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

Xinyu Feng, *Student Member, IEEE,* Mohammed EL-Hajjar, *Senior Member, IEEE,* Chao Xu, *Senior Member, IEEE,* and Lajos Hanzo, *Life Fellow, IEEE*.

***Abstract*—Index Modulation (IM) is a flexible transmission scheme capable of striking a flexible performance, throughput, diversity and complexity trade-off. The concept of Multi-dimensional IM (MIM) has been developed to combine the benefits of IM in multiple dimensions, such as space and frequency. Furthermore, Compressed Sensing (CS) can be beneficially combined with IM in order to increase its throughput. However, having accurate Channel State Information (CSI) is essential for reliable MIM, which requires high pilot overhead. Hence, Joint Channel Estimation and Detection (JCED) is harnessed to reduce the pilot overhead and improve the detection performance at a modestly increased estimation complexity. We then circumvent this by proposing Deep Learning (DL) based JCED for CS aided MIM (CS-MIM) of significantly reducing the complexity, despite reducing the pilot overhead needed for Channel Estimation (CE). Furthermore, we conceive training-aided Soft-Decision (SD) detection. We first analyze the complexity of the conventional joint CE and SD detection followed by proposing our reduced-complexity learning-aided joint CE and SD detection. 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.**

***Index Terms*—Index Modulation, Compressed Sensing-Aided Multi-Dimensional Index Modulation (CS-MIM), joint channel estimation and detection, Soft Detection, Machine learning, Neural Network.**

NOMENCLATURE

Acronym

Approx-Log-MAP	Approximate Logarithm MAP
BER	Bit-Error Ratio
CE	Channel Estimation
CP	Cyclic Prefix
CS	Compressed Sensing
CS-MIM	CS-aided MIM
CS-SIM-OFDM	Compressed Sensing-aided Sparse Index Modulation-Orthogonal Frequency Division Multiplexing
CSI	Channel State Information
DCMC	Discrete-Input Continuous-Output Memoryless Channel
DL	Deep Learning
DNN	Deep Neural Network
FC	Fully-Connected
FD	Frequency Domain
HD	Hard-Decision
IFFT	Inverse Fast Fourier Transform
IM	Index Modulation
JCED	Joint Channel Estimation and data Detection
LLR	Log-Likelihood Ratio
LSCE	Least squared CE
LSTM	Long Short-Term Memory
MIM	Multi-dimensional Index Modulation
MIMO	Multiple-In and Multiple-Out
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
OFDM-IM	OFDM with Index Modulation
RA	receiver antenna
Relu	Rectified linear unit
RF	Radio Frequency
RSC	Recursive Systematic Convolutional
SD	Soft-Decision
SIM	Subcarrier-Index Modulation
SIM-OFDM	Subcarrier-Index Modulated OFDM
SM	Spatial Modulation
SpD	Spatial Domain
STSK	Space-Time Shift Keying
TA	transmit antenna
TD	Time Domain

Symbols

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 in FD
N_v	number of subcarriers 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
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
K	number of activated index in each subcarrier group in VD
Q	number of dispersion matrices for STSK mapping

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: xf2u18,meh,cx1g08,lh@ecs.soton.ac.uk

L	number of constellation realizations for STSK mapping
$\boldsymbol{X}$	STSK codeword
$\boldsymbol{S}$	Space-time symbol block of each subcarrier group in VD
$\boldsymbol{s}$	Space-time symbol sequence of each subcarrier group in VD
$\boldsymbol{A}$	CS measurement matrix
$\boldsymbol{s}^{FD}$	Space-time symbol sequence after CS of each subcarrier in FD
$\boldsymbol{Y}$	Received signal
$\boldsymbol{H}_\alpha$	channel model in FD
$\boldsymbol{I}_{AC}$	selection pattern of active TAs for each subcarrier group
$\bar{\boldsymbol{S}}$	modulated CS-MIM symbol at transmitter
$\boldsymbol{h}_\alpha$	channel model in TD
$\boldsymbol{W}$	Additive White Gaussian noise
$\boldsymbol{S}^{FD}$	Space-time symbol block after CS of each subcarrier in FD
$\boldsymbol{I}_{SI}$	subcarrier index selection pattern for each subcarrier group
$\boldsymbol{Y}_p$	Received pilot symbol
$\hat{\boldsymbol{H}}_\alpha$	estimated CSI in the FD
$\bar{\boldsymbol{S}}_{\boldsymbol{p}}$	pilot modulated CS-MIM symbol
$\boldsymbol{R}_{\boldsymbol{H}}$	the channel's correlation matrix of channel $\boldsymbol{H}$
L_e	output extrinsic LLR after soft demodulation
L_a	de-interleaved LLR sequence of L_e
$\hat{\boldsymbol{b}}_r$	detected data bits
$\hat{\boldsymbol{C}}$	detected coded data bits
$\boldsymbol{W}_{\boldsymbol{n}}$	weights of n-th neuron
$\boldsymbol{\theta}_{\boldsymbol{n}}$	bias of n-th neuron
B	sample size of current iteration of DNN training phase

I. Introduction

INDEX Modulation (IM) constitutes a cost- and energy-efficient technique in the face of escalating throughput requirements [1] [2] [3]. The concept of IM has evolved from the idea of space-shift keying proposed by Chau *et al.* [4] in 2001, which maps the information to the indices of the activated Transmit Antennas (TAs). Then, Spatial Modulation (SM) was proposed, which transmits the classic amplitude-phase modulated symbols over the activated TA [5] [6]. To eliminate the influence of Channel State Information (CSI), differential SM is proposed by Bian *et al.* [7]. As a further advance, the concept of IM has been devised by harnessing the philosophy of SM in several single dimensions, which was finally further developed to activating multiple of these dimensions [8] [3].

To elaborate further, the SM is first applied to Orthogonal Frequency Division Multiplexing (OFDM) transmission to avoid inter-channel interference [9]. As Subcarrier-IM combined with OFDM (SIM-OFDM) exploits the IM concept in the Frequency Domain (FD) [10], where extra information can be delivered by the index of the activated subcarriers. Then Wen *et al.* [11] also investigate the IM-aided OFDM (OFDM-IM), which split the whole available OFDM spectrum into groups and Iqbal *et al.* [12] extend OFDM-IM in Multiple-In-Multiple-Out (MIMO) scheme. Although Tsonev *et al.* [13] and Basar *et al.* [14] investigated enhanced OFDM-IM for increasing the spectral efficiency, the presence of inactive subcarriers resulted in throughput reduction compared to classical OFDM. Hence, Zhang *et al.* proposed a novel Compressed Sensing (CS) [15] aided SIM-OFDM [16] for exploiting the sparsity of subcarriers to improve the performance, despite also reducing the detection complexity [17].

As a further evolved arrangement, Space-Time Shift Keying (STSK) is a multi-functional MIMO technique in the IM family that utilizes both the Time Domain (TD) and Spatial Domain (SpD) to strike a flexible diversity vs multiplexing trade-off [18]. The information bits in STSK are used for selecting one or several dispersion matrices from a set of Q dispersion matrices, which spread the signal over T time slots and M TA elements in the SpD. By the careful design of dispersion matrices, an improved Bit Error Ratio (BER), throughput and complexity design trade-off can be struck [19].

