed in [99] and has been extensively used in wireless systems [100]. However, the complexity of the optimal maximum *a posteriori* probability MIMO detector rapidly becomes prohibitive upon increasing the modulation order and the number of antennas [98]. This motivates the design of sub-optimal SD detectors. In light of this motivation, we propose learning aided SD detection, which is capable of providing near-optimal performance at a reduced complexity.

3.3.2.1 Conventional SD Detection

A CS-MIM ISD scheme is shown in Fig. 3.7. At the transmitter side of Fig. 3.7, the information bit sequence b is encoded by a RSC encoder. The coded bit sequence c is interleaved to generate the interleaved stream u, which is entered into the CS-MIM modulator of Fig. 3.2.

The LLR is defined as the probability ratio of the bit being '1' and '0', which can be written as $L(b) = log\frac{p(b=1)}{p(b=0)}$. The conditional probability $p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})$ of receiving the group signal $\mathbf{Y}$ is given by [145]

$$\begin{aligned}&p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) \\ &= \frac{1}{(\pi \boldsymbol{N}_0)^{NT}} exp(-\frac{||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2}{N_0}),\end{aligned} \tag{3.14}$$

where $\mathcal{X}_{\gamma,\beta,\varphi}$ represents the STSK codewords at the β-th realization of active subcarriers, which are transmitted through the γ-th realization of the active TA. Furthermore, N_0 is the noise power, where we have $\sigma_n^2 = N_0/2$. The equivalent received signal $\boldsymbol{Y}$ per subcarrier group carries $\boldsymbol{B}$ channel-coded bits $\boldsymbol{u} = [u_1, u_2, ..., u_B]$ and the extrinsic LLR of bits $u_l (l = 1, 2, \cdots, B)$ is expressed by (3.15) [19]. In (3.15), $\mathcal{X}^l{}_1$ and $\mathcal{X}^l{}_0$ represent a subset of the legitimate equivalent signal $\mathcal{X}$ corresponding to bit u_l when $u_l = 1$ and $u_l = 0$, respectively, yielding $X_1^l \equiv \{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X} : u_l = 1\}$ and

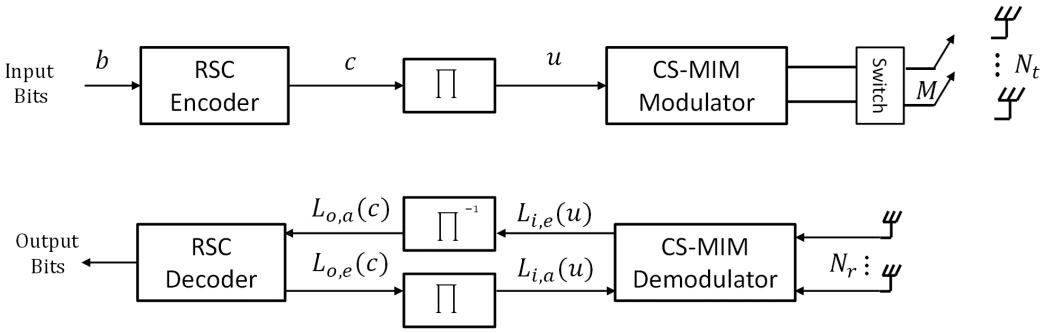

FIGURE 3.7: The transceiver architecture of SD aided CS-MIM employing iterative extrinsic information exchange between the CS-MIM detector and the RSC decoder.

$X_0^l \equiv \{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X} : u_l = 0\}$. Therefore, the extrinsic LLR of bit $u_l (i = 1, 2..., B)$ can be expressed by (3.15). The variable $L_a()$ in (3.15) represents the *a priori* LLR fed back from the RSC decoder to the demodulator.

First of all, the Approx-Log-MAP of (3.17) can be simplified by the Maximum Log-MAP (Max-Log-MAP) algorithm [99]

$$\boldsymbol{L}_e(u_l) = max_{\mathcal{X}_{\gamma,\beta,\varphi \in \mathcal{X}_1^l}}(\lambda_{\gamma,\beta,\varphi}) - max_{\mathcal{X}_{\gamma,\beta,\varphi \in \mathcal{X}_0^l}}(\lambda_{\gamma,\beta,\varphi}). \tag{3.16}$$

The expression in (3.15) can be further simplified by the Approximate Log Maximum A Posteriori (Approx-Log-MAP) algorithm based on the Jacobian Maximum operation [146] [147] as

$$\boldsymbol{L}_e(u_l) = \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi \in \mathcal{X}_1^l}}(\lambda_{\gamma,\beta,\varphi}) - \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi \in \mathcal{X}_0^l}}(\lambda_{\gamma,\beta,\varphi}), \tag{3.17}$$

where jac$(.)$ denotes the Jacobian maximum operation and the intrinsic metric of $\lambda_{\gamma,\beta,\varphi}$ is

$$\begin{aligned}\lambda_{\gamma,\beta,\varphi} = &(-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2)/N_0 \\ &+ \sum_{j \neq l} u_j L_a(u_j).\end{aligned} \tag{3.18}$$

At the receiver, the soft demodulator evaluates the probability of each bit being logical '1' as well as '0', and applies the Approx-Log-MAP algorithm for obtaining the extrinsic LLR of the coded bits. This has a complexity order of $\mathcal{O}[2^{(c_g)}(N_{AC}N_{SI}(Q\mathcal{L})^K)]$, where c_g represents the coded bits after the RSC encoder and interleaver. This complexity will be reduced by our proposed DNN based SD.

3.3.2.2 Single-Stage DNN-Aided SD Detector

Given the high computation complexity of the CS-MIM SD detection, in this section we propose a reduced-complexity SD detector using NNs. In the conventional iterative soft-detector described in Section 3.3.2.1, the extrinsic LLRs $L_{i,e}(u)$ are updated in

every iteration by utilizing the *a priori* LLRs $L_{i,a}(u)$. By modifying (3.17) and (3.18), the extrinsic LLR of every iteration can be expressed as [19]

$$\boldsymbol{L}_e^n(u_l) = jac_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_1^l}(\lambda_{\gamma,\beta,\varphi}^n) - jac_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_0^l}(\lambda_{\gamma,\beta,\varphi}^n), \tag{3.19}$$

where $n_{it} = 0, 1, 2, ..., I_t$ is the iteration index and I_t is the maximum affordable number of iterations. The intrinsic metric of each iteration is

$$\begin{aligned}\lambda_{\gamma,\beta,\varphi}^n = &(-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2)/N_0 \\ &+ \sum_{j\neq l} u_j L_a^n(u_j).\end{aligned} \tag{3.20}$$

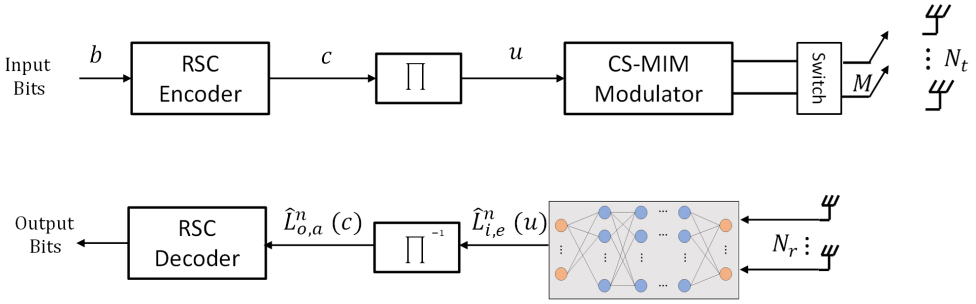

FIGURE 3.8: Structure of the learning-aided SD detector of CS-MIM for single-stage DNN-aided SD detection

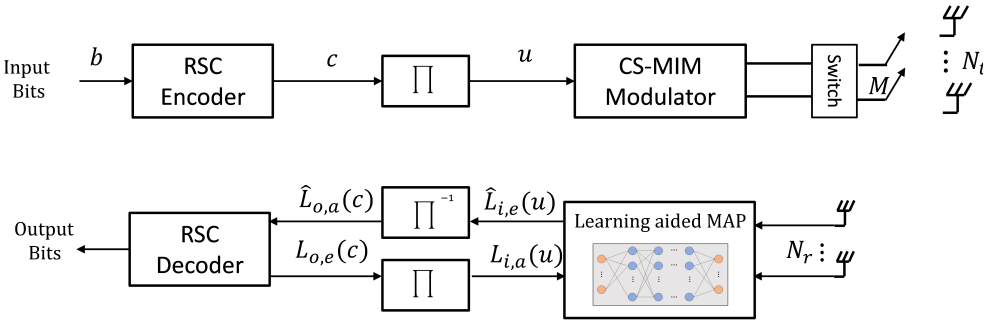

FIGURE 3.9: Structure of the learning-aided SD detector of iterative DNN-aided MAP SD detection

Firstly, we consider non-iterative DNN-aided SD detection, where there is no information exchange between the RSC decoder and the CS-MIM demodulator. As shown in Fig. 3.8, the DNN architecture is employed for replacing the CS-MIM demodulator. The output of the NN model should be $L_e(u) = \boldsymbol{L}_e^0(u)$, which denotes the first LLRs after the soft demodulator without any *a priori* feedback information. Following this, the SD is characterized as follows. The received signal $\mathbf{Y}$ undergoing the same pre-processing as in Section 3.3.1.5 forms the input of the NN and the aim here is to accurately estimate the value of the extrinsic LLR $L_{i,e}(u)$. Hence, we harness the classic regression model of [148] formulated as:

$$\begin{aligned}\hat{L}_e(u) =& \boldsymbol{W_n}...f_{Relu}(\boldsymbol{W_2}(f_{Relu}(\boldsymbol{W_1}f_{Relu}(\boldsymbol{y}) + \boldsymbol{\theta_1}) \\ &+ \boldsymbol{\theta_2}) + ... + \boldsymbol{\theta}_n),\end{aligned} \tag{3.21}$$

where $\boldsymbol{W}_n$ and $\boldsymbol{\theta}_n$, $(n = 1, 2, ..., N)$ represent the weights and bias, respectively, of the n-th layer of the NNs and $\hat{L}_e(u)$ represents the estimated extrinsic LLR information at the output of the demodulator.

As discussed in Section 3.3.1.5, the NN updates the weights $\boldsymbol{W_n}$ and $\boldsymbol{\theta_n}$ of each layer by back propagation in each iteration of the training process, to minimize the error between the given LLR and the predicted LLR. Here, we use the following loss function:

$$\mathcal{L}(\boldsymbol{L}_i, \hat{\boldsymbol{L}}_{\boldsymbol{i}}; \boldsymbol{W}_i, \boldsymbol{\theta_i}) = \frac{1}{B}\sum_{i=1}^{S} \|\boldsymbol{L_i} - \hat{\boldsymbol{L}}_i\| + \sum_{i=1} \rho_n \|\boldsymbol{W}_n\|_2^2, \tag{3.22}$$

where B is the number of training samples, $\boldsymbol{L_i}$ and $\hat{\boldsymbol{L}}_{\boldsymbol{i}}$ represent the target LLR of (3.17 and the predicted LLR, respectively of the i-th training sets and the regularization factors ρ_n are introduced for avoiding a over-fitting [111].

The assumption of slow fading channel which is applied in the HD DNN training phase is also considered here, so that the learning based detector could learn the mapping of the input data and predict the LLR value directly from the received signal Y without any channel information in the training phase. Hence, the NN-aided detector is a blind detector, as in the HD case.

When using iterative extrinsic information exchange between the RSC decoder and our MIMO detector, the LLRs $L_e^n(u_l)(n = 1, 2, 3, \cdots)$ can be typically improved over iterations. We consider to use the updated extrinsic LLR $L_{i,e}^n$ after several iterations as our training target, with the received signal Y employing the same pre-processing as used for the non-iterative method for its input. The test model predicts the LLRs gleaned from the received signal $\mathbf{Y}$ in the conventional iterative soft-detector after several iterations.

Fig. 3.8 also shows the architecture of this model, which has a similar structure to that of the non-iterative SD model, but the output HDs are replaced by extrinsic LLRs.

3.3.2.3 Iterative DNN-Aided MAP SD Detection

In this section, we activate and optimize the DNN for each iteration in the ISD assisted channel coded system.

In order to incorporate both (3.16) and (3.17), the extrinsic LLRs can be partitioned according to the legitimate bits '1' $\boldsymbol{L}_e^1$ and '0' $\boldsymbol{L}_e^0$ corresponding to an intrinsic metric

$\lambda_{\gamma,\beta,\varphi}$. Then $\boldsymbol{L}_e^1$ and $\boldsymbol{L}_e^0$ can be extended into the equivalent MAP function $f_{map}(.)$, which is expressed as:

$$\boldsymbol{L}_e(u_l) = \boldsymbol{L}_{\boldsymbol{e}}^{\boldsymbol{1}} - \boldsymbol{L}_{\boldsymbol{e}}^{\boldsymbol{0}} = f_{map}^1(\lambda_{\gamma,\beta,\varphi}) - f_{map}^2(\lambda_{\gamma,\beta,\varphi}), \tag{3.23}$$

where $f_{map}^m(.)(m = 1, 2)$ represents the Max-Log-MAP algorithm [19] or the Approx-Log-MAP relying on the simplified Jacobian Maximum operation. With the aid of (3.20), the intrinsic metric in the equivalent MAP can be extended as:

$$f_{map}^m(\lambda_{\gamma,\beta,\varphi}) = \bar{f}_m(\boldsymbol{y} + \bar{\boldsymbol{L}}_{\boldsymbol{a}}), \tag{3.24}$$

where $\bar{\boldsymbol{L}}_{\boldsymbol{a}}$ denotes the equivalent *a priori* LLR gleaned from the decoder after modification, $\bar{f}_m(.)(m = 1, 2)$ represents the modified MAP function from (3.23) and $\boldsymbol{y}$ is the received signal.

Our proposed learning method aims for replacing the $\bar{f}_m(.)$ function, where we use a pair of DNN models for replacing the functions of $\bar{f}_1(\boldsymbol{y} + \bar{\boldsymbol{L}}_{\boldsymbol{a}})$ and $\bar{f}_2(\boldsymbol{y} + \bar{\boldsymbol{L}}_{\boldsymbol{a}})$, which can be expressed as:

$$\boldsymbol{L}_e(u_l) = \boldsymbol{L}_{\boldsymbol{e}}^{\boldsymbol{1}} - \boldsymbol{L}_{\boldsymbol{e}}^{\boldsymbol{0}} = NN_1(\boldsymbol{y}, \boldsymbol{L}_{\boldsymbol{a}}) - NN_2(\boldsymbol{y}, \boldsymbol{L}_{\boldsymbol{a}}), \tag{3.25}$$

where $NN_m(.)(m = 1, 2)$ represent the DNN model replacing the modified MAP function from (3.24).

As shown in Fig. 3.9, the received signal $\boldsymbol{y}$ and *a priori* LLR $L_{i,a}(u)$ feedback from the RSC decoder after interleaving will become the input data of the DNN model.

In this case, the input data have three labels, which include the real part of the received signal $Real(\boldsymbol{y})$, the imaginary part $Imag(\boldsymbol{y})$ and the *a priori* information L_a. In the training phase, both NN models, which are designed to replace the Approx-Log-MAP process of evaluated extrinsic LLR $L_{i,e}$, are fed with the *a priori* LLR $L_{i,a}$ of different iterations and the received signal.

For the input data sets $\boldsymbol{\epsilon} = \boldsymbol{y}, \boldsymbol{L}_{\boldsymbol{a}}$, the output of the proposed DNN architecture can be expressed as:

$$\hat{L}_e(u) = \boldsymbol{W_n}...f_{Relu}(\boldsymbol{W_2}(f_{Relu}(\boldsymbol{W_1}f_{Relu}(\boldsymbol{\epsilon}) + \boldsymbol{\theta_1}) + \boldsymbol{\theta_2}) + ... + \boldsymbol{\theta}_n). \tag{3.26}$$

Similarly, the normalized loss function L_2 can be applied for both NN_1 and NN_2:

$$\mathcal{L}(\boldsymbol{L}_i^m, \hat{\boldsymbol{L}}_{\boldsymbol{i}}; \boldsymbol{W}_i, \boldsymbol{\theta_i}) = \frac{1}{B}\sum_{i=1}^{S} \|\boldsymbol{L}_i^m - \boldsymbol{L}_i^{\hat{m}}\| + \sum_{i=1} \rho_n \|\boldsymbol{W}_n\|_2^2, \tag{3.27}$$

where B is the number of training samples, $(m = 1, 2)$ denote the index of the two NN models, $\boldsymbol{L}_i^m$ and $\hat{\boldsymbol{L}_i^m}$ representing the target LLR of (3.17) and the predicted LLR,

respectively of the i-th training sets. The regularization factors ρ_n are introduced for avoiding over-fitting [111].

The assumption of slow fading channel which is applied in the HD DNN training phase is also considered here, so that the learning based detector could learn the mapping of the input data and predict the LLR value directly from the received signal Y without any channel information in the training phase. Hence, the NN-aided detector is a blind detector, as in the HD case.

3.4 Performance Analysis

In this section, we characterize the performance of our learning-aided CS-MIM system relying on both HD and SD. For all learning-aided models, we assume that the system's signalling rate is $100MBaud$ and the maximum Doppler frequency is $100Hz$, which corresponds to a normalised Doppler frequency f_m of 10^{-6}. The performance results of the conventional model-based solutions relying on practical CE will be portrayed in this section as benchmarks to the proposed schemes. The simulation parameters shared by CE-aided and blind-detection-aided CS-MIM schemes are summarized in Table VI, while the simulation parameters of the learning-based blind detection scheme are specified in Table 3.4. In summary, the following six schemes are compared in this section:

1. **Scheme 1**: ML HD aided CS-MIM system based on
   (a) idealistic perfect CSI;
   (b) practical MMSE-CE.
2. **Scheme 2**: HD aided CS-MIM system relying on perfect CSI using
   (a) S-MP based RC detector.
   (b) AMP aided detector.
3. **Scheme 3**: HD DNN-aided blind detector employed for the CS-MIM system. No CSI is required for the detection.
4. **Scheme 4**: Conventional SD detector employed for the CS-MIM system relying on RSC channel coding. Similarly, we also consider different CSI conditions at the receiver as follows:
   (a) non-iterative (0 iteration) SD detector of CS-MIM under perfect CSI at the receiver.
   (b) non-iterative (0 iteration) SD detector of CS-MIM under imperfect CSI at the receiver.

(c) ISD detector of CS-MIM using 1-5 iterations under perfect CSI at the receiver.

(d) ISD detector of CS-MIM using 1 iteration under imperfect CSI at the receiver.

5. **Scheme 5**: Single-stage SD- DNN-aided detector employed for the CS-MIM system. In this case, the information gleaned from the channel decoder will be fed back to the soft demodulator as *a priori* LLRs.

   (a) Single-stage SD- DNN-aided detector of CS-MIM using non-iterative (0-iteration) soft demodulated LLRs as training target.

   (b) Single-stage SD- DNN-aided detector of CS-MIM without ISD iterations but using the soft demodulated LLRs obtained from 1-5 ISD iterations as training target.

6. **Scheme 6**: DNN-aided MAP SD detector of CS-MIM with 1-3 ISD iterations.

TABLE 3.4: CS-MIM system simulation parameters.

Parameters	Values
Multi-carrier System	OFDM
Number of subcarriers, N_c	128
CP length	16
Number of subcarrier groups, G	16
Number of subcarrier/group, M_g	8
Number of available indices/group, N_a	16
Number of active indices/group, K	2
STSK, $(M, N, T, Q, \mathcal{L})$	(2,2,2,2,2)
TAs, N_t	8
RAs, N_r	8
RSC code, (n, k, K)	(2,1,3)

The Discrete-Input Continuous-Output Memoryless Channels (DCMC) capacity of CS-MIM based on the equivalent received signal model of (3.7) can be expressed as [132]

$$C^{DCMC}(SNR) = \max_{p(\bar{\boldsymbol{S}}^i|_{i=1}^I} \frac{1}{N_f} \sum_{i=1}^{I} \int p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^i) \log_2 \frac{p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^i)}{\sum_{j=1}^{I} p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^j) p(\bar{\boldsymbol{S}}^j)} d\boldsymbol{Y}, \tag{3.28}$$

where $I = N_{AC} N_{SI} (QL)^K$ denotes the total number of realizations in each subcarrier groups including the active TAs, active subcarriers and K STSK codewords at the activated subcarrier. The DCMC capacity of (3.29) is maximized, when the realizations of each subcarrier group are equiprobable with $\{p(\bar{\boldsymbol{S}}^i) = \frac{1}{I}\}_{i=1}^{I}$. For fading channels, contaminated by the noise, we have

$$p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^i) = \frac{1}{(_o)^{N_r N_f}} (-\frac{||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{S}}^i||^2}{N_0}). \tag{3.29}$$

TABLE 3.5: Configuration of the system presented in Section 3.3.1

No.	Scheme	Detection condition	iterations
HD		throughput,R_t	1.333
1	ML	Perfect CSI at receiver Imperfect CSI at receiver	
2	S-MP based RC detector AMP aided detector	Perfect CSI at receiver Perfect CSI at receiver	
3	DNN-aided	Blind	
SD		throughput,R_t	0.6667
4	ML	Perfect CSI at receiver Imperfect CSI at receiver Perfect CSI at receiver Imperfect CSI at receiver	0 0 1-5 1
5	Single-stage DNN-aided	Blind Blind	0 1-5
6	DNN-aided MAP detector	Blind	1-3

TABLE 3.6: Training configuration for learning-aided method

Setting	Hard-decision	Soft-decision
Hidden layers	LSTM+FC+FC	LSTM+FC+FC+FC
Activation function	ReLu	
Output layer	softmax	\
Number neurons	50	50
Input size	4x(128x2)	
Output size	1	12
Maximum training epoch	400	1000
Initial learning rate	0.001	
Target SNR for training	0dB-20dB	-5dB to 5dB
Mini batch size	1000	200 to 500
Optimizer	Adam	
Training data size	50000	
Validation data ratio	0.1	

Then we can further simplify the DCMC capacity of (3.29) to:

$$C^{DCMC}(SNR) = \frac{1}{IN_f}\sum_{i=1}^{I}\mathbb{E}\{\log_2[\frac{I.p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^{\boldsymbol{i}})}{\sum_{j=1}^{I}p(\boldsymbol{Y}|\bar{\boldsymbol{S}}^{\boldsymbol{j}})}]\} \tag{3.30}$$

$$= R - \frac{1}{IN_f}\mathbb{E}\{\log_2[\sum_{j=1}^{I}\exp\left(\Psi_{i,j}\right)]\}, \tag{3.31}$$

where we have $\Psi_{i,j} = \frac{-||\boldsymbol{H}(\bar{\boldsymbol{S}}^i-\bar{\boldsymbol{S}}^j)+\boldsymbol{W}||^2}{N_0}$. The system has the throughput of $R = \frac{log_2 I}{N_f}$ approaching the DCMC capacity as the SNR is increased. Given the system parameters of Table 3.4, the achievable rate of CS-MIM is $R = \frac{b_g G}{N_c+L_{CP}} = 1.333$ bits/sec/Hz. Fig. 3.10 characterizes the theoretical DCMC capacity curves for both the proposed HD NN

detector and conventional ML detector. Then the maximum rate of this system will be achieved at 1.42dB of SNR. To acquire more accurate results, the pilot symbols may impose 1% to 10% throughput loss. By contrast, when considering our proposed DNN aided HD detector of Scheme 3, no CSI is needed, hence no pilot overhead is imposed. Fig. 3.11 also shows the BER of the conventional HD detector of Scheme 1 as well as Scheme 2 and the proposed HD NN based detector which is Scheme 3, as shown in Table 3.5.

As mentioned in Section 3.2.2, the classic MMSE-CE is applied for estimating the CSI and deliver to ML detector at the receiver. As shown in Fig. 3.11, a high overhead provides the detector with more accurate CSI, but the achievable throughput is reduced owing to the increased pilot overhead. As shown in Fig. 3.11, the proposed DNN aided HD detector achieves a BER of 10^{-5} at 7.9 dB SNR, while the ML detector relying on the idealized simplifying assumption of having perfect CSI knowledge requires an SNR of about 3.6 dB at the same BER. There is a 4.5 dB SNR difference between the conventional ML method and DNN aided method. The performance is slightly reduced, when considering practical CE methods such as MMSE-CE. More explicitly, as shown in Fig. 3.11, although the MMSE-CE aided ML detector could achieve a BER of 10^{-5} at a low SNR loss of 0.7 dB, 10% overhead is required to estimate the channel. Furthermore, the benchmarks of S-MP and AMP are capable of mitigating the excessive complexity of the full-search-based ML method at the costs of substantially eroded performance, as evidenced by Fig. 3.11. The DNN method is capable of achieving better performance without overhead compared to both RC methods, as shown in Fig. 3.11. Furthermore, we quantify the computational complexity order of each method, as shown in Table 3.7, observe that the DNN-aided method exhibits lower complexity than both S-MP and AMP.

We also compare the HD learning based method with ML detection under slow fading channels. The normalized maximum Doppler frequency f_m is used to adjust the speed of fading variation. In Fig 3.12, we consider $f_m = 10^{-6}$ as a very slow fading channel, $f_m = 10^{-5}$ corresponding to a slow fading channel and $f_m = 10^{-4}$ as a fast fading channel. For the DNN detector, we train separate models with the aid of the received signal under different channel conditions for more accurate detection performance. In this case, we have to introduce more training samples for faster fading channels to outperform the conventional methods. Fig. 3.12,Fig. 3.13 and Fig. 3.14 demonstrate that the performance of the coherent scheme relying on the same pilot percentage degrades as f_m increases. By contrast, the learning-assisted HD detector that dispenses with CE is capable of achieving a robust performance against f_m and always achieves better performance than both the S-MP and AMP methods, as confirmed by Fig. 3.12, Fig. 3.13 and Fig. 3.14. Nonetheless, Fig. 3.14 demonstrates that when the fading channel varies

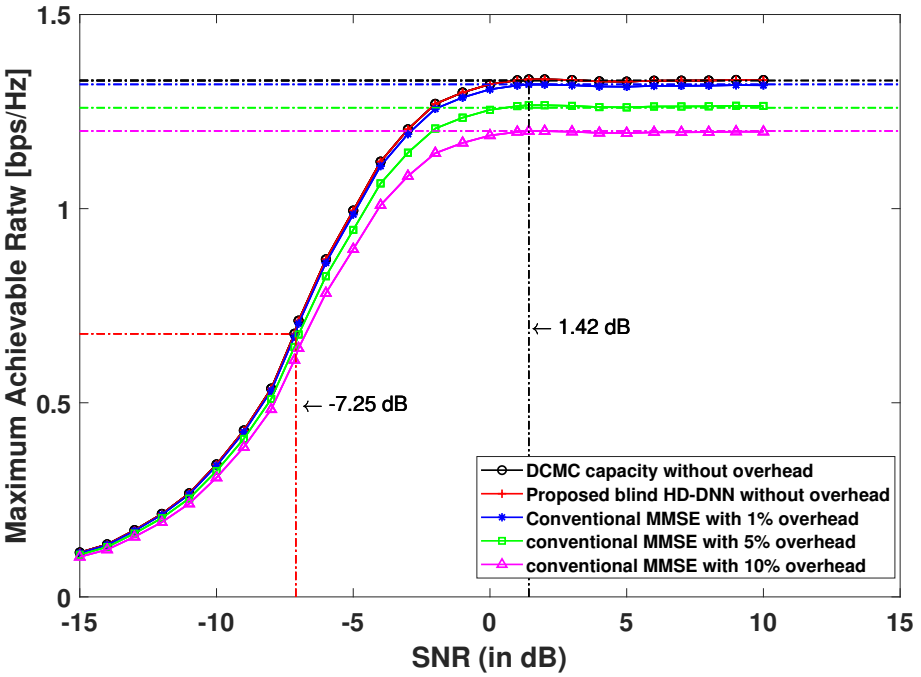

FIGURE 3.10: The DCMC capacity of the proposed design, and of the conventional design at 1%, 5% and 10% pilot overheads. Our simulation parameters are shown in Table 3.4 and Table 3.6.

rapidly, both the coherent and blind-detection-aided schemes suffer from a severe performance loss. Thus, in a slow-fading channel, both the detection performance and detection complexity of our DNN-based detector are better than that of the conventional S-MP and AMP detectors, while the performance of conventional detectors depend on the resources allocated to the pilots for CE. However, harnessing a massive amount of training data used for the training phase can result in slight improvements in detection performance for blind DNN detectors. Furthermore, SD detection in combination with channel coding is needed to mitigate the channel impairments.

As discussed in Section 3.3.1, the ML detector applies an exhaustive search having a complexity order of $\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$, while the reduced complexity method [19] has a complexity order of $\mathcal{O}[N_{AC}N_{SI} + Q\mathcal{L}K]$. By contrast, the complexity of the NN is determined by its size, weight and bias without considering the complexity of the training phase, where we have the NN complexity order of $\mathcal{O}[n_i n_h] + \mathcal{O}[n_h^2] + \mathcal{O}[n_h n_o]$ [3]. The computational complexity associated with the parameters of Table 3.4 and Table 3.6 is shown in Table 3.7, which confirms that the proposed DNN method achieves the lowest detection complexity.

Let us now consider the performance of SD detection, where we employ a $RSC(2,1,3)$ channel codec. For a typical $RSC(n,k,K)$ codeword, n denotes the output per input

[3]Complexity order of NN only used to compare the ML detection, while there is no search complexity associated with the NN-aided detection. n_i and n_o denote the neuron size of input and output layer, $n_{h_i}(i = 1, 2, \cdots)$ denote the neuron size of hidden layer between input and output.

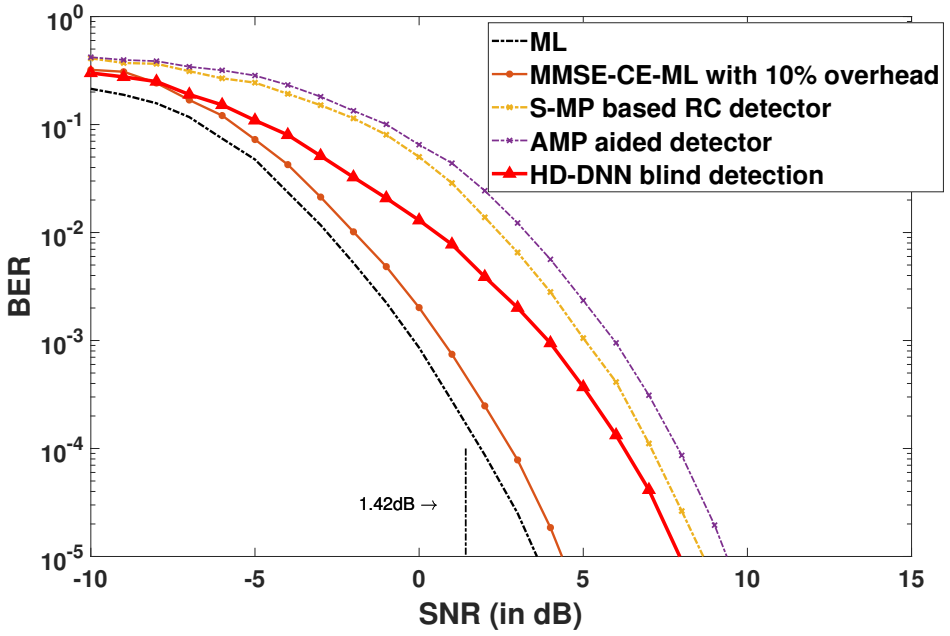

FIGURE 3.11: BER performance comparison of CS-MIM using HD detection which Scheme 1(c), 2, 3 is considered. Our simulation parameters are shown in Table 3.4 and Table 3.6.

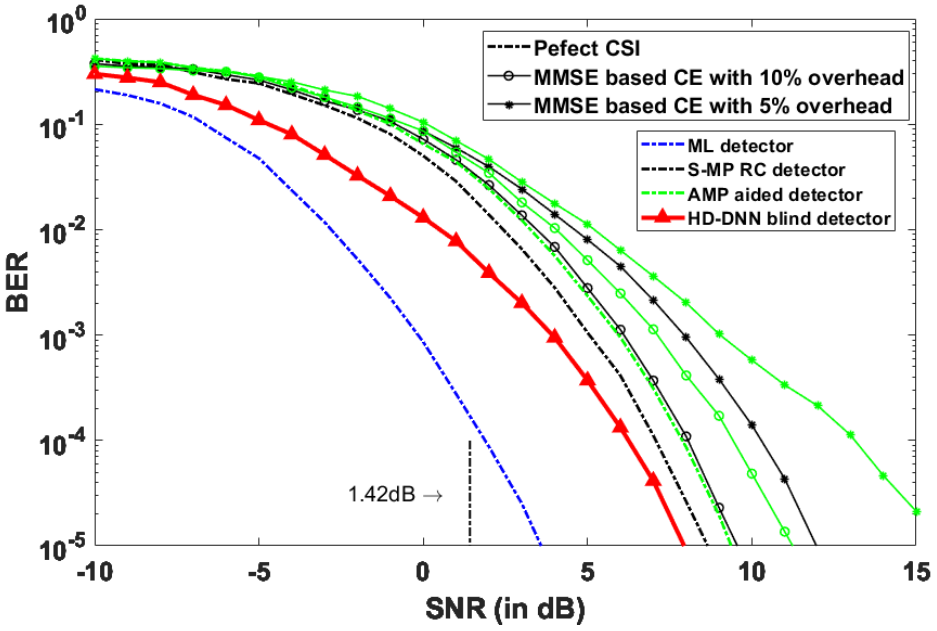

FIGURE 3.12: BER performance comparison of CS-MIM using HD detection under fading channel with normalized Doppler frequency of 10^{-6}. which Scheme 1, 2, 3 is considered. Our simulation parameter are shown in Table 3.4 and Table 3.6.

cycle in the convolutional encoder, k represents the number of input bits and K denotes the constraint length of the encoder. Explicitly, we applied a half-rate RSC encoder as shown in Table 3.4. The corresponding maximum achievable rate is $R_t = 0.66667$ bits/sec/Hz. In [19], it was shown that a half-rate RSC coded CS-MIM system with

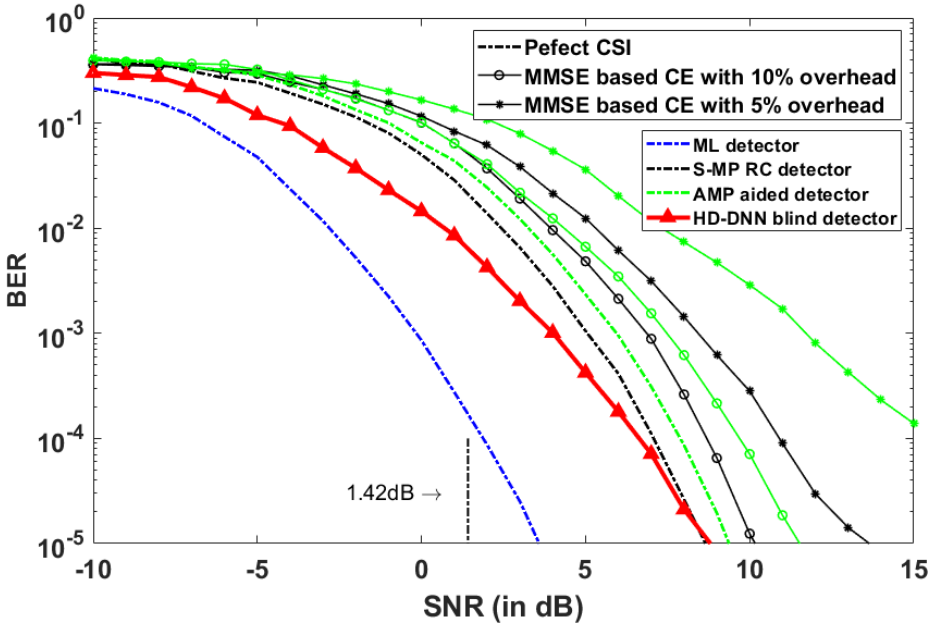

FIGURE 3.13: BER performance comparison of CS-MIM using HD detection under fading channel with normalized Doppler frequency of 10^{-5}. which Scheme 1, 2, 3 is considered . Our simulation parameter are shown in Table 3.4 and Table 3.6.