Multi-dimensional Index Modulation (MIM) was conceived by Shamasundar *et al.* [20]. This scheme enhances the degrees of freedom in IM designs by exploiting its advantages across multiple domains without requiring additional hardware resources, such as extra Radio Frequency (RF) chains or increased transmission power. As a further development, Zhang *et al.* [16] proposed the concept of Compressed Sensing-aided Sparse Index Modulation-Orthogonal Frequency Division Multiplexing (CS-SIM-OFDM). Briefly, this scheme employs Compressed Sensing (CS) [15] to capitalize on the inherent sparsity of symbols in the Frequency Domain (FD), thereby improving the system's throughput [17]. Furthermore, Lu et al. [21] proposed a method that combines CS techniques with STSK and OFDM-IM. This integrated approach seeks to garner the collective benefits of both STSK and OFDM-IM. Further refinements incorporating SM were discussed in their subsequent treatise [2]. Additionally, Hemadeh *et al.* introduced a multi-functional layered SM paradigm in [3]. This concept aims for maximizing the flexibility in dimension combinations, optimizing the trade-offs among performance, hardware costs, and power consumption.

Since the MIM conveys information in several dimensions, Maximum Likelihood (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 [22]. In [2], CS-aided MIM (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 [23]. Although SM exhibits energy savings 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 [24], 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.* [25] investigated a novel Channel Estimation (CE) scheme by exploiting the channel correlation, which significantly reduced the pilot overhead. Acar *et al.* [26] 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 CE complexity increases with the number of antennas [27].

Historically speaking, Abuthinien *et al.* [28] 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.* [29] 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.* [30] applied JCED in a SM scheme, while Acar *et al.* [31] 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 algorithm investigated by Cozzo *et al.* [32], while employing a reduced number of pilot symbols.

Deep Learning (DL) has been attracting increasing attention in wireless communications [38] [39]. For instance, a Deep Neural Network (DNN) was used for detecting MIMO signals in [40] [41]. In [42], a DNN was employed for estimating the channel of OFDM systems and data detection. Recently, Qing *et al.* [43] 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 Minimum Mean Square Error (MMSE) arrangement. In [35], 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 Soft-Decisions (SD) in [36]. Additionally, both CE and detection have been performed using neural network in [37], 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 Hard-Decision (HD) as well as SD detection.

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

- 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 the paper 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.

II. SYSTEM MODEL

In this section, we introduce the transceiver model of the CS-MIM system employing N_t TAs and N_r Receiver Antennas (RAs). Fig. 1 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 Virtual Domain (VD) [2]. The FD signal is attained by compressing the VD signal using CS as detailed in [2], 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. 1. Then, after transmission over the wireless channel, the receiver estimates the channel and detects the signal. In the following, we present the details of the processing stages at the transmitter and the receiver.

A. Transmitter

As shown in Fig. 1, b bits are split into G groups, where b_g bits, $(g = 1, 2, 3...G)$ of each group are split into three parts

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

[2]VD is the actual domain, where subcarrier index modulation is applied before the CS process as shown in Fig. 1. This concept was firstly introduced in [16] to illustrate the CS techniques in IM system to improve the spectral efficiency.

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

Contribution	proposed*	[16]	[33]	[2]	[34]	[30]	[31]	[35]	[36]	[37]
Index modulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multi-dimensional index modulation	✓			✓					✓	
CS at transmitter	✓	✓		✓						
Channel estimation	✓		✓			✓	✓	✓		✓
Joint channel estimation and detection (JCED)	✓					✓	✓			✓
Learning aided detector	✓		✓		✓			✓	✓	✓
Soft Decision detector	✓		✓				✓		✓	
Learning aided JCED	✓									✓

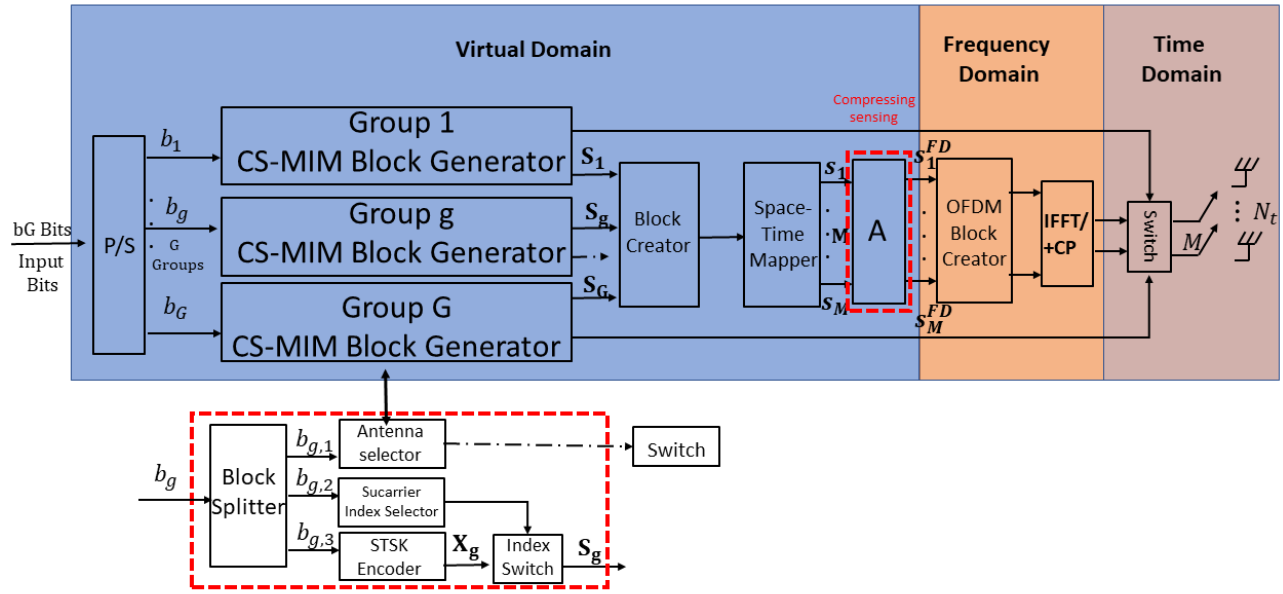

Fig. 1: CS-MIM system transmitter block diagram.

by the block splitter: $b_{g,1}$ bits for SM, $b_{g,2}$ bits for frequency index modulation, $b_{g,3}$ bits for STSK to form the space-time symbols. In the following we detail the different blocks of the CS-MIM transmitter in Fig. 1.

1) Subcarrier Index Selection: The bit sequence $b_{g,2}$ is applied in the subcarrier index selector to activate a subcarrier in each group, as shown in Fig. 1. 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, where we consider the example of $K = 2$ active subcarrier out of the $N_v = 4$ available subcarriers in each group. This results in 4 possible subcarrier index combinations in total. Table II shows an example of the subcarrier selection, where K_1, K_2 represents the active subcarriers and 0 represents the inactive subcarriers. Explicitly, when the input bits sequence is $b_{g,2} = [00]$, the first and second subcarriers are activated, as shown in Table II. Then the selected active subcarrier combination is populated by K space-time symbols, where the STSK codewords are fenerated by the STSK scheme for each group.

b_2	Indices	Allocation
[0 0]	(1)	$[K_1\ K_2\ 0\ 0]$
[0 1]	(2)	$[0\ K_1\ K_2\ 0]$
[1 0]	(3)	$[0\ 0\ K_1\ K_2]$
[1 1]	(4)	$[K_1\ 0\ 0\ K_2]$

TABLE II: An example of subcarrier selection for a CS-MIM system having $K = 2, N_v = 4$

2) STSK Encoding: The bit sequences $b_{g,3}$ of size $K \log_2(QL)$ are fed into the STSK encoder of Fig. 1 to output K STSK codewords $\{\boldsymbol{X}_i, \ldots, \boldsymbol{X}_i, \ldots, \boldsymbol{X}_K\}$, where the dispersion matrix spreads the information both over M TAs and over T time slots in each subcarrier and each space-time codeword $\boldsymbol{X}[i] \in \mathbb{C}^{M \times T}$ is generated by spreading a conventional $\mathcal{L}$-ary constellation symbol by a specific dispersion matrix selected from Q available dispersion matrices. The STSK encoder is characterized by the parameters (M, N, T, Q, L), where M, N, T represent the number of TAs, RAs and time slots, while Q, L are the number of dispersion matrices and of L-ary constellation symbols. Then, the space-time symbol $\boldsymbol{S}$ is generated by mapping the K generated STSK codewords to the K selected active subcarriers decided by the subcarrier index selector, while other subcarriers remain inactive and are set to zero. Considering $b_{g,2} = [00]$ in the example shown in Table II, in this case, we assume STSK$(M, N, T, Q, L) = (2, 2, 2, 2)$, which have codewords $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4\}$. Then, given the assumption that $b_{g,3} = [0\ 0\ 0\ 1]$, the space-time symbols become $\boldsymbol{S} = [\boldsymbol{X_1}, \boldsymbol{X_2}, 0, 0]$, where we assume that $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are the STSK codewords generated based on the bit sequence $b_{g,3}$.

3) Space-Time Symbol Formation and Application of Compressed Sensing: The G groups of space-time symbols $\boldsymbol{S}$ are assembled by the block creator of Fig. 1 to form a long space-time frame, which is processed by the space-time mapper to output a symbol for transmission over multiple TAs and time-slots, as shown in Fig. 1. Equivalently, the space-time symbols $\boldsymbol{S}$ of each subcarrier group are mapped to M TAs during T time slots, which have MT symbol sequences $\{\boldsymbol{s}_{1,1}, ..., \boldsymbol{s}_{M,T}\}$ for spreading the M TA's signals during T time slots.

These symbol sequences $\{\boldsymbol{s}_{1,1}, ..., \boldsymbol{s}_{M,T}\}$ are then compressed by a CS measurement matrix $\boldsymbol{A} \in \mathbb{C}^{N_f \times N_v}$ selected from the N_v-dimensional $\boldsymbol{s}_{m,t} (m = 1, 2, ..., M)(t = 1, 2, 3, \cdots, T)$ in the VD into the N_f-dimensional form in the signal ${s_{m,t}}^{FD}$. The FD vector $\boldsymbol{s}_m^{FD}$ after CS is then mapped to the OFDM subcarriers, which can be written as: $s_{m,t}^{FD} = \boldsymbol{A} s_{m,t}$. Similar to conventional OFDM, the FD symbol

per time slot will be transformed into TD symbols to be transmitted by their corresponding TAs and then a Cyclic Prefix (CP) will be added.

b_1	Indices	Allocation
[0]	(1,2)	$[TA^1TA^2\ 0\ 0]$
[1]	(3,4)	$[0\ 0\ TA^3TA^4]$

TABLE III: A look up table example of Antenna selection in the CS-MIM system having $M = 2, N_t = 4$.

4) Antenna Selection: After Inverse Fast Fourier Transform (IFFT) and CP addition, the TD symbols are transmitted by the activated TAs specified by the antenna selector of Fig. 1. Explicitly, $b_{g,1}$ bits are conveyed by the antenna selector of Fig. 1, which selects M antennas from the N_t available TAs, where we have N_{AC} antenna combinations in total. To avoid the correlation caused by sharing the same TA elements among different antenna combinations, the Distinct Antenna Combination scheme of [44] is used to decide upon the index N_{AC}, with $b_1 = [\log_2(N_{AC})]$. To elaborate further, let us consider an example using $M = 2, N_t = 4$ and $N_{AC} = 2$. As shown in Table III, when the input bit is $b_1 = [0]$, then the first and second TAs will be activated to transmit the modulated symbols in a specific subcarrier block, while the other two TAs remain inactive. Similarly, if the incoming bit sequence is $b_1 = [1]$, then the third and fourth TAs will be selected to transmit the symbols. More specifically, for $b_{g,2} = [0\ 0]$ and $b_{g,3} = [0\ 0]$ along with STSK $(2, 2, 2, 2, 2)$ we can have the space-time block formulated as $\boldsymbol{S} = [\boldsymbol{X_1}, 0, 0, 0]$ and after CS the FD $\boldsymbol{S}^{FD}$ may be expressed as

$$\boldsymbol{S}^{FD} = \begin{bmatrix} \boldsymbol{s}_{1,1}^{FD} & s_{2,1}^{FD} \\ \boldsymbol{s}_{1,2}^{FD} & s_{2,2}^{FD} \end{bmatrix} \tag{1}$$

Then, we assume 4 TAs for transmission and $b_{g,1} = [1]$. As shown in Table III, we can have the CS-MIM modulated symbol $\bar{\boldsymbol{S}}$ formulated as

$$\bar{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{s}_{1,1}^{FD} & s_{2,1}^{FD} \\ \boldsymbol{s}_{1,2}^{FD} & s_{2,2}^{FD} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{2}$$

B. Receiver Structure

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 Fast Fourier Transform (FFT), as shown in Fig. 2. The space-time demapper collects the FD symbols received from N_r RAs over T time slots to recover the space-time symbols, which are then split into G groups by the Block Splitter of Fig. 2. 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 space-time symbol received at the α-th subcarrier of each subcarrier group, respectively.