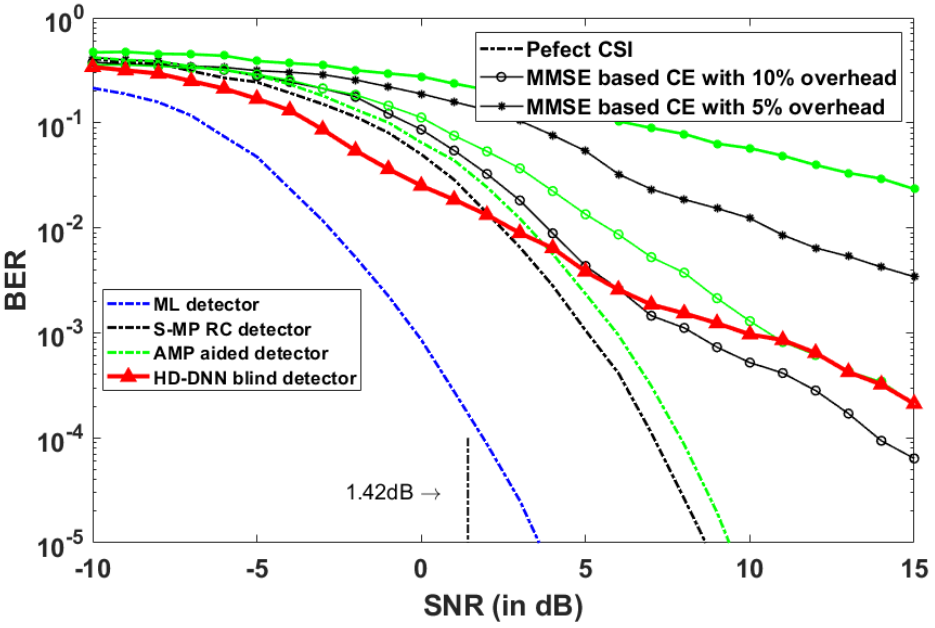

FIGURE 3.14: BER performance comparison of CS-MIM using HD detection under fading channel with normalized Doppler frequency of 10^{-4}. which Scheme 1, 2, 3 is considered. Our simulation parameter are shown in Table 3.4 and Table 3.6.

the same parameters as in Table 3.4 achieves the theoretical maximum rate of 0.667 bit-s/sec/Hz at -7.25 dB in terms of its DCMC, which is also indicated in Fig. 3.10. We first examine the performance of Scheme 4(a) and 4(b) for the non-iterative SD detector. As shown in Fig. 3.15, Scheme 4(a) achieves a BER of 10^{-5} at -1.7 dB. Naturally,

Scheme 4(b) that employs imperfect CSI degrades the performance. By contrast, our proposed SD blind detection DNN model is capable of achieving similar performance to the conventional SD detection, when the pilot overhead is 5% and even better, when the pilot overhead is 1%. Furthermore, note that our proposed learning aided method of Scheme 5 and Scheme 6 are blind and hence have no pilot overhead.

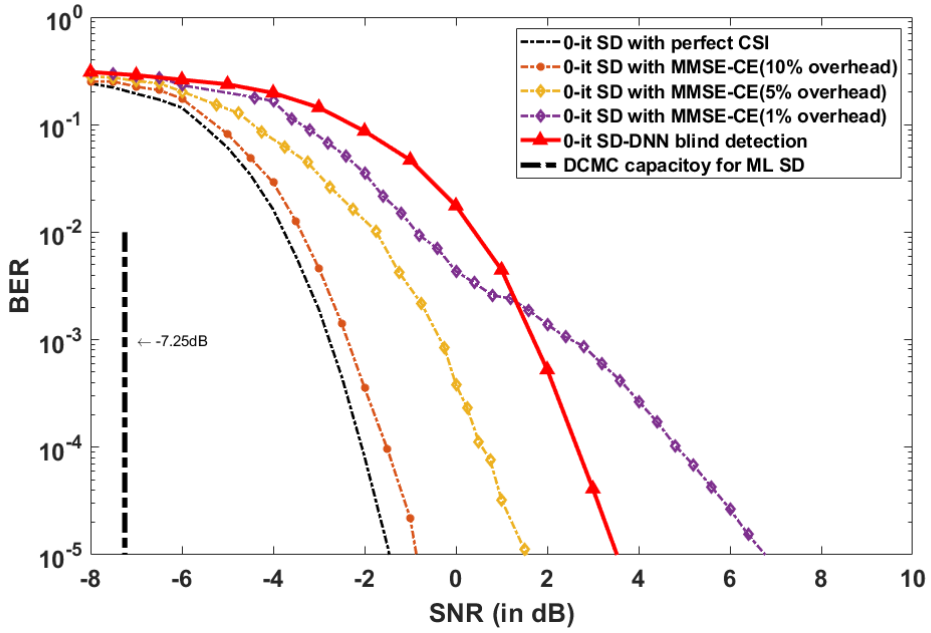

FIGURE 3.15: BER performance comparison of the conventional SD detector with the proposed SD- DNN detector in 0 iteration which are the comparison of Scheme 4(a), 4(b) and 5(a) as shown in Table 3.5. The simulation parameters are shown in Table 3.4 and 3.6.

The performance of SD can be further improved upon iteratively exchanging extrinsic information between the decoder and the demodulator. Fig. 3.16 shows the iteration gain of the conventional iterative detection Scheme 4. However, realistic CE results in 3dB performance loss compared to the perfect CSI case at BER of 10^{-5}. Fig. 3.16 and Fig. 3.17 characterize two different learning aided ISD detection methods applied to our CS-MIM system. The first method, which is Scheme 5, namely the single-stage DNN- SD detector is about 5.2 dB worse than the conventional ISD relying on perfect CSI with 1 iteration. For higher number of iterations, the NN model will have an improved performance, where the BER difference is reduced to 3.4 dB after 5 iterations. However, the proposed learning method has a complexity order of $O[\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)]$ compared to $O[n_{it} 2^{c_g}((N_r N_t M N_f^2) + (N_r N_t M N_f^3 + N_r M^2 N_f^2 N_v + N_r N_f M^2 N_v K + N_r N_f M K T) N_{AC} N_{SI} (Q\mathcal{L})^K]$ for the conventional scheme, where n_{it} denotes the number of iterations.

Fig. 3.17 characterizes the second method, which is Scheme 6 of the learning-aided MAP method. In our simulations, we used the LLRs of simplified Jacobian operation

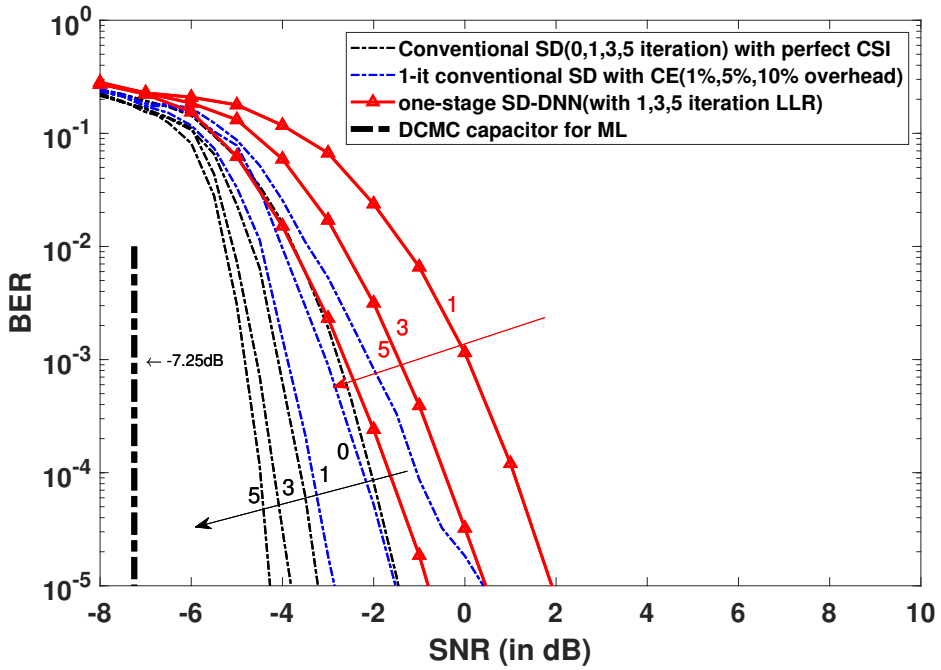

FIGURE 3.16: BER performance comparison of the ISD detector with NN detector of Scheme 4 and 5 as shown in Table 3.5. The simulation parameters are shown in Table 3.4 and Table 3.6.

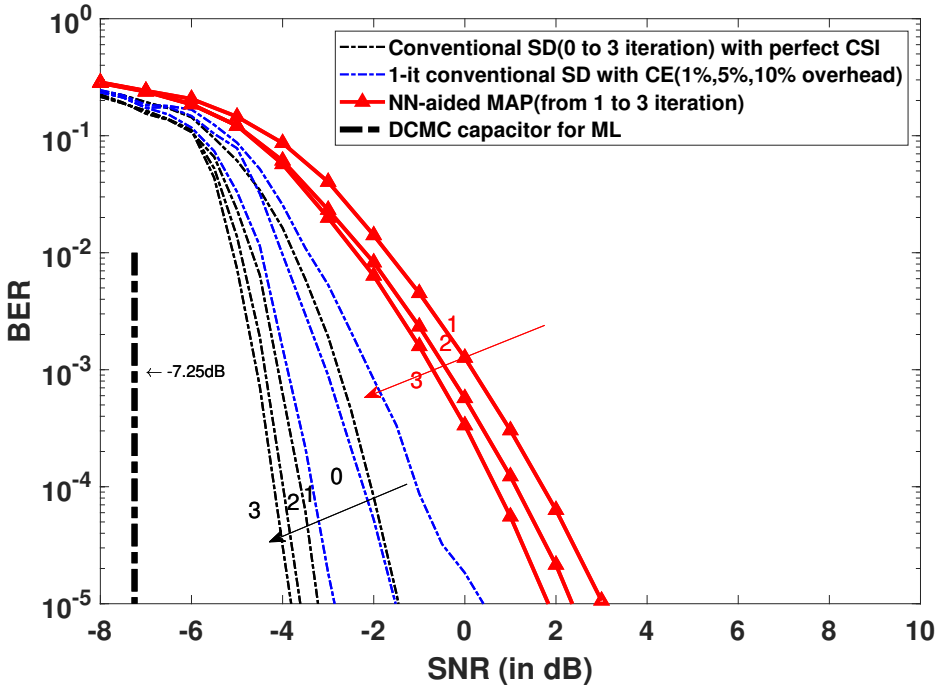

FIGURE 3.17: BER performance comparison of the ISD detector with NN detector of Scheme 4 and 6 as shown in Table 3.5. The simulation parameters are shown in Table 3.4 and Table 3.6.

to train the learning model. When the number of ISD iterations is $n_{it} = 1 - 3$, the

TABLE 3.7: Simulation results and complexity analysis of each Scheme.

Scheme index		SNR at BER of 10^{-5}	Complexity Order	Complexity
HD Detection				
Scheme 1	a)	3.6	$\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$	8.51×10^6
	b)	4.5	$\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$	8.51×10^6
Scheme 2	a)	8.1	$\mathcal{O}(N_{AC}N_{SI} + Q\mathcal{L}K)$	1.64×10^5
	b)	9.7	$\mathcal{O}[N_{AC} + N_{SI}(Q\mathcal{L})^K]$	1.2×10^5
Scheme 3		7.9	$\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$	9.2×10^3
SD Detection				
Scheme 4	a)	-1.8	$O[2^{c_g}(N_{AC}N_{SI}(Q\mathcal{L})^K)]$	-
	b)	0.3		
	c)	-3.2(1 it) -3.9 (3 it) -4.2 (5 it)	$O[n_{it}2^{c_g}(N_{AC}N_{SI}(Q\mathcal{L})^K)]$	-
	d)	-3(1 it)		
Scheme 5	a)	3.8	$\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$	-
	b)	2(1 it) 0.3(3 it)-0.9(5 it)		
Scheme 6		3(1 it) 2.2(2 it)1.9 (3it)	$2n_{it}(\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o))$	-

proposed detector achieves a BER of 10^{-5} at 3.05 dB for $n_{it} = 1$, while the conventional ISD associated with perfect and imperfect CSI requires -3.3 dB and -0.3 dB for the single-iteration scheme. This BER performance gap is expected, because the model-based conventional detectors exhibit higher complexity and have a pilot overhead, leading to reduced throughput that cannot be compensated upon increasing the SNR, as previously evidenced by Fig. 3.10.

TABLE 3.8: Runtime comparison between proposed methods and conventional SD of CS-MIM

scheme	iteration	Runtime (in seconds)	iteration	Runtime (in seconds)
Scheme 4	0	0.735	1	3.43
	2	4.65	3	6.25
Scheme 5	1	0.0064	3	0.0062
	5	0.0074		
Scheme 6 MAP	1	0.1044	2	0.2533
	3	0.4754		

The complexity order of the conventional SD detector and of the proposed learning methods are shown in Table 3.7. Then we also calculate the computational complexity of each HD schemes which derived from simplification of matrix calculation and algorithm of the simulation model. Then the run time comparisons between conventional SD detection and the proposed methods are shown in Table 3.8. Here, the run time is measured for detection of a data sample at the receiver. For fair comparison, we measure the runtime of all schemes in MATLAB run on the same computer.

As shown in Fig. 3.16, Fig. 3.17 and Table 3.7, it can be observed that the proposed one-stage DNN-aided and iterative DNN-aided MAP detectors have a somewhat eroded performance, but impose an extremely low complexity compared to the conventional SD detection.

3.5 Summary

Blind learning-aided detection of CS-MIM communicating over fading channels using both HD and SD was proposed. For HD, we demonstrated the proposed learning aided schemes are capable of outperforming their ML, S-MP and AMP counterparts relying on pilot overheads in terms of their DCMC capacity. Meanwhile, the proposed DNN methods exhibit the lowest complexity. Then for SD without iterations, the proposed DNN methods are capable of approaching the performance of their coherent CS-MIM counterparts, even though their computational complexity is substantially reduced. Furthermore, ISD is devised for the proposed DNN methods in order to benefit from iteration gains. In summary, our methods are the first blind detection solutions in the literature that can (I) eliminate pilot overhead and CE complexity; (II) substantially reduce the ISD complexity.

Chapter 4

Reduced Complexity Learning-Based Joint Channel Estimation and Detection of Compressed Sensing-aided Multi-Dimensional Index Modulation

4.1 Introduction

Learning-based techniques can be applied in CE-aided detection of the IM schemes as discussed in Chapter 2. Chapter 3 introduced a direction of how the learning-aided detection applied in the CS-MIM system, which is blind detection both in HD and SD detection. In this chapter, we further investigate the combination of learning method for the JCED scheme as shown in Fig. 4.1.

Since the MIM conveys information in several dimensions, ML detection is theoretically capable of detecting the multi-dimensional signal jointly, albeit at an escalating complexity upon increasing the degrees of freedom or dimensions [6]. In [19], CS-MIM was proposed, where multiple detection stages were harnessed for recovering data from the CS, STSK and OFDM-IM domains, again, at an extremely high complexity.

On the other hand, coherent detection requires the knowledge of CSI, which is estimated by transmitting pilots to the receiver [63]. Although SM exhibits energy savings

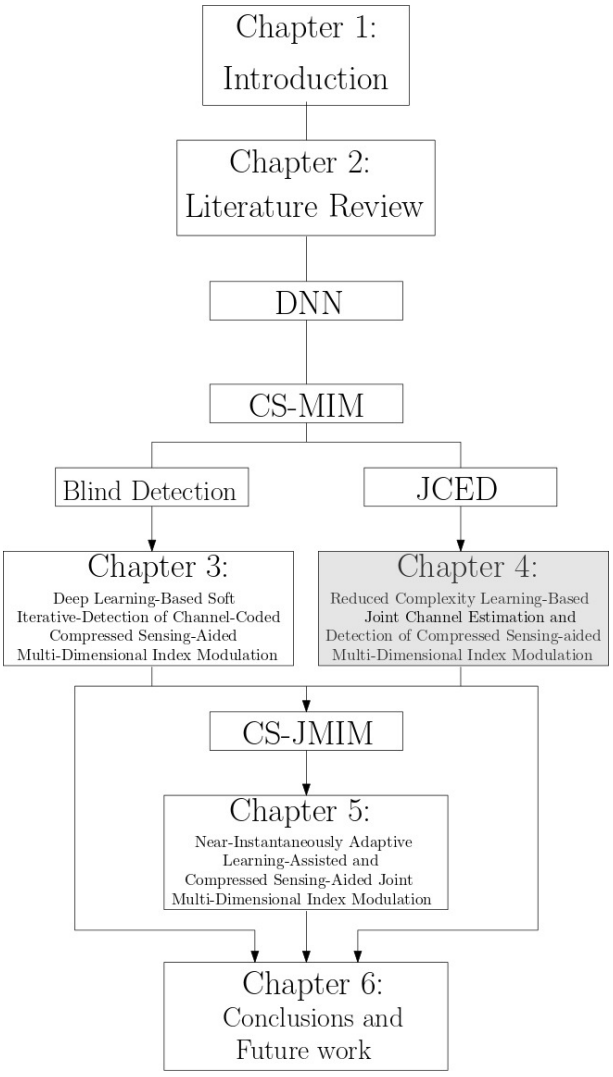

FIGURE 4.1: The organization of this thesis-Chapter3

by only employing a single RF chain, the pilot based Channel Estimator can only obtain the active TAs' CSI, hence it requires more time to estimate the whole MIMO channel. In [149], Faiz *et al.* proposed recursive least-squares-based adaptive channel estimator for SM under the assumption that the MIMO channel experienced block fading. Then, Wu *et al.* [150] investigated a novel CE scheme by exploiting the channel correlation, which significantly reduced the pilot overhead. Acar *et al.* [141] employed a systematical pilot insertion method to estimate the SM- MIMO channel. However, the pilot overhead reduces the payload at a given bandwidth efficiency and the MIMO complexity increases with the number of antennas [131].

Historically speaking, Abuthinien *et al.* [151] proposed a semi-blind ML-based Joint Channel Estimation and data Detection (JCED) scheme for MIMO systems at a minimum pilot overhead. Furthermore, Chen *et al.* [152] designed an iterative JCED for STSK systems, which imposes reduced complexity, while maintaining a high throughput and attaining near-optimal BER performance. Similarly, Sugiura *et al.* [142] applied

JCED in a SM scheme, while Acar *et al.* [153] proposed a similar iterative JCED for coded SM- OFDM.

The complexity of ML sequence detection in Gaussian channels may be deemed practical for one-dimensional IM, but the complexity of MIM detection increases significantly. It is therefore of interest to explore sub-optimal receivers that can approach the performance of the ML detector at a reduced complexity, such as the Expectation-Maximization (EM) algorithm investigated by Cozzo *et al.* [154], while employing a reduced number of pilot symbols.

DL has been attracting increasing attention in wireless communications [136] [155]. For instance, a DNN was used for detecting MIMO signals in [117] [119]. In [114], a DNN was employed for estimating the channel of OFDM systems and data detection. Then in [156], Boloursaz Mashhadi*et al.* proposed learning based pilots design and CE to significantly reduce the pilot overhead in MIMO sysyem. Recently, Qing *et al.* [157] proposed an effective CE and detection scheme relying on a sophisticated neural network models for achieving a similar or possible even better detection performance than the conventional MMSE arrangement. In [124], Satyanarayana *et al.* proposed a DNN-aided semi-blind detector for drastically reducing the pilot overhead needed for CE of a multi-set STSK scheme, which was also extended to SD in [125]. Additionally, both CE and detection have been performed using neural network in [158], where Xiang *et al.* proposed a DNN-based iterative JCED for SM systems.

However, the JCED techniques presented in the literature have been designed for single-dimensional IM systems, Furthermore, there is a paucity of research exploiting DL-based SD detection. Hence, in order to narrow the knowledge gap, we design a DNN-based JCED for CS-MIM systems, which can harness both HD as well as SD detection.

TABLE 4.1: Boldly and explicitly contrasting our contributions to the literature

Contribution	proposed*	[17]	[105]	[19]	[120]	[142]	[153]	[124]	[125]	[158]
IM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MIM	✓			✓					✓	
CS at transmitter	✓	✓		✓						
CE	✓		✓			✓	✓	✓		✓
JCED	✓					✓	✓			✓
Learning aided detector	✓		✓		✓			✓	✓	✓
SD detector	✓		✓				✓		✓	
Learning aided JCED	✓									✓

Table 4.1 boldly contrasts the novelty of this chapter to the literature. Against the above background, the detailed contributions of this chapter are summarized as follows:

- In this chapter, we propose a reduced complexity JCED for HD CS-MIM, employing a data driven DNN. The proposed learning aided JCED method is capable of attaining near- ML performance at a low pilot overhead and complexity.

- We then further extend this DNN-aided JCED CS-MIM scheme for producing soft information, where we combine our system with channel coding in order to attain an improved BER performance.
- Our simulation results demonstrate that the proposed MMSE-based JCED DNN model is capable of outperforming the conventional MMSE-CE and detection scheme in different channel environments at a reduced pilot overhead, while also approaching the performance of conventional JCED, despite its reduced complexity and pilot overhead.

The rest of this chapter is organized as follows. In Section II, the system model of CS-MIM is presented. In section III, we design JCED techniques for our CS-MIM system relying on our proposed learning-aided detector along with its complexity analysis. Finally, in Section IV and Section V, we analyze the results and conclude, respectively.

4.2 System Model

In this section, we introduce the model of the CS-MIM system employing N_t TAs and N_r RAs. In Chapter 3, Fig. 3.2 shows the block diagram of the CS-MIM system considered , where an OFDM symbol has N_c subcarriers, which are then equally divided into G groups. Each group has $N_f = N_c/G$ subcarriers in the FD[1], while N_v subcarriers of each group are applied for the CS-MIM system in the VD [2]. The FD signal is attained by compressing the VD signal using CS as detailed in [19], where N_f is set lower than N_v to increase the throughput. The CS-aided OFDM symbols will then be transmitted from the activated TAs decided by the antenna selector of Fig. 3.2. Then, after transmission over the wireless channel, the receiver estimates the channel and detects the signal.

The same system model is applied in both Chapter 3 and Chapter 4. In the following, we only present the details of the receiver. We consider a receiver employing N_R RAs. The signal arriving from the transmitter is assumed to be transmitted over a frequency-selective Rayleigh fading channel and the CSI is acquired by CE, as discussed in Section III-C.

The CP is removed and then the received signal is transformed to the FD signals by using the FFT, as shown in Fig. 3.5. The ST demapper collects the FD symbols received from N_r RAs over T time slots to recover the ST symbols, which are then split into G

[1] FD is the OFDM symbol domain after CS processing, as shown in Fig. 3.2.

[2] VD is the actual domain, where subcarrier index modulation is applied before the CS process as shown in Fig. 3.2. This concept was firstly introduced in [17] to illustrate the CS techniques in IM system to improve the spectral efficiency.

groups by the Block Splitter of Fig. 3.5. Afterwards, the symbols received by each subcarrier group are represented as $\boldsymbol{Y} = \{\boldsymbol{Y}[1]^T, \ldots, \boldsymbol{Y}[\alpha]^T, \ldots, \boldsymbol{Y}[N_f]^T\}^T$ with $\boldsymbol{Y} \in \mathbb{C}^{N_r N_f \times T}$ and $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T}$ characterizing the ST structure per group and the ST symbol received at the α-th subcarrier of each subcarrier group, respectively.

Let the FD channel matrix be represented as $H_\alpha \in \mathbb{C}^{N_r \times N_t}$ for $\alpha = 1, \ldots, N_f$. Then the signal $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T} (\alpha = 1, \ldots, N_f)$ received during T TSs for each subcarrier group can be expressed as [19]

$$\boldsymbol{Y}[\alpha] = \boldsymbol{H}_\alpha \bar{\boldsymbol{S}}[\alpha] + \boldsymbol{W}[\alpha] = \boldsymbol{H}_\alpha \boldsymbol{I}_{AC} \boldsymbol{S}^{FD}[\alpha] + \boldsymbol{W}[\alpha], \tag{4.1}$$

where $\bar{\boldsymbol{S}}$ represents the modulated signal after SM at transmitter and $\boldsymbol{S}^{FD}[\alpha] \in \mathbb{C}^{M \times T}$ denotes the ST symbols at α subcarriers transmitted from M TAs over T TSs and $\boldsymbol{W}[\alpha] \in \mathbb{C}^{N_r \times T}$ represents the AWGN obeying the distribution of $\mathcal{CN}(0, \sigma_N^2)$, and σ_N^2 is the noise variance. Furthermore, $\boldsymbol{I}_{AC} \in \mathbb{C}^{N_t \times M}$ denotes the $(N_t \times M)$-element submatrix, which describes the selection pattern of active TAs for each subcarrier group at the transmitter. For high-integrity detection, accurate channel information is required, which is attained by employing CE techniques relying on known pilots in practical model-based solutions. In the next section, we will discuss CE techniques suitable for CS-MIM and characterize the JCED method.

4.3 CE and Detection for CS-MIM

Given the received signal $\boldsymbol{Y}$ of (4.1), the receiver infers the information bits of the STSK codewords, the bits embedded into the activated subcarrier indices and the bits mapped to the active TAs. This detection process requires the channel state information, which can be acquired by channel estimation. In the following, we consider both separate channel estimation and detection and JCED, where we propose a deep learning aided JCED technique capable of reducing both the complexity as well as the pilot overhead without substantially eroding the performance.

The signal received at the α-th subcarrier during a TS, can be represented as

$$\begin{bmatrix} \boldsymbol{Y}_1^\alpha \\ \vdots \\ \boldsymbol{Y}_r^\alpha \\ \vdots \\ \boldsymbol{Y}_{N_r}^\alpha \end{bmatrix} = \begin{bmatrix} h_{1,1}^\alpha & h_{1,2}^\alpha & \cdots & h_{1,N_t}^\alpha \\ h_{2,1}^\alpha & h_{2,2}^\alpha & \cdots & h_{2,N_t}^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r,1}^\alpha & h_{N_r,2}^\alpha & \cdots & h_{N_r,N_t}^\alpha \end{bmatrix} \boldsymbol{I}_{AC} \boldsymbol{S}^{\alpha(FD)} + \boldsymbol{W}^\alpha, \tag{4.2}$$

where $h_{r,t}^\alpha$ is the CSI between the r-th RA and the t-th TA for the α-th subcarrier for subcarrier group g. Additionally, $\boldsymbol{S}[\alpha]^{FD}$ can be extended as $\{\boldsymbol{S}_\alpha^1, \cdots, \boldsymbol{S}_\alpha^M\}$ for a single

TS. Then, the channel matrix $\boldsymbol{H}$ corresponding to N_f ST signals of each subcarrier group can be expressed in a diagonal structure of size $(N_r N_f \times N_t N_f)$ as

$$\boldsymbol{H} = diag\{\boldsymbol{H_1}, \boldsymbol{H_2}, \cdots, \boldsymbol{H}_{N_f}\}, \tag{4.3}$$

where $\boldsymbol{H}_\alpha(\alpha = 1, 2..., N_f)$ represents the corresponding CSI at the α-th subcarrier.

Similarly, the antenna selection pattern matrix associated with N_f subcarriers of each group $\bar{\boldsymbol{I}}_{AC} \in \mathbb{C}^{N_t N_f \times M N_f}$ has the structure of

$$\bar{\boldsymbol{I}}_{AC} = diag\{\boldsymbol{I_{AC}}, \boldsymbol{I_{AC}}, \cdots, \boldsymbol{I_{AC}}\}. \tag{4.4}$$

The received signal $\boldsymbol{Y}$ contains N_f space-time symbols at N_f subcarriers in the FD of each subcarrier group. Given the received signal model $\boldsymbol{Y}[\alpha] \in \mathbb{C}^{N_r \times T}, \alpha = 1, \ldots, N_f$, we can write $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\boldsymbol{S}^{FD} + \boldsymbol{W}. \tag{4.5}$$

The FD space-time signal can be represented as

$$\boldsymbol{S^{FD}} = \bar{\boldsymbol{A}}\boldsymbol{S} = \bar{\boldsymbol{A}}\boldsymbol{I_{SI}}\boldsymbol{X}, \tag{4.6}$$

where $\bar{\boldsymbol{A}} \in \mathbb{C}^{MN_f \times MN_v}$ is the equivalent measurement matrix $\boldsymbol{A}$ used for compressing the VD vector and $\boldsymbol{S} \in \mathbb{C}^{MN_v \times T}$ denotes the VD space-time symbol. Then, $\boldsymbol{S}$ can be expanded as $\boldsymbol{S} = \boldsymbol{I_{SI}}\boldsymbol{X}$, where $\boldsymbol{X} \in \mathbb{C}^{MK \times T}$ represents K STSK codewords and $\boldsymbol{I_{SI}} \in \mathbb{C}^{MN_v \times MK}$ is the subcarrier index selection pattern.

Hence, (4.5) can be rewritten as:

$$\boldsymbol{Y} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W}. \tag{4.7}$$

In the following, we first present the conventional CE and HD detection for the CS-MIM system considered, followed by the conventional JCED. Then, we introduce both the conventional SD detection and the SD- JCED scheme of the CS-MIM system. Afterwards, we present our proposed NN aided HD- JCED, where the neural network replaces the exhaustive search with a learned classification model in order to significantly reduce the computational complexity, followed by the neural network aided SD- JCED.

4.3.1 Conventional CE and HD Detection

In this section we present the conventional channel estimation and detection designed for the MIM system, followed by the JCED to output both HD as well as SD values.

4.3.1.1 CE Methods

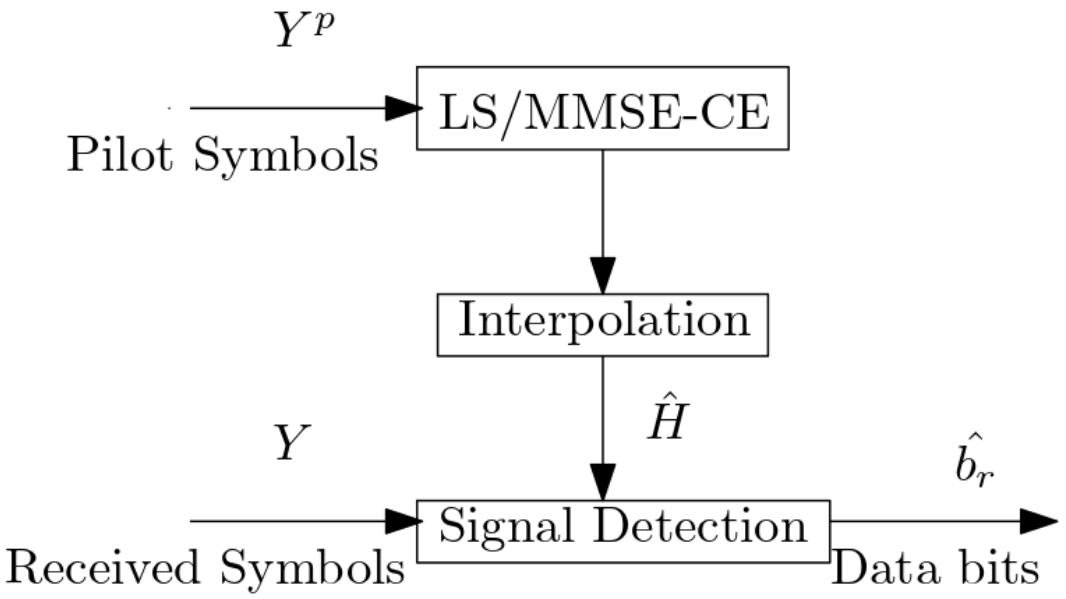

FIGURE 4.2: Conventional CE flow chart.

As shown in Fig. 3.5, we use the CE scheme for acquiring the CSI used for detection. Conventional pilot based CE, which inserts pilots in each symbol may become inefficient in this context due to randomly activating both the subcarriers and TAs [141]. We circumvent this problem by constructing a dedicated pilot frame for estimating the CSI by the channel estimator for our CS-MIM receiver, as shown in Fig. 3.5. This mitigates the challenge of randomness caused by the TA index selection. The pilot frame has the same size as the information frame, where only a single TA is activated for each subcarrier group. In this case, the number of subcarrier groups G is higher than or equal to that of the TAs N_t. Furthermore, each of the N_t TAs can be activated more than once in each frame. Then the CSI of every single TA and subcarrier group can be estimated by the channel estimator. Afterwards, we can obtain the estimated CSI matrix $\hat{\boldsymbol{H}}$ of the equivalent subcarrier group by linear interpolation techniques [63].

Fig.4.2 shows the flow chart of the conventional CE and detection. Firstly, the pilot symbol $\boldsymbol{Y}^p$ is input to the channel estimator. Then, with the aid of the appropriate CE method, the estimated CSI $\hat{H}$ may be acquired by the detector and then used for recovering the information bits.

Let us model the received space-time pilot symbol based on (4.5) as

$$\boldsymbol{Y}_p = \boldsymbol{H}\bar{\boldsymbol{S}}_p + \boldsymbol{W}, \tag{4.8}$$

where the space-time pilot symbol is $\bar{\boldsymbol{S}}_p = diag\{\bar{\boldsymbol{S}}_{p,1}, \bar{\boldsymbol{S}}_{p,2}, \cdots, \bar{\boldsymbol{S}}_{p,M}\}$.

Then the LSCE is given by

$$\hat{\boldsymbol{H}}_{LS} = \boldsymbol{Y}_p \bar{\boldsymbol{S}}_p^H (\bar{\boldsymbol{S}}_p \bar{\boldsymbol{S}}_p^H)^{-1}. \tag{4.9}$$

In this case, we can calculate the complexity of LSCE, as shown in (4.9). To elaborate, the complexity of LSCE is dominated by the CSI matrix inversion and multiplication. Then we can characterize the complexity of LSCE by the complexity order of $\mathcal{O}_{LSCE}[N_r N_t M T N_f^2]$.

To minimize the estimation MSE of $\boldsymbol{H}$, the popular MMSE-CE formulated as

$$\hat{\boldsymbol{H}}_{MMSE} = \boldsymbol{Y}_p(\bar{\boldsymbol{S}}_p^H \boldsymbol{R}_H \bar{\boldsymbol{S}}_p + N_0 N_f \boldsymbol{I})^{-1} \bar{\boldsymbol{S}}_p^H \boldsymbol{R}_H, \tag{4.10}$$

where $\boldsymbol{R}_H$ represents the channel's correlation matrix [142]. The MMSE-CE requires the calculation of $\boldsymbol{R}_H$ and CSI matrix inversion. Then, we can characterize the computational complexity as $\mathcal{O}_{MMSE-CE}[N_r N_t T(MN_f^2 + N_f^3)]$.

To track the channel, piecewise linear interpolation is used for acquiring the CSI, which can be formulated as:

$$\boldsymbol{H_n} = \hat{\boldsymbol{H}}_{n_p} + (\hat{\boldsymbol{H}}_{n_{p+1}} - \hat{\boldsymbol{H}}_{n_p})(\frac{n - n_p}{D}), for\ n_p \leq n \leq n_{p+1}, \tag{4.11}$$

where $\hat{\boldsymbol{H}}_{n_p}$ and $\boldsymbol{H_n}$ are the estimated CSI matrix at the pilot symbol position and D denotes the pilot insertion spacing.

4.3.1.2 ML Detection

The ML detector makes a joint decision on the TA index of the STSK codewords and of the subcarrier using an exhaustive search, which can be formulated as

$$\langle \hat{\gamma}, \hat{\beta}, \hat{\varphi} \rangle = \arg\min_{\gamma,\beta,\varphi} \|\boldsymbol{Y} - \boldsymbol{H}\bar{\mathcal{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\bar{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)\|^2, \tag{4.12}$$

where $\hat{\gamma}, \hat{\beta}$ and $\hat{\varphi}$ represent the estimates of the activated TAs index, the activated subcarrier index and the index of K STSK codewords in each subcarrier group, respectively [19].

At the receiver, the ML detector carriers out a full search for evaluating all possible candidates, which has a complexity order of $\mathcal{O}[N_{AC} N_{SI} (Q\mathcal{L})^K]$ per subcarrier group. Then, the total computational complexity of the ML detector relying on perfect CSI can be expressed as $\mathcal{O}_{ML}[(N_r N_t M N_f^3 + N_r M^2 N_f^2 N_v + N_r N_f M^2 N_v K + N_r N_f M K T) N_{AC} N_{SI} (Q\mathcal{L})^K]$. With the aid of the LSCE/ MMSE-CE relying on ML, we can have the total complexity order of CE-aided ML detection formulated as $\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML}$

4.3.1.3 JCED

To further improve the detection performance, data detection based iterative JCED is considered. Fig.4.3 shows the flow chart of the JCED, which starts using the same

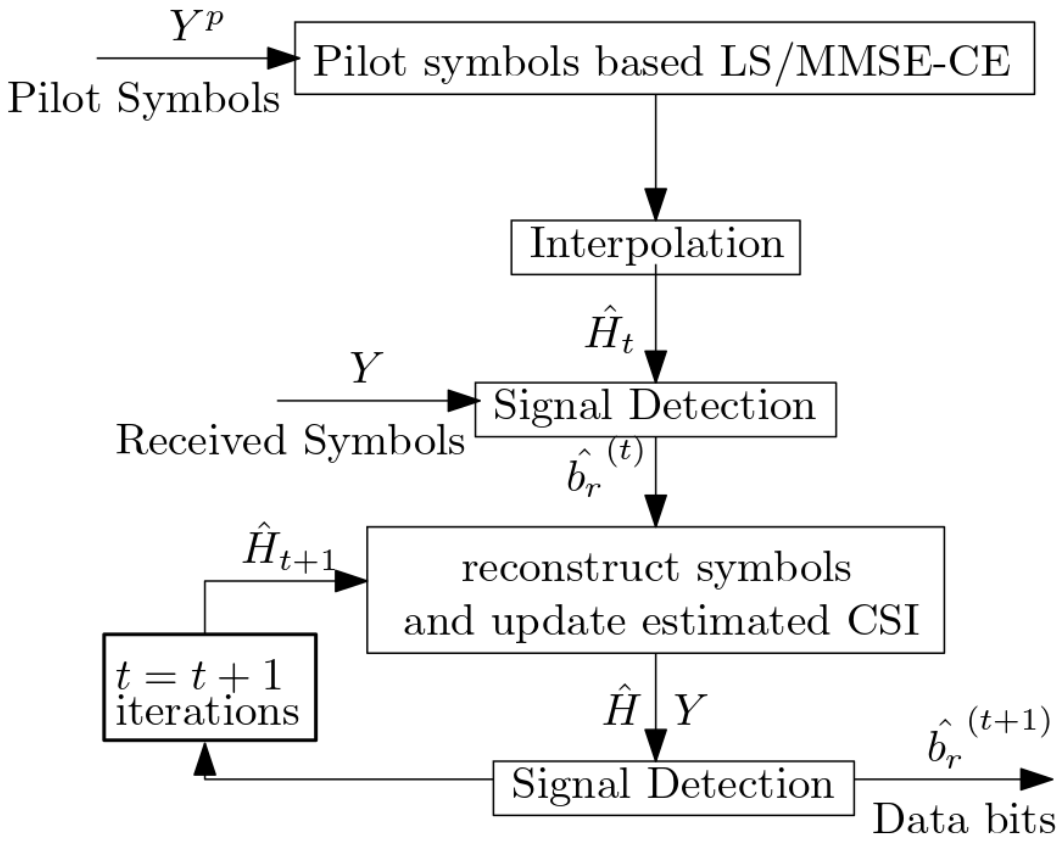

FIGURE 4.3: JCED flow chart.

procedure as the conventional CE, where the estimated CSI is acquired by the channel estimator with the aid of pilot symbols. Afterwards, we recover the bits from the received signal and the estimated CSI. Then, the remodulated symbols created from the recovered bits are used for updating the CSI.' This process is then repeated for several iterations to improve the estimated CSI accuracy. By exploiting the remodulated symbols, the JCED can increase the CE accuracy and hence increase the detection performance without increasing the pilot overhead. Based on [152] and [142], the JCED of CS-MIM is described by **Algorithm 1**.

As illustrated in Algorithm 1, there are two thresholds, which are used for terminating the update loop. First, we set a maximum number of iterations, I_{max}, which progressively enhances the CE and detection performance. This allows for an adjustable algorithmic complexity based on the number of iterations. The second approach introduces a termination constant, β, which controls the accuracy of the CE. Based on the theoretical results of the MMSE-CE-aided and ML-based detection, we can determine the MSE gap between conventional CE-based detection and ML detection, assuming perfect CSI. Consequently, the constant β can be selected within this gap and should be sufficiently low. In this scenario, the algorithm's complexity solely hinges

Algorithm 1: LS/MMSE-CE based HD-JCED of CS-MIM

Input: Pilot symbol $\boldsymbol{Y}^p$, Received signal $\boldsymbol{Y}$

Output: Detected data $\hat{\boldsymbol{b}_r}^{(t+1)}$

$t := 0$ `// initial iteration index`

The initial CSI estimated by LSCE or MMSE-CE to achieve $\hat{\boldsymbol{H}}_{\boldsymbol{t}} = \hat{\boldsymbol{H}}_{(LS/MMSE)}$ with $\boldsymbol{Y}^p$ and $\boldsymbol{Y}$

while $t \leq I_{max}$ **do**

  $t = t + 1$

  Detect the data $\hat{\boldsymbol{b}_r}^{(t)} = [\hat{\boldsymbol{b}_r}^{(t)}(1)\ \hat{\boldsymbol{b}_r}^{(t)}(2)\ \hat{\boldsymbol{b}_r}^{(t)}(3)\ ...\ \hat{\boldsymbol{b}_r}^{(t)}(\tau)]$ and then remodulated the detected data as $\hat{\boldsymbol{S}}^{(t)} = [\hat{\boldsymbol{S}}^{(t)}(1)\ \hat{\boldsymbol{S}}^{(t)}(2)\ \hat{\boldsymbol{S}}^{(t)}(3)\ ...\ \hat{\boldsymbol{S}}^{(t)}(\tau)]$

  Update the estimate CSI with LS estimator with $\boldsymbol{H}_{LS}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)})^H(\boldsymbol{S}^{(t)}(\boldsymbol{S}^{(t)})^H)^{-1}$ or MMSE estimator with $\boldsymbol{H}_{MMSE}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)} + N_0 N \boldsymbol{I})^{-1}\boldsymbol{S}^{(t)H}\boldsymbol{R}_{\boldsymbol{H}}$

  Achieve the updated detected data $\hat{\boldsymbol{b}_r}^{(t+1)}$ with $\boldsymbol{H}^{(t+1)}$ and residual of each iteration $\Delta = ||\hat{\boldsymbol{H}}^{(t+1)} - \hat{\boldsymbol{H}}^{(t)}||^2$

  if $\Delta \leq \beta$ **then**

    quit `// given a threshold to end the loop`

  end

end

on the CSI condition, which can be unpredictable. In general, a suitable termination threshold is chosen to strike an appropriate performance vs. complexity trade-off. Alternatively, both threshold may be harnessed for maximizing the algorithm's efficiency. In this case, we can represent the complexity order of the HD-JCED as $\mathcal{O}_{HD-JCED} = \mathcal{O}_{LSCE/MMSE-CE} + \log(\beta)(\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML})$, if the number of iterations is smaller than I_{max}. In a nutshell, the total complexity oder can be expressed as $\mathcal{O}_{HD-JCED} = \mathcal{O}_{LSCE/MMSE-CE} + I_{max}(\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML})$.

4.3.2 SD Detection

SD detection is employed for attaining near-capacity performance when combined with channel coding by exchanging soft values between the MIMO detector and the channel decoder. However, the complexity of the optimal maximum *a posteriori* probability MIMO detector rapidly becomes prohibitive upon increasing the modulation order and the number of TAs [98]. In the following, we will present the conventional SD detector of CS-MIM, followed by our LS/MMSE-CE based SD- JCED aided CS-MIM system.

4.3.2.1 Conventional SD Detection

A channel coded CS-MIM scheme is shown in Fig. 4.4, which was proposed in [19] for achieving near-capacity performance. The information bit sequence b is encoded by a Recursive Systematic Convolutional (RSC) encoder. Then, the coded bit sequence c is interleaved to generate the interleaved stream u, which is entered into the CS-MIM modulator of Fig. 3.2.

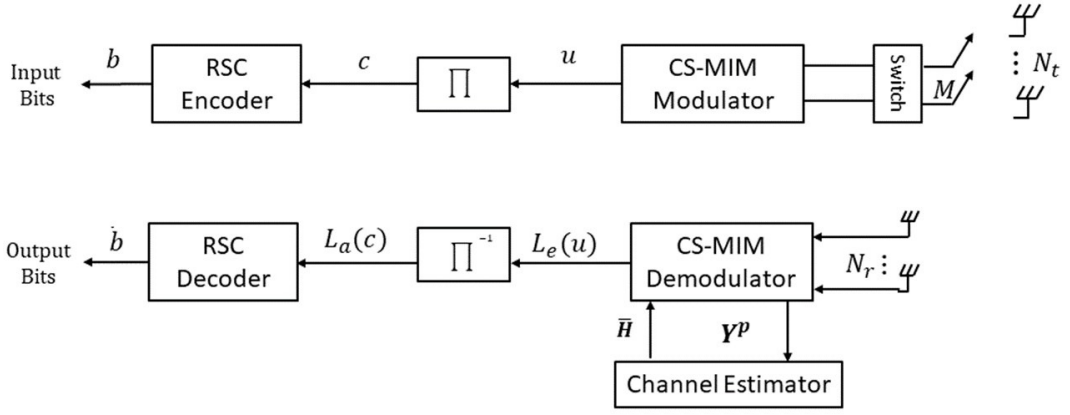

FIGURE 4.4: The transceiver architecture of channel-coded CS-MIM.

At the receiver side of Fig. 4.4, the pilot data is processed first for estimating the channel, where the estimated channel $\bar{\boldsymbol{H}}$ is entered into the soft CS-MIM receiver that outputs Log-Likelihood Ratio (LLRs). The LLRs output from the demodulator are then passed to the de-inteleaver and the RSC decoder performs soft decoding. In Fig. 4.4, $L(\cdot)$ represents the LLRs of the bit sequences, where $L_e(u)$ is the extrinsic LLR output after soft demodulation and $L_a(c)$ is the de-interleaved LLR sequence of $L_e(u)$, which constitutes the *a priori* information for the RSC decoder.

The LLR of a bit is defined as the ratio of probabilities associated with the logical bits '1' and '0', which can be written as $L(b) = log\frac{p(b=1)}{p(b=0)}$. The conditional probability $p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})$ of receiving the signal $\boldsymbol{Y}$ of a subcarrier group defined in (4.1)is given by [111]

$$\begin{aligned} &p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) \\ &= \frac{1}{(\pi \boldsymbol{N}_0)^{NT}} exp(-\frac{||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2}{N_0}), \end{aligned} \tag{4.13}$$

where $\mathcal{X}_{\gamma,\beta,\varphi}$ represents the STSK codewords at the β-th realization of active subcarriers, which are transmitted through the ϕ-th realization of an active TA. Furthermore, N_0 is the noise power, where we have $\sigma_n^2 = N_0/2$ with $N_0/2$ representing the double-sided noise power spectral density.

Hence, we can formulate the LLR of bit u_i as

$$\begin{aligned} L_e(u_i) &= \ln \frac{p(\boldsymbol{y}|u_i=1)}{p(\boldsymbol{y}|u_i=0)} \\ &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_1^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_0^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})}, \end{aligned} \tag{4.14}$$

where $\mathcal{X}_1^l$ and $\mathcal{X}_0^l$ represent a subset of the legitimate equivalent signal $\mathcal{X}$ corresponding to bit u_l, when $u_l = 1$ and $u_l = 0$, respectively, yielding $X_1^l \equiv \{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X} : u_i = 1\}$ and $X_0^l \equiv \{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X} : u_i = 0\}$.

Upon using (4.13) and (4.14) we obtain the LLR $L(b_i)$ of the bit sequence conveyed by the received signal $\mathbf{Y}$. To simplify the LLR calculation, the Approximate Log-MAP (Approx-Log-MAP) algorithm based on the Jacobian Maximum operation [146] is used [147], which is given by

$$\boldsymbol{L}_e(u_l) = \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_1^l}(\lambda_{\gamma,\beta,\varphi}) - \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_0^l}(\lambda_{\gamma,\beta,\varphi}), \tag{4.15}$$

where jac$(.)$ denotes the Jacobian maximum operation and the intrinsic metric of $\lambda_{\gamma,\beta,\varphi}$ is

$$\lambda_{\gamma,\beta,\varphi} = -||\boldsymbol{Y} - \boldsymbol{H}\tilde{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\tilde{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2/N_0. \tag{4.16}$$

At the receiver, the soft demodulator evaluates the probability of each bit being logical '1' and '0'. Then it applies the Approx-Log-MAP algorithm for obtaining the extrinsic LLRs of the coded bits, which has a complexity order of $\mathcal{O}[2^{(c_g)}(N_{AC}N_{SI}(Q\mathcal{L})^K)]$, where c_g represents the numbers of coded bits after the RSC encoder and interleaver. Then, we can have total complexity of $\mathcal{O}_{MAP}[c_g(N_rN_tMN_f^3 + N_rM^2N_f^2N_v + N_rN_fM^2N_vK + N_rN_fMKT)N_{AC}N_{SI}(Q\mathcal{L})^K + c_g2^{c_g}N_{AC}N_{SI}(Q\mathcal{L})^K]$

4.3.2.2 SD-JCED

Then we can also apply the same JCED algorithm for SD CS-MIM and the resultant procedure is described in **Algorithm 2**. Similarly, we can represent the complexity of the SD-JCED as $\mathcal{O}_{HD-JCED} = \mathcal{O}_{LSCE/MMSE-CE} + \log(\beta)(\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML})$ or $\mathcal{O}_{HD-JCED} = \mathcal{O}_{LSCE/MMSE-CE} + I_{max}(\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML})$.

However, both HD and SD JCED impose excessive complexity upon updating the CSI of each symbol. In the following, we propose DNN-based MIM detectors for reducing the complexity.

Algorithm 2: LS/MMSE-CE based SD-JCED of CS-MIM

Input: Pilot symbol $\boldsymbol{Y}^p$, Received signal $\boldsymbol{Y}$

Output: Detected data $\hat{\boldsymbol{b}_r}^{(t+1)}$

$t := 0$ `// initial iteration index`

The initial CSI estimated by LSCE or MMSE-CE to achieve $\hat{\boldsymbol{H}}^{(t)} = \hat{\boldsymbol{H}}_{(LS/MMSE)}$ with $\boldsymbol{Y}^p$ and $\boldsymbol{Y}$

while $t \leq I_{max}$ **do**

  $t = t + 1$

  Detect the received signal LLR based on eq.(4.14)(4.15) as $\hat{\boldsymbol{L}}_e^{(t)} = [\hat{\boldsymbol{L}}_e^{(t)}(1)\hat{\boldsymbol{L}}_e^{(t)}(2)\hat{\boldsymbol{L}}_e^{(t)}(3)...\hat{\boldsymbol{L}}_e^{(t)}(\tau)]$ and the detected coded data as $\hat{\boldsymbol{C}}^{(t)} = [\hat{\boldsymbol{C}}^{(t)}(1)\hat{\boldsymbol{C}}^{(t)}(2)\hat{\boldsymbol{C}}^{(t)}(3)...\hat{\boldsymbol{C}}^{(t)}(\tau)]$

  Recover the uncoded data by RSC decoder as $\hat{\boldsymbol{b}_r}^{(t)} = [\hat{\boldsymbol{b}_r}^{(t)}(1)\hat{\boldsymbol{b}_r}^{(t)}(2)\hat{\boldsymbol{b}_r}^{(t)}(3)...\hat{\boldsymbol{b}_r}^{(t)}(\tau)]$

  re-encode the data and remodulated the re-coded data as $\hat{\boldsymbol{S}}^{(t)} = [\hat{\boldsymbol{S}}^{(t)}(1)\hat{\boldsymbol{S}}^{(t)}(2)\hat{\boldsymbol{S}}^{(t)}(3)...\hat{\boldsymbol{S}}^{(t)}(\tau)]$

  Update the estimate CSI with LS estimator with $\boldsymbol{H}_{LS}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)})^H(\boldsymbol{S}^{(t)}(\boldsymbol{S}^{(t)})^H)^{-1}$ and MMSE estimator with $\boldsymbol{H}_{MMSE}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)} + N_0 N\boldsymbol{I})^{-1}\boldsymbol{S}^{(t)H}\boldsymbol{R}_{\boldsymbol{H}}$

  $\Delta = ||\hat{\boldsymbol{H}}^{(t+1)} - \hat{\boldsymbol{H}}^{(t)}||^2$

  Achieve the updated detected data $\hat{\boldsymbol{b}_r}^{(t+1)}$ with $\boldsymbol{H}^{(t+1)}$ $\Delta = ||\hat{\boldsymbol{H}}^{(t+1)} - \hat{\boldsymbol{H}}^{(t)}||^2$

  if $\Delta \leq \beta$ **then**

    quit `// given a threshold to end the loop`

  end

end

4.3.3 Proposed Learning based Channel Estimation and Detection

In this section, we first introduce the DNN-aided HD detection of CS-MIM. Then, we propose an iteratively updated DNN model for JCED of CS-MIM. Afterwards, we extend the proposed DNN-based JCED model to SD CS-MIM systems.

4.3.3.1 Conventional DNN-aided CE and HD detection

The DNN architecture of Fig. 4.5 can be harnessed for replacing the conventional HD data detector of Section III-A2). As shown in Fig. 4.5, the pilot symbols $\boldsymbol{Y_p}$ and the received symbols $\boldsymbol{Y}$ constitute the inputs of the L-layer fully-connected network. The channel $\hat{\boldsymbol{H}}$ is estimated from the pilot symbols $\boldsymbol{Y_p}$ by the DNN model during the training phase. Then the output bits $\hat{\boldsymbol{u}}$ can be obtained using the estimated channel and the received signal, yielding the output of

$$\begin{aligned}\hat{u} =& f_{sigmoid}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{Y}) + \boldsymbol{\theta_1}]) \\ &+ \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n),\end{aligned} \tag{4.17}$$

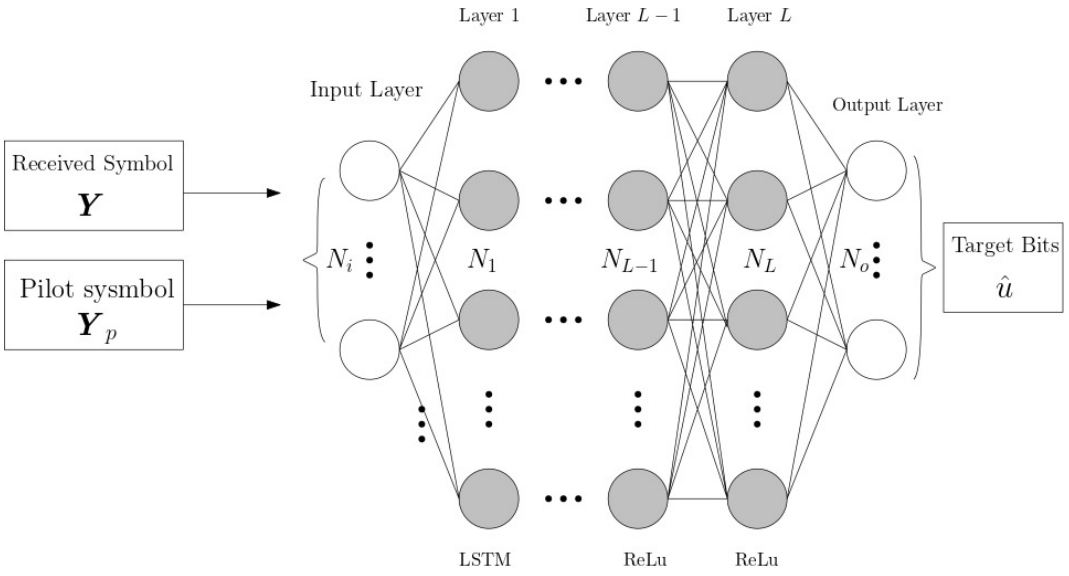

FIGURE 4.5: Fully-connected DNN model for CS-MIM channel estimation and data detection.

where $\boldsymbol{W_n}$ and $\boldsymbol{\theta_n}$, $n = 1, \cdots, L$ represent the weights and biases, respectively. A Long Short-Term Memory (LSTM) layer is employed as the initial layer to capture the non-linear relationships between the transmitted signals and the CSI. The LSTM layer can be mathematically represented as

$$\{C_k, z_k\} = \text{LSTM}(C_{k-1}, z_{k-1}, x_k; \phi_{k-1}), \tag{4.18}$$

where C_k is commonly referred to as the cell state [144], which represents the information flow over time. Additionally, x_k and z_k denote the input and output at the k-th symbol instant, respectively. The term z_{k-1} represents the output at the $k-1$-st instant, and ϕ_{k-1} denotes the LSTM layer's parameters. These parameters are stored in the cell state for subsequent iterations and are shared across them. Then in (4.17), the Rectified linear unit (Relu) function of $f_{Relu}(s) = max(0, s)$ is employed for activating the DNN during the training phase, and the sigmoid function of $f_{sigmoid}(s) = \frac{1}{1+e^{-s}}$ is used to obtain the detected bits $\hat{u}$.

Furthermore, the complexity of the Neural Network (NN) is governed by the operations involved in forward and backward propagation between each neuron. Generally, the complexity order of an NN can be expressed as $\mathcal{O}[n_i n_1] + L\mathcal{O}[n_l n_{l-1}] + \mathcal{O}[n_L n_o]$ [125], where n_i and n_o represent the sizes of the input and output layers, respectively, and $n_l (l = 1, 2, \ldots, L)$ denotes the numbers of the hidden layers between them. The equation of the sigmoid layer is formulated as $f_{sigmoid}(s) = \frac{1}{1+e^{-s}}$, which has the evaluation complexity order of $\mathcal{O}[1]$ and the LSTM has the complexity order of $\mathcal{O}[n_l(n_d + n_l)]$, where n_d is the neural dimension of the input layer of the LSTM. Then we have the total computational complexity of Fig.4.5 characterized as $\mathcal{O}[4n_l(N_f + 2 + n_l) + \sum_l^{L-1}(2n_{l+1}n_l - n_l) + 2n_{L-1}]$.

The raw input data represented in the complex-valued matrix form obtained from the received signal $\boldsymbol{Y}$ has to be vectorized first. We rearrange the complex values by separately extracting the real as well as imaginary parts and then merging them into a real-valued vector.

In the training phase, we employ randomly generated data, which are transmitted over a frequency selective Rayleigh fading channel using MIM. Then, both the received pilot and data symbols are employed as the input data of the DNN. In this case, we use a high pilot overhead for simulating a high-performance CE scenario. To maximize the performance of the trained learning-based CE and detection, different pilot overheads are applied for considering sufficiently diverse scenarios. The number of training samples required is selected based on experimentation by gradually increasing the training size until acceptable MSE values are achieved. In this case, the MSE loss function of the DNN used for the training is

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W}_n, \boldsymbol{\theta}_n) = \frac{1}{B} \sum_{i=1}^{B} \|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{4.19}$$

where B is the sample size of the current iteration. A stopping criterion can be defined either by the number of iterations or by an MSE threshold. Then, the parameter sets $\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\}$ can be updated in each training iteration based on our learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\} \leftarrow \{\boldsymbol{W}_n, \boldsymbol{\theta}_n\} - \alpha \nabla \boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\}),$$

where $\alpha > 0$ is the learning rate and $\nabla \boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W}_n, \boldsymbol{\theta}_n\})$. In our proposed NN aided detection, we use $\alpha = 0.001$.

After the training phase, the DNN model learns the mapping from the received signal and stores both the weight as well as the bias information, which will be used for producing the desired outputs based on the input data in the testing phase. The statistical properties of the input/output data have to remain the same as those used in training.

4.3.3.2 Separate DNN-aided CE and detection

To further reduce the effect of CE error, we propose the two-part DNN models of Fig. 4.6 and Fig. 4.7 for CE and detection, respectively. Firstly, the fully connected NN of Fig. 4.6 is used for estimating the channel using the current received symbol $\boldsymbol{Y}_{\boldsymbol{\tau-1}}$ and next received a symbol $\boldsymbol{Y}_{\boldsymbol{\tau}}$ as input and then it outputs the estimated CSI $\hat{\boldsymbol{H}}_\tau$, where $\hat{\boldsymbol{H}}_\tau = \{\boldsymbol{H}_\tau^1, \cdots, \boldsymbol{H}_\tau^s, \cdots, \boldsymbol{H}_\tau^{N_t}\}$.

In this case, the first received symbol is $\boldsymbol{H}_{\tau-1} = \boldsymbol{H}_p$, where $\boldsymbol{H}_p$ is the pilot symbol and the fully connected layer is used as output layer to learn the CSI. A variety of different

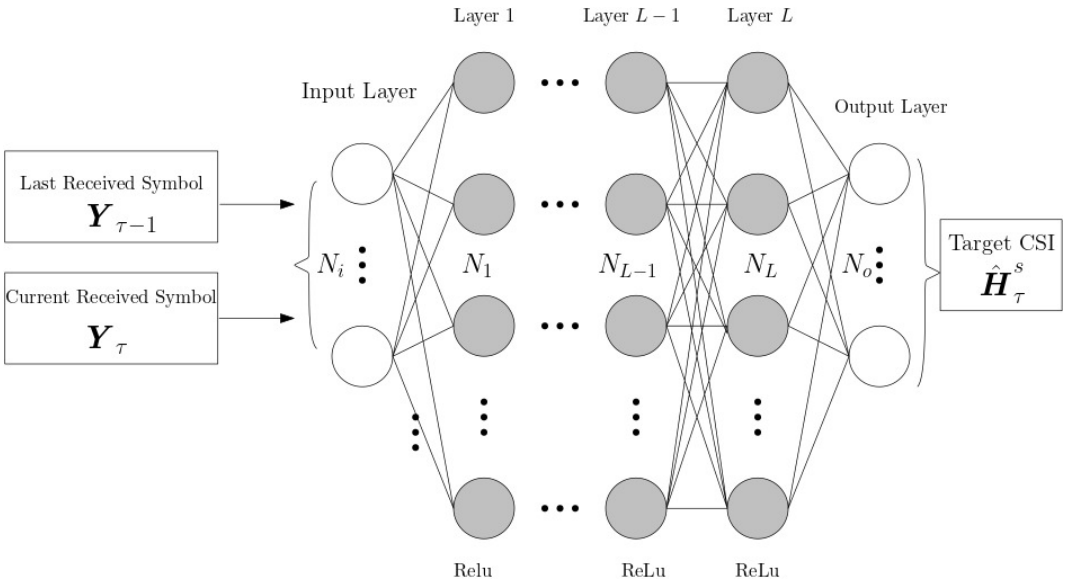

FIGURE 4.6: Separate fully-connected DNN model in CE of CS-MIM systems

pilot overheads are considered in the training phase to enhance CE performance of the trained model in diverse channel conditions. Then we can obtain the output of the DNN-aided CE as

$$\begin{aligned}\hat{H}^s_{\tau} =& \boldsymbol{W}_{N_1} ... f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1^1}(\boldsymbol{Y}_{\tau}) + \boldsymbol{b}_1]) \\ &+ \boldsymbol{b}_2\} + ... + \boldsymbol{b}_{N_1}).\end{aligned} \tag{4.20}$$

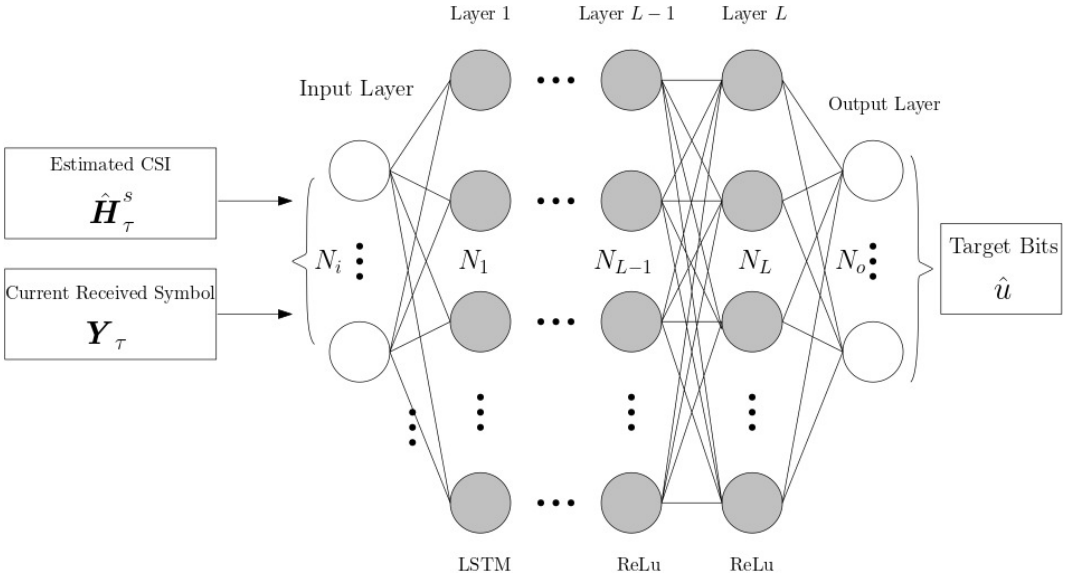

FIGURE 4.7: Separate fully-connected DNN model for detection in CS-MIM systems

In this case, we can have the complexity of CE NN as $\mathcal{O}_{CE-NN}[\sum_{l-1}^{L}(2n_l n_{l-1} - n_l) + 2n_{L-1}]$.

The training process optimizes the network weights θ by minimizing the loss function based on the MSE between the estimated CSI $\hat{\boldsymbol{H}}_{\tau}^{S}$ of antenna S and current real CSI $\boldsymbol{H}_{\tau}^{S}$. In this case, the MSE loss function used for the training is

$$\mathcal{L}(\boldsymbol{H}_{\tau}^{s}, \hat{\boldsymbol{H}}_{\tau}^{s}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B}\sum_{i=1}^{B} \|\hat{\boldsymbol{H}}_{\tau}^{s} - \boldsymbol{H}_{\tau}^{s}\|^2, \tag{4.21}$$

where B is the sample size of the current iteration.

Fig. 4.7 shows the DNN employed for detection, which is performed after completing the CSI estimation using the DNN of Fig. 4.6. The output of the first DNN model of Fig. 4.6, which is the CSI $\boldsymbol{H}_{\tau}^{s}$ of a specific activated TA s, and the received symbol $\boldsymbol{Y_{\tau}}$ are used as input for the NN of harnessed for signal detection. The output of the DNN of Fig. 4.7 corresponds to the output bits $\hat{u}$, which is formulated as:

$$\begin{aligned}\hat{u} =& f_{sigmoid}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{Y}) + \boldsymbol{\theta_1}]) \\ &+ \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n).\end{aligned} \tag{4.22}$$

Afterwards, the total complexity of two NN is $\mathcal{O}_{CE} + \mathcal{O}_{detection}$, where $\mathcal{O}_{detection}$ have the same form with the conventional NN.

In this case, the MSE loss function used for the training is

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B}\sum_{i=1}^{B} \|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{4.23}$$

where B is the sample size of the current iteration.

4.3.3.3 Proposed DNN-aided JCED

Then, in the following we propose a NN model for performing the entire JCED process, as opposed to harnessing thr pair of NNs presented in Fig. 4.6 and Fig. 4.7. The proposed DNN architecture is shown in Fig. 4.8. In this case, we have the received signal $\boldsymbol{Y}_{\tau}$ and we represent the estimated CSI of each symbol as $\mathcal{O}[\ \{\hat{\boldsymbol{H}}_0, \hat{\boldsymbol{H}}_1, \hat{\boldsymbol{H}}_2, \cdots, \hat{\boldsymbol{H}}_{\tau}\}$, which is estimated using the pilot symbols $\boldsymbol{Y}_p$

As shown in Fig. 4.8, the input of the DNN model is the estimated CSI of the previous symbol $\boldsymbol{Y}_{\tau-1}$ and the current received signal $\boldsymbol{Y}_{\tau}$. Then, the target output is the detected bit string of the symbol $\hat{u}$ and the updated CSI of the current time slot $\hat{\boldsymbol{H}}_{\tau}^{s}$, where s represents the activated TA for the current transmitted symbols.

More specifically, both the estimated CSI obtained by the DNN model of the previous symbol and the current received data are entered into the model, which requires an input layer having $[2N_tN_rN_f + 2N_rN_f]$-nodes. As shown in Fig. 4.8, the proposed DNN

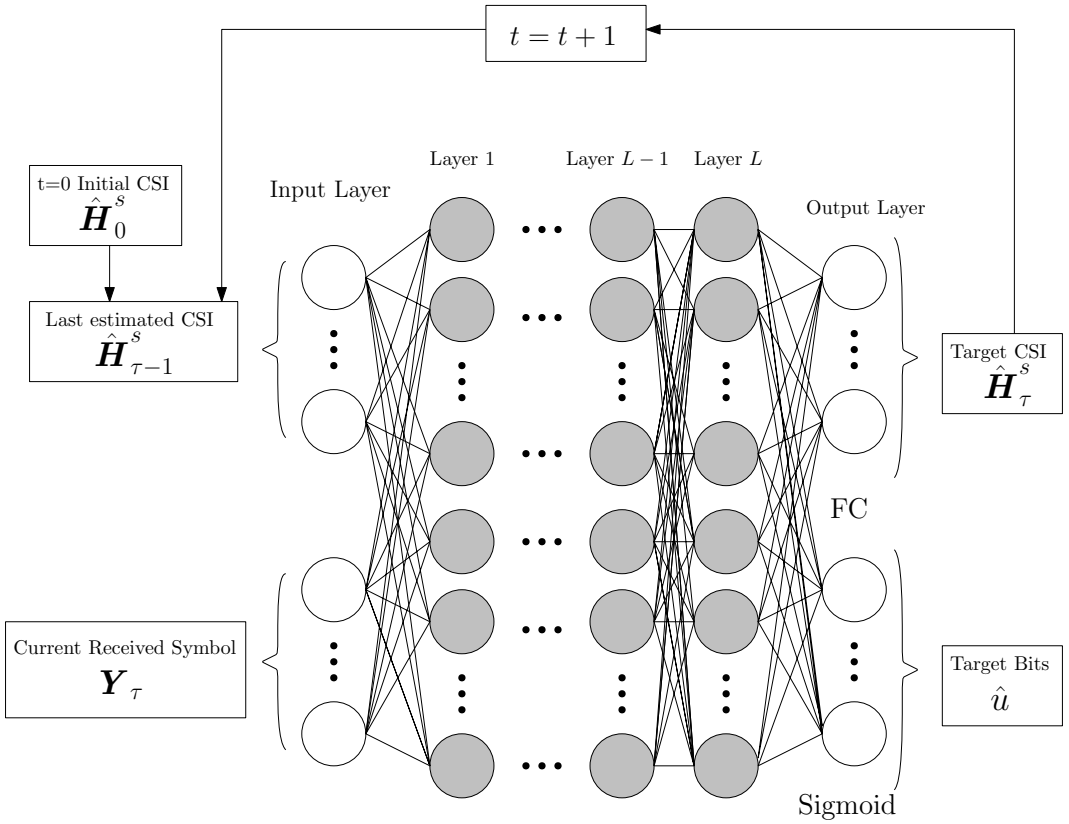

FIGURE 4.8: Fully-connected DNN model for CS-MIM JCED system

model can be split into two subgroups. The first subgroup utilizes the information of the received data and the estimated CSI to update the estimated CSI of the next symbol, while the second subgroup detects the transmitted bits of the current symbol. The proposed DNN-JCED procedure is described in **Algorithm 3**.

For HD- JCED, we consider the subgroup of the detection as a multi-label classification problem, where both the pre-processed symbols and the estimated CSI are input to a NN, which outputs the corresponding classification based candidates of each bits. For the upper subgroup of Fig. 4.8, the DNN will update the CSI using the trained weights of each layer.