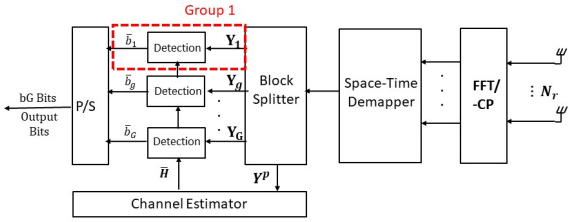

Fig. 2: CS-MIM system receiver block diagram

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 time slots for each subcarrier group can be expressed as [2]

$$\boldsymbol{Y}[\alpha] = \boldsymbol{H}_\alpha \bar{\boldsymbol{S}}[\alpha] + \boldsymbol{W}[\alpha] = \boldsymbol{H}_\alpha \boldsymbol{I}_{AC} \boldsymbol{S}^{FD}[\alpha] + \boldsymbol{W}[\alpha], \tag{3}$$

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 space-time symbols at α subcarriers transmitted from M TAs over T time slots and $\boldsymbol{W}[\alpha] \in \mathbb{C}^{N_r \times T}$ 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. Furthermore, $\boldsymbol{I}_{AC} \in \mathbb{C}^{N_t \times M}$ denotes the $(N_t \times M)$-element sub-matrix, which describes the selection pattern of 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.

III. Channel Estimation and Detection for CS-MIM

Given the received signal $\boldsymbol{Y}$ of (3), the receiver infers the information bits of the STSK codewords, the bits embedded into the activated the 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 time slot, can be represented as

$$\begin{bmatrix} \boldsymbol{Y}_1^\alpha \\ \vdots \\ \boldsymbol{Y}_r^\alpha \\ \vdots \\ \boldsymbol{Y}_{N_r}^\alpha \end{bmatrix} = \begin{bmatrix} h_{1,1}^\alpha & h_{1,2}^\alpha & \cdots & h_{1,N_t}^\alpha \\ h_{2,1}^\alpha & h_{2,2}^\alpha & \cdots & h_{2,N_t}^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r,1}^\alpha & h_{N_r,2}^\alpha & \cdots & h_{N_r,N_t}^\alpha \end{bmatrix} \boldsymbol{I}_{AC} \boldsymbol{S}^{\alpha(FD)} + \boldsymbol{W}^\alpha, \tag{4}$$

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 time slot. 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

$$\boldsymbol{H} = diag\{\boldsymbol{H_1}, \boldsymbol{H_2}, \cdots, \boldsymbol{H}_{N_f}\}, \tag{5}$$

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

$$\bar{\boldsymbol{I}}_{AC} = diag\{\boldsymbol{I_{AC}}, \boldsymbol{I_{AC}}, \cdots, \boldsymbol{I_{AC}}\}. \tag{6}$$

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

$$\boldsymbol{Y} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\boldsymbol{S}^{FD} + \boldsymbol{W}. \tag{7}$$

The FD space-time signal can be represented as

$$\boldsymbol{S^{FD}} = \bar{\boldsymbol{A}}\boldsymbol{S} = \bar{\boldsymbol{A}}\boldsymbol{I_{SI}}\boldsymbol{X}, \tag{8}$$

where $\bar{\boldsymbol{A}} \in \mathbb{C}^{M N_f \times M N_v}$ is the equivalent measurement matrix $\boldsymbol{A}$ used for compressing the VD vector and $\boldsymbol{S} \in \mathbb{C}^{M N_v \times T}$ 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 $\boldsymbol{I_{SI}} \in \mathbb{C}^{M N_v \times MK}$ is the subcarrier index selection pattern.

Hence, (7) can be rewritten as:

$$\boldsymbol{Y} = \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W}. \tag{9}$$

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.

A. Conventional Channel Estimation and 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.

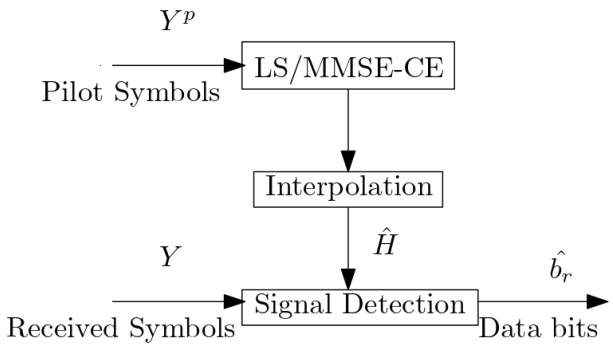

Fig. 3: Conventional CE flow chart.

1) Channel Estimation: As shown in Fig. 2, 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 [26]. 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. 2. 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 [23].

Fig.3 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 (7) as

$$\boldsymbol{Y}_p = \boldsymbol{H}\bar{\boldsymbol{S}}_p + \boldsymbol{W}, \tag{10}$$

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 Least Squared CE (LSCE) is given by

$$\hat{\boldsymbol{H}}_{LS} = \boldsymbol{Y}_p \bar{\boldsymbol{S}}_p^H (\bar{\boldsymbol{S}}_p \bar{\boldsymbol{S}}_p^H)^{-1}. \tag{11}$$

In this case, we can calculate the complexity of LSCE, as shown in (11). 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 Mean Square Error (MSE) of $\boldsymbol{H}$, the popular MMSE-CE formulated as

$$\hat{\boldsymbol{H}}_{MMSE} = \boldsymbol{Y}_p (\bar{\boldsymbol{S}}_p^H \boldsymbol{R}_H \bar{\boldsymbol{S}}_p + N_0 N_f \boldsymbol{I})^{-1} \bar{\boldsymbol{S}}_p^H \boldsymbol{R}_H, \tag{12}$$

where $\boldsymbol{R}_H$ represents the channel's correlation matrix [30]. The MMSE-CE requires the calculation of $\boldsymbol{R}_H$ and CSI matrix inversion. Then, we can characterize the computational complexity as $\mathcal{O}_{MMSE-CE}[N_r N_t T (M N_f^2 + N_f^3)]$.

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

$$\boldsymbol{H_n} = \hat{\boldsymbol{H}}_{n_p} + (\hat{\boldsymbol{H}}_{n_{p+1}} - \hat{\boldsymbol{H}}_{n_p})(\frac{n - n_p}{D}), for\ n_p \leq n \leq n_{p+1}, \tag{13}$$

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.

2) Maximum Likelihood 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

$$\langle \hat{\gamma}, \hat{\beta}, \hat{\varphi} \rangle = \arg \min_{\gamma, \beta, \varphi} \|\boldsymbol{Y} - \boldsymbol{H}\bar{\mathcal{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\bar{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)\|^2, \tag{14}$$

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 [2].

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 $\mathcal{O}_{ML}[(N_rN_tMN_f^3 + N_rM^2N_f^2N_v + N_rN_fM^2N_vK + N_rN_fMKT)N_{AC}N_{SI}(Q\mathcal{L})^K]$. 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 $\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML}$

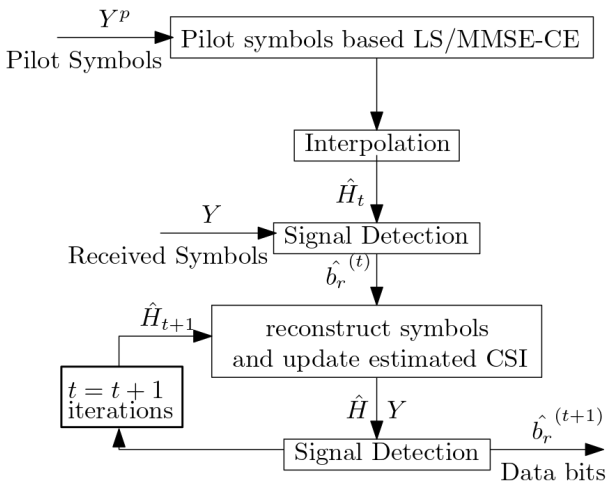

Fig. 4: JCED flow chart.