Then, sigmoid activation is used for the output layer of the proposed subgroup DNN to generate dependent probabilities at the output layer of our classification problem. Hence, the output of the DNN model can be expressed as

$$\begin{aligned}\hat{H} =& \boldsymbol{W}_{N_1}^1 ... f_{Relu}\{\boldsymbol{W_2^1}(f_{Relu_1}[\boldsymbol{W_1^1}(\boldsymbol{H}_{\tau-1}) + \boldsymbol{b}_1^1]) \\ &+ \boldsymbol{b}_2^1\} + ... + \boldsymbol{b}_{N_1}^1),\end{aligned} \tag{4.24}$$

$$\begin{aligned}\hat{u} =& f_{sigmoid}(\boldsymbol{W}_{N_2}^2 ... f_{Relu}\{\boldsymbol{W}_2^2(f_{Relu_1}[\boldsymbol{W}_1^2(\boldsymbol{Y}_\tau) + \boldsymbol{b}_1^2]) \\ &+ \boldsymbol{b}_2^2\} + ... + \boldsymbol{b}_{N_2}^2),\end{aligned} \tag{4.25}$$

Algorithm 3: DNN model for JCED of CS-MIM

Input: Estimated CSI $\hat{\boldsymbol{H}}_0$, received symbol $\boldsymbol{Y}_\tau$
Output: Target bits $\hat{u}$; updated estimated CSI $\hat{\boldsymbol{H}}_\tau^s$

1 $t := 0$ `// initial iteration index`
2 **for** $\tau = 1, 2, \cdots, T$ **do**
3 Input the $\boldsymbol{Y}_\tau$ and $\hat{\boldsymbol{H}}_{\tau-1}$ to the proposed DNN model
4 Use the DNN model with the trained weights to obtain the detected bits $\hat{u}$ and updated estimated CSI, $\hat{\boldsymbol{H}}_\tau^s$
5 Update the estimate CSI with LS estimator with $\boldsymbol{H}_{LS}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)})^H(\boldsymbol{S}^{(t)}(\boldsymbol{S}^{(t)})^H)^{-1}$ and MMSE estimator with $\boldsymbol{H}_{MMSE}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)} + N_0 N\boldsymbol{I})^{-1}\boldsymbol{S}^{(t)H}\boldsymbol{R}_{\boldsymbol{H}}$
6 Update the corresponding activated s-TA CSI based on the estimate CSI with LS estimator with $\hat{\boldsymbol{H}}_\tau^s$. `// go back to step (3) for next symbol.`
7 **end**

where $\boldsymbol{W}_{n_1}^1$ and $\boldsymbol{b}_{n_1}^1$,$n_1 = 1, \cdots, N_1$, represent the weights and biases of the subgroup layers used for updating the channel estimate, while $\boldsymbol{W}_{n_2}^2$ and the bias $\boldsymbol{b}_{n_2}^2$,$n_2 = 1, \cdots, N_2$, are the weights and biases of the layers employed for detecting the information bits. Then, we have the weight sets of

$$\theta^1 = \{\boldsymbol{W}_1^1, \boldsymbol{b}_1^1, \boldsymbol{W}_2^1, \boldsymbol{b}_2^1, \cdots, \boldsymbol{W}_{N_1}^1, \boldsymbol{b}_{N_1}^1\} \tag{4.26}$$

and

$$\theta^2 = \{\boldsymbol{W}_1^2, \boldsymbol{b}_1^2, \boldsymbol{W}_2^2, \boldsymbol{b}_2^2, \cdots, \boldsymbol{W}_{N_2}^2, \boldsymbol{b}_{N_2}^2\} \tag{4.27}$$

As the number of the first layer nodes depends on the input data size, the appropriate number of nodes should be selected for the hidden layers which is sufficiently high for attaining an enhanced BER performance, at reduced detection complexity. In this case, we designed 3 hidden layers having 64 nodes used for both subgroups.

In the training phase, we use randomly generated data, transmitted over the wireless channel using MIM as the input data and perfect CSI for training the model weights θ^1 and θ^2. In this case, the MSE loss function used for the training is

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W}_n, \boldsymbol{\theta}_n) = \frac{1}{B}\sum_{i=1}^{B} \|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{4.28}$$

where u represents the target labels, $\hat{u}$ denotes the detected bits and B is the sample size of the current iteration. Using (4.17) and (4.25), we can obtain the loss function of

this DNN model as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}^1,\boldsymbol{\theta}^2) = &\frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{H}}_\tau^s - \boldsymbol{H}_\tau^s\|_2^2 \\ &+ \frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{u}}_\tau - \boldsymbol{u}_\tau\|_2^2.\end{aligned} \tag{4.29}$$

We can define a stopping criterion, which can be either the number of iterations or an MSE threshold. Then, the parameter sets $\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}$ can be updated in each training iteration based on the learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} \leftarrow \{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} - \alpha\nabla\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}),$$

where $\alpha > 0$ is the learning rate and $\nabla\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$. In our proposed NN aided detection, we use $\alpha = 0.001$.

Then, during the training phase, the model learns the mapping from the received signal and stores both the weight and bias information, followed by outputting the predicted results that are expected to approximate the desired input data having similar statistical properties to those of the training.

In this model, the pair of inputs exhibit independent input connection complexity, which is characterized by $\mathcal{O}[n_{i1}n_1 + n_{i2}n_1]$. The complexity of the hidden layers and of the output layer is identical to that of the conventional NN. More specifically, we can have the computational complexity of $\mathcal{O}[2n_{i1}n_1 + 2n_{i2}n_1 + \sum_l^{L-1}(2n_{l+1}n_l - n_l) + 2n_{L-1}]$

For our SD-JCED system, we also consider a similar DNN architecture to that of [125], but we have a different output for the model. Since the conventional SD detector will obtain the LLRs of received signal after the CS-MIM soft demodulator, we replace the detected bits $\hat{\boldsymbol{u}}$ by the extrinsic LLR L_e at the output. Then the output of the SD DNN model can be expressed as

$$\begin{aligned}\hat{\boldsymbol{H}} = &\boldsymbol{W}_{N_1}^1 ... f_{Relu}\{\boldsymbol{W_2^1}(f_{Relu_1}[\boldsymbol{W_1^1}(\boldsymbol{H}_{\tau-1}) + \boldsymbol{b}_1^1]) \\ &+ \boldsymbol{b}_2^1\} + ... + \boldsymbol{b}_{N_1}^1),\end{aligned} \tag{4.30}$$

$$\begin{aligned}\hat{\boldsymbol{L}}_e = &\boldsymbol{W}_{N_2}^2 ... f_{Relu}\{\boldsymbol{W}_2^2(f_{Relu_1}[\boldsymbol{W}_1^2(\boldsymbol{Y}_\tau) + \boldsymbol{b}_1^2]) \\ &+ \boldsymbol{b}_2^2 + ... + \boldsymbol{b}_{N_2}^2),\end{aligned} \tag{4.31}$$

and the corresponding loss function is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}^1,\boldsymbol{\theta}^2) = &\frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{H}}_\tau^s - \boldsymbol{H}_\tau^s\|_2^2 \\ &+ \frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{L}}_e(\tau) - \boldsymbol{L}_e(\tau)\|_2^2.\end{aligned} \tag{4.32}$$

4.4 Performance Analysis

In this section, we characterize the learning-aided CS-MIM system proposed in Section III relying on both HD and SD. The performance of the conventional detector will also be presented for comparison with the proposed methods. We also consider systems having $N_t = 4, 8$ with 2 RF chains. More specifically, only the bits for antenna selection $b_{g,3}$ is changeable. Furthermore, we also investigate the performance of the proposed methods in different channel conditions. To characterize the channel conditions, we adjust the normalized maximum Doppler frequency f_m in order to emulate both slow- and fast-fading channels. We assume that the system's signalling rate is $100MBaud$ and the maximum Doppler frequency is $100Hz$, which corresponds to a normalised Doppler frequency f_m of 10^{-6}.

The resultant BER performance is evaluated by Monte-Carlo simulations. Using the parameters summarized in Table 4.2 and the parameters used by the learning models outlined in Table 4.3, we investigate a set of five schemes for $N_t = 4, 8$, respectively, which are summarised as follows:

1. **Scheme 1**: HD- ML-based Detection of CS-MIM system with TAs and RAs $N_t = N_r = 4, 8$.
   (a) perfect CSI at receiver.
   (b) MMSE CE and ML detection.
   (c) MMSE-aided-JCED with $I_{max} = 3$ iteration.
2. **Scheme 2**: HD DNN-aided CE and detection of CS-MIM system with TAs and RAs $N_t = N_r = 4, 8$.
   (a) Conventional DNN-aided CE and Detection.
   (b) DNN-aided JCED with $I_{max} = 3$ iteration.
3. **Scheme 3**: SD Detection of convolutional coded CS-MIM system with TAs and RAs $N_t = N_r = 4, 8$.
   (a) perfect CSI at receiver.
   (b) MMSE-aided-JCED with $I_{max} = 3$ iteration.
4. **Scheme 4**: SD DNN-aided JCED of convolutional coded CS-MIM system with TAs and RAs $N_t = N_r = 4, 8$.
   (a) SD conventional DNN-aided CE and Detection.
   (b) SD DNN-aided JCED with 3 iteration.

TABLE 4.2: CS-MIM system simulation parameters.

Parameters	Values
Multi-carrier System	OFDM
Number of subcarriers, N_c	128
Cyclic prefix	16
Number of subcarrier groups, G	16
Number of subcarrier/group, M_g	8
Number of available indices/group, N_a	16
Number of active indices/group, K	2
STSK, $(M, N, T, Q, \mathcal{L})$	(2,2,2,2,2)
TAs, N_t	4/8
RAs, N_r	4/8
RSC code, (n, k, K)	(2,1,3)

TABLE 4.3: Training configuration for Conventional DNN detection with CSI-aided

Setting	Hard-decision	Soft-decision
Input layer	LSTM	
Activation function	ReLu	
Number of FC layers in NN	3	3
Number of neurons in each FC	[128,256,128]	[256,512,256]
Input	Received Symbol+Pilot Symbol	
output	Detected Bits	LLR
Activation Function for output layer	Sigmoid	ReLu
Initial learning rate	0.001	
Target SNR for training	25dB	10dB
Training data size	100000	
Validation data ratio	0.1	

TABLE 4.4: Training configuration for our learning-aided JCED methods

Setting	Hard-decision	Soft-decision
Input Layer	FC	
Activation Layer	ReLu	
Number of FC layers in NN	3	3
Number of Neurons in each FC	[128,256,128]	[256,512,256]
Input	Received Symbol+ estimated CSI	
Output	Detected Bits Estimated CSI	LLR Estimated CSI
Activation function for output layer	Sigmoid ReLu	ReLu ReLu
Initial learning rate	0.001	
Target SNR for training	25dB	10dB
Training data size	80000	
Validation data ratio	0.15	

In the following, we present various schemes considered for comparative analysis in our simulations. Table 4.5 characterizes the performance vs. complexity relationship of

TABLE 4.5: Simulation results and complexity analysis of each Scheme.

Scheme index			overhead	SNR at BER of 10^{-4}	Throughput(bits/s/Hz)	Complexity
HD Detection						
Scheme 1	a)	$N_t = 4$	-	4.2	1.222	2.1×10^6
		$N_t = 8$	-	2.0	1.333	8.5×10^6
	b)	$N_t = 4$	2%	6.0	1.198	1.2×10^6
			5%	3.8	1.161	1.2×10^6
			10%	2.1	1.100	2.1×10^6
		$N_t = 8$	2%	9.6	1.306	8.5×10^6
			5%	6.1	1.266	8.5×10^6
			10%	4.4	1.200	8.5×10^6
	c)	$N_t = 4$	2%	4.3	1.198	4.5×10^6
		$N_t = 8$	2%	2.0	1.306	4.1×10^7
Scheme 2	a)	$N_t = 4$	2%	5.7	1.198	2.1×10^4
		$N_t = 8$	2%	3.4	1.306	2.2×10^4
	b)	$N_t = 4$	2%	5.1	1.198	3.2×10^4
		$N_t = 8$	2%	2.3	1.306	3.8×10^4
SD Detection						
Scheme 3	a)	$N_t = 4$	-	1.7	0.611	3.2×10^8
		$N_t = 8$	-	-1.9	0.667	1.2×10^9
	b)	$N_t = 4$	1.1%	1.7	0.599	5.6×10^9
		$N_t = 8$	2%	-1.8	0.653	3.6×10^9
Scheme 4	a)	$N_t = 4$	2%	2.9	0.599	5×10^4
		$N_t = 8$	2%	0	0.653	5.3×10^4
	b)	$N_t = 4$	2%	2.1	0.599	2.5×10^5
		$N_t = 8$	2%	-1.1	0.653	3.1×10^5

diverse schemes. Initially, we demonstrate the performance of JCED and conventional CE as displayed in **Scheme 1** with $N_t = 8$ and $N_t = 4$. Then we also show the benefit of the DNN-based CE and detection as well as proposed JCED-DNN as listed in **Scheme 2**.

Given the system parameters of Table 4.2, the achievable rate is $\frac{bG}{N_c+L_{CP}} = 1.333$ bits/sec/Hz for $N_t = 8$ and $R_t = 1.222$ bits/sec/Hz for $N_t = 4$. Fig. 4.10 shows the BER of the **Scheme 1** and **Scheme 2**. Besides, Fig. 4.10 characterizes the theoretical maximum rate of CS-MIM in Discrete-Input Continuous-Output Memoryless Channels (DCMC) for both the neural network model and conventional CE methods. For CS-MIM system with 8 TAs, both at transmitter and receiver, the **Scheme 1a)** achieves about 1.95 dB at the BER of 10^{-4} under the assumption of perfect CSI knowledge at the receiver. In this case, we can achieve highest throughput as shown in Fig. 4.9 which is $R_t = 1.333$ bits/sec/Hz. However, in more realistic situations, pilots are required to be deployed by the CE techniques and result in pilot overhead. Then, 1% pilot overhead indicates that every 100 symbols require 1 pilot symbol. As shown in Fig. 4.9 and Fig. 4.10, **Scheme 1b)** is capable of achieving an improved performance, but at an increased pilot overhead. **Scheme 1b)** associated with 10% pilot overhead is capable of achieving similar results to those of **Scheme 1c)**. Furthermore, **Scheme 1b)** associated with 2% overhead and 5% overhead exhibit a 4dB and 1.7dB discrepancy with respect to the ideal **Scheme 1c)** at a BER of 10^{-4}, respectively. When **Scheme 1c)** of JCED is applied at the receiver,

it can significantly reduce the pilot overhead and yet obtain a near-ML performance. More specifically, we consider the JCED under 3 iteration updating and achieve BER of 10^{-4} only 0.1 dB SNR worse than **Scheme 1a)** of ML detector with very few pilot.

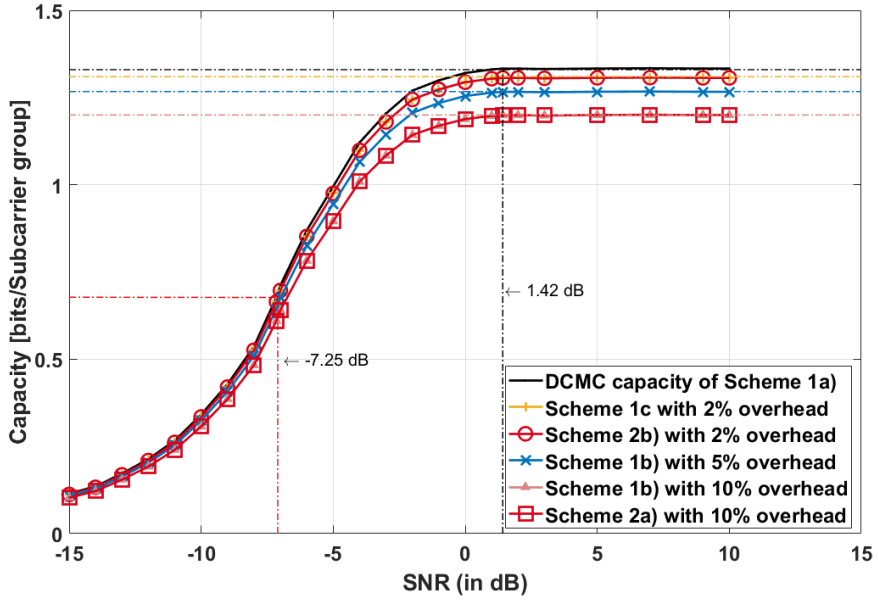

FIGURE 4.9: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

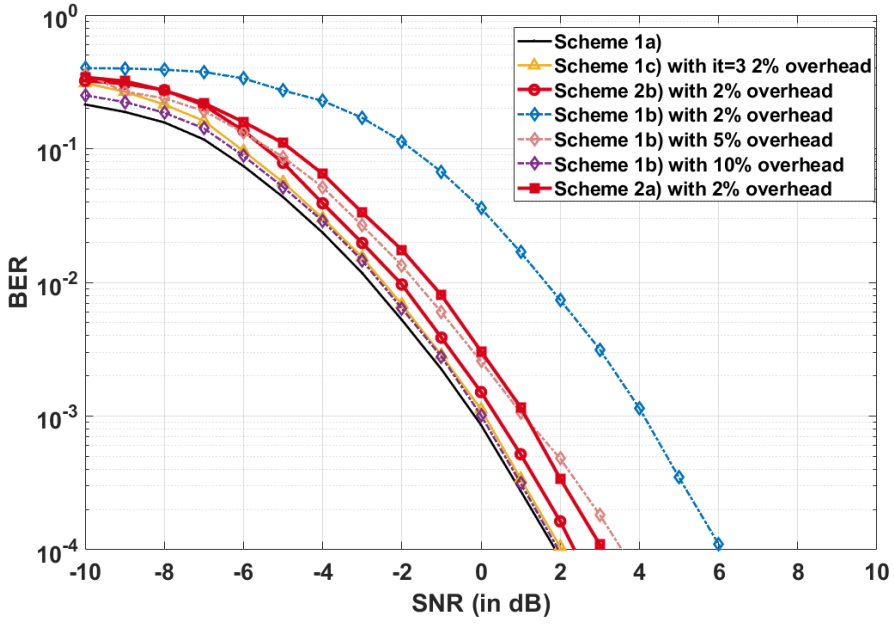

FIGURE 4.10: The DCMC capacity for HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

We also analysis the detection performance of the system with 4 TAs. Fig. 4.11 also shows the DCMC of **Scheme 1** and **Scheme 2**. With less antennas, firstly, the performance of CS-MIM is reduced due to reduction of space sparsity. **Scheme 1a)** having

$N_t = 4$ TAs achieves a BER of 10^{-4} at 4.3 dB SNR, as shown in Fig.4.12. Similarly, along with $N_T = 8$ TAs **Scheme 1b)** also requires a higher pilot overhead for achieving a high performance. In conjunction with a 2% pilot overhead **Scheme 1b)** is about 5 dB worse than **Scheme 1a)** and their gap is reduced to 1.6 dB for 5% pilot overhead. However, **Scheme 1c)** still succeeds in achieving near-capacity performance, as shown in Fig. 4.12.

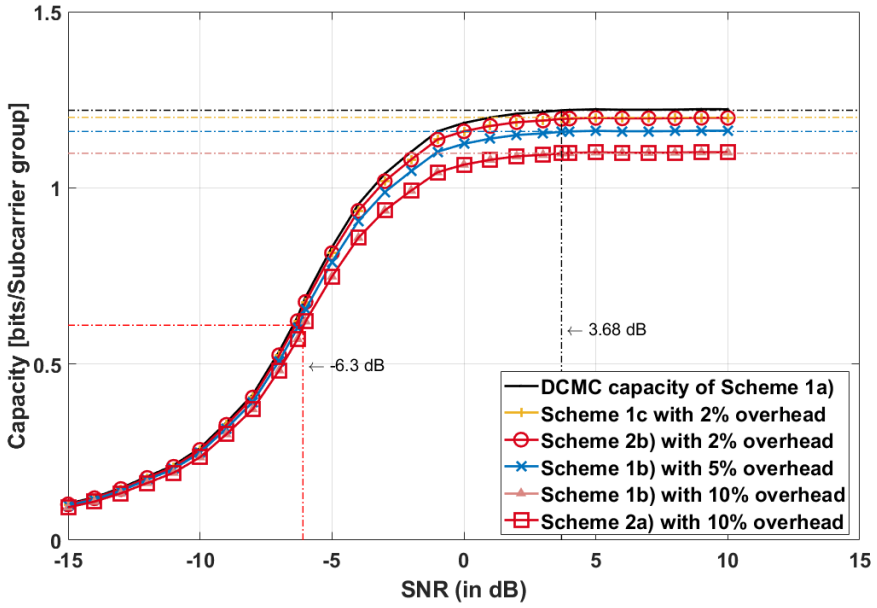

FIGURE 4.11: The DCMC capacity for HD detector of **Schemes 1,2** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

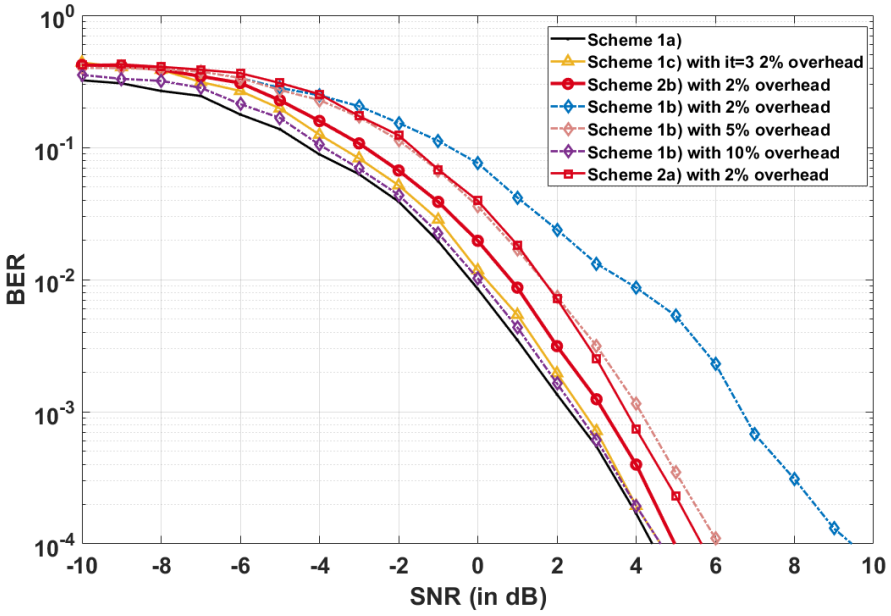

FIGURE 4.12: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

As we discussed in Section III, the ML detector applies an exhaustive search having complexity order of $\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$. On the other hand, the complexity of the Neural network is determined by the forward and backward propagation, where we have the NN complexity order of $\mathcal{O}[n_i n_{h_i} + n_{h_i} n_{h_{i+1}} + \cdots + n_{h_I} n_o]$[3]. In this case, we assume that the network have I layers and each layer have nerual size of $n_{h_i}(i = 1, 2, 3, \cdots, I)$ and n_i, n_o represents the neural size of input and output layers. Although the DNN-based JCED model require at least 3 iteration, which means 3 times complexity than conventional DNN model, to achieve near-ML performance, it is several magnitude less of complexity compared to the conventional JCED either CE-ML detection method.

We also conducted simulations for two variants of **Scheme 2**. Leveraging a high pilot overhead based estimated CSI from **Scheme 1b**, the model can be efficiently trained to achieve improved detection performance, even with a reduced pilot overhead in challenging channel conditions. As depicted in Fig. 4.9, **Scheme 2a** exhibits a performance that is approximately 2 dB inferior to **Scheme 1a**. By employing the iteratively updated CE model, **Scheme 2b** further minimizes the estimated CE error, resulting in a mere 0.9 dB loss at a BER of 10^{-4}. Notably, **Scheme 2b** achieves a nearly 1 dB improvement over **Scheme 2a** at a computational complexity of roughly 3×10^4. This increase in complexity may be deemed acceptable, especially when compared to the complexity of **Scheme 1b** (1.2×10^6) and to that of **Scheme 1a** (8.5×10^6) over three iterations. We also investigate the system associated with $N_t = 4$ TAs. Then the performance of **Scheme 2** is slightly degraded owing to is eroded diversity gain. **Scheme 2b)** attains a BER of 10^{-4} at SNR of 5.1 dB, while the conventional CE-aided DNN **Scheme 2a)** performs 0.8 dB worse than **Scheme 2b)**.

Additionally, we compare the performance for varying Doppler frequency values. Specifically, we modulate the normalized Doppler frequency f_m to emulate channels ranging from slow to fast variations.In fig. 4.13 we consider a channel with normalised Dople frequency of$f_m = 2 \times 10^{-6}$, while we used $fm = 10^{-6}$ in the results in Fig. 4.10. **Scheme 1a** maintains consistent results as observed in Fig. 4.9, while **Scheme 1b** with a 10% overhead incurs a 0.7 dB loss at a BER of 10^{-4}, compared to the 0.1 dB difference in Fig. 4.10. In this context, **Scheme 1c** with a 5% overhead demonstrates superior CE accuracy compared to **Scheme 1b**, necessitating a higher overhead to achieve near-optimal performance relative to **Scheme 1c** in Fig. 4.10. Similarly, despite the increased overhead in **Scheme 2a** and **Scheme 2b** aiming for enhanced CE accuracy and detection performance, they exhibit losses of 0.3 dB and 1.3 dB, respectively, at a BER of 10^{-4} when compared to their counterparts in Fig. 4.10. This suggests that **Scheme 1c** and **Scheme 2b** offer some resilience against rapidly varying channels. As illustrated in Fig. 4.14 with $f_m = 10^{-5}$, **Scheme 2a** is 2.6 dB inferior to **Scheme 1b** in Fig. 4.10, while **Scheme 2b** lags by 1.2 dB compared to **Scheme 1c** in Fig. 4.10.

[3]Complexity order of NN only used to compares the ML detection, while there is no search complexity associated with the NN-aided detection.

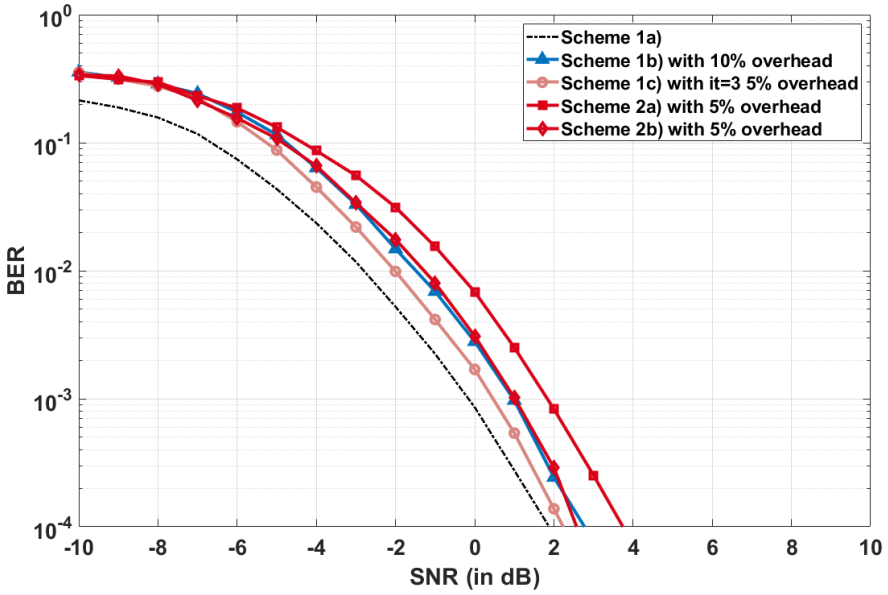

FIGURE 4.13: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 2 \times 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

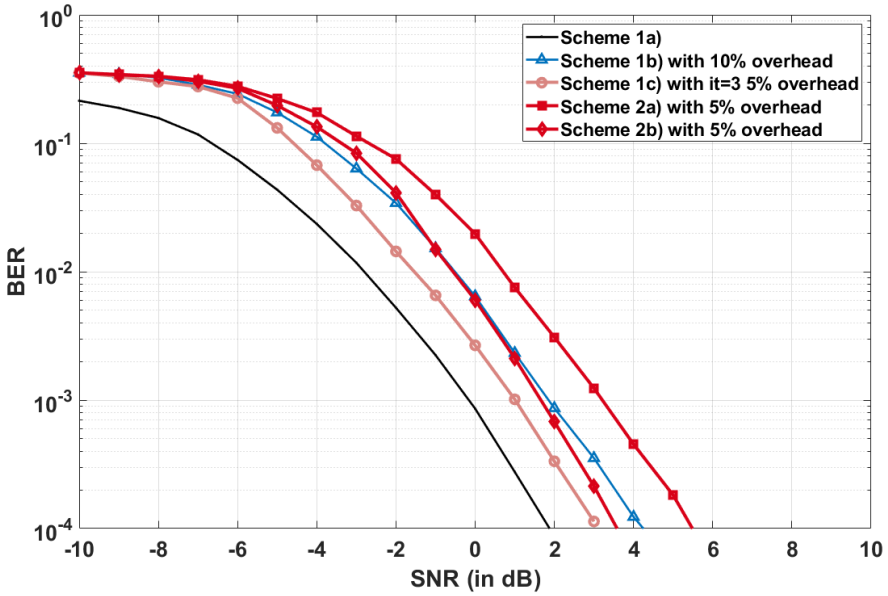

FIGURE 4.14: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-5}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

Let us now consider the performance of SD detection, where we employ a half-rate RSC encoder as shown in Table 4.2. Then we can calculate the maximum achievable rate is $R_t = 0.66667$ bits/sec/Hz for system which $N_t = 8$ and $R_t = 0.61111$ bits/sec/Hz for system which $N_t = 4$. As shown in Fig. 4.17, **Scheme 3a)** could achieve a BER of 10^{-4} at -1.83 dB with perfect CSI acquired at receiver. In practical situation, CE is required with highly pilot overhead. Naturally,the **scheme 3b)** of JCED detection could achieve near-ML performance which achieve $-1.8dB$ SNR at 10^{-4} of BER with few pilot and

moderate complexity mounting. For $N_t = 4$, the JECD could achieve less difference with ML detector which is $0.14dB$ worse than the ML detector. For NN-based CE and detection, the conventional model **Scheme 4a)** leads to about 2 dB gap of 10^{-4} BER compared with **Scheme 3b)** and **scheme 3b)**. With the assist of DNN-based JCED, we can narrow the gap to 1 dB with 3 iteration updating. For system with $N_t = 4$, the performance of DNN-JCED is more effective with the significant reduce in TA and RA number.

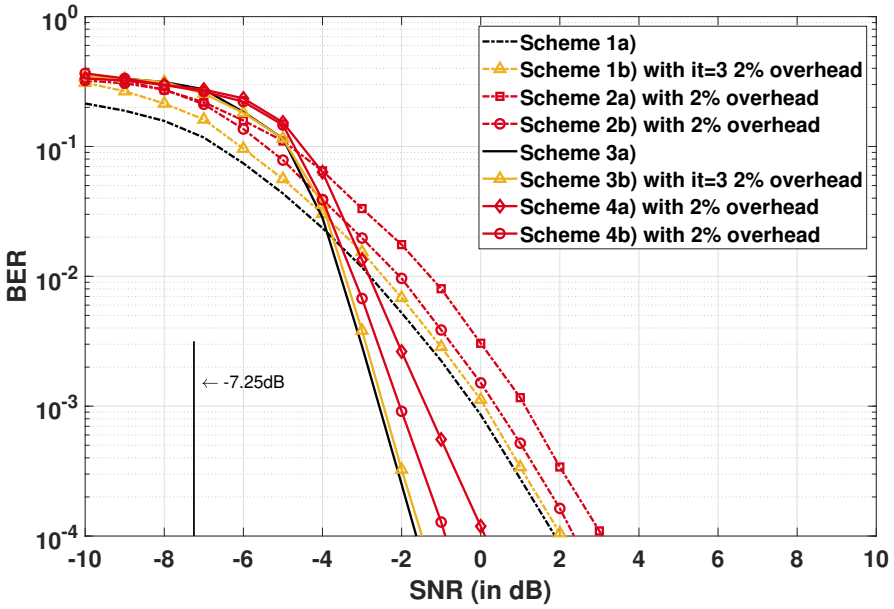

FIGURE 4.15: BER performance comparison of HD and SD detector of **Schemes 1-4** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

In Fig. 4.15 and Fig. 4.16 , we compare the performance of the HD and SD. In Fig.4.15 the benefit of the SD is clearly visible, because it provides a sharp BER reduction at an SBR of about -3 dB, while **Scheme 3a)** requires -2.2 dB SNR at at a BER of 10^{-4}. Both **Scheme 1b)** and **Scheme 1c)** are capable of achieving near-optimal results. For learning-based CE and detection, **Scheme 3a)** and **Scheme 3b)** perform slightly worse than the conventional **Scheme 1**, while the SD scheme attains a 3.1 dB and 2.9 dB gain compared to the HD-aided **Scheme 2a)** and **Scheme 2b)**. As expected, the performance improvement of SD is worse for $N_t = 4$ than for $N_t = 8$, as shown in Fig.4.16, where the gap between **Scheme 1a)** and **Scheme 3a)** is about 2.4 dB at a BER of 10^{-4}, compared to a discrepancy of 3.6 dB in Fig.4.15.

Fig. 4.17 also characterizes the learning aided JCED-SD detection methods applied to our CS-MIM system. The NN based JCED method is about 0.3 dB worse than ML detector with the perfect CSI acquired at receiver. With more antenna for transmitting, the performance is slightly degrading. However, the complexity of DNN based JCED is far small than conventional JCED method with $N_t = 8$ system. For

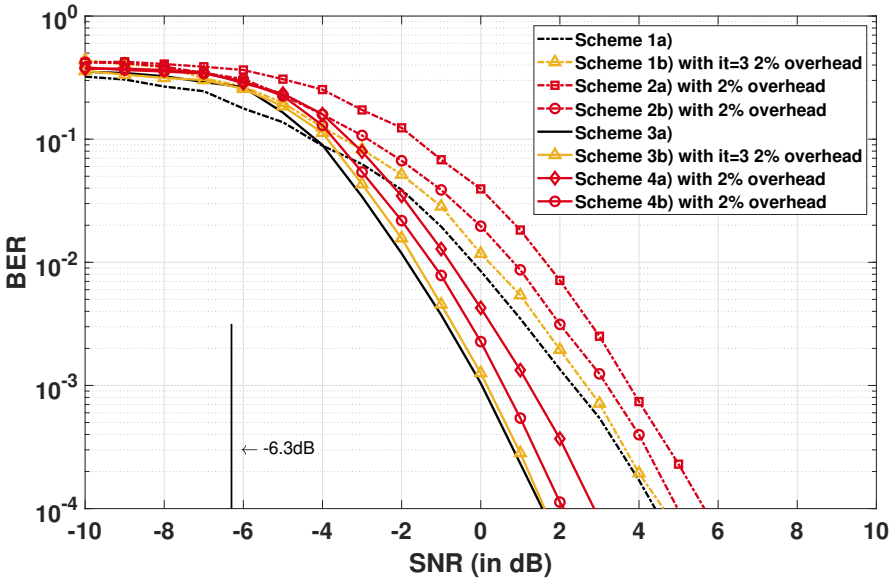

FIGURE 4.16: BER performance comparison of HD and SD detector of **Schemes 1-4** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table 4.2 and Table 4.3.

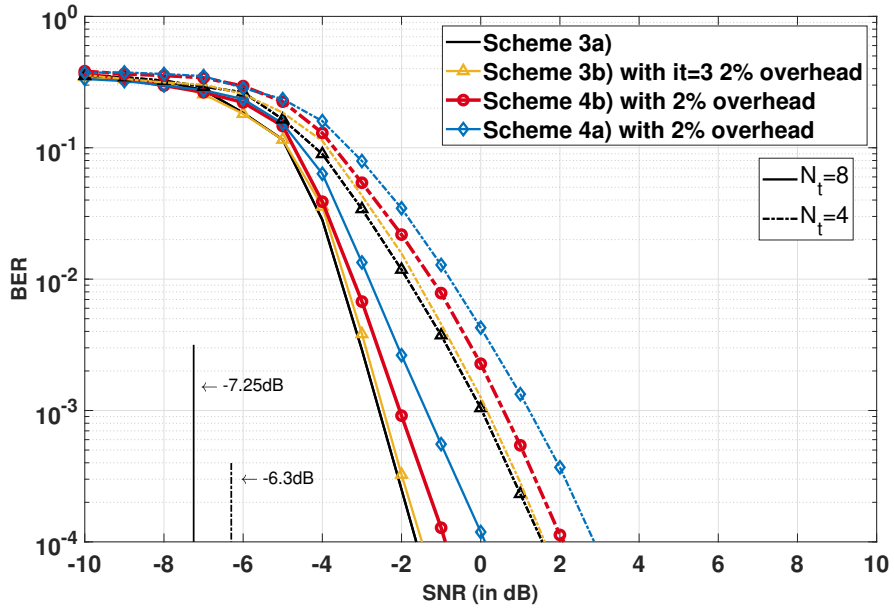

FIGURE 4.17: BER performance comparison of SD detector with **Scheme 3,4** under $f_m = 10^{-6}$. The simulation parameter are shown in Table 4.2 and 4.3

higher number of iterations update, the NN model will have an improved performance. However, the proposed learning method has a complexity order of $\mathcal{O}[2n_{i1}n_1 + 2n_{i2}n_1 + \sum_l^{L-1}(2n_{l+1}n_l - n_l) + 2n_{L-1}]$ compared to $\mathcal{O}_{MAP}[c_g(N_rN_tMN_f^3 + N_rM^2N_f^2N_v + N_rN_fM^2N_vK + N_rN_fMKT)N_{AC}N_{SI}(Q\mathcal{L})^K + c_g 2^{c_g}N_{AC}N_{SI}(Q\mathcal{L})^K]$ for the conventional scheme, where I_{it} denotes the number of iterations. The complexity is shown in Table 4.5

4.5 Summary

Both conventional and learning-assisted JCED of CS-MIM was proposed relying on HD and SD. Our analysis shows that JCED was the potential of reducing the pilot overhead and yet improve the detection performance compared to the separate CE and detection. In simulation, we have first used the conventional HD JCED of CS-MIM systems communicating over Rayleigh fading channels and the learning-aided JCED is capable of achieving similar performance while decrease the complexity of JCED. Then, a DNN model with subgroups has been designed for SD JCED in CS-MIM systems, which are capable of approaching the performance of conventional SD CS-MIM system with reduced computational complexity. In summary, our studies and simulation results have shown that the conventional JCED is capable of achieving a similar BER performance to the ML detector with idealized CSI.