3) Joint Channel Estimation and Detection: To further improve the detection performance, data detection based iterative JCED is considered. Fig.4 shows the flow chart of the JCED, which starts using the same 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 [29] and [30], the JCED of CS-MIM is described by **Algorithm 1**.

As illustrated in Algorithm 1, there are two thresholds, which are used for terminate 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

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 $\boldsymbol{H}_{MMSE}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)} + N_0N\boldsymbol{I})^{-1}\boldsymbol{S}^{(t)H}\boldsymbol{R}_{\boldsymbol{H}}$
  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

hinges 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 $\mathcal{O}_{HD-JCED} = \mathcal{O}_{LSCE/MMSE-CE} + I_{max}(\mathcal{O}_{LSCE/MMSE-CE} + \mathcal{O}_{ML})$.

B. Soft Decision 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 [45]. 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.

1) Conventional Soft Decision Detection: A channel coded CS-MIM scheme is shown in Fig. 5, which was proposed in [2] 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. 1.

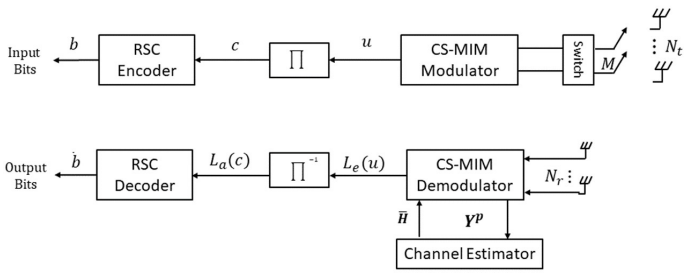

Fig. 5: The transceiver architecture of channel-coded CS-MIM.

At the receiver side of Fig. 5, 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. 5, $L(\cdot)$ 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 (3)is given by [46]

$$p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) = \frac{1}{(\pi \boldsymbol{N}_0)^{NT}} exp(-\frac{||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2}{N_0}), \tag{15}$$

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

$$\begin{aligned} L_e(u_i) &= \ln \frac{p(\boldsymbol{y}|u_i=1)}{p(\boldsymbol{y}|u_i=0)} \\ &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_1^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_0^l} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})}, \end{aligned} \tag{16}$$

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 (15) and (16) 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 [47] is used [48], which is given by

$$\boldsymbol{L}_e(u_l) = \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_1^l}(\lambda_{\gamma,\beta,\varphi}) - \text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi} \in \mathcal{X}_0^l}(\lambda_{\gamma,\beta,\varphi}), \tag{17}$$

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

$$\lambda_{\gamma,\beta,\varphi} = -||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2/N_0. \tag{18}$$

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.(16)(17) 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

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 $\mathcal{O}_{MAP}[c_g(N_rN_tMN_f^3 + N_rM^2N_f^2N_v + N_rN_fM^2N_vK + N_rN_fMKT)N_{AC}N_{SI}(Q\mathcal{L})^K + c_g 2^{c_g}N_{AC}N_{SI}(Q\mathcal{L})^K]$

2) Soft Decision Joint Channel Estimation and Detection: 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 follow-

ing, we propose DNN-based MIM detectors for reducing the complexity.

C. 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.

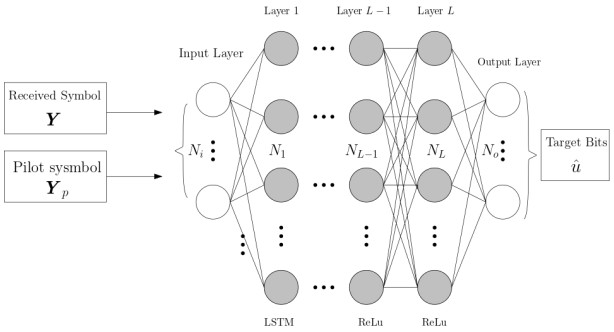

Fig. 6: Fully-connected DNN model for CS-MIM channel estimation and data detection.

1) Conventional DNN-aided CE and HD detection: The DNN architecture of Fig. 6 can be harnessed for replacing the conventional HD data detector of Section III-A2). As shown in Fig. 6, 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

$$\begin{aligned}\hat{u} =& f_{sigmoid}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{Y})+\boldsymbol{\theta_1}])\\ &+\boldsymbol{\theta_2}\}+...+\boldsymbol{\theta}_n),\end{aligned} \tag{19}$$

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 nonlinear relationships between the transmitted signals and the CSI. The LSTM layer can be mathematically represented as

$$\{C_k, z_k\} = \text{LSTM}(C_{k-1}, z_{k-1}, x_k; \phi_{k-1}), \tag{20}$$

where C_k is commonly referred to as the cell state [49], 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 (19), 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 $\mathcal{O}[n_i n_1] + L\mathcal{O}[n_l n_{l-1}] + \mathcal{O}[n_L n_o]$ [36], 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.6 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

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B}\sum_{i=1}^{B} \|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{21}$$

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.

2) Separate DNN-aided CE and detection: To further reduce the effect of CE error, we propose the two-part DNN models of Fig. 7 and Fig. 8 for CE and detection, respectively. Firstly, the fully connected NN of Fig. 7 is used for estimating the channel using the current received symbol $\boldsymbol{Y_{\tau-1}}$ and next received a symbol $\boldsymbol{Y_\tau}$ 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 pilot overheads are considered in the training phase to enhance

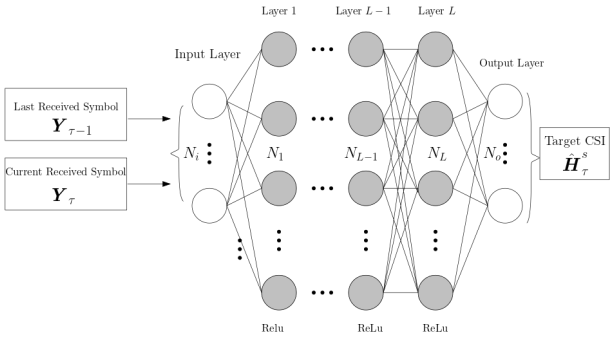

Fig. 7: Separate fully-connected DNN model in CE of CS-MIM systems

CE performance of the trained model in diverse channel conditions. Then we can obtain the output of the DNN-aided CE as

$$\begin{aligned}\hat{H}^s_{\tau} =&\boldsymbol{W}_{N_1}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1^1}(\boldsymbol{Y}_{\tau})+\boldsymbol{b}_1]) \\ &+\boldsymbol{b_2}\}+...+\boldsymbol{b}_{N_1}).\end{aligned} \tag{22}$$

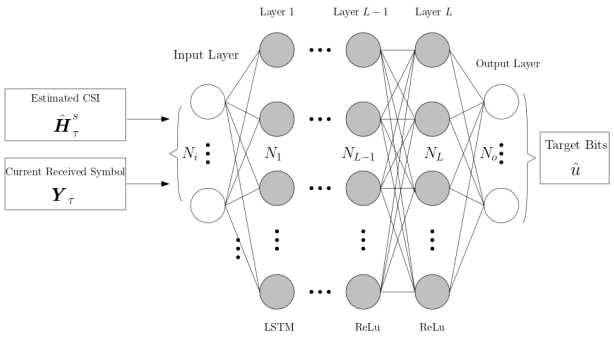

Fig. 8: Separate fully-connected DNN model for detection in CS-MIM systems

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 $\hat{\boldsymbol{H}}^S_{\tau}$ of antenna S and current real CSI $\boldsymbol{H}^S_{\tau}$. In this case, the MSE loss function used for the training is

$$\mathcal{L}(\boldsymbol{H}^s_{\tau},\hat{\boldsymbol{H}}^s_{\tau};\boldsymbol{W_n},\boldsymbol{\theta_n})=\frac{1}{B}\sum_{i=1}^{B}\|\hat{\boldsymbol{H}}^s_{\tau}-\boldsymbol{H}^s_{\tau}\|^2, \tag{23}$$

where B is the sample size of the current iteration.

Fig. 8 shows the DNN employed for detection, which is performed after completing the CSI estimation using the DNN of Fig. 7. The output of the first DNN model of Fig. 7, which is the CSI $\boldsymbol{H}^s_{\tau}$ 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. 8 corresponds to the output bits $\hat{u}$, which is formulated as:

$$\begin{aligned}\hat{u} =&f_{sigmoid}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{Y})+\boldsymbol{\theta_1}]) \\ &+\boldsymbol{\theta_2}\}+...+\boldsymbol{\theta}_n).\end{aligned} \tag{24}$$

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

$$\mathcal{L}(\boldsymbol{u},\hat{\boldsymbol{u}};\boldsymbol{W_n},\boldsymbol{\theta_n})=\frac{1}{B}\sum_{i=1}^{B}\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|^2, \tag{25}$$

where B is the sample size of the current iteration.

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. 7 and Fig. 8. The proposed DNN architecture is shown in Fig. 9. 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 $\boldsymbol{Y}_p$

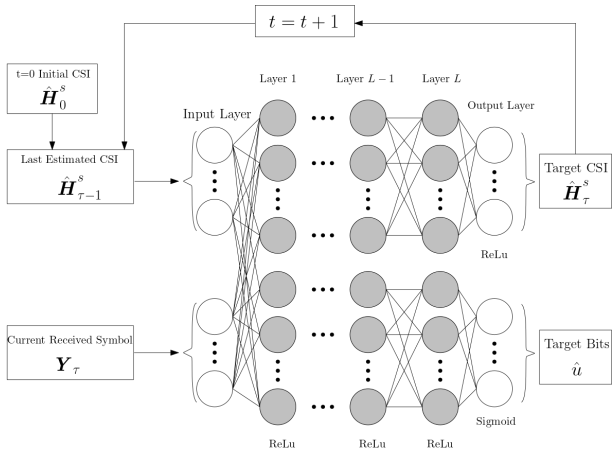

Fig. 9: Fully-connected DNN model for CS-MIM JCED system

As shown in Fig. 9, 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 $\hat{\boldsymbol{H}}^s_{\tau}$, 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. 9, the proposed DNN 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. 9, the DNN will update the CSI using the trained weights of each layer.

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 $\hat{\boldsymbol{H}}_\tau^s$

$t := 0$ // initial iteration index
for $\tau = 1, 2, \cdots, T$ **do**
  Input the $\boldsymbol{Y}_\tau$ and $\hat{\boldsymbol{H}}_{\tau-1}$ to the proposed DNN model
  Use the DNN model with the trained weights to obtain the detected bits $\hat{u}$ and updated estimated CSI, $\hat{\boldsymbol{H}}_\tau^s$
  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 $\boldsymbol{H}_{MMSE}^{(t+1)} = \boldsymbol{Y}(\boldsymbol{S}^{(t)} + N_0 N\boldsymbol{I})^{-1}\boldsymbol{S}^{(t)H}\boldsymbol{R}_{\boldsymbol{H}}$
  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.
end

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

$$\begin{aligned}\hat{H} =&\boldsymbol{W}_{N_1}^1 ... f_{Relu}\{\boldsymbol{W}_{\boldsymbol{2}}^{\boldsymbol{1}}(f_{Relu_1}[\boldsymbol{W}_{\boldsymbol{1}}^{\boldsymbol{1}}(\boldsymbol{H}_{\tau-1}) + \boldsymbol{b}_1^1]) \\ &+ \boldsymbol{b}_2^1\} + ... + \boldsymbol{b}_{N_1}^1),\end{aligned} \tag{26}$$

$$\begin{aligned}\hat{u} =&f_{sigmoid}(\boldsymbol{W}_{N_2}^2 ... f_{Relu}\{\boldsymbol{W}_2^2(f_{Relu_1}[\boldsymbol{W}_1^2(\boldsymbol{Y}_\tau) + \boldsymbol{b}_1^2]) \\ &+ \boldsymbol{b}_2^2\} + ... + \boldsymbol{b}_{N_2}^2),\end{aligned} \tag{27}$$

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 $\theta^1 = \{\boldsymbol{W}_1^1, \boldsymbol{b}_1^1, \boldsymbol{W}_2^1, \boldsymbol{b}_2^1, \cdots, \boldsymbol{W}_{N_1}^1, \boldsymbol{b}_{N_1}^1\}$ and $\theta^2 = \{\boldsymbol{W}_1^2, \boldsymbol{b}_1^2, \boldsymbol{W}_2^2, \boldsymbol{b}_2^2, \cdots, \boldsymbol{W}_{N_2}^2, \boldsymbol{b}_{N_2}^2\}$

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

$$\mathcal{L}(\boldsymbol{u}, \hat{\boldsymbol{u}}; \boldsymbol{W_n}, \boldsymbol{\theta_n}) = \frac{1}{B}\sum_{i=1}^{B}\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|^2, \tag{28}$$

where u represents the target labels, $\hat{u}$ denotes the detected bits and B is the sample size of the current iteration. Using (19) and (27), we can obtain the loss function of this DNN model as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}^1, \boldsymbol{\theta}^2) = &\frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{H}}_\tau^s - \boldsymbol{H}_\tau^s\|_2^2 \\ &+ \frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{u}}_\tau - \boldsymbol{u}_\tau\|_2^2.\end{aligned} \tag{29}$$