Chapter 5

Near-Instantaneously Adaptive Learning-Assisted and Compressed Sensing-Aided Joint Multi-Dimensional Index Modulation

5.1 Introduction

In Chapter 3 and chapter 4, learning-aided blind detection and JCED have been applied in the joint CS and MIM and both HD and SD were proposed using learning aided methods to reduce the complexity compared with the classical detection methods. However, the MIM concept discussed in Chapter 2, IM of multiple dimensions are associated, while in Chapter 3 and 4 , only TD and SpD are jointly designed. In this chapter, we consider a tighter integration of different IM dimensions to achieve further trade-off benefits and introduce one possibility of MIM combination in this Chapter as shown in Fig. 5.1.

To further increase the overall performance, Datta *et al.* proposed the concept of Generalized SIM (GSIM) and proved that Generalized Space-and-Frequency IM (GSFIM) achieves better performance than MIMO-OFDM. Their solution conveyed extra information in the SM part compared to GSIM [57]. However, the detection complexity of GSFIM escalates. Hence, Chakrapani *et al.* [159] proposed a message passing based low-complexity detection method for reducing the complexity of GSFIM detection. Furthermore, inspired by the SM and Quadrature SM (QSM) concepts [41], Quadrature Space-Frequency IM (QSF-IM) was proposed in [160], which applies dual antenna

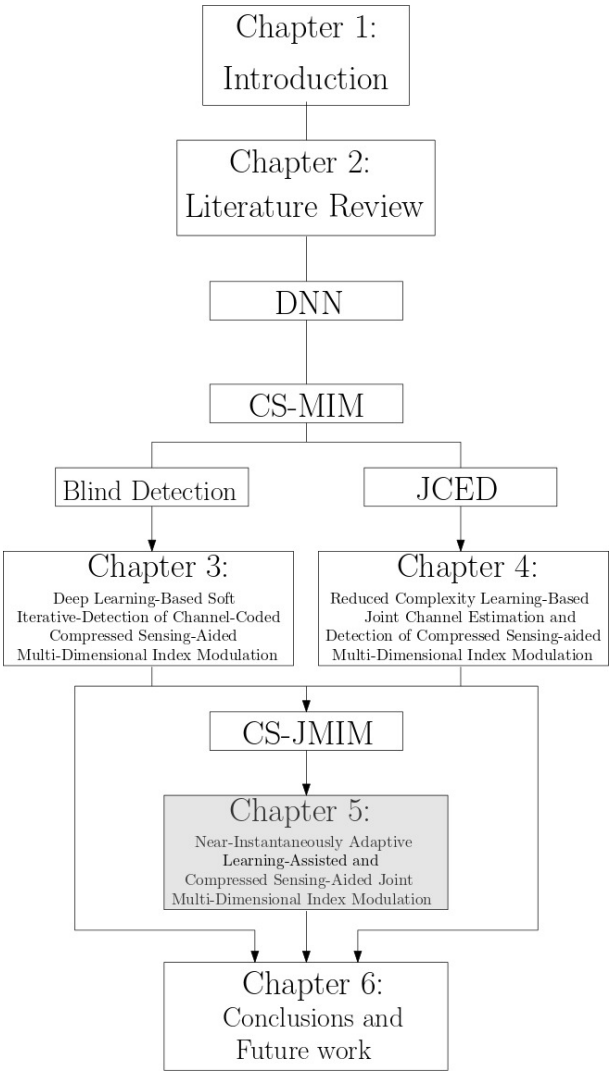

FIGURE 5.1: The organization of this thesis-Chapter5

constellation for the in-phase and quadrature-phase transmission, in order to increase the throughput without extra energy consumption. Hence this solution struck a compelling SE, EE and CoE trade-off.

Furthermore, several researchers considered the design of multi-dimensional index modulation relying on both the spatial domain and FD. For example, Space-Frequency Shift Keying (SFSK) [58] relies on an SFSK dispersion matrix (DM), which achieves beneficial transmit diversity in rapidly time-varying channels. Space-Time Shift Keying (STSK) constitutes another multi-functional MIMO [50] technique in the family of Multi-Dimensional Index Modulation (MIM). It combines the time domain (TD) and the spatial domain and it is capable of striking a beneficial diversity versus multiplexing trade-off [51]. More specifically, in STSK, Q DMs are designed for spreading the signal over T Time Slots (TSs) and M Transmit Antennas (TA) in the TD and the spatial domain, respectively. Furthermore, the IM design activates one out of the Q DMs

TABLE 5.1: Contrasting our contributions to the literature

Contribution	proposed*	[17]	[105]	[19]	[120]	[142]	[153]	[124]	[125]	[158]
Index modulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CS at transmitter	✓	✓		✓						
Learning aided detector	✓		✓		✓			✓	✓	✓
Soft decision detector	✓		✓				✓		✓	
Adaptive design for index modulation	✓									✓
Multi-dimensional index modulation	✓			✓					✓	
Joint index mapping design	✓			✓					✓	
3-Dimensional joint index modulation	✓									

for transmission, hence $\log_2 Q$ extra IM bits may be conveyed. By appropriately adjusting these parameters, improved Bit Error Ratio (BER), throughput and complexity trade-offs may be struck [161].

Additionally, the concept of Multi-dimensional IM (MIM) was proposed in [18], which is capable of improving the degrees of freedom, hence achieving all the benefits of the IM concept in several domains without introducing extra deployment costs, such as extra RF chains or transmission power. Furthermore, Lu *et al.* [59] proposed to combine CS techniques with STSK and OFDM-IM, which is an MIM system concept that inherits the benefits of both STSK and OFDM-IM. As a further advance, SM was also integrated into this MIM scheme for TA selection in [19]. In [90], the concept of multi-functional layered SM was proposed, which offers flexible trade-offs in terms of performance, hardware cost and power dissipation.

However, in previous MIM schemes, the index selection was performed separately in each dimension. By contrast, in this Chapter, we extend this concept to a Joint MIM system, where we jointly designs the IM in several dimensions. More specifically, the degrees of freedom of the IM mapping design is increased by harnessing multiple dimensions, which leads to a more flexible trade-off between the throughput, power efficiency, and cost. In this case, both SFSK and STSK can be considered as special cases of the proposed joint multi-dimensional index modulation (JMIM) family. JMIM may also be combined with CS techniques for increasing the spectral efficiency.

Table 5.1 boldly contrasts the novelty of this chapter to the literature. More explicitly, the contributions of this chapter can be further detailed as follows:

1. We propose the CS-JMIM system concept and present several JMIM mapping matrix designs. Then, we demonstrate that the proposed JMIM mapping design is capable of striking an attractive trade-off between diversity and throughput.

2. We propose a DL-based HD detection aided CS-JMIM system that can achieve near-ML performance, while imposing significantly reduced complexity. Furthermore, we propose a DNN-aided SD detector for the proposed CS-JMIM that is capable of achieving near-capacity performance.

3. Both a *K*-nearest neighbour (KNN) algorithm based and a DL-assisted adaptive modulation scheme is proposed for CS-JMIM. We demonstrate that the learning-assisted adaptive CS-JMIM scheme is capable of selecting more appropriate CS-JMIM mapping design for transmission than its conventional threshold-based adaptive counterparts. Hence it can obtain a significant throughput gain over the conventional threshold-based adaptive method.

4. Our simulation results demonstrate that the proposed learning-based detector is capable of approaching the performance of the conventional coherent detection techniques at a reduced detection complexity. We also provide the associated capacity and throughput analysis, for characterising the trade-off between each mapping matrix and the benefits of the learning-assisted adaptive method.

The rest of the chapter is organized as follows. In Section 5.2, the system model of CS-JMIM is presented. In Section 5.3, we characterize both HD and SD based learning-aided detectors. Then, in Section 5.4 we present our proposed adaptive system design. In Section 5.5, we present our simulation results, while our conclusions are offered in Section 5.6.

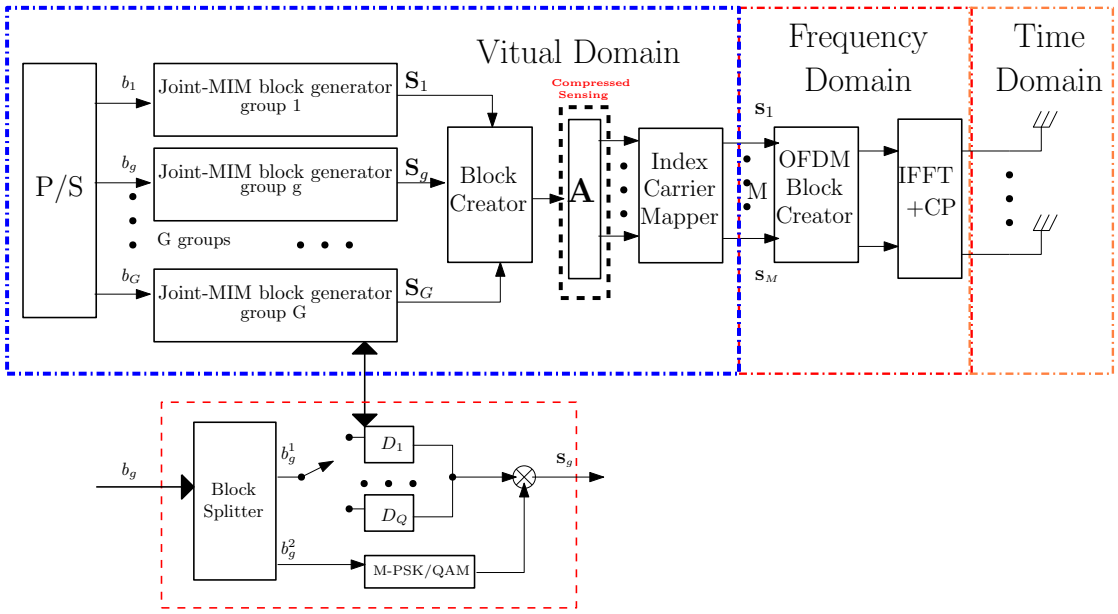

FIGURE 5.2: CS-JMIM system transmitter block diagram.

5.2 System Model

In this section, we introduce the transceiver model of the proposed CS-JMIM system employing N_t TAs and N_r Receive Antennas (RAs). Fig. 5.2 shows the block diagram of the CS-JMIM system considered, where b bits are equally divided into G groups. We consider OFDM having Nc subcarriers, which are then split into G groups and

each group has $N_f = N_c/G$ subcarriers in the FD[1], while N_{vt} TAs and N_v subcarriers of each group are applied for the CS-JMIM system in the Virtual Domain (VD) [2]. To be more specific, in each subcarrier group, there are N_v available subcarrier indices within the VD, where the dimension N_v of the VD is larger than the dimension N_f of the FD. Similarly, N_{vt} antennas in the VD are larger than the N_t antennas of the SpD. For each group of b bits as $b_g(g = 1, 2, \cdots, G)$, b_g^1 bits are used for generating K Phase Shift Keying/Quadrature Amplitude Modulation PSK/QAM symbols, while the remaining b_g^2 bits are mapped to the JMIM mapping matrix selector, which chooses a specific mapping matrix out of Q JMIM matrices. Then, these K PSK/QAM codewords and the selected JMIM mapping DM are combined to generate a Space-Time (ST) block $\boldsymbol{S}$. Afterwards, the block creator of Fig. 5.2 collects all codewords from the G groups for forming a frame, which is mapped to multiple index domains by the carrier index mapper, followed by the CS method and OFDM modulation, as shown in Fig. 5.2. Then, after transmission over the wireless channel, the receiver estimates the channel and detects the signal. At the receiver side, the signal is transformed back to the subcarrier symbols and each JMIM group signal is detected separately.

In the following, we present the details of the processing stages at the transmitter and the receiver. In this case, we only focus our attention on a single group instead of G groups, since the same procedure is applied to all groups, as shown in Fig.5.2. The transmitter model is introduced in Section 5.2.1, followed by the receiver model in Section II-B.

5.2.1 Transmitter

As shown in Fig. 5.2, b bits are split into G groups, where the b_g bits, $(g = 1, 2, 3...G)$ of each group are split into two parts by the block splitter: b_g^1 bits are used for JMIM mapping matrix selection and b_g^2 bits for the classic PSK/QAM. In the following we explain in detail the Joint Index Mapping (JIM) part of the CS-JMIM transmitter of Fig. 5.2.

5.2.1.1 JIM

As shown in Fig.5.2, the N_c subcarriers of the OFDM symbol are divided into G groups of size N_f, with $N_f = N_c/G$. For each b_g group of bits, the first part b_g^1 is used for selecting the active DM from the Q candidates $\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_q, \cdots, \boldsymbol{D}_Q$ with $\boldsymbol{D}_q \in \mathbb{C}^{N_v \times N_{vt}}, q = 1, 2, \cdots, Q$. The second part is used for determining the constellation

[1]FD is the OFDM symbol domain after CS processing, as shown in Fig. 5.2.

[2]VD is the actual domain. This concept was firstly introduced in [17] to illustrate the CS techniques in IM systems to improve the spectral efficiency.

symbol, which is employed for modulating the active DM. The classic constellation symbol is then selected from a M-ary PSK or QAM constellation χ.

Let us denote the selected DM and the selected constellation symbol, respectively, by $\boldsymbol{D}_i, i \in \{1, \cdots, Q\}$ and $x, x \in \chi$. Then the combined signal in group g can be expressed by

$$\boldsymbol{S}_g = x\boldsymbol{D}_i, g = 1, \cdots, G. \tag{5.1}$$

In the following, we introduce three designs of the DMs. Firstly, to leverage the multi-dimensionality of MIM systems, the design of IM encompasses all dimensions. Then, the activation of the corresponding indices is guided by the coordinates of these joint dimensions, which is detailed in the following section in the context of a DM design referred to as General JIM. Secondly, to strike a design trade-off between the throughput and diversity gain attained, we can further split the joint multi-dimensional matrix into sub-group matrices, where different general JIM DMs can be selected for each sub-group matrix. We refer to this mapping design as Grouped JIM, which is further detailed in the following sections. Additionally, we introduce a coded DM design for achieving a high diversity gain, which is detailed in the following sections. Furthermore, we start a discussion considering the Space-Frequency (SF) dimensions and then we present a 3-dimensional mapping design for the Time-Space-Frequency (TSF) dimensions of JMIM.

General JIM As JIM, first we consider joint SF DM design. The index is selected based on both dimensions' coordinates. We assign N_{vt} TAs and N_v subcarriers to a specific group, which results in $N_{vt}N_v$ possible active positions and to a total of $C(N_tN_{vt}, K)$ legitimate realizations. As an example, let us consider having $K = 2$ active subcarriers and $N_{vt} = 2, N_v = 2$ for each group. Then, we have $b_g^1 = [\log_2 C(N_{vt}N_v, K)] = [\log_2 C(4, 2)] = 2$ bits for selecting $K = 2$ active subcarriers out of 4 available subcarriers in each group, since we have $2^2 = 4$ legitimate combinations which equivalent to $Q = 4$ DMs, as shown in Table 5.2. Fig.5.3 shows a block diagram of the general JIM example presented in Table 5.2, where the activated index is then combined with the QAM symbol by the multiplier to form the combined symbol S. *Furthermore, when compared to the CS-aided separate MIM system, CS-JMIM can attain comparable throughput as CS-MIM with significant sparsity.*

Grouped JIM Given a substantial number of TAs, subcarriers, and a limited quantity of active index elements K in each group, most elements in the DM remain inactive, leading to diminished SE. To address this, we propose grouped JIM, which divides the DM matrix into smaller sub-group matrices, each adopting a general JIM. Furthermore, striking a trade-off between throughput and diversity involves choosing either the same or different DMs across groups. To elaborate further, applying the same DM

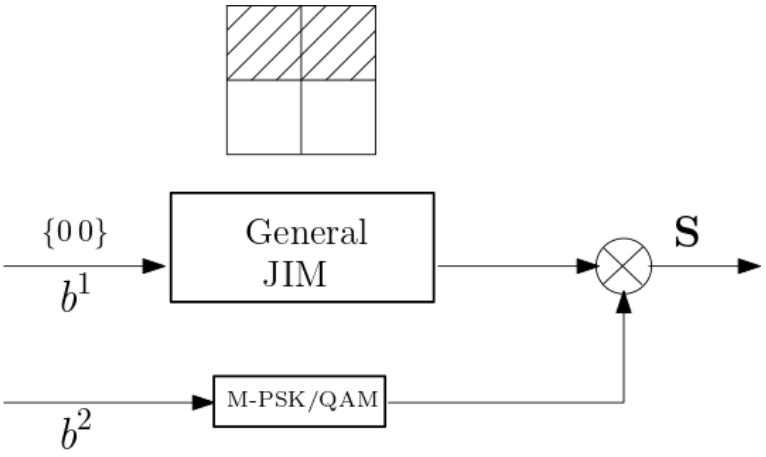

FIGURE 5.3: Block diagram of the general JIM example in Table 5.2 with b_1=[0 0].

TABLE 5.2: An example selection procedure of joint SF index selection in a CS-JMIM system having $K = 2, N_v = N_{vt} = 2$

b_2	matrix No.	Indices	Allocation
[00]	$\boldsymbol{D}_1$	$(1,2)$	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$
[01]	$\boldsymbol{D}_2$	$(1,3)$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$
[10]	$\boldsymbol{D}_3$	$(1,4)$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
[11]	$\boldsymbol{D}_4$	$(2,3)$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

across all groups results in multiple copies of the information bits, which produces a diversity gain. On the other hand, employing different DMs for each group improve the throughput.

For example upon assuming $N_{vt} = 4, N_v = 4$ and $K = 2$ for each groups DM results in $\boldsymbol{D}_q \in \mathbb{C}^{N_{vt} \times N_v}$. Then, we further split $\boldsymbol{D}_q$ into four equal sub-matrices expressed as

$$\boldsymbol{D}_q = \begin{bmatrix} \boldsymbol{D}_q^{1,1} & \boldsymbol{D}_q^{2,1} \\ \boldsymbol{D}_q^{1,2} & \boldsymbol{D}_q^{2,2} \end{bmatrix}, \tag{5.2}$$

where we have the sub-matrix $\boldsymbol{D}_q^i \in \mathbb{C}^{N_{vt}/2 \times N_v/2}, (i = 1,2,3,4)$. For each sub-matrix $\boldsymbol{D}_q^{i,j}$, $(i = 1,2,3...gsx)$, $(j = 1,2,3...gsy)$ general JIM can be applied. Here, gsx and gsy represent the number of sub-group's in the FD and SpD, respectively. In the above example, we can have a total of $gs = gsx \times gsy = 4$ sub-groups and $b_g^1 = [\log_2 C(4,2)] = 2$ bits for each sub-groups matrix. To maximize the throughput, four different sub-matrices can be aggregated to one DM $\boldsymbol{D}_q$ to obtain 8 bits in total. Fig.5.4 shows the block diagram of the grouped JIM, where we have four sub-groups of smaller general JMIM matrix. For a small general JMIM matrix we can apply $Q = 4$ DMs in total, where we can assign 4×2 bits for all sub-groups. On the other hand, if four repeated

sub-matrices are used, we can achieve similar structure of coded JMIM which will be discussed below.

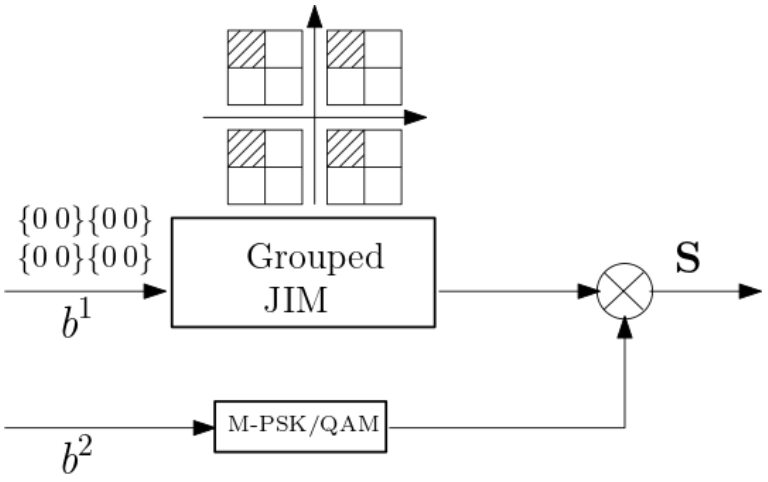

FIGURE 5.4: Block diagram of a grouped JIM example with b_1=[0 0 0 0 0 0 0 0]

Subsequently, the grouped JIM optimally utilizes the available space of the VD matrix, albeit at the expense of sparsity. By adjusting the index mapping of each sub-group, it offers significant throughput or diversity gains. However, this leads to a substantial increase in detection complexity for conventional methods, such as the ML detector.

Coded JIM Another way of further increasing the transmit diversity is to employ coded index mapping, where we use a circular shift based design of the DMs, which was proposed for SFSK in [58]. In this method, the number of active subcarriers in each column is n_q, with $N_q - n_q$ inactive subcarriers, where N_q is the column length of $\boldsymbol{D}_q$. Then, the second column is the circular down shift of the first column by one position. Similarly, other columns can be obtained based on the previous column distribution.

To elaborate a little further, using a 'toy' example, for $N_q = N_{vt} = 4, n_q = 2$, we can have $Q = C(N_q, n_q) = 6$ possible combinations, yielding $b_g^1 = \lfloor \log_2 C(N_q, n_q) \rfloor = 2$

bits. The following is an example of a circular shifting based DM:

$$D_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, D_2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$
$$D_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, D_4 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$
$$D_5 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, D_6 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Given $b_g^1 = 2$ bits, then $2^2 = 4$ DMs are selected for the CS-JMIM system.

Fig.5.5 shows a block diagram of the coded JIM, where we can apply the first JIM DM for b1=[0 0] based on the code book used. *In this scenario, coded JIM offers the maximum diversity in the design of coded DMs, enabling reliable detection even in highly noisy environments.*

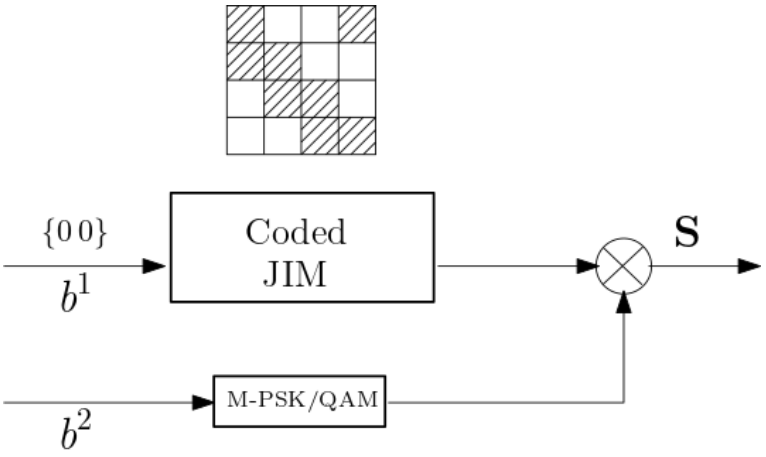

FIGURE 5.5: Block diagram of a coded JIM example with b_1=[0 0]

3-Dimensional JIM-DM Design In this design, the TD is introduced as an extra dimension for the JIM. We assume that T_v TSs are applied in the VD and T TSs are used in the TD, while we have $T_v > T$. Then, we can assign three-dimensional DMs $\boldsymbol{D}_q \in \mathbb{C}^{N_v \times N_{vt} \times T_v}$. In this case, the above-mentioned three mapping techniques can be applied.

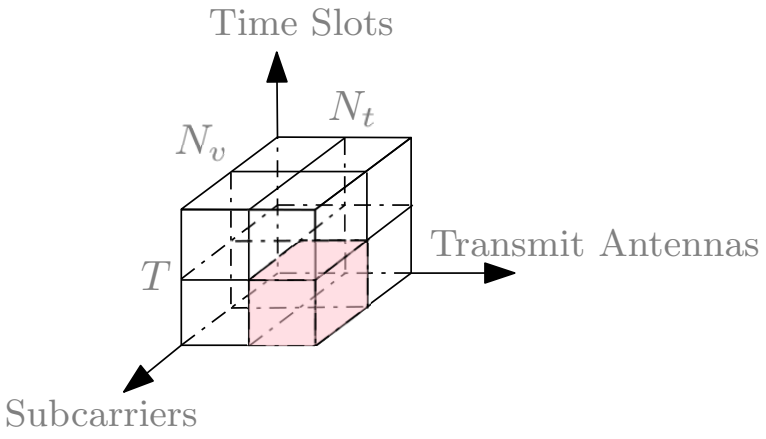

FIGURE 5.6: Illustration of the Structure for general JIM DM in time-space-frequency domain, while having $K = 1$,$N_{vt} = 2$ TAs, $T_v = 2$ TSs, $N_v = 2$ subcarriers.

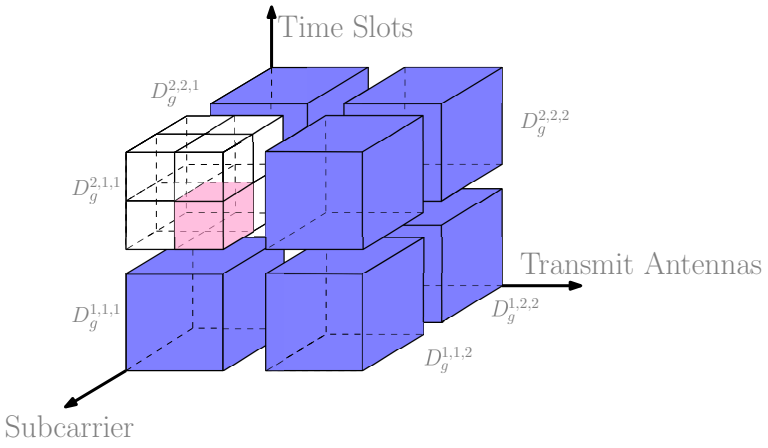

FIGURE 5.7: Illustration of the Structure for group JIM DM in time-space-frequency domain, while having $K = 1$,$N_{vt} = 4$ TAs, $T_v = 4$ TSs, $N_v = 4$ subcarriers with 4 sub-group.

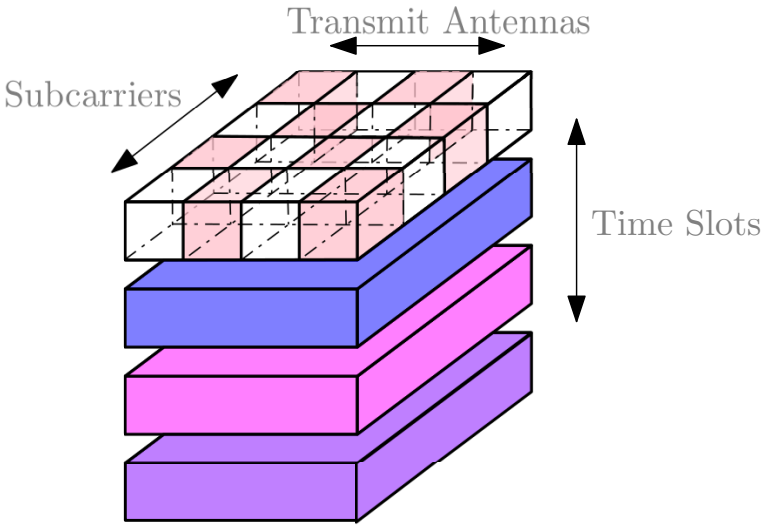

FIGURE 5.8: Illustration of the Structure for coded JIM DM in time-space-frequency domain, while having $Q = 6$,$N_{vt} = 4$ TAs, $T_v = 4$ TSs, $N_v = 4$ subcarriers.

Specifically, for the general JIM we may consider the following example for further illustration. Let $K = 1$ and $N_{vt} = N_v = T_v = 2$ as shown in Fig. 5.6 and $b_g^1 = [001]$. More specifically, the three-dimensional matrix can be expressed in the coordinate form of (n_v, n_{vt}, t_v). In this case, given the IM bits $b_g^1 = [010]$, we activate the fourth element in a set of 8 elements in this three-dimensional matrix with the coordinates $(2, 2, 1)$ as shown in Fig. 5.6. Then, the number of bits of this JMIM applied for the DM selection becomes $b_g^1 = \lfloor \log_2 C(N_{vt} N_v T_v, K) \rfloor = \lfloor \log_2 C(8, 1) \rfloor = 3$ bits.

Fig. 5.7 shows the structure of the grouped JIM applied in three dimensions. Similar to the SF matrix, the TSF matrix can be split into several equal sub-groups. As shown in Fig. 5.7, we assume $N_{vt} = N_v = T_v = 4$ and $K = 1$ for each group's DM, which results in $\boldsymbol{D}_q \in \mathbb{C}^{N_v \times N_{vt} \times T_v}$. Then, we further split $\boldsymbol{D}_q$ into 8 equal sub-matrices. Each sub-group DM can be expressed in the form of $\boldsymbol{D}_g^{gsx,gsy,gsz}$, where gsx, gsy, gsz represents the split size in the FD, the SpD and the TD, respectively. For each sub-matrix $D_g^{gsx,gsy,gsz}$, general JIM can be applied within a set of $gs = gsx \times gsy \times gsz = 8$ sub-group matrices. Then, we can have $\lfloor \log_2 C(8, 1) \rfloor = 3$ bits for each sub-matrix. To maximize the throughput, we can also assign different information to each sub-group and then the 8 sub-matrices can be aggregated to form a single DM $\boldsymbol{D}_q$ to obtain $b_g^1 = gs \lfloor \log_2 C((N_v/gsx)(N_t/gsy)(T/gsz), K) \rfloor = 8 \lfloor \log_2 C(8, 1) \rfloor = 24$ bits for the JMIM design. Compared to the same DM size used in the general JIM, which has $b_g^1 = \lfloor \log_2 C(64, 1) \rfloor = 6$ bits, the grouped JIM can provide a significant gain in the spectral efficiency. On the other hand, in order to attain a diversity gain, the sub-matrices can achieve maximum diversity gain, when all 8 sub-groups have the same active index.

Furthermore, for the coded JIM matrix design in three dimensions, the same method is applied for the first TS of the space-frequency matrix. Then, circular shifting is applied to the entire SF matrix to generate the next TS matrix with shifting by one position. As shown in Fig. 5.8, upon assuming $N_{vt} = N_v = T_v = 4$ for the DM size, as well as $N_q = n_q = 2$ for the activated subcarriers and $b^1 = [01]$, then the corresponding circular shifting based DM D_2 presented in the previous section is applied to the first TS of the 3D matrix. Then, we can generate each TS index mapping with the aid of a single position shifting, which can be represented as:

$$D_{t1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, D_{t2} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix},$$
$$D_{t3} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, D_{t4} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

5.2.1.2 CS and Block Assembly

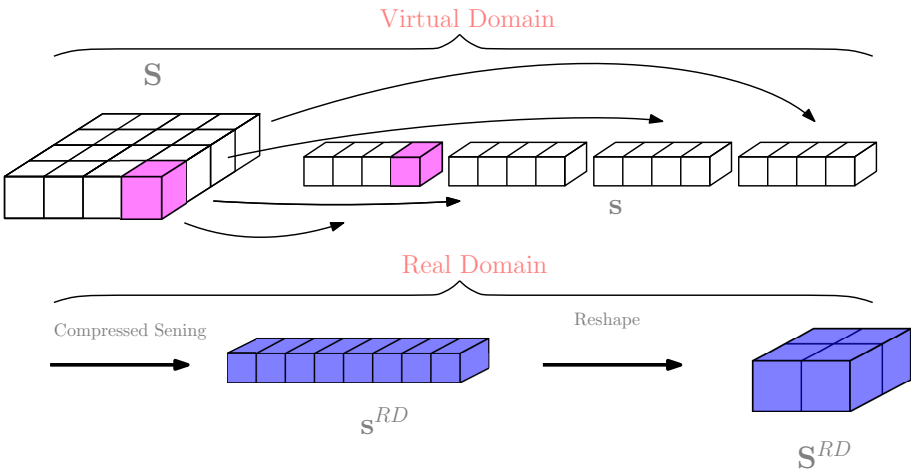

FIGURE 5.9: Illustration of the process for compressing the JMIM DM in the SF domain with $N_{vt} = N_v = 4, K = 1$. Note that this example applies the general JIM with $b_g^1 = [0100]$.

In order to exploit the sparsity of the JIM DM, CS is applied to all the dimensions of the joint multi-dimensional matrix symbol created by the block assembled to increase the throughput. As shown in Fig. 5.9, a matrix $\boldsymbol{S}_g$ associated with $N_{vt} = N_v = 4$ will be transformed from the matrix $\boldsymbol{S}, \boldsymbol{S} \in \mathbb{C}^{N_{vt} \times N_v}$ into the vector $\boldsymbol{s}, \boldsymbol{s} \in \mathbb{C}^{N_{vt}N_v \times 1}$.

The symbol vector $\boldsymbol{s}$ is then compressed by a CS measurement matrix $\boldsymbol{A} \in \mathbb{C}^{N_f N_t \times N_v N_{vt}}$ from the $N_v N_{vt}$-dimensional $\boldsymbol{s}$ in the VD into the $N_f N_t$-dimensional form in the Real Domain (RD)[3] denoted as $\boldsymbol{s}^{(RD)}$, which can be written as: $\boldsymbol{s}^{RD} = \boldsymbol{A}\boldsymbol{s}$. The RD vector $\boldsymbol{s}^{RD}$ after CS is then transferred into a compressed joint multidimensional symbol matrix $\boldsymbol{S}^{(RD)}$, where $\boldsymbol{S}^{(RD)} \in \mathbb{C}^{N_t \times N_f}$. Then, the index carrier mapper maps the corresponding joint multidimensional symbol elements to the OFDM subcarriers and the TAs to form the SF symbols. Afterwards, G groups of SF symbols $\boldsymbol{S}$ are assembled by the OFDM creator to a long SF symbol frame, as shown in Fig. 5.2. The RD SF symbol can be

[3]RD is the joint dimension of DM after the CS process. For instance, the SF-based JMIM signal conveys more bits in the VD than in the RD.

separated into N_t FD symbols, which means that N_t FD symbols are transmitted by N_t TAs. Similar to conventional OFDM, the FD symbol will be transformed into TD symbols to be transmitted by their corresponding TAs and then a Cyclic Prefix (CP) will be added. The G groups of SF symbols $\boldsymbol{S}$ are assembled by the block creator of Fig. 5.2 to form a long ST frame, which is processed by the ST mapper to output a symbol for transmission over multiple TAs and TSs, Equivalently, the ST symbols $\boldsymbol{S}$ of each subcarrier group are mapped to N_t TAs during T TSs, which have N_t symbol sequences $\{\boldsymbol{s}_1, ..., \boldsymbol{s}_{N_t}\}$ for transmission from the N_t TAs during each TS.