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 [36], 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

$$\begin{aligned}\hat{\boldsymbol{H}} =&\boldsymbol{W}_{N_1}^1 ... f_{Relu}\{\boldsymbol{W}_{\boldsymbol{2}}^{\boldsymbol{1}}(f_{Relu_1}[\boldsymbol{W}_{\boldsymbol{1}}^{\boldsymbol{1}}(\boldsymbol{H}_{\tau-1}) + \boldsymbol{b}_1^1]) \\ &+ \boldsymbol{b}_2^1\} + ... + \boldsymbol{b}_{N_1}^1),\end{aligned} \tag{30}$$

$$\begin{aligned}\hat{\boldsymbol{L}}_e =&\boldsymbol{W}_{N_2}^2 ... f_{Relu}\{\boldsymbol{W}_2^2(f_{Relu_1}[\boldsymbol{W}_1^2(\boldsymbol{Y}_\tau) + \boldsymbol{b}_1^2]) \\ &+ \boldsymbol{b}_2^2 + ... + \boldsymbol{b}_{N_2}^2),\end{aligned} \tag{31}$$

and the corresponding loss function is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}^1, \boldsymbol{\theta}^2) = &\frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{H}}_\tau^s - \boldsymbol{H}_\tau^s\|_2^2 \\ &+ \frac{1}{BT}\sum_{i=1}^{B}\sum_{t=1}^{T}\|\hat{\boldsymbol{L}}_e(\tau) - \boldsymbol{L}_e(\tau)\|_2^2.\end{aligned} \tag{32}$$

IV. Simulation results and 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 IV and the parameters used by the learning models outlined in Table VI, 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.
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.
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 IV: 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 V: 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 VI: 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. 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 IV, 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. 11 shows the BER of the **Scheme 1** and **Scheme 2**. Besides, Fig. 11 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. 10 which is $R_t = 1.333$ bits/sec/Hz. However, in more realistic situation, pilot required to deploy CE techniques and cause pilot overhead. Generally, in the simulation, pilot symbols are designed and applied. Then, 1% pilot overhead indicates that every 100 symbols require 1 pilot symbol. As shown in Fig. 10 and Fig. 11, **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 discrepance 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.

We also analysis the detection performance of the system with 4 TAs. Fig. 12 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.13. 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

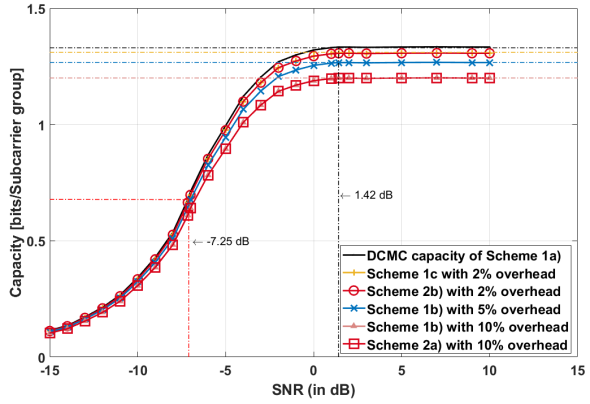

Fig. 10: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

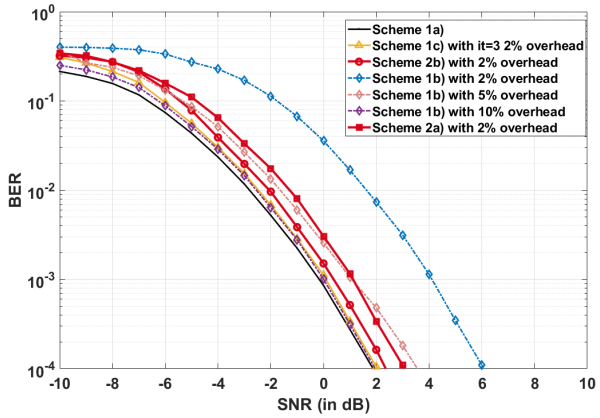

Fig. 11: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

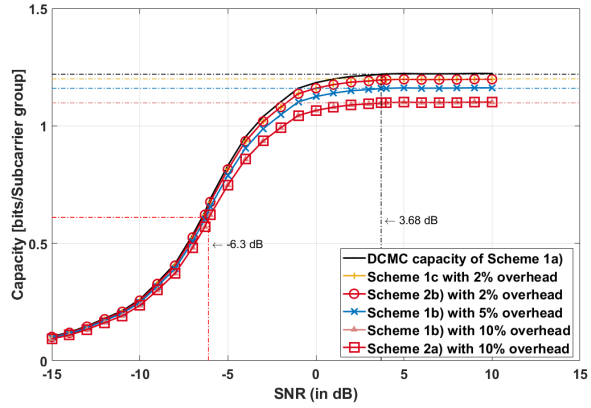

Fig. 12: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

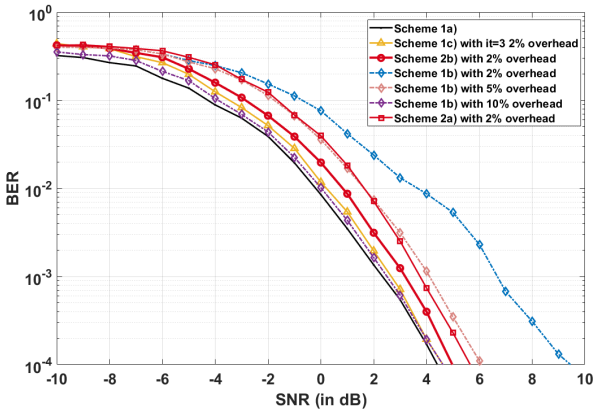

Fig. 13: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

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. 13.

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.

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

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. 10, **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. 14 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. 11. **Scheme 1a** maintains consistent results as observed in Fig. 10, 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. 11. 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. 11. 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. 11. This suggests that **Scheme 1c** and **Scheme 2b** offer some resilience against rapidly varying channels. As illustrated in Fig. 15 with $f_m = 10^{-5}$, **Scheme 2a** is 2.6 dB inferior to **Scheme 1b** in Fig. 11, while **Scheme 2b** lags by 1.2 dB compared to **Scheme 1c** in Fig. 11.

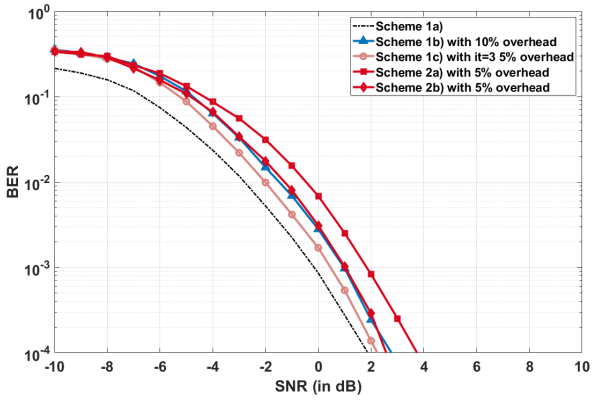

Fig. 14: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 2 \times 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

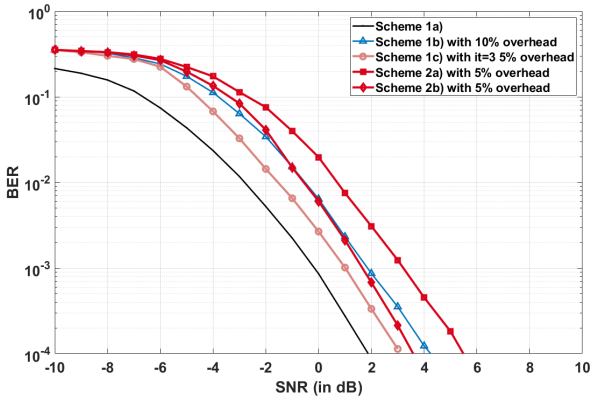

Fig. 15: BER performance comparison of HD detector of **Schemes 1,2** with $N_t = 8$ under $f_m = 10^{-5}$. Our simulation parameter are shown in Table IV and Table VI.