For the three-dimensional JMIM, utilizing the TSF dimensions, the TD is also compressed by CS for improving the throughput, where T_v TSs are introduced in the VD for IM, complemented by T TSs in the TD. Specifically, for the general JMIM scheme, the TD is introduced for increasing the sparsity and for incorporating extra embedded information bits. As shown in Fig. 5.10, we apply CS to the TSF JMIM, where all the three dimensions are compressed for increasing the throughput. Specifically, a $(4 \times 4 \times 4)$-sized DM in the VD will be compressed to a $(2 \times 2 \times 2)$-sized DM of the RD. For example, when we have $T_v = N_{vt} = N_v = 4$, $b_g^1 = [000100]$ and $K = 1$, the element at the fourth subcarrier, fourth TA and first TS is activated, corresponding to the coordinate of $(4, 4, 1)$.

As for the coded JMIM scheme, additionally the TD is harnessed for further increasing the diversity gain, where CS is not considered for the TD. We assign either the same or different symbols in a sub-group matrix of the grouped JMIM scheme, which leads to a different CS approach. Given the different sub-group matrix symbols, the TD is exclusively harnessed for carrying extra copies of the symbol without CS. The design objective of this scheme is to increase the diversity gain.

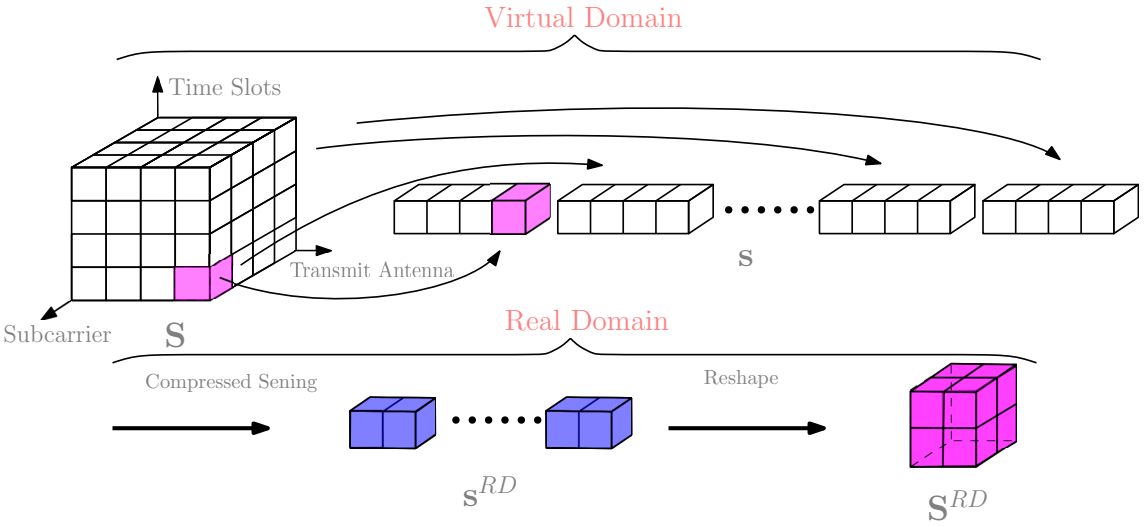

FIGURE 5.10: Illustration of the process for compressing the JMIM DM in the TSF domain with $T = N_t = N_v = 4K = 1$. Additionally, the example presented applies the general JIM for $b_g^1 = [000100]$.

5.2.2 Receiver Processing

As shown in Fig. 5.11, a receiver having N_r antennas is employed, where we assume that the transmitted signals are conveyed over a frequency-selective Rayleigh fading channel and the CSI is perfectly acquired at the receiver side. The G groups of signal are received by the receiver over N_r antennas and then the CP part of the received signals is removed. Finally, the processed signal is transformed into the FD by using the Fast Fourier Transform (FFT), as shown in Fig. 5.11.

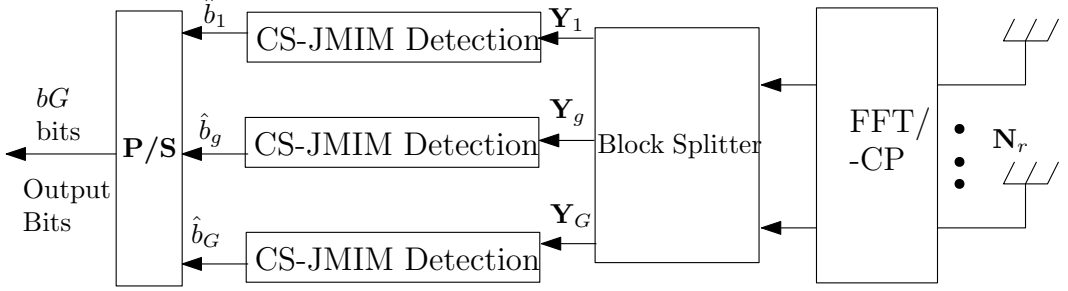

FIGURE 5.11: CS-JMIM system receiver block diagram

The channel model can be expressed as $\boldsymbol{h}_\alpha \in \mathbb{C}^{N_r \times N_t}$, which represents the TD CSI between the N_t TAs and the N_r RAs. Then, the FD channel matrix can be expressed as $\boldsymbol{H}_\alpha \in \mathbb{C}^{N_r \times N_t}$ for $\alpha = 1, \ldots, M$, which are then split into G groups by the Block Splitter of Fig. 5.11. The symbols received by each subcarrier group are represented as $\boldsymbol{Y} = \{\boldsymbol{Y}[1], \ldots, \boldsymbol{Y}[\alpha], \ldots, \boldsymbol{Y}[N_f]\}$, with $\boldsymbol{Y} \in \mathbb{C}^{N_r \times N_f}$ and $\alpha = 1, 2, \cdots, N_f$.

As for the three-dimensional signal, the transmitted signal is mapped ST symbols, which are also collected by the receiver and split into G groups by the Block Splitter of Fig. 5.11. Afterwards, the symbols received in the three dimensions by each subcarrier group $\boldsymbol{Y} \in \mathbb{C}^{N_r \times M \times T}$ may be expressed as

$$\boldsymbol{Y} = \left\{ \begin{bmatrix} \boldsymbol{Y}^1_{1,1} & \ldots & \boldsymbol{Y}^1_{1,N_f} \\ \ldots & \ldots & \ldots \\ \boldsymbol{Y}^{N_r}_{1,1} & \ldots & \boldsymbol{Y}^{N_r}_{1,N_f} \end{bmatrix}_1, \ldots, \begin{bmatrix} \boldsymbol{Y}^1_{T,1} & \ldots & \boldsymbol{Y}^1_{T,N_f} \\ \ldots & \ldots & \ldots \\ \boldsymbol{Y}^{N_r}_{T,1} & \ldots & \boldsymbol{Y}^{N_r}_{T,N_f} \end{bmatrix}_T \right\}. \tag{5.3}$$

The received symbol of the t-th TS can be represented as $\boldsymbol{Y_t} = \{\boldsymbol{Y_t}[1], \ldots, \boldsymbol{Y_t}[\alpha], \ldots, \boldsymbol{Y_t}[N_f]\}$, with $\boldsymbol{Y_t} \in \mathbb{C}^{N_r \times T}$ and $\alpha = 1, 2, \ldots, N_f, t = 1, 2, \ldots, T$ characterizing the ST structure per group and the ST symbol received at the α-th subcarrier of each subcarrier group, respectively. Since the index is jointly decided in the multi-dimensional space, we can transform the ST symbol into a vectorial form $\boldsymbol{y}$ associated with $\boldsymbol{y} \in \mathbb{C}^{N_r N_f T \times 1}$.

Let the FD channel be $\mathbf{H}_\alpha \in \mathbb{C}^{N_r \times T}$ for $\alpha = 1, \ldots, N_f$. Then the signal $\boldsymbol{Y}_t[\alpha] \in \mathbb{C}^{N_r \times T} (\alpha = 1, \ldots, N_f)$ received during the T TSs for each subcarrier group can be expressed as [19]

$$\boldsymbol{Y}[\alpha] = \boldsymbol{H}_\alpha \boldsymbol{S}^{(RD)}[\alpha] + \boldsymbol{W}[\alpha], \tag{5.4}$$

where $\boldsymbol{S}^{RD}[\alpha] \in \mathbb{C}^{N_r \times T}$ denotes the ST symbols at the subcarrier α transmitted from the N_t TAs in the RD. Furthermore, $\boldsymbol{W}[\alpha] \in \mathbb{C}^{N_r \times T}$ represents the Additive White Gaussian noise (AWGN) obeying the distribution of $\mathcal{CN}(0, \sigma_N^2)$, and σ_N^2 is the noise variance.

5.3 CS-JMIM Signal Detection

Given the received signal model $\boldsymbol{Y}$ in (5.4), the receiver detects the information bits of the JMIM mapping matrix, which jointly conveys the index of the active subcarrier, the active TA and TS in the VD. Firstly, we reshape the received signal into a vectorial form $\boldsymbol{y}$ associated with $\boldsymbol{y} \in \mathbb{C}^{N_r N_f T \times 1}$.

The received signal $\boldsymbol{y}$ contains N_f ST symbols at N_f subcarriers in the FD of each subcarrier group. Then, we can rewrite $\boldsymbol{y}$ with the aid of (5.4) in the following form:

$$\boldsymbol{y} = \boldsymbol{H}\bar{\boldsymbol{A}}\bar{\boldsymbol{s}} + \boldsymbol{w}, \tag{5.5}$$

where $\bar{\boldsymbol{A}}$ is the equivalent measurement matrix $\boldsymbol{A}$ used for compressing the $\boldsymbol{s}$ VD vectors. In our three-dimensional CS-JMIM system, $\bar{\boldsymbol{A}}$ also compresses the TD, where we have $\bar{\boldsymbol{A}} \in \mathbb{C}^{N_v t N_v \times N_t N_f}$. Furthermore, $\boldsymbol{s} \in \mathbb{C}^{N_v N_v t T_v \times 1}$ denotes the vector of DM combined with the PSK/QAM symbol. In this case, we could rewrite $\bar{\boldsymbol{s}}$ in a matrix $\bar{\boldsymbol{S}}$ associated with $\bar{\boldsymbol{S}} = x\bar{\boldsymbol{D}}$, where $\bar{\boldsymbol{D}} \in \mathbb{C}^{N_v \times N_v t \times T_v}$ denotes the realization of the JMIM DM in each subcarrier group.

Conventional exhaustive search based maximum likelihood (ML) detection can be applied at the receiver, albeit this may lead to excessive complexity [3]. Furthermore, in the soft detection scenario, the received signal is converted into probability values, which are referred to as Log Likelihood Ratios (LLR) that are fed into the channel decoder for obtaining a near-capacity performance [140].

In the following section we present the conventional ML-based HD detector, followed by our proposed DNN aided HD detector, where the neural network replaces the exhaustive search by a learning-based classification model in order to significantly reduce the complexity. Afterwards, we discuss the SD detector, where we first present the conventional SD detectors followed by our learning-aided SD receiver.

5.3.1 HD Detection

Again, we commence with the conventional ML-based detection of the CS-JMIM system, followed by the DNN-based detector.

5.3.1.1 ML Detection

As shown in Fig. 5.11, we detect each group;s signal separately. In the CS-JMIM detector, according to the receiver model of (5.5), we have the modified joint JMIM and PSK/QAM symbol, which can be expressed as $\bar{\boldsymbol{S}} = \mathbf{x}\bar{\boldsymbol{D}}$. Here $\bar{\boldsymbol{D}}$ represents a specific realization of the selected JMIM DM and $\mathbf{x}$ represents K STSK PSK/QAM symbols. To detect the specific realization, we use $\bar{\mathcal{D}}(\beta)$ $(\beta = 1, 2, ..., N_{JMIM})$ to denote all the possible realizations of the JMIM DM. Furthermore, as there are $N_x = (X)^K$ realizations of $\mathbf{x}$, $\bar{\mathcal{X}}(\gamma)$ $(\gamma = 1, 2, ..., N_x)$ denotes all the possible realizations of the selected PSK/QAM symbol.

The ML detector makes a joint decision concerning the JMIM DM and PSK/QAM with the aid of exhaustive search, which can be modelled as

$$\langle \hat{\gamma}, \hat{\beta} \rangle = \arg\min_{\gamma,\beta} \|\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{A}}\mathcal{X}_(\gamma)\bar{\mathcal{I}}_D(\beta)\|^2, \tag{5.6}$$

where $\hat{\gamma}$ and $\hat{\beta}$ represent the estimates of the selected DM and the corresponding PSK/QAM constellation in each subcarrier group, respectively.

The excessively high search complexity of considering all possible candidates by the ML detector is given by $\mathcal{O}[N_{JMIM}(\mathcal{X})^K]$ per subcarrier group.

5.3.1.2 DNN-based Detection

To reduce the complexity of the ML detector, learning based detection is considered in this section, where a DNN based model is proposed for detecting the received CS-JMIM signal.

Detection may also be considered as a classification problem, where the corresponding bits of the harnessed CS-JMIM DM and PSK/QAM symbol constitute the DNN output. Under the assumption of perfect CSI at the receiver side, we use the received signal and the CSI as the input of the DNN model. The proposed DNN structure is shown in Fig. 5.12, where both the CSI $\boldsymbol{H}$ at the receiver and the received symbols $\boldsymbol{Y}$ constitute the inputs of the L-layer Fully-Connected (FC) network. Then, the output bits $\hat{\boldsymbol{u}}$ can be modelled as

$$\begin{aligned} \hat{u} =& f_{sigmoid}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{Y}) + \boldsymbol{\theta_1}]) \\ & + \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n), \end{aligned} \tag{5.7}$$

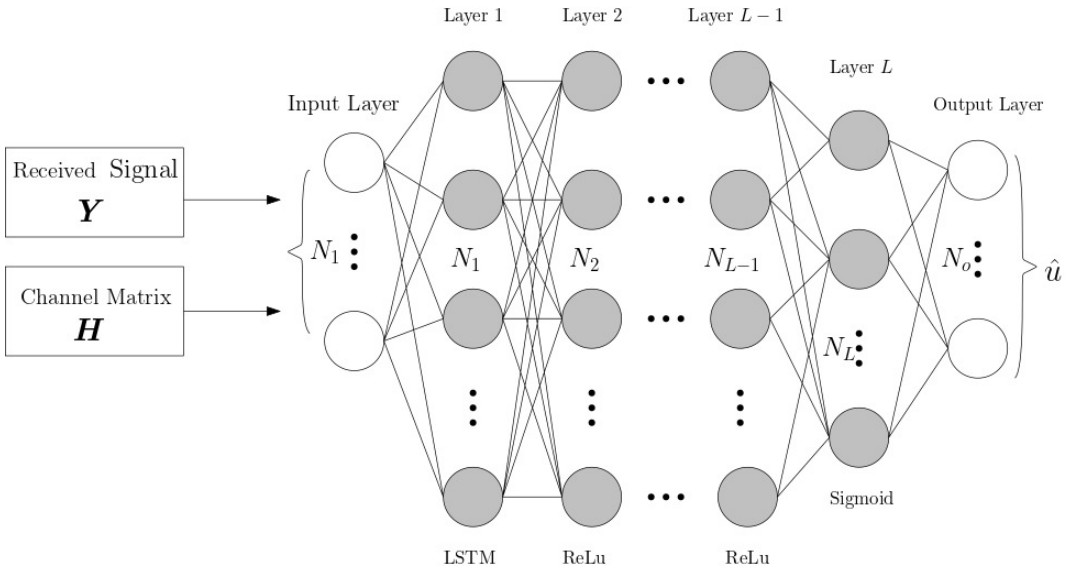

FIGURE 5.12: Fully-connected DNN model for CS-JMIM HD detection.

where $\boldsymbol{W_n}$ and $\boldsymbol{\theta_n}$, $n = 1, \cdots, L$ represent the weights and biases, respectively. In (5.7), the Rectified linear unit (Relu) function of $f_{Relu}(s) = max(0, s)$ is employed for activating the DNN during the training phase, while the sigmoid function of $f_{sigmoid}(s) = \frac{1}{1+e^{-s}}$ is used to obtain the detected bits $\hat{u}$. The raw input data represented in the complex-valued matrix form obtained from the received signal $\boldsymbol{Y}$ is vectorized first and then we rearrange the complex values by separately extracting the real as well as the imaginary parts and then merging them into a real-valued vector.

In the training phase, we employ randomly generated received signals, which are transmitted over a frequency selective Rayleigh fading channel for CS-JMIM. Afterwards, both the CSI and the received symbols are employed as the input data of the DNN. The number of training samples required is selected based on experimentation by gradually increasing the training size until acceptable mean square error (MSE) values are achieved. In this case, the MSE loss function of the DNN used for the training is

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B} \sum_{i=1}^{B} \|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{5.8}$$

where B is the sample size of the current iteration. A stopping criterion can be defined either by the number of iterations or by an MSE threshold. Then, the parameter sets $\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}$ can be updated in each training iteration based on our learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} \leftarrow \{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} - \alpha \nabla \boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}),$$

where $\alpha > 0$ is the learning rate and $\nabla \boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$. In our proposed network aided detection, we use $\alpha = 0.001$.

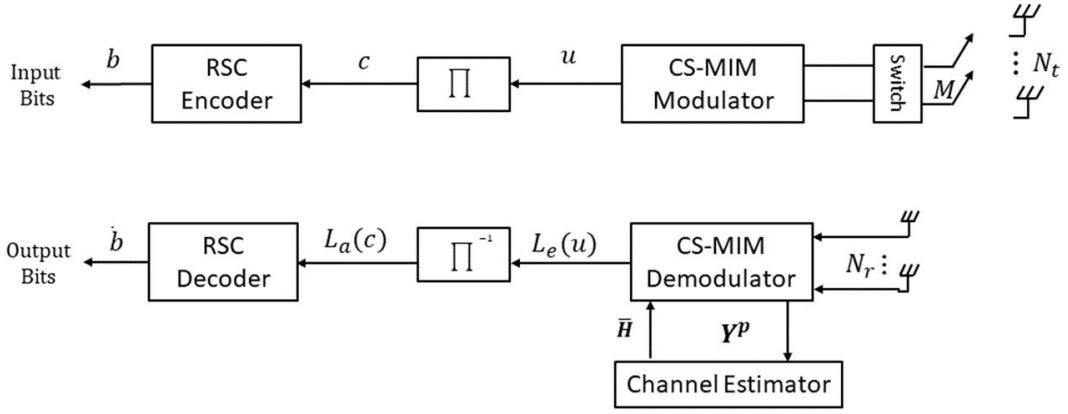

FIGURE 5.13: The transceiver architecture of channel-coded CS-MIM.

By the end of the training phase, the DNN has learnt the mapping from the received signal and stores both the weight as well as the bias information, which will be used for producing the desired outputs based on the input data in the testing phase. The statistical properties of the input/output data have to remain the same as those used during training.

The detection complexity of the learning algorithm is dominated by the calculation of the layer weights and biases, which may be considered to be of the order of $\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$ [125], with n representing the number of neurons in each layer. Hence, the DNN complexity order is significantly lower than that of the ML detector.

5.3.2 SD Detection

SD detection is employed for attaining near-capacity performance, when combined with channel coding. As the computational complexity of the maximum *a posteriori* probability in SD detector rapidly increases upon increasing the modulation order and the number of dimensions [98], the complexity of CS-JMIM rapidly becomes prohibitive, owing to the joint detection of JMIM signal in multiple dimensions. In the following, we present the conventional SD detector of CS-JMIM, followed by the correspond learning aided SD detector.

5.3.2.1 Conventional SD Detection

A channel coded CS-JMIM scheme is shown in Fig. 5.13, which was derived from the CS-MIM model of [19, 126] for achieving near-capacity performance. A Recursive Systematic Convolutional (RSC) encoder encodes the information bit sequence **b** followed by an interleaver, where the coded bit sequence **c** is interleaved to generate the stream **u** of Fig. 5.13. Then, the stream **u** is modulated in the CS-JMIM modulator of Fig. 5.2.

At the receiver side of Fig. 5.13, the received signal $\mathbf{Y}$ and CSI $\bar{\boldsymbol{H}}$ are input to the soft CS-JMIM that outputs LLRs. The LLRs output from the demodulator are then passed to the de-inteleaver and the RSC decoder performs soft decoding. In Fig. 5.13, $L(\cdot)$ represents the LLRs of the bit sequences, where $L_e(u)$ is the output extrinsic LLR after soft demodulation and $L_a(c)$ is the de-interleaved LLR sequence of $L_e(u)$.

The LLR of a bit is defined as the ratio of probabilities associated with the logical bits '1' and '0', which can be written as $L(b) = log\frac{p(b=1)}{p(b=0)}$. The conditional probability $p(\boldsymbol{Y}|\mathcal{X}_{\beta,\gamma})$ of receiving the group signal $\boldsymbol{Y}$ is given by [111]

$$p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta}) = \frac{1}{(\pi \boldsymbol{N}_0)^{NT}} \exp(-\frac{||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{A}}x(\gamma)\bar{\boldsymbol{D}}(\beta)||^2}{N_0}), \tag{5.9}$$

where $\mathcal{X}_{\gamma,\beta}$ represents the PSK/QAM symbol at the β-th CS-JMIM DM. Furthermore, N_0 is the noise power, where we have $\sigma_n^2 = N_0/2$ with $N_0/2$ representing the double-sided noise power spectral density.

Hence, we can formulate the LLR of bit u_i as

$$L_e(u_i) = \ln \frac{p(\boldsymbol{y}|u_i = 1)}{p(\boldsymbol{y}|u_i = 0)} = \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta} \in \mathcal{X}_1^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta})}{\sum_{\mathcal{X}_{\gamma,\beta} \in \mathcal{X}_0^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta})}, \tag{5.10}$$

where $\mathcal{X}_1^l$ and $\mathcal{X}_0^l$ represent a subset of the legitimate equivalent signal $\mathcal{X}$ corresponding to bit u_l, when $u_l = 1$ and $u_l = 0$, respectively, yielding $X_1^l \equiv \{\mathcal{X}_{\gamma,\beta} \in \mathcal{X} : u_i = 1\}$ and $X_0^l \equiv \{\mathcal{X}_{\gamma,\beta} \in \mathcal{X} : u_i = 0\}$.

Upon using (5.9) and (5.10) we obtain the LLR $L(b_i)$ of the bit sequence conveyed by the received signal $\mathbf{Y}$. To simplify the calculation, the Approximate Log-MAP (Approx-Log-MAP) algorithm based on the Jacobian Maximum operation can be used, which is given by [146, 147]

$$\boldsymbol{L}_e(u_l) = \text{jac}_{\mathcal{X}_{\gamma,\beta} \in \mathcal{X}_1^l}(\lambda_{\gamma,\beta}) - \text{jac}_{\mathcal{X}_{\gamma,\beta} \in \mathcal{X}_0^l}(\lambda_{\gamma,\beta}), \tag{5.11}$$

where jac(.) denotes the Jacobian maximum operation and the intrinsic metric of $\lambda_{\gamma,\beta}$ is

$$\lambda_{\gamma,\beta} = -||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{A}}x(\gamma)\bar{\boldsymbol{D}}(\beta)||^2/N_0. \tag{5.12}$$

At the receiver, the soft demodulator evaluates the probability of each bit being logical '1' and '0'. Then it applies the approx-log-MAP algorithm for obtaining the extrinsic LLR of the coded bits, which has a complexity order $\mathcal{O}[2^{(c_g)}(N_{JMIM}(\mathcal{X})^K)]$, where c_g represents the number of coded bits after the RSC encoder and interleaver, and N_{JMIM} represents the number of possible realizations of JMIM.

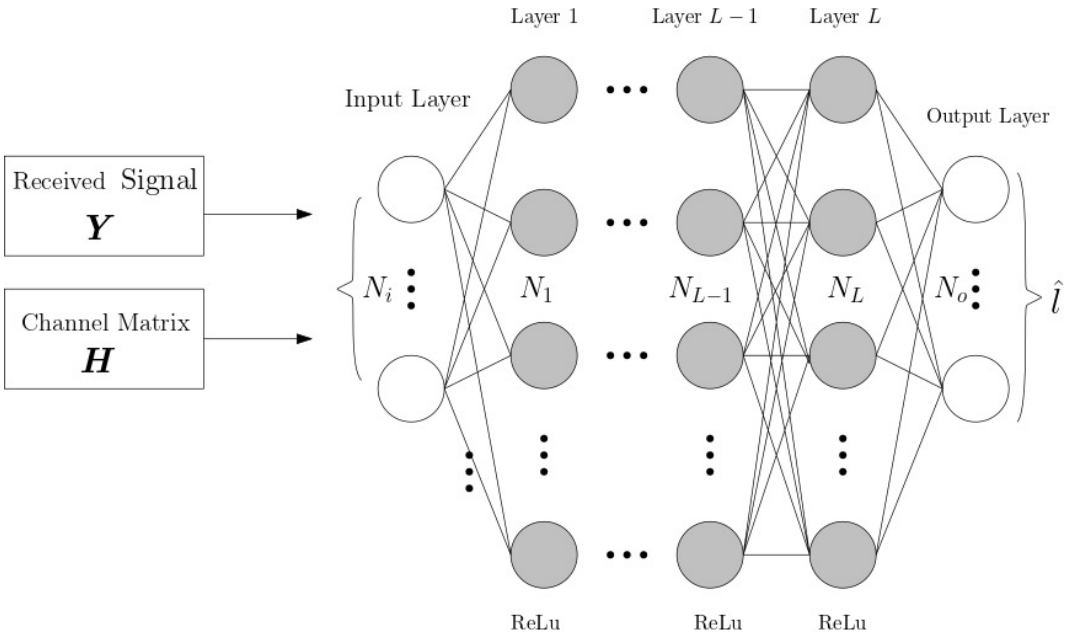

FIGURE 5.14: Fully-connected DNN model for CS-JMIM SD detection.

5.3.2.2 DNN-based SD Detection

In this section, we propose a reduced-complexity SD detector using DNN, which considers a similar DNN architecture to that of [125]. Since the conventional SD detector obtains the LLRs of the received signal after the CS-MIM soft demodulator, we replace the detected bits $\hat{\boldsymbol{u}}$ output by the DNN in Fig. 5.12 with the extrinsic LLR L_e at the output of the DNN, as shown in Fig. 5.14. Then, the output of the SD DNN model can be expressed as

$$\begin{aligned}\hat{\boldsymbol{L}}_e =& \boldsymbol{W}_{N_2} ... f_{Relu}\{\boldsymbol{W}_2(f_{Relu_1}[\boldsymbol{W}_1(\boldsymbol{Y}_\tau)+\boldsymbol{b}_1]) \\ &+\boldsymbol{b}_2+...+\boldsymbol{b}_{N_2}),\end{aligned} \tag{5.13}$$

and the corresponding loss function is

$$\mathcal{L}(\boldsymbol{\theta},) = \frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{L}}_e(\tau)-\boldsymbol{L}_e(\tau)\|_2^2. \tag{5.14}$$

We can also define a stopping criterion, which can be either the number of iterations or meeting a maximum MSE threshold. Then, the parameter sets $\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}$ can be updated in each training iteration based on the learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} \leftarrow \{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} - \alpha\nabla\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}),$$

where $\alpha > 0$ is the learning rate and $\nabla\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$.

TABLE 5.3: Configuration of the modes presented in Fig. 5.15

Mode	Mapping Type	Q	N_t	$N_v t$	N_f	N_v	K	R_t
1	Coded	4	4	8	4	8	2	1.333
2	General	4	4	8	4	8	3	4.666
3	Grouped	4	4	8	4	8	1	5.333

In our proposed neural network aided detection, we use $\alpha = 0.001$. Similar to the HD DNN detector described above, the model learns the parameters in the training phase and then outputs the LLR information.

The detection complexity of the learning algorithm is dominated by the calculation of the layer weights and biases, which may be considered to be of the order $\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$ [125], with n representing the number of neurons in each layer.

5.4 Adaptive design

Since the proposed CS-JMIM design provides flexibility in the design of the JMIM DM, we can design appropriate JMIM DMs for different channel conditions that can provide either an improved BER performance or an increased throughput. Furthermore, in our system, the transmitter can adapt both the JMIM DM $\boldsymbol{D}$ and the modulation order Q of PSK/QAM. Then, the system throughput may be adapted by appropriately adjusting the above parameters, while maintaining a target BER performance.

In the following two subsections, we highlight the classic threshold-based adaptive modulation, followed by its learning-aided counterpart. More specifically, both the KNN and DNN based adaptive model are applied for the proposed system.

5.4.0.1 Conventional Threshold-based Adaptive Design

In our adaptive scheme, we can adapt both the configuration of JMIM DM and of the PSK/QAM mode. We can define the different configurations as $Mode1$, $Mode2$, $Mode3$, ..., which can attain different BER performance and throughput. Based on the different modes, the parameters N_v, N_t, T and A of JMIM DM can be selected according to the SNR calculated at the receiver, where the SNR threshold values are selected for the different modes to satisfy a specific target BER [162, 163]. In the following, we present the scenario, where the different adaptive modes P refer to different configurations of the JMIM DM for characterising its design flexibility[4].

[4]Note that the modulation scheme such as PSK/QAM can also be adapted, but in this design example, we aim to show the flexibility of the proposed CS-JMIM design.

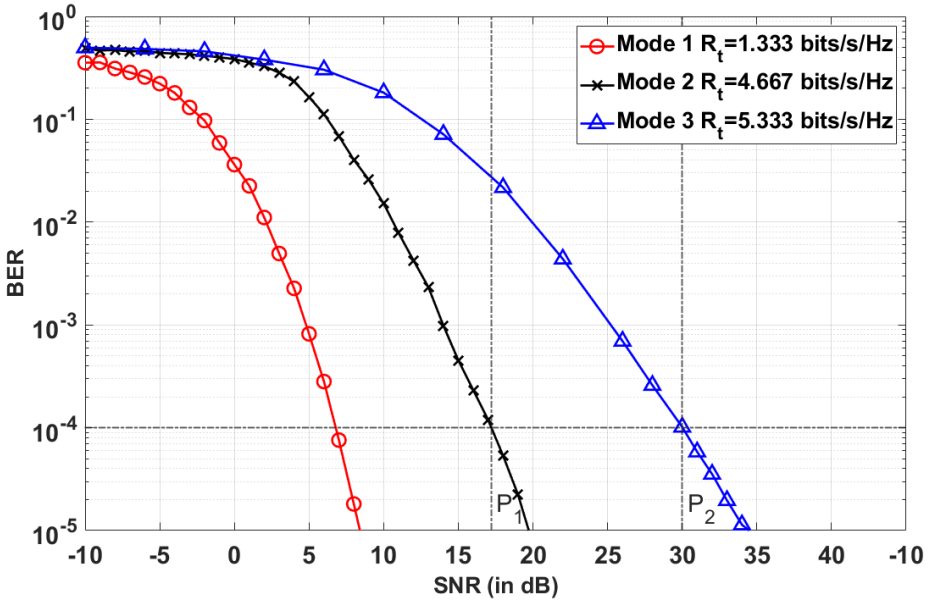

FIGURE 5.15: BER *vs.* SNR performance of the CS-JMIM system for different mapping modes shown in Table 5.3.

As an example, Fig.5.15 shows the BER performance of three different CS-JMIM mapping modes. The corresponding parameters and data rates provided by these modes are shown in Table 5.3. For a target BER of 10^{-3}, as shown in Fig. 5.15 the SNR values of mode transition points P_1 and P_2 can be selected as the thresholds for operating the appropriate modes. Specifically, *Mode*1 is applied at low SNR values until the specific SNR reaches P_1. Then, the mode is changed to *Mode*2 to provide higher throughput, when the SNR range spans from P_1 to P_2. Finally, *Mode*3 is selected at SNRs higher than P_2, which has the highest throughput among the three modes.

For adaptive modulation, the receiver has to confidently infer the choice of the most appropriate transmission mode by comparing the instantaneous SNR of the received symbol against the Mode-switching threshold values. Then, the decision is fed back to the transmitter and applied for the next frame to be transmitted. Generally, with more available operation modes as well as faster and more accurate SNR feedback to the transmitter, we can obtain an increased throughput compared to non-adaptive designs. However, threshold-based adaptive modulation design ignores many of the hardware imperfections when deciding upon the threshold values, which results in sub-optimal performance of the adaptive system [162, 163]. Hence, in the next subsection, we propose the learning-based adaptive modulation scheme for our CS-JMIM system to further improve the adaptive system's performance.

5.4.0.2 Learning aided adaptive modulation

The adaptive modulation can be modelled as a classification problem, which can be solved using learning-based methods. The SNR of the received signal, which is evaluated at the receiver side, can be fed back to the transmitter and then given the SNR information, which also corresponds to the current channel state information, the transmitter can select a specific mode from a range of candidates to achieve the highest throughput, which still maintain the target BER. Therefore, for a given channel condition, adaptive modulation selects the most suitable mode to achieve the highest throughput, under the constraint of achieving the target BER. In this chapter, both the KNN and DNN techniques are investigated in the context of adaptive modulation.

Before the training phase, the input data should be pre-processed to improve the learning efficiency. First, we randomly generate the training data of each mode under different instantaneous SNR values at the receiver. Then, the corresponding switching SNRs that can maintain a BER lower than the target BER are stored. Given these training data, we can use learning models to find the mode switching thresholds in the training phase. After training, the trained model becomes capable of predicting the next mode, given the knowledge of the SNR. In the following, we first employ KNN for our adaptive modulation scheme and then we propose a DNN-based adaptive model for further improving the performance.

KNN-based Adaptive Design KNN is a popular classification techniques relying on low-complexity implementation and yet providing a good performance [31]. Yang *et al.* [164] developed KNN-assisted adaptive modulation schemes for SM, while Liu *et al.* [162] further developed DNN aided adaptive modulation to millimeter wave communication. To elaborate briefly on the KNN process, we define the training sets as

$$\mathcal{T}^{(i)} = [\xi_1^{(i)}, \cdots, \xi_n^{(i)}, \cdots, \S_{N_p}^{(i)}]^T, \tag{5.15}$$

where ξ represents the SNR value of a symbol with a BER lower than the target BER value, with $i = 1, 2, \cdots, \mathcal{I}$ representing the adaptive mode index and N_p is the total number of instantaneous SNR values with BER under the target. Then, the total training set of each mode can be formulated as

$$\mathcal{T} = [\mathcal{T}^{(1)}, \cdots, \mathcal{T}^{(i)}, \cdots, \mathcal{T}^{(I)}]^T. \tag{5.16}$$

During runtime, for a given new data point, which corresponds to the instantaneous SNR ξ, the KNN model finds k nearest neighbours in the training set $\mathcal{T}$, using a distance metric $d(.)$, which can be expressed as

$$d(\xi_n^{(i)}), \xi_{new}) = ||\xi_n^{(i)}) - \xi_{new}||^2. \tag{5.17}$$

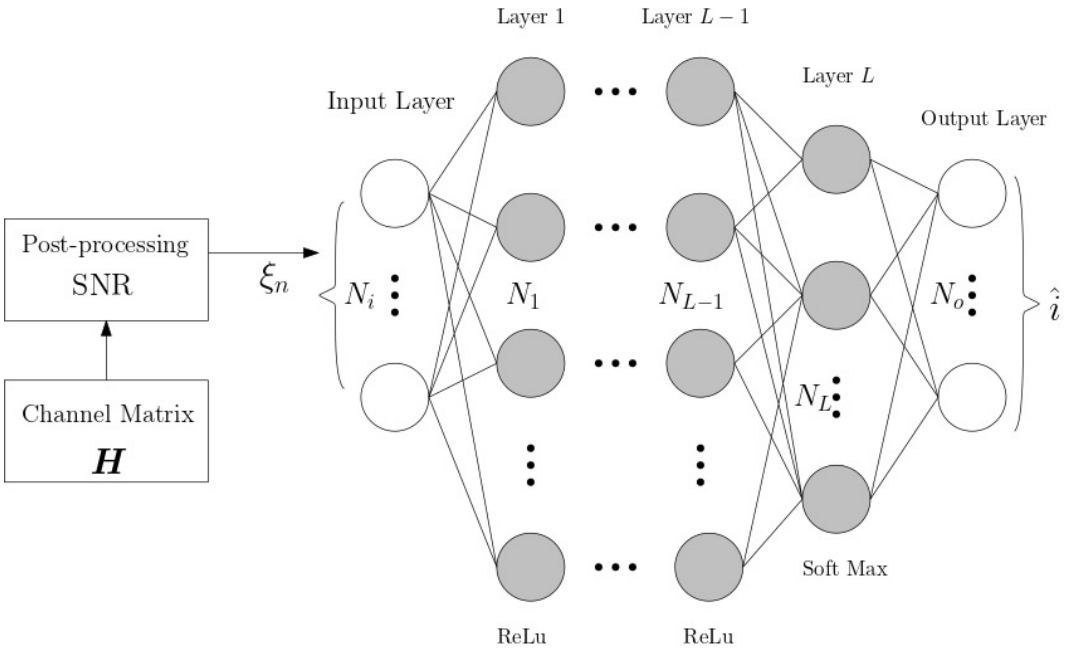

FIGURE 5.16: Fully-connected DNN model for CS-MIM adaptive modulation selection.