Let us now consider the performance of SD detection, where we employ a half-rate RSC encoder as shown in Table IV. 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. 18, **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 3c)** 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 3c)**. 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.

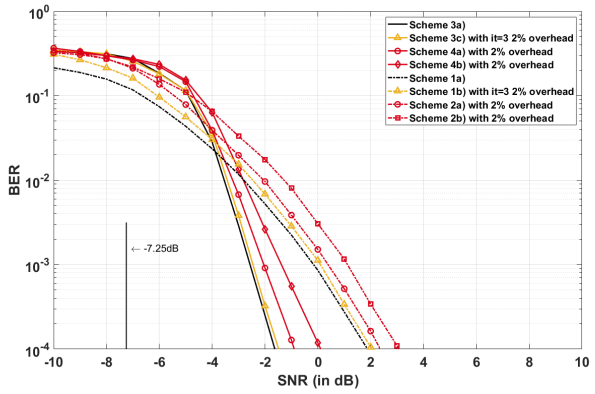

Fig. 16: BER performance comparison of HD and SD detector of **Schemes 1-4** with $N_t = 8$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

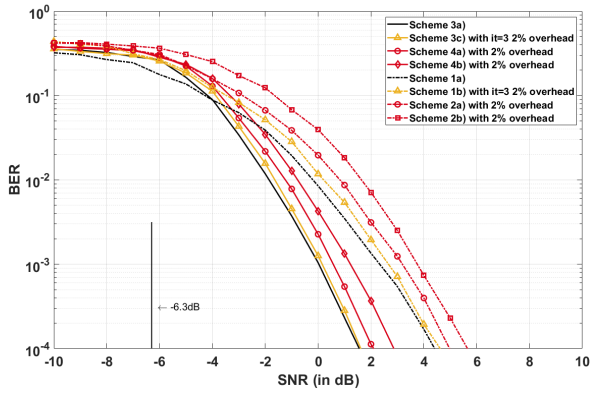

Fig. 17: BER performance comparison of HD and SD detector of **Schemes 1-4** with $N_t = 4$ under $f_m = 10^{-6}$. Our simulation parameter are shown in Table IV and Table VI.

In Fig. 16 and Fig. 17 , we compare the performance of the HD and SD. In Fig.16 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.17, 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.16.

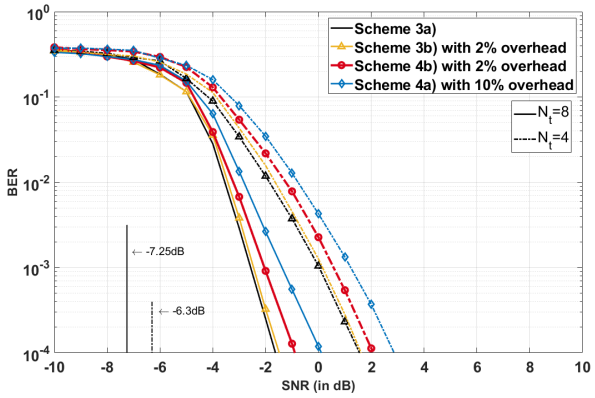

Fig. 18: BER performance comparison of SD detector with **Scheme 3,4** under $f_m = 10^{-6}$. The simulation parameter are shown in Table IV and VI

Fig. 18 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 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 $\mathcal{O}_{MAP}[c_g(N_rN_tMN_f^3 + N_rM^2N_f^2N_v + N_rN_fM^2N_vK + N_rN_fMKT)N_{AC}N_{SI}(Q\mathcal{L})^K + c_g2^{c_g}N_{AC}N_{SI}(Q\mathcal{L})^K]$ for the conventional scheme, where I_{it} denotes the number of iterations.

V. Conclusion

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.

References

[1] 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, 2017.

[2] 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.

[3] I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Hierarchical multi-functional layered spatial modulation," *IEEE Access*, vol. 6, pp. 9492–9533, 2018.

[4] Y. Chau and S.-H. Yu, "Space modulation on wireless fading channels," vol. 3, pp. 1668–1671 vol.3, 2001.

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

[6] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: from classic rf to visible light communications and data storage," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1905–1938, 2018.

[7] Y. Bian, X. Cheng, M. Wen, L. Yang, H. V. Poor, and B. Jiao, "Differential spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3262–3268, 2015.

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

[9] S. Ganesan, R. Mesleh, H. Ho, C. W. Ahn, and S. Yun, "On the performance of spatial modulation ofdm," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 1825–1829.

[10] R. Abu-alhiga and H. Haas, "Subcarrier-index modulation ofdm," pp. 177–181, 2009.

[11] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the achievable rate of ofdm with index modulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1919–1932, 2016.

[12] Z. Iqbal, F. Ji, J. Li, M. Wen, and X. Qi, "Multiple-input multiple-output multiple-mode ofdm with index modulation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13 441–13 446, 2022.

[13] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (sim) ofdm," pp. 728–732, 2011.

[14] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[15] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, apr 2006.

[16] 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.

[17] 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.

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

[19] 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.

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

[21] 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.

[22] 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.

[23] M. Ergen, S. Coleri, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems ," *IEEE TRANSACTIONS ON BROADCASTING*, vol. 48, no. 3, p. 223, 2002.

[24] 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.

[25] 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.

[26] Y. Acar, H. Dogan, and E. Panayırcı, "Channel estimation for spatial modulation orthogonal frequency division multiplexing systems," pp. 382–385, 2015.

[27] 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.

[28] 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.

[29] 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.

[30] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 805–808, 2012.

[31] 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.

[32] 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.

[33] 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.

[34] J. Liu and H. Lu, "Imnet: A learning based detector for index modulation aided mimo-ofdm systems," pp. 1–6, 2020.

[35] S. Katla, L. Xiang, Y. Zhang, M. El-Hajjar, A. A. Mourad, and L. Hanzo, "Deep learning assisted detection for index modulation aided mmWave systems," *IEEE Access*, vol. 8, pp. 202 738–202 754, 2020.

[36] 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.

[37] 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.

[38] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, dec 2017, pp. 563–575.

[39] 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.

[40] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, vol. 2017-July, 12 2017, pp. 1–5.

[41] N. Samuel and T. Diskin, "Learning to detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, May 2019.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning."

[47] 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.

[48] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *Electronics Letters*, vol. 32, no. 10, pp. 887–888, 1996.

[49] 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.