Then, the mode is decided by the majority mode of the k nearest neighbours to the input test point. With the possibility of several modes having the same number in the k nearest neighbours, the mode with the highest throughput will be selected.

The performance of KNN significantly depends on its parameters and on the value of k, where the best value of k can be selected empirically. In this adaptive system, the best value of k is determined by considering the trade-off between the BER and throughput. Furthermore, KNN results in a high computational complexity for the nearest neighbour search in addition to requiring a large memory for storing the training. Hence, in the following we present a DNN based design alternative.

DNN-aided Adaptive Design In this section, we present the DNN-based adaptive modulation regime of Fig. 5.16. Similarly to KNN, we randomly generate the training data and then store the mode index and SNR value pairs, which have BERs lower than the target value. Then, the training set $\mathcal{T}$ constitutes the estimated SNR ξ of a symbol associated with a BER lower than the target BER. We use the DNN-based classification model, where the input corresponds to the instantaneous SNR and the output corresponds to the mode index of adaptive modulation.

The output mode index $\hat{i}$ of the DNN can be expressed as

$$\begin{aligned}\hat{i} = & f_{softmax}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}\xi + \boldsymbol{\theta_1}]) \\ & + \boldsymbol{\theta_2}\} + ... + \boldsymbol{\theta}_n),\end{aligned} \tag{5.18}$$

where $\boldsymbol{W_n}$ and $\boldsymbol{\theta_n}$, $n = 1, \cdots, L$ represent the weights and biases, respectively. Relu is also employed for activating the DNN during the training phase, and the softmax function is used to obtain the mode index $\hat{i}$, which is

$$f_{softmax}(s) = \frac{e^s}{\sum_{c=1}^{C} e^{s_c}}. \tag{5.19}$$

The number of training samples required is selected based on experimentation by gradually increasing the training size until acceptable MSE values are achieved. In this case, the MSE loss function of the DNN used for the training is

$$\mathcal{L}(\xi, \hat{\xi}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B} \sum_{i=1}^{B} \|\xi - \hat{\xi}\|^2, \tag{5.20}$$

where B is the sample size of the current iteration.

A stopping criterion can be defined either by the number of iterations or by the maximum tolerable MSE threshold. Then, the parameter sets $\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}$ can be updated in each training iteration based on our learning algorithm using gradient descent, which is formulated as

$$\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} \leftarrow \{\boldsymbol{W_n}, \boldsymbol{\theta_n}\} - \alpha \nabla \boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\}),$$

where $\alpha > 0$ is the learning rate and $\nabla \boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$ represents the gradient of $\boldsymbol{L}(\{\boldsymbol{W_n}, \boldsymbol{\theta_n}\})$. In our proposed DNN-aided detection, we use $\alpha = 0.001$.

5.5 Performance Analysis

In this section, we characterize the performance of the proposed CS-JMIM system, where conventional detection will be used for benchmarking the proposed learning aided detection methods. Furthermore, we consider the system employing SF CS-JMIM and TSF CS-JMIM. The BER performance is evaluated by Monte-Carlo simulations, where we use the simulation parameters summarized in Table 5.4The parameters used by the learning models are presented in Table 5.6. In our simulations, we assume that the receiver has perfect channel knowledge, while in practice this is estimated using channel estimation techniques.

In the following, we present the different schemes considered in our simulations for comparison purposes. Firstly, we compared CS-aided separate multi-dimensional IM with CS-JMIM. More specifically, for our SF domain system, we compared CS-aided Generalized Subcarrier Index Modulation with SM (CS-GFIM-SM). These are termed as **Scheme 1, 3**, with CS-JMIM as **Scheme 2, 4**. Then, for the TSF domain, the CS-JMIM of **Scheme 5** is compared to **Scheme 6**, which represents the CS-MIM [19] [126]. Secondly, we compared the performance of different parameters in the context of **Schemes 2, 4,**

TABLE 5.4: CS-MIM system simulation parameters.

Parameters	Scheme 1	Scheme 2	Scheme 3	Scheme 4	Scheme 5	Scheme 6	Scheme 7	Scheme 8	Scheme 9
Scheme type	CS-GFIM-SM	CS-JMIM	CS-GFIM-SM	CS-JMIM	CS-MIM	CS-JMIM			
Detection type	HD							SD	
Multi-carrier System	OFDM								
Number of subcarriers, N_c	128								
Cyclic prefix	16								
Num of subcarrier group, G	64		32		64				
Num of active indices/gp, K	1,2		1,2,3		1,2		2		
Receiver antennas, N_r	2		4		8		2		
RSC code, (n, k, K)	-							(2,1,3)	
Real Domain									
Num of subcarrier/group, N_f	2		4		2				
Transmit antennas, N_t	2		4		8		2		
Activated antennas, N_{at}	1	-	1	-	2				
Time Slots,T	2								
Virtual Domain									
Num of available subcarrier/group, N_v	8	4	16	8	8		4		
Transmit antennas, N_{vt}	-	4	-	8	-		2		
Time Slots,T_v	4								
STSK codeword, (m, n, t, q, l)	-				(2,2,2,2,4)		-		

TABLE 5.5: Simulation results and complexity analysis of each Scheme.

Scheme index			SNR at BER of 10^{-5}	Throughput(bits/s/Hz)	Complexity
HD Detection					
Scheme 1		K=1	20.8	2.667	1.4×10^5
		K=2	26.5	4	5.6×10^5
Scheme 2	a)	K=1	30.3	2.667	9.5×10^5
		K=2	30.2	4.444	3.8×10^6
	b)		34.9	7.111	5.1×10^8
	c)		22.4	1.778	1.8×10^5
Scheme 3		K=1	16.6	1.778	8.6×10^6
		K=2	23.4	2.667	3.4×10^7
		K=3	28.1	3.778	1.4×10^8
Scheme 4	a)	K=1	8.2	1.778	5.3×10^7
		K=2	13.4	3.111	2.1×10^8
		K=3	19.4	4.667	8.3×10^8
	b)		34.6	5.333	2.2×10^{10}
	c)		8.3	1.333	7.2×10^6
Scheme 5		K=1	9.6	3.556	1.2×10^7
		K=2	13.3	5.333	4.9×10^7
Scheme 6	a)	K=1	-0.4	3.556	4.1×10^7
		K=2	4.9	6.222	6.5×10^9
	b)		15.7	17.778	5.4×10^{11}
	c)		1.5	1.778	1.1×10^7
Scheme 7	a)		5.6	3.556	2.2×10^5
	b)		18.7	17.778	1.7×10^6
	c)		1.8	1.778	6.6×10^4
SD Detection					
Scheme 8	a)		1.1	1.778	2.2×10^{13}
	b)		6.2	8.889	3.2×10^{14}
	c)		0.1	0.889	3.4×10^{12}
Scheme 9	a)		4.3	1.778	1.3×10^6
	b)		8.9	8.889	8.3×10^6
	c)		4.1	0.889	1.2×10^5
Adaptive Modulation					
Scheme 10	a)		-	-	-
	b)		-	-	5.2×10^6
	c)		-	-	1.22×10^5

6. Thirdly, we characterized the performance of DNN-aided CS-JMIM both in HD and SD in **Schemes 6-9**. We also quantified the complexity and compared it to conventional ML detection. Finally, we also exploited the adaptation of CS-JMIM between different JMIM methods in **Scheme 10**. To elaborate:

1. **Scheme 1**: applies ML HD detection for the CS-GFIM-SM, which activated one of 2 TAs, 2 RAs, and 2 subcarriers per group, while considering 8 subcarriers per group in the VD and $K = 1, 2$ activated subcarriers.

2. **Scheme 2**: applies maximum likelihood hard decision detection for the CS-JMIM system in the SF domain along with 2 TAs, 2 RAs, and 2 subcarriers per group in

TABLE 5.6: Training configuration for learning-aided detection method of Scheme 7,9

Setting	Hard-decision	Soft-decision
Maximum training epoch	400	1000
Initial learning rate	0.001	
Target SNR for training	0dB-20dB	-10dB to 5dB
Mini batch size	1000	200 to 500
Optimizer	Adam	
Training data size	50000	
Validation data ratio	0.1	

TABLE 5.7: Training configuration for adaptive modulation of Scheme 10

Setting	value
Number of Channel realizations for training	100000
Number of Channel realizations for testing	20000
Target SNR for training	0dB-30dB
Number of neighbors in KNN searchin k	15
Number of FC layers in DNN	3
Number of neurons in each FC layer	(128,256,128)
Number of output layer size	3
Activation function for output layer	Soft Max

the RD, while considering 4 antennas and 4 subcarriers per group in the VD. In this scheme, we consider the following mappings:

(a) General JMIM with $K = 1, 2$.

(b) Grouped JMIM with $gs = 4$ subgroups, and each subgroup applies general JMIM in conjunction with $K = 1$ (In this case, we can consider that both the FD and SpD is split into two sub groups, which have $gsx = gsy = 2$.).

(c) Coded JMIM with $n_q = 2$.

3. **Scheme 3**: applies ML HD detection for the CS-GFIM-SM, which activated one antenna out of 4 TAs, 4 RAs, and 4 subcarriers per group, while considering 16 subcarriers per group in the VD and $K = 1, 2, 3$ activated subcarriers.

4. **Scheme 4**: applies maximum likelihood hard decision detection for the CS-JMIM system in the SF domain along with 4 TAs, 4 RAs, and 4 subcarriers per group in this RD, with 8 antennas and 8 subcarriers per group in the VD. In this scheme, we consider the following mappings:

(a) General JMIM with $K = 1, 2, 3$.

(b) Grouped JMIM with $gs = 4, gsx = gsy = 2$ subgroups, with each subgroup applying the general JMIM along with $K = 1$.

(c) Coded JMIM with $n_q = 4$

5. **Scheme 5**: applies ML HD detection for the CS-MIM system in the TSF domain with 8 TAs, 8 RAs, 2 subcarriers per group and 2 TSs, while having 8 subcarriers per group in the VD. For the Space-Time-Shift-Keying (STSK) codeword $STSK(M, N, T, Q, L)$ used in CS-MIM [19], STSK(2,2,2,2,4) is applied. In this case, we have 2 activated antennas out of 8 and $K = 1, 2$ activated subcarrier out of 8 subcarrier in the VD.

6. **Scheme 8**: applies conventional SD detection for the CS-JMIM system in the TSF domain, while using RSC channel coding RSC(2,1,3). Here, we consider 2 TAs, 2 RAs, 2 subcarriers per group, and 2 TSs in the RD, while using 4 antennas, 4 subcarriers per group and 4 TSs in the VD. In this scheme, we consider the following mappings:

   (a) General JMIM with $K = 2$.

   (b) Grouped JMIM with $gs = 8, gsx = gsy = gsz = 2$ subgroups, each subgroup applied general JMIM with $K = 1$.

   (c) Coded JMIM with $n_q = 2$.

7. **Scheme 9**: applies DNN-based SD detection for the CS-JMIM system in the TSF domain, while using RSC channel coding RSC(2,1,3). Here, we consider 2 TAs, 2 RAs, 2 subcarriers per group, and 2 TSs in the RD, while using 4 antennas, 4 subcarriers per group and 4 TSs in the VD. In this scheme, we consider the following mappings:

   (a) General JMIM with $K = 2$.

   (b) Grouped JMIM with $gs = 8, gsx = gsy = gsz = 2$ subgroups, each subgroup applied general JMIM with $K = 1$.

   (c) Coded JMIM with $n_q = 2$.

8. **Scheme 10**: Adaptive HD-CS-JMIM system based on the TSF domain with 2 TAs, 2 RAs, 2 subcarriers per group, 2 TSs in RD and 4 antennas, 4 subcarriers per group and 4 TSs in VD. The details of the DNN based adaptive system design are shown in Table 5.7. In this system, we consider the following adaptation schemes:

   (a) Conventional adaptation.

   (b) KNN-based adaptation.

   (c) DNN-based adaptation.

As shown in Fig. 5.17, we compared the CS-aided separate MIM - namely the CS-GFIM-IM in this case - to CS-JMIM, which applied the general JMIM method of Section 5.2.1.1). In this case, based on the transmission rate calculation formula $\frac{bG}{N_c+L_{CP}}$, we have the transmission rate of the CS-GFIM-IM associated with $K = 1$ in **Scheme 1**

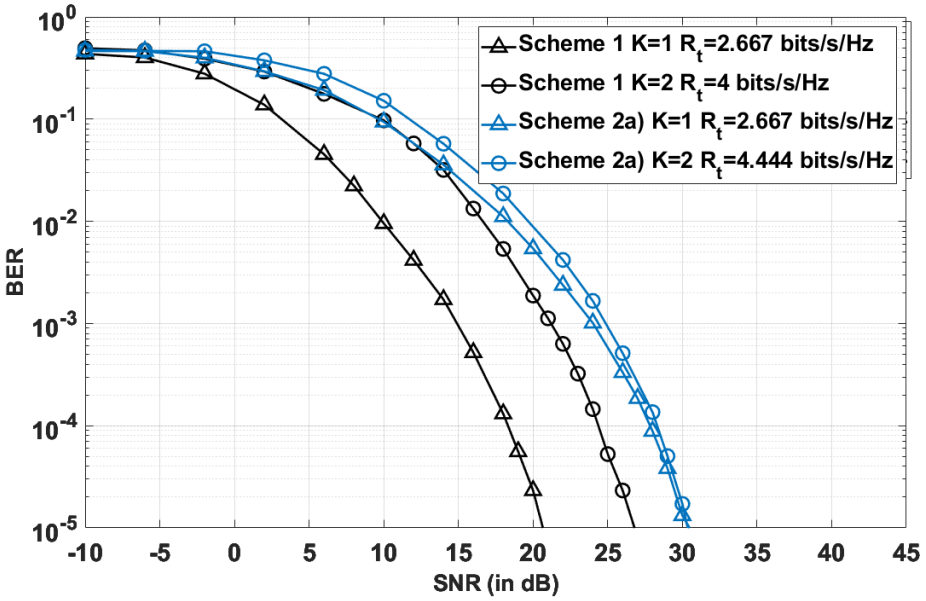

FIGURE 5.17: BER performance comparison of Scheme 1 and Scheme 2a). Our simulation parameters are shown in Table 5.4.

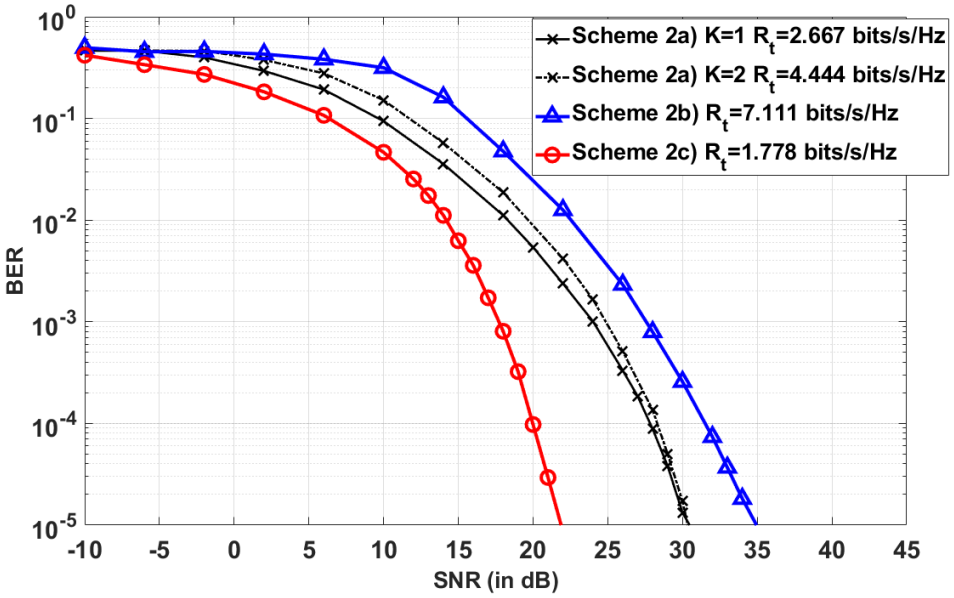

FIGURE 5.18: BER performance comparison of CS-JMIM Scheme 2. Our simulation parameters are shown in Table 5.4.

as $R_t^{k=1} = 2.667$ bits/s/Hz. This is the same as the CS-JMIM associated with $K = 1$ in **Scheme 2a)** under identical hardware configuration. However, the performance of **Scheme 2a)** is almost 10 dB worse than that of **Scheme 1** at a BER of 10^{-5}. Hence CS-JMIM is unattractive in this situation. For more activated index entities of both CS-JMIM and CS-GFIM-IM, the throughput of **Scheme 1** is increased to $R_t^{1,k=2} = 4$ bits/s/Hz and **Scheme 2a** has $R_t^{2,k=2} = 4.444$ bits/s/Hz. In this case, **Scheme 2a** of $K = 2$ has a 3.6 dB better performance than **Scheme 1** of $K = 2$ at a BER of 10^{-5}.

Fig. 5.18 shows the performance of the proposed CS-JMIM **Scheme 2** for different JMIM methods. Observe that for a small index space of $N_t = N_f = 2$, the detector cannot beneficially exploit the sparsity. The transmission rate of **Scheme 2** is either $R_t^{k=1} = 2.667$ bits/s/Hz, or $R_t^{k=2} = 4.444$ bits/s/Hz and we have $R_t^b = 7.111$ bits/s/Hz, $R_t^c = 1.778$ bits/s/Hz. As shown in Fig. 5.18, **Scheme 2a)** associated with $K = 1, 2$ has a similar BER performance, while **Scheme 2a)** of $K = 2$ has a higher throughput. Additionally, **Scheme 2b)** has almost 4 times the transmission rate compared to **Scheme 2c)**, but the latter has an increased diversity gain. Hence the BER performance of **Scheme 2c)** is 12dB better than that of **Scheme 2c)**.

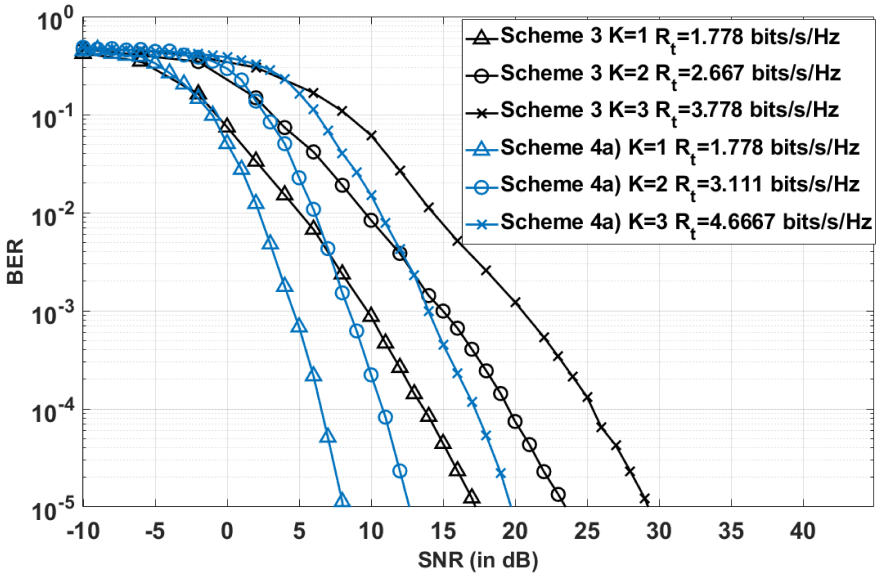

FIGURE 5.19: BER performance comparison of Scheme 3 and Scheme 4a). Our simulation parameters are shown in Table 5.4.

To further exploit the sparsity of CS-JMIM, we also consider larger SF dimensions applied to the JMIM method, as shown in Fig. 5.19. We assume that both schemes have the same number of TAs and subcarriers per group along with an adjustable number of VD subcarriers. For $N_t = 4, N_f = 4$, the CS-JMIM of **Scheme 4a)** achieves better performance than the separate MIM in **Scheme 3** with the same K value. Specifically, both schemes have $R_t^{k=1} = 1.777$ bits/s/Hz and **Scheme 3** associated with $K = 1$ obtains 5 dB SNR gain over **Scheme 4a)** with $K = 1$ at BER of 10^{-5}. When relying on a higher K, CS-JMIM is capable of providing higher throughput as well as improved detection performance. With $K = 2, 3$, the throughput of **Scheme 3** is $R_t^{k=2} = 2.667$ bits/s/Hz and $R_t^{k=3} = 3.333$ bits/s/Hz, respectively, while **Scheme 4a)** could achieve $R_t^{k=2} = 3.111$ bits/s/Hz and $R_t^{k=3} = 4.667$ bits/s/Hz.

Fig. 5.20 shows the BER performance of **Scheme 4**. A higher VD index mapping DM size allows for more flexible K value selection in **Scheme 4a)**. Observe that **Scheme**

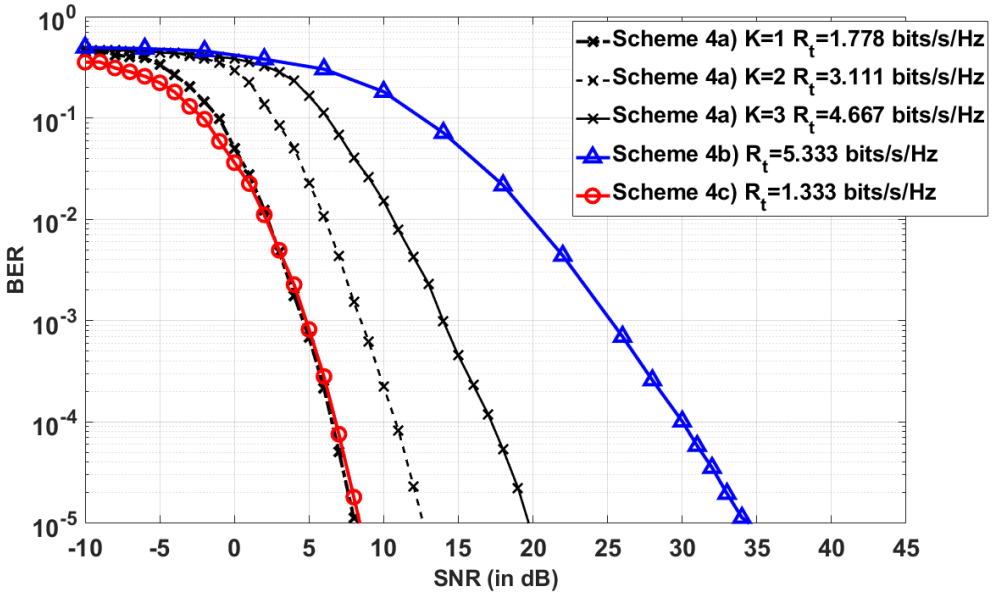

FIGURE 5.20: BER performance comparison of CS-JMIM Scheme 4. Our simulation parameters are shown in Table 5.4.

4a) with $K = 1$ achieves a similar performance to **Scheme 4c)**, where **Scheme 4a)** with $K = 1$ has $R_t = 1.778$ bits/s/Hz and **Scheme 4c)** has $R_t = 1.333$ bits/s/Hz.

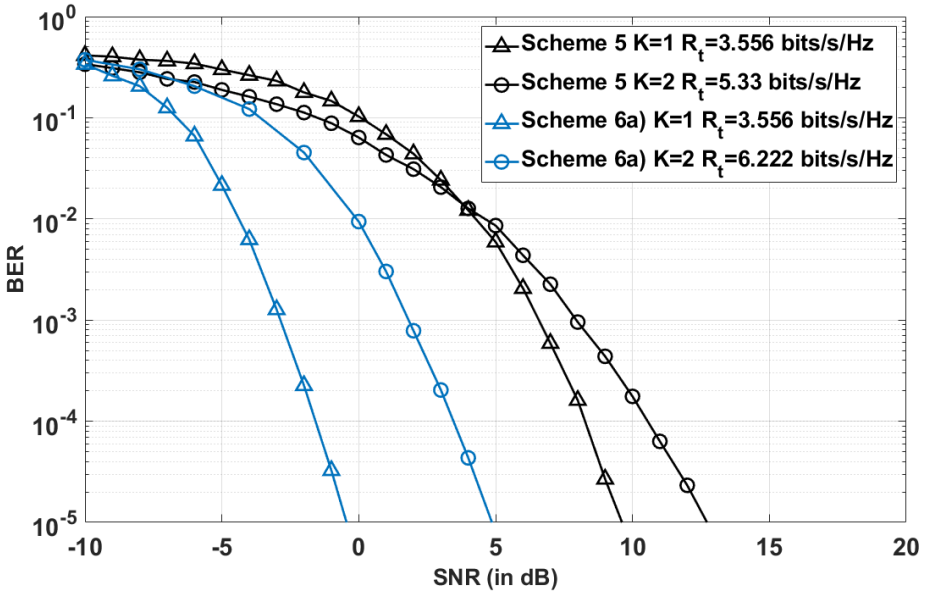

FIGURE 5.21: BER performance comparison of Scheme 5 and Scheme 6a). Our simulation parameters are shown in Table 5.4.

For the TSF domain system of Fig.5.6, we consider a separate model termed as CS-MIM [19]. This model applied SIM and STSK in the FD and CS is applied for the FD. Then the symbol after IFFT is modulated using SM and transmitted by the activated antennas. The CS-MIM scheme is simulated using the parameters of Table 5.4 for **Scheme 5**. In this case, to achieve the same throughput as **Scheme 5** and **Scheme 6a)** at $K = 1$, for

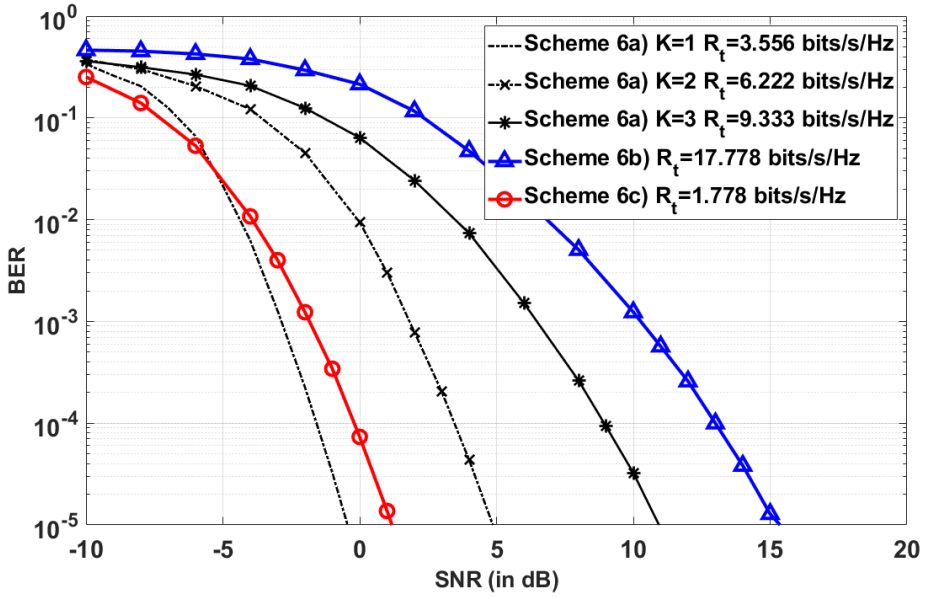

FIGURE 5.22: BER performance comparison of CS-JMIM Scheme 6. Our simulation parameters are shown in Table 5.4.

Scheme 5, we deliver the signals over 8 TAs with the aid of 2 RF chains. Then both **Scheme 5** and **Scheme 6a)** can have a throughput of $R_t^{K=1} = 3.556$ bits/s/Hz with $K = 1$. Then, we can observe in Fig.5.21 that **Scheme 6a)** achieves a BER of 10^{-5} at -0.1 dB while **Scheme 5** requires about 9.8 dB at the same BER. For $K = 2$, **Scheme 5** requires 13.5 dB SNR at 10^{-5} BER for $R_t^{K=2} = 5.333$ bits/s/Hz and **Scheme 6a)** requires 7.5 dB lower SNR than **Scheme 5** for $R_t^{K=2} = 6.222$ bits/s/Hz.

In Fig. 5.22, the TSF domains are considered for the CS-JMIM using **Scheme 6**. As shown in Fig. 5.22, **Scheme 6a)** with $K = 1$ attains the best performance among all types in **Scheme 6**. Quantitatively, at a BER of 10^{-5}, it requires an SNR of -0.3 dB and has a throughput of $R_t = 3.555 bits/s/Hz$. **Scheme 6c)** achieves a BER of 10^{-5} at an SNR of 1.1 dB. When higher dimensions are introduced, both the general JMIM and grouped JMIM can provide a high throughput as well as a good BER performance, albeit at the cost of a huge detection complexity. In Fig. 5.22, **Scheme 6b)** represents the grouped JMIM associated with 8 sub-groups. When $K = 1$ and the general JMIM DM is applied, we have $R_t = 17.778 bits/s/Hz$. This scheme attains a BER of 10^{-5} at an SNR of 15.1 dB. **Scheme 6a)** with $K = 3$ has $R_t = 9.333 bits/s/Hz$ and achieves a BER of 10^{-5} at an SNR of 11 dB. Hence, for higher dimensions, the grouped JMIM outperforms the other two JMIM methods. However, the complexity of grouped JMIM is exponentially increasing. Specifically, the detection complexity order of the grouped JMIM can be expressed as $\mathcal{O}[(N_{JMIM}(\mathcal{X}^K)^{N_{sub}}]$ for the TSF domain CS-JMIM system. This can be simplified to $\mathcal{O}[((N_v N_{vt} T_v/(g_s))(M^K)^{N_{sub}}]$, where N_{sub} represents the number of sub-groups. On the other hand, the detection complexity order of the general JMIM is $\mathcal{O}[(N_v N_{vt} K T_v M^K)]$. Furthermore, the coded JMIM complexity order

can be $\mathcal{O}[(N_q - n_q)n_q M]$. Then we can formulate the computational complexity order of ML for **Scheme 7a)** as $\mathcal{O}_{ML}[N_r N_f N_t T(N_r N_f N_t T N_{vt}^2 N_v^2 T_v^2 + N_{vt} N_v T_v M^2 N_f N_t T + MK)(N_{JMIM}(\mathcal{X}^K))]$. For **Scheme 7b)**, the sub-groups must be considered in each rounds, which have a complexity of $\mathcal{O}_{ML}[(N_{sub} N_r N_f N_t T/gs)(N_r N_f N_t T N_{vt}^2 N_v^2 T_v^2/(gs^2) + N_{vt} N_v T_v M N_f N_t T M/gs + MK)(N_{JMIM}(\mathcal{X}^K))^{N_{sub}}]$. For **Scheme 7c)**, we have a reduced complexity order of $\mathcal{O}_{ML}[N_r N_f N_t T N_{vt} N_v T_v M N_f N_t T M(N_q - n_q)n_q M]$ due to having multiple bit copies. Then we can calculate the computational complexity based on Table 5.4, as shown in Table 5.5.

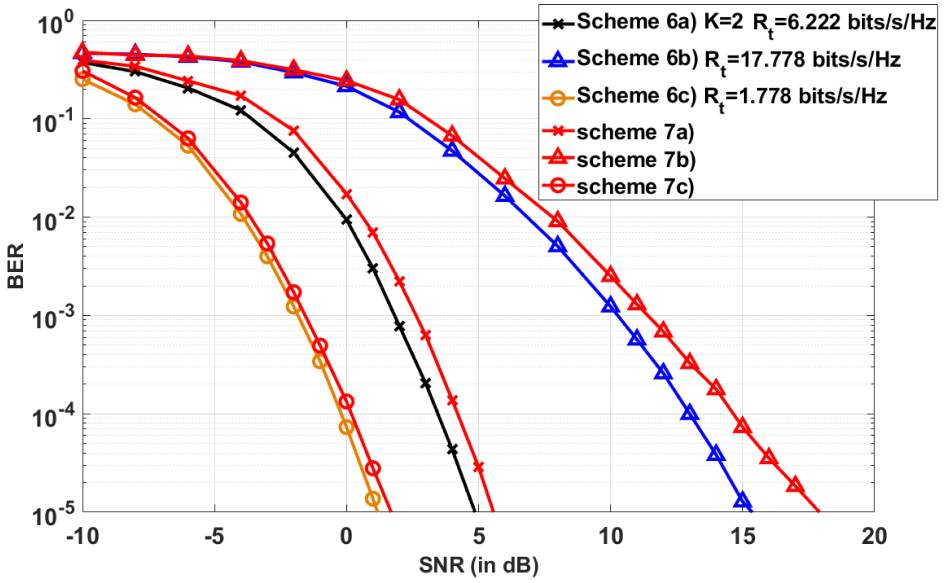

FIGURE 5.23: BER performance comparison of CS-JMIM Scheme 6 and Scheme 7. Our simulation parameters are shown in Table 5.4.

Upon increasing the throughput excessive detection complexity is imposed by conventional ML detection. To reduce the detection complexity, we have to accept a performance *vs.* complexity trade-off. In this context, we compare our DNN-based detector of the TSF based CS-JMIM system to conventional maximum likelihood detection by comparing **Scheme 6** and **Scheme 7** in Fig 5.23. Observe that the DNN-assisted HD detector achieves a similar performance to the ML detector. Furthermore, the complexity of the NN is determined by that of the forward and backward propagation, where we have the general DNN complexity order of $\mathcal{O}[n_i n_l n_{l+1} n_{h_L} n_o]$. Here n_i and n_o denote the number of neurons in the input and output layers, $n_l(l = 1, 2, \cdots)$ is the number of neurons in the hidden layer between the input and output. Then we can analyse each DNN model in **Scheme 8**. For a classification neural network, we have the LSTM layer as the activation layer of the input layer, which has the complexity of $\mathcal{O}_{LSTM}[4n_l(n_d + 2 + n_l)]$, where n_d is the number of neurons in the input layer and the popular sigmoid function is used as the activation layer of the output layer. The associated complexity is $\mathcal{O}[2n_L n_{L-1} - n_{L-1} + 2n_{L-1}]$. The complexity of the FC layer with

the ReLu function is given by $\mathcal{O}[2n_l n_{l-1} - n_{l-1}+$. Then we have the computational complexity order of $\mathcal{O}[4n_l(n_1+2+n_l)+\sum^{L}-1_l(2n_{l+1}n_l-n_l)+2n_{L-1}]$. Now we can also summarize the computational complexity of the DNN methods in Table 5.5.

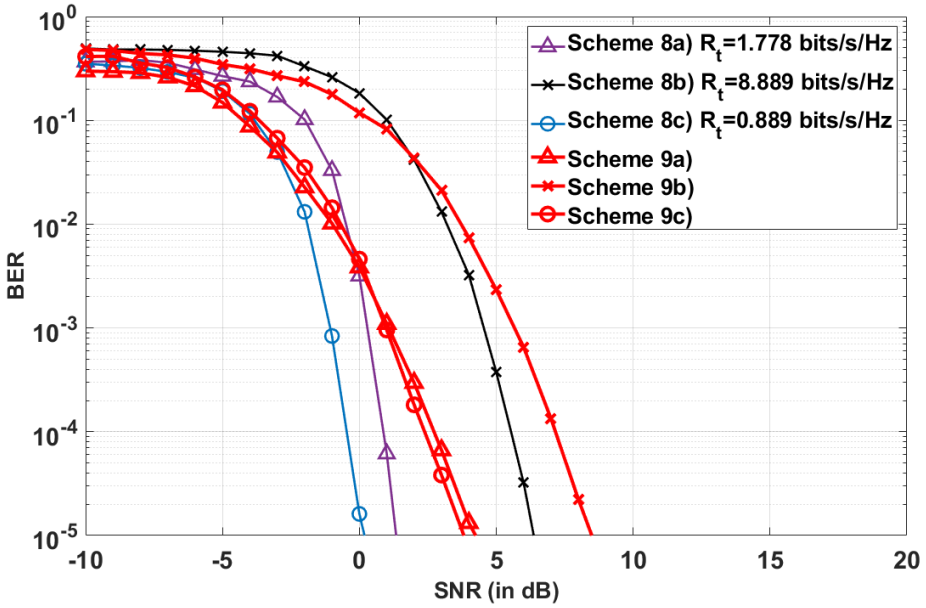

FIGURE 5.24: BER performance comparison of CS-JMIM Scheme 8 and Scheme 9. Our simulation parameters are shown in Table 5.4.

Furthermore, we extend the DNN-assisted detector to the SD of the TSF domain CS-JMIM system in **Scheme 8** and **Scheme 9**, while using the half-rate RSC encoder codeward RSC(2,1,3), having a memory of 3. As shown in Fig. 5.24, with the aid of channel coding, the performance of CS-JMIM can be further increased, as seen for **Scheme 8**. By comparing **Scheme 8** of Fig. 5.24 and **Scheme 6** of Fig. 5.23, the detection performance is 1dB better for **Scheme 8c)** than for **Scheme 6c)** at the BER of 10^{-5}. Furthermore, **Scheme 8a)** requires an SNR of 6.2 dBs at BER=10^{-5}, while **Scheme 6a)** necessitates SNR=1.6 dB. **Scheme 8b)** has the best performance, outperforming **Scheme 6b)** by about 8 dB at a BER of 10^{-5}. Fig.5.24 also shows the performance of DNN based detection for TSF CS-JMIM, where **Scheme 9a)** and **Scheme 9c)** exhibit similar performance. Quantitatively, they require about 4 and 3.2 dB at a BER of 10^{-5}. **Scheme 9b)** requires 3 dB higher SNR than the conventional SD detector, but it is still about 6 dB better than **Scheme 7b)**. The proposed learning method has a complexity order of $O[\mathcal{O}(n_i n_l)+\mathcal{O}(n_l^2)+\mathcal{O}(n_l n_o)]$ compared to $O[2^{c_g}(T_v N_t N_{vt}(Q\mathcal{X})^K]$ for the conventional scheme, where c_g denotes the RSC-coded number of bits in a transmitted symbol.

Finally, we present the performance of **Scheme 10** in Fig. 5.25. For the sake of a fair comparison, we use the data sets of the same size for both training and testing the KNN and DNN-based systems. Table 5.8 presents the configurations of the three modes of operation used in the adaptive system simulated. The switching thresholds for the conventional adaptive modulation are set as $P_1 = 1.85$ dB and $P_2 = 10.3$ dB, as shown in

TABLE 5.8: Configuration of mode used in conventional adaption with TSF domain CS-JMIM

No	Type	Scheme	R_t
1	Coded	Scheme 7a)	1.778
2	General	Scheme 7b)	6.222
3	Grouped	Scheme 7c)	17.778

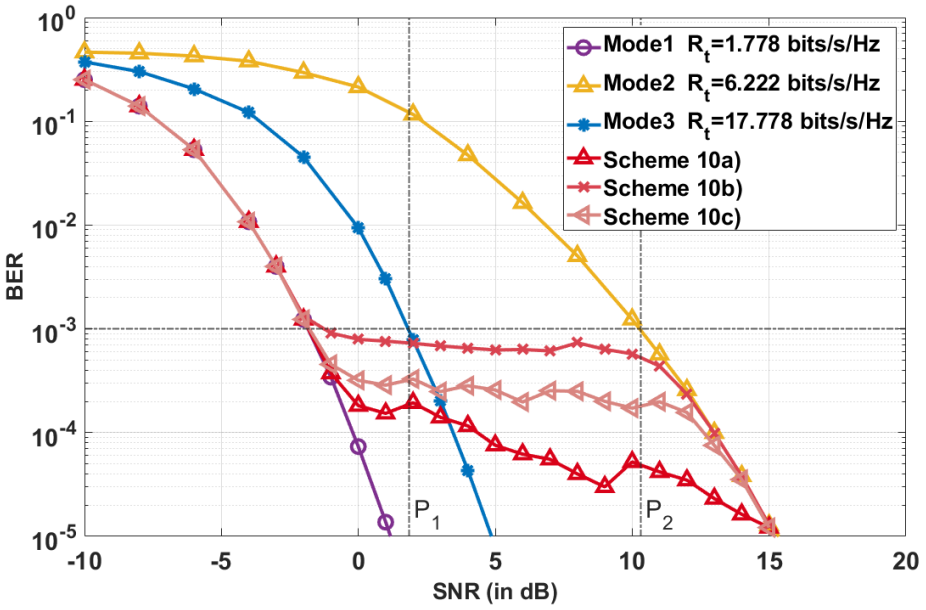

FIGURE 5.25: Adaptive modulation performance comparison of CS-JMIM Scheme 7. Our simulation parameters are shown in Table 5.4.

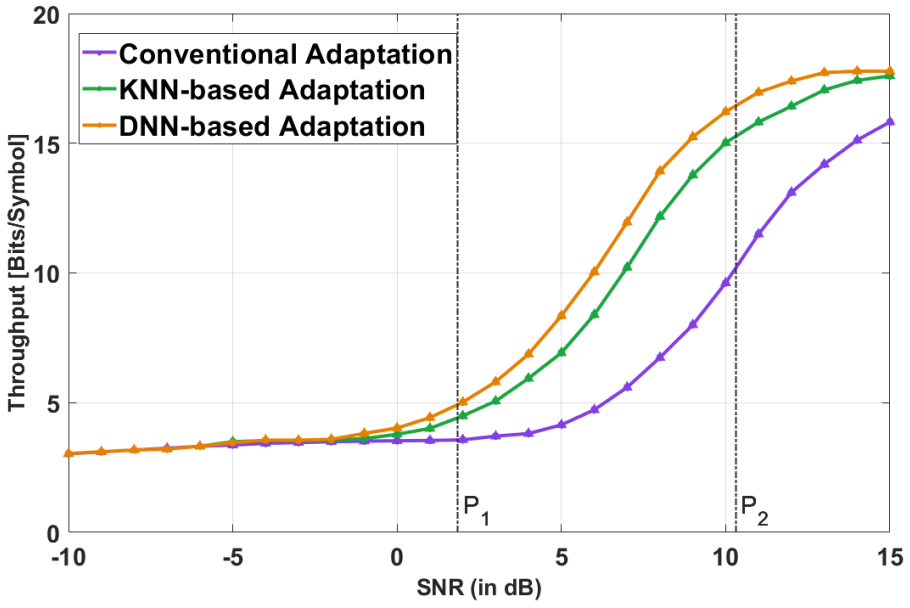

FIGURE 5.26: Adaptive modulation performance comparison of CS-JMIM Scheme 7. Our simulation parameters are shown in Table 5.4.

Fig. 5.25. Specifically, the conventional adaptive modulation scheme characterized in Fig. 5.25 uses $Mode_1$ when $SNR < P_1$, and $Mode_2$ for $P_2 > SNR$. After the instantaneous SNR becomes higher than P_2, $Mode_3$ is selected. Again, our KNN-based and DNN-based mode-selection algorithms are used in Fig. 5.25. Observe that the DNN based adaptive system attains a BER closer to the target of 10^{-3} than the KNN based adaptive system. Then we can further analyse the throughput of each mode selection scheme in Fig. 5.26. Observe that the DNN-based adaptive modulation scheme achieves a higher throughput than the KNN-based one, because more accurate decisions can be made by the DNN classifier than by the KNN classifier. Clearly, the learning assisted adaptive schemes are capable of selecting the best possible mode, while the conventional adaptive modulation uses the predefined average SNR-based thresholds for mode selection.

5.6 Summary

A CS-JMIM system was proposed and DL-aided detection using both HD and SD was conceived for reducing the detection complexity. We demonstrated that the proposed JMIM system is capable of outperforming its individual domain based counterpart, striking more flexible trade-offs between the BER performance and throughput. The learning method constructed is capable of approaching the performance of the maximum likelihood detector at a significantly reduced complexity. Furthermore, we showed that adaptive modulation can be applied for the selection of the JMIM DM design. We demonstrated that the CS-JMIM can flexibly adjust the transmission mode for accommodating time-variant channel conditions. We presented both KNN and DNN based adaptive schemes. Our simulation results showed that both the KNN and DNN-based approaches outperform the conventional threshold-based adaptive modulation. We also demonstrated that the DNN based adaptive design has a lower computational complexity and higher throughput than the KNN based approach.

Chapter 6

Conclusions and Future work

In this chapter, the conclusions and future research emerging from this thesis are discussed. More specifically, Section 6.1 summarizes the conclusions of this thesis, and Section 6.2 presents potential ideas for future research.

6.1 Conclusion

The evolution of wireless communication systems has led to the exploration of sophisticated techniques like the CS-MIM systems of Chapter 3. These systems face challenges in detection over fading channels, particularly in balancing the detection performance *vs.* computational complexity. Recent advances have shown the benefits of machine learning, especially DL, in enhancing the detection capabilities while reducing complexity. This trend inspired blind learning-aided detection in Chapter 3 and JCED in Chapter 4 for CS-MIM systems, highlighting their efficiency in both HD and SD contexts. Furthermore, a more flexible system termed as CS-JMIM was proposed in [165], where HD and SD based learning methods were also investigated in the context of learning-based adaptive modulation.

Blind learning-aided detection has emerged as a reduce complexity detection approach in CS-MIM systems. As seen in Chapter 3, our proposed learning-aided schemes outperform the traditional HD- ML, S-MP and AMP methods in terms of DCMC, despite their lower complexity. In the SD detection, the proposed DNN methods may approach the performance of their coherent CS-MIM counterparts without iterations, while substantially reducing their computational complexity. Furthermore, ISD is integrated into these DNN methods leverage for attaining iteration gains. Overall, these methods represent the first in the literature to eliminate the pilot overhead, while reducing the ISD complexity.

Then in Chapter 5, more practical transmission channel conditions were considered, which require CE to acquire CSI. By exploring both conventional and learning-assisted JCED in CS-MIM systems, our analysis revealed that JCED has the potential to reduce the pilot overhead while improving the detection performance compared to separate CE and detection. Our simulations demonstrate that conventional HD- JCED used in CS-MIM systems for transmission over Rayleigh fading channels achieves similar performance to learning-aided JCED at a lower complexity. Additionally, a subgroup-based DNN model conceived for SD JCED in CS-MIM systems approaches the performance of conventional SD CS-MIM systems at a reduced computational complexity. Our studies confirmed that conventional JCED achieves comparable BER performance to the ML detector relying on idealized CSI.

To further explore the benefits of MIM, in Chapter 5 a CS-JMIM system iwas proposed, coupled with DL-aided detection for both HD and SD, which aims to reduce the detection complexity. This system outperforms its individual domain-based counterparts, offering flexible trade-offs between the BER performance and throughput. The learning method constructed approaches the performance of the maximum likelihood detector at significantly reduced complexity. Moreover, adaptive modulation was applied for selecting the JMIM DMs design, allowing the CS-JMIM to adjust the transmission modes for time-variant channel conditions. Both KNN and DNN-based adaptive schemes were presented, with the simulation results showing their superiority over conventional threshold-based adaptive modulation. Notably, the DNN-based adaptive design imposes reduced computational complexity, despite its higher throughput compared to the KNN-based approach.

6.2 Future work

6.2.1 Graph Neural Network (GNN)-Based Detection of Multi-User (MU)-IM

MU IM have many issues to be solved due to increased users and dynamic environment. The influence of the channel becomes more severe than in single-user systems due to the inter-user interference. As DNN can work properly in solve MIMO system, we also consider apply DNN in MU-MIMO system. Hence it is promising to investigate an AMP-based GNN detector to detect the signal of CS-aided IM and evaluate its performance compared to conventional MMSE and ML detectors.

6.2.2 Downlink Communication in MU- IM Systems

For the downlink transmit, pre-processing may be used at the base station by exploiting the individual CSIs of the users reported back to the base station. Then learning-aided transmit pre-processing may be harnessed by the base station for eliminating the inter-user interference by the MU transmitter, so that low-complexity single-user detectors might be harnessed at the receiver. Furthermore, a jointly designed precoder and decoder pair may be optimized by learning tools.

6.2.3 Learning-Based Detection in Channel Coded IM Systems

The channel coding schemes of Chapter 3 may be further improved with the aid of near-capacity Extrinsic Information Transfer Chart (EXIT) chart-aided irregular coding, as discussed for example, in [166]. Furthermore, sophisticated turbo-equalization techniques may be designed relying on [100]. These schemes may also be extended to cooperative multi-cell processing-aided MU systems.

6.2.4 Multi-cell and Stacked Intelligent Meta-surfaces

Multi-cell systems should also be investigated [167]. The main issue is to solve the cell-edge problems [168]. Furthermore, at the system-level, we can consider grant-free [169] scenarios with IM. Stacked Intelligent Meta-surfaces and Holographic MIMO [170]may also be combined with IM.

References

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[2] C. Xu, N. Ishikawa, R. Rajashekar, S. Sugiura, R. G. Maunder, Z. Wang, L.-L. Yang, and L. Hanzo, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178 246–178 299, 2019.

[3] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 168–175, 7 2016.

[4] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index Modulation Techniques for Next-Generation Wireless Networks," *IEEE Access*, vol. 5, pp. 16 693–16 746, aug 2017.

[5] M. Wen, X. Cheng, and L. Yang, *Index modulation for 5G wireless communications*. Springer, 2017, vol. 52.

[6] X. Cheng, M. Zhang, M. Wen, and L. Yang, "Index modulation for 5G: Striving to do more with less," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 126–132, 2018.

[7] R. Y. Mesleh, H. Haas, S. Sinanović, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 7 2008.

[8] L. Xiao, P. Yang, Y. Xiao, S. Fan, M. Di Renzo, W. Xiang, and S. Li, "Efficient compressive sensing detectors for generalized spatial modulation systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1284–1298, feb 2017.

[9] R. Fan, Y. J. Yu, and Y. L. Guan, "Generalization of orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5350–5359, 2015.

[10] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (SIM) OFDM," *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 728–732, 2011.

[11] Y. Naresh and A. Chockalingam, "A low-complexity maximum-likelihood detector for differential media-based modulation," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2158–2161, 2017.

[12] M. Nakao, T. Ishihara, and S. Sugiura, "Dual-mode time-domain index modulation for nyquist-criterion and faster-than-nyquist single-carrier transmissions," *IEEE Access*, vol. 5, pp. 27 659–27 667, 2017.

[13] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3219–3230, nov 2010.

[14] A. K. Khandani, "Media-based modulation: A new approach to wireless transmission," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 3050–3054.

[15] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: a survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 315–348, 2019.

[16] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on signal processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[17] H. Zhang, L.-L. Yang, and L. Hanzo, "Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, 2016.

[18] B. Shamasundar, S. Bhat, S. Jacob, and A. Chockalingam, "Multidimensional index modulation in wireless communications," *IEEE Access*, vol. 6, pp. 589–604, 2018.

[19] S. Lu, I. A. Hemadeh, M. El-hajjar, and L. Hanzo, "Compressed sensing-aided multi-dimensional index modulation," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4074–4087, 2019.

[20] P. Yang, Y. Xiao, Y. L. Guan, M. Di Renzo, S. Li, and L. Hanzo, "Multi-domain index modulation for vehicular communications: A survey," *IEEE Vehicular Technology Magazine*, 2018.

[21] R. Chen and J. Zheng, "Index-modulated mimo-ofdm: Joint space-frequency signal design and linear precoding in rapidly time-varying channels," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 7067–7079, 2018.

[22] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Spatial modulation: optimal detection and performance analysis," *IEEE Communications Letters*, vol. 12, no. 8, pp. 545–547, 2008.

[23] Y. Xiao, Z. Yang, L. Dan, P. Yang, L. Yin, and W. Xiang, "Low-complexity signal detection for generalized spatial modulation," *IEEE Communications Letters*, vol. 18, no. 3, pp. 403–406, 2014.

[24] L. Wei, J. Zheng, and Q. Liu, "Approximate message passing detector for index modulation with multiple active resources," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 972–976, 2019.

[25] Z. Sui, S. Yan, H. Zhang, L.-L. Yang, and L. Hanzo, "Approximate message passing algorithms for low complexity ofdm-im detection," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9607–9612, 2021.

[26] C. S. Yeh and Y. Lin, "Channel estimation using pilot tones in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 45, no. 4, pp. 400–409, 1999.

[27] S. M. Lee, D. H. Lee, and H. J. Choi, "Performance comparison of space-time codes and channel estimation in OFDM systems with transmit diversity for wireless LANs," *APCC 2003 - 9th Asia-Pacific Conference on Communications, in conjunction with 6th Malaysia International Conference on Communications, MICC 2003, Proceedings*, vol. 1, pp. 406–410, 2003.

[28] T. Gruber, S. Cammerer, J. Hoydis, and S. T. Brink, "On deep learning-based channel decoding," in *2017 51st Annual Conference on Information Sciences and Systems, CISS 2017*. Institute of Electrical and Electronics Engineers Inc., may 2017.

[29] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5g wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 214–222, 2020.

[30] X. Jin and H. N. Kim, "Parallel Deep Learning Detection Network in the MIMO Channel," *IEEE Communications Letters*, vol. 24, no. 1, pp. 126–130, 2020.

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[32] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive mimo is a reality—what is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[33] M. Zamanipour, "A survey on deep-learning based techniques for modeling and estimation of massivemimo channels," *arXiv preprint arXiv:1910.03390*, 2019.

[34] L. Liang, H. Ye, and G. Y. Li, "Toward intelligent vehicular networks: A machine learning framework," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 124–135, 2019.

[35] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648–664, 12 2018.

[36] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[37] Y. Chau and S.-H. Yu, "Space modulation on wireless fading channels," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001. Proceedings (Cat. No.01CH37211)*, vol. 3, 2001, pp. 1668–1671 vol.3.

[38] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Generalized space shift keying modulation for mimo channels," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2008, pp. 1–5.

[39] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[40] C. C. Cheng, H. Sari, S. Sezginer, and Y. T. Su, "New signal designs for enhanced spatial modulation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7766–7777, 2016.

[41] R. Mesleh, S. S. Ikki, and H. M. Aggoune, "Quadrature spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2738–2742, 2015.

[42] C.-C. Cheng, H. Sari, S. Sezginer, and Y. T. Su, "Enhanced spatial modulation with multiple signal constellations," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2237–2248, 2015.

[43] R. Abu-alhiga and H. Haas, "Subcarrier-index modulation OFDM," *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 177–181, 2009.

[44] E. Başar, "Multiple-input multiple-output ofdm with index modulation," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2259–2263, 2015.

[45] E. Basar, "On multiple-input multiple-output ofdm with index modulation for next generation wireless networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3868–3878, 2016.

[46] E. Başar, "Ofdm with index modulation using coordinate interleaving," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 381–384, 2015.

[47] S. Doğan Tusha, A. Tusha, E. Basar, and H. Arslan, "Multidimensional index modulation for 5g and beyond wireless networks," *Proceedings of the IEEE*, vol. 109, no. 2, pp. 170–199, 2021.

[48] S. Jacob, T. L. Narasimhan, and A. Chockalingam, "Space-time index modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[49] A. F. Naguib and R. Calderbank, "Space-time coding and signal processing for high data rate wireless communications," *Wireless communication technologies: new multimedia systems*, pp. 23–59, 2002.

[50] S. Sugiura, S. Chen, and L. Hanzo, "Space-time shift keying: A unified mimo architecture," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 2010, pp. 1–5.

[51] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, Z. Wang, and L. Hanzo, ""Near-perfect" ginite-cardinality generalized space-time shift keying," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 2146–2164, 2019.

[52] S. Sugiura, S. Chen, and L. Hanzo, "A unified mimo architecture subsuming space shift keying, ostbc, blast and ldc," Tech. Rep., 2010.

[53] H. A. Ngo, C. Xu, S. Sugiura, and L. Hanzo, "Space-time-frequency shift keying for dispersive channels," *IEEE Signal Processing Letters*, vol. 18, no. 3, pp. 177–180, 2011.

[54] M. I. Kadir, S. Chen, K. Hari, K. Giridhar, and L. Hanzo, "Ofdm-aided differential space–time shift keying using iterative soft multiple-symbol differential sphere decoding," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 4102–4108, 2014.

[55] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Multi-Set Space-Time Shift-Keying with Reduced Detection Complexity," *IEEE Access*, vol. 4, pp. 4234–4246, 2016.

[56] M. Driusso, F. Babich, M. I. Kadir, and L. Hanzo, "Ofdm aided space-time shift keying for dispersive downlink channels," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, 2012, pp. 1–5.

[57] T. Datta, H. S. Eshwaraiah, A. Chockalingam, and S. Member, "Generalized space-and-frequency index modulation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 4911–4924, 2016.

[58] Z. Li and J. Zheng, "Space-frequency shift keying in rapidly time-varying MIMO OFDM channels," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–5.

[59] S. Lu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Compressed-sensing-aided space-time frequency index modulation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6259–6271, 2018.

[60] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, apr 2006.

[61] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[62] Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 144–153, 2018.

[63] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, pp. 223–229, 2002.

[64] W. G. Jeon, K. H. Paik, and Y. S. Cho, "An efficient channel estimation technique for OFDM systems with transmitter diversity," *IEICE Transactions on Communications*, vol. E84-B, no. 4, pp. 967–974, 2001.

[65] V. K. Jones and G. G. Raleigh, "Channel estimation for wireless OFDM systems," *Conference Record / IEEE Global Telecommunications Conference*, vol. 2, pp. 980–985, 1998.

[66] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," *IEEE Vehicular Technology Conference*, vol. 3, pp. 1639–1643, 1997.

[67] H. Tang, K. Y. Lau, and R. W. Brodersen, "Interpolation-based maximum likelihood channel estimation using OFDM pilot symbols," *Conference Record / IEEE Global Telecommunications Conference*, vol. 2, pp. 1860–1864, 2002.

[68] Y. Sun, C. Pandana, X. Wang, and K. J. Liu, "A joint channel estimation and unequal error protection scheme for image transmission in wireless OFDM systems," *Proceedings of 2002 IEEE Workshop on Multimedia Signal Processing, MMSP 2002*, pp. 380–383, 2002.

[69] J. Ran, R. Grunheid, H. Rohling, E. Bolinth, and R. Kern, "Decision-directed channel estimation method for ofdm systems with high velocities," in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring.*, vol. 4, 2003, pp. 2358–2361 vol.4.

[70] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "A study of channel estimation in OFDM systems," *IEEE Vehicular Technology Conference*, vol. 56, no. 2, pp. 894–898, 2002.

[71] Y. Acar, H. Dogan, and E. Panayirci, "Channel estimation for spatial modulation orthogonal frequency division multiplexing systems," *9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings*, pp. 382–385, nov 2015.

[72] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 67–75, 2002.

[73] ——, "Optimum training sequences for OFDM systems with multiple transmit antennas," *Conference Record / IEEE Global Telecommunications Conference*, vol. 3, pp. 1478–1482, 2000.

[74] A. Huang and Y. Zhao, "Estimating channel response from pilot subcarrier pairs for OFDM systems," *Midwest Symposium on Circuits and Systems*, vol. 2, pp. 774–777, 1997.

[75] A. Dowler, A. Doufexi, and A. Nix, "Performance evaluation of channel estimation techniques for a mobile fourth generation wide area ofdm system," in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 4, 2002, pp. 2036–2040 vol.4.

[76] J. Rinne and M. Renfors, "Pilot spacing in Orthogonal Frequency Division Multiplexing systems on practical channels," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 4, pp. 959–962, 1996.

[77] C. R. Athaudage and A. D. Jayalath, "Low-complexity channel estimation for wireless OFDM systems," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 1, pp. 521–525, 2003.

[78] H. Yu, M. S. Kim, and S. K. Lee, "Channel estimation and equalization for high speed mobile OFDM systems," *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 693–697, 2003.

[79] S. G. Kang, Y. M. Ha, and E. K. Joo, "A comparative investigation on channel estimation algorithms for OFDM in mobile communications," *IEEE Transactions on Broadcasting*, vol. 49, no. 2, pp. 142–149, jun 2003.

[80] X. Wang and K. J. Ray Liu, "OFDM channel estimation based on time-frequency polynomial model of fading multipath channels," *Conference Record / IEEE Global Telecommunications Conference*, vol. 1, pp. 212–216, 2001.

[81] M. X. Chang and Y. T. Su, "2D regression channel estimation for equalizing OFDM signals," *IEEE Vehicular Technology Conference*, vol. 1, pp. 240–244, 2000.

[82] M.-X. Chang and Y. T. Su, "Model-based channel estimation for ofdm signals in rayleigh fading," *IEEE Transactions on Communications*, vol. 50, no. 4, pp. 540–544, 2002.

[83] R. Mesleh, H. Haas, C. W. Ahn, and S. Yun, "Spatial modulation - A new low complexity spectral efficiency enhancing technique," *First International Conference on Communications and Networking in China, ChinaCom '06*, 2006.

[84] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 805–808, 2012.

[85] E. Karami and M. Shiva, "Decision-directed recursive least squares mimo channels tracking," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, pp. 1–10, 2006.

[86] X. Wu, H. Claussen, M. Di Renzo, and H. Haas, "Channel estimation for spatial modulation," *IEEE Transactions on Communications*, vol. 62, no. 12, pp. 4362–4372, dec 2014.

[87] Y. Acar, S. Aldirmaz Çolak, and E. Başar, "Channel estimation for OFDM-IM systems," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 3, pp. 1908–1921, 2019.

[88] L. He, J. Wang, and J. Song, "On massive spatial modulation mimo: Spectral efficiency analysis and optimal system design," pp. 1–6, 2016.

[89] S. Lu, M. El-Hajjar, and L. Hanzo, "Two-dimensional index modulation for the large-scale multi-user mimo uplink," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7904–7918, 2019.

[90] I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Hierarchical Multi-Functional Layered Spatial Modulation," *IEEE Access*, vol. 6, pp. 9492–9533, 2018.

[91] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity iterative-detection-aided generalized space-time shift keying," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3656–3664, 2012.

[92] J. Tropp, A. G. I. T. on information Theory, and undefined 2007, "Signal recovery from random measurements via orthogonal matching pursuit," *ieeexplore.ieee.org*.

[93] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on signal processing*, vol. 60, no. 12, pp. 6202–6216, 2012.

[94] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Communications of the ACM*, vol. 53, no. 12, pp. 93–100, dec 2010.

[95] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[96] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2986–3001, 2014.

[97] J. Hagenauer and P. Hoeher, "A viterbi algorithm with soft-decision outputs and its applications," in *1989 IEEE Global Telecommunications Conference and Exhibition'Communications Technology for the 1990s and Beyond'*. IEEE, 1989, pp. 1680–1686.

[98] H. V. Nguyen, C. Xu, S. X. Ng, and L. Hanzo, "Near-capacity wireless system design principles," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1806–1833, 2015.

[99] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on information theory*, vol. 42, no. 2, pp. 429–445, 1996.

[100] L. Hanzo, J. P. Woodard, and P. Robertson, "Turbo decoding and detection for wireless applications," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1178–1200, 2007.

[101] P. Zhang, S. Chen, and L. Hanzo, "Near-capacity joint channel estimation and three-stage turbo detection for MIMO systems," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 61, no. 5, pp. 3986–3991, 2013.

[102] ——, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE transactions on communications*, vol. 61, no. 5, pp. 1902–1913, 2013.

[103] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[104] D. J. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 33, no. 6, pp. 457–458, 1997.

[105] J. Zheng and Q. Liu, "Low-complexity soft-decision detection of coded OFDM with index modulation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7759–7763, aug 2018.

[106] "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, sep 1977.

[107] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1266–1270, aug 2003.

[108] H. Zhang, W. Li, and Y. Xu, "EM-based channel estimation with index modulation for high speed train communication," *Proceedings - 2019 International Conference on Communications, Information System, and Computer Engineering, CISCE 2019*, pp. 390–394, jul 2019.

[109] S. Chen, S. Sugiura, and L. Hanzo, "Semi-blind joint channel estimation and data detection for space-time shift keying systems," *IEEE Signal Processing Letters*, vol. 17, no. 12, pp. 993–996, 2010.

[110] E. Alpaydin, *Machine learning*. Mit Press, 2021.

[111] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[112] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, and L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, apr 2017.

[113] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[114] H. Ye, G. Y. Li, and B. H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2018.

[115] A. Al-Baidhani and H. H. Fan, "Learning for Detection: A Deep Learning Wireless Communication Receiver Over Rayleigh Fading Channels," *2019 International Conference on Computing, Networking and Communications, ICNC 2019*, pp. 6–10, 2019.

[116] T. Erpek, T. J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep Learning for Wireless Communications," *Studies in Computational Intelligence*, vol. 867, pp. 223–266, 2020.

[117] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, vol. 2017-July. Institute of Electrical and Electronics Engineers Inc., 12 2017, pp. 1–5.

[118] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[119] N. Samuel and T. Diskin, "Learning to Detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, may 2019.

[120] J. Liu and H. Lu, "Imnet: A learning based detector for index modulation aided mimo-ofdm systems," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2020.

[121] T. V. Luong, Y. Ko, N. A. Vien, D. H. N. Nguyen, and M. Matthaiou, "Deep learning-based detector for OFDM-IM," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1159–1162, 2019.

[122] T. Wang, F. Yang, J. Song, and Z. Han, "Deep convolutional neural network-based detector for index modulation," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1705–1709, oct 2020.

[123] M. Turhan, E. Öztürk, and H. A. Çırpan, "Deep convolutional learning-aided detector for generalized frequency division multiplexing with index modulation,"

in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–6.

[124] S. Katla, L. Xiang, Y. Zhang, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Deep learning assisted detection for index modulation aided mmwave systems," *IEEE Access*, vol. 8, pp. 202 738–202 754, 2020.

[125] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, P. Pietraski, and L. Hanzo, "Soft-decoding for multi-set space-time shift-keying mmWave systems: A deep learning approach," *IEEE Access*, vol. 8, pp. 49 584–49 595, 2020.

[126] X. Feng, M. EL-Hajjar, C. Xu, and L. Hanzo, "Deep learning-based soft iterative-detection of channel-coded compressed sensing-aided multi-dimensional index modulation," *IEEE Transactions on Vehicular Technology*, pp. 1–16, 2023.

[127] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep Learning Based Channel Estimation for Massive MIMO with Mixed-Resolution ADCs," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1989–1993, nov 2019.

[128] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. Naderializadeh, "Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 989–1000, 2019.

[129] E. Balevi and J. G. Andrews, "Two-Stage Learning for Uplink Channel Estimation in One-Bit Massive MIMO," *Conference Record - Asilomar Conference on Signals, Systems and Computers*, vol. 2019-November, pp. 1764–1768, nov 2019.

[130] L. Xiang, Y. Liu, T. V. Luong, R. G. Maunder, L. L. Yang, and L. Hanzo, "Deep-learning-aided joint channel estimation and data detection for spatial modulation," *IEEE Access*, vol. 8, pp. 191 910–191 919, 2020.

[131] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[132] C. Xu, S. Sugiura, S. X. Ng, P. Zhang, L. Wang, and L. Hanzo, "Two decades of mimo design tradeoffs and reduced-complexity mimo detection in near-capacity systems," *IEEE Access*, vol. 5, pp. 18 564–18 632, 2017.

[133] S. Dorner, S. Cammerer, J. Hoydis, and S. T. Brink, "Deep Learning Based Communication over the Air," *IEEE Journal on Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, feb 2018.

[134] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. Murthy, and M. van der Schaar, "Machine Learning in the Air," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2184–2199, apr 2019.

[135] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a Reality – What is Next? Five Promising Research Directions for Antenna Arrays," *Digital Signal Processing: A Review Journal*, vol. 94, pp. 3–20, feb 2019.

[136] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[137] C. Xu, T. Van Luong, L. Xiang, S. Sugiura, R. G. Maunder, L.-L. Yang, and L. Hanzo, "Turbo detection aided autoencoder for multicarrier wireless systems: Integrating deep learning into channel coded systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 600–614, 2022.

[138] O. Shental and J. Hoydis, ""machine llrning": Learning to softly demodulate," pp. 1–7, 2019.

[139] L. Hanzo, B. Choi, T. Keller *et al.*, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2005.

[140] W. Koch and A. Baier, "Optimum and sub-optimum detection of coded data disturbed by time-varying intersymbol interference (applicable to digital mobile radio receivers)," in *[Proceedings] GLOBECOM '90: IEEE Global Telecommunications Conference and Exhibition*, 1990, pp. 1679–1684 vol.3.

[141] Y. Acar, H. Dogan, and E. Panayırcı, "Channel estimation for spatial modulation orthogonal frequency division multiplexing systems," pp. 382–385, 2015.

[142] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 805–808, 2012.

[143] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[144] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[145] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[146] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 127–153, mar 2014.

[147] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *Electronics Letters*, vol. 32, no. 10, pp. 887–888, 1996.

[148] T. Diskin, G. Draskovic, F. Pascal, and A. Wiesel, "Deep robust regression," *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2017*, vol. 2017-December, pp. 1–5, mar 2018.

[149] M. M. U. Faiz, S. Al-Ghadhban, and A. Zerguine, "Recursive least-squares adaptive channel estimation for spatial modulation systems," in *2009 IEEE 9th Malaysia International Conference on Communications (MICC)*, 2009, pp. 785–788.

[150] X. Wu, M. Di Renzo, and H. Haas, "Channel estimation for spatial modulation," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 306–310.

[151] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for mimo systems," *IEEE Signal Processing Letters*, vol. 15, pp. 202–205, 2008.

[152] S. Chen, S. Sugiura, and L. Hanzo, "Semi-blind joint channel estimation and data detection for space-time shift keying systems," *IEEE Signal Processing Letters*, vol. 17, no. 12, pp. 993–996, 2010.

[153] Y. Acar, M. M. Leblebici, H. Doğan, and E. Panayirci, "Data detection based iterative channel estimation for coded SM-OFDM systems," in *2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2016, pp. 1–4.

[154] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1266–1270, 2003.

[155] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, 2019.

[156] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for mimo-ofdm systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6315–6328, 2021.

[157] C. Qing, L. Dong, L. Wang, J. Wang, and C. Huang, "Joint model and data-driven receiver design for data-dependent superimposed training scheme with imperfect hardware," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 3779–3791, 2022.

[158] L. Xiang, Y. Liu, T. Van Luong, R. G. Maunder, L.-L. Yang, and L. Hanzo, "Deep-learning-aided joint channel estimation and data detection for spatial modulation," *IEEE Access*, vol. 8, pp. 191 910–191 919, 2020.

[159] B. Chakrapani, T. L. Narasimhan, and A. Chockalingam, "Generalized space-frequency index modulation: Low-complexity encoding and detection," in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.

[160] P. Patcharamaneepakorn, C.-X. Wang, Y. Fu, E.-H. M. Aggoune, M. M. Alwakeel, X. Tao, and X. Ge, "Quadrature space-frequency index modulation for energy-efficient 5G wireless communication systems," *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 3050–3064, 2018.

[161] S. Sugiura, S. Chen, and L. Hanzo, "Generalized space-time shift keying designed for flexible diversity-, multiplexing- and complexity-tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1144–1153, 2011.

[162] H. Liu, S. Lu, M. El-Hajjar, and L.-L. Yang, "Machine learning assisted adaptive index modulation for mmWave communications," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1425–1441, 2020.

[163] H. Liu, Y. Zhang, X. Zhang, M. El-Hajjar, and L.-L. Yang, "Deep learning assisted adaptive index modulation for mmWave communications With channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9186–9201, 2022.

[164] P. Yang, Y. Xiao, M. Xiao, Y. L. Guan, S. Li, and W. Xiang, "Adaptive spatial modulation mimo based on machine learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 2117–2131, 2019.

[165] X. **Feng**, M. EL-Hajjar, C. Xu, and L. Hanzo, "Near-instantaneously adaptive learning-assisted and compressed sensing-aided joint multi-dimensional index modulation," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 893–912, 2023.

[166] M. El-Hajjar and L. Hanzo, "Exit charts for system design and analysis," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 127–153, 2014.

[167] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

[168] L. He, J. Wang, J. Song, and L. Hanzo, "On the multi-user multi-cell massive spatial modulation uplink: How many antennas for each user?" *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1437–1451, 2017.

[169] A. T. Abebe, J. Lee, M. Rim, and C. G. Kang, "Multi-cell performance of grant-free and non-orthogonal multiple access," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017, pp. 1–6.

[170] J. An, C. Xu, D. W. K. Ng, G. C. Alexandropoulos, C. Huang, C. Yuen, and L. Hanzo, "Stacked intelligent metasurfaces for efficient holographic mimo communications in 6g," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2380–2396, 2023.