Deep Learning-Based Soft Iterative-Detection of Channel-Coded 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*—The concept of Index Modulation (IM) has been actively researched as a benefit of its flexible trade-off between performance, achievable rate, energy efficiency, hardware cost and complexity. In order to fully exploit its the degrees of freedom, the concept of Multi-dimensional IM (MIM) has been developed in literature, where Compressed Sensing (CS) is often utilized to exploit the sparsity of the multi-dimensional transmitted signals. However, this flexibility and performance gains are attained at the cost of a substantially increased detection complexity. In this paper, we propose Deep Learning (DL) based detection for CS-aided MIM (CS-MIM), where both Hard-Decision (HD) as well as Soft-Decision (SD) detection combined with iterative decoding are conceived. More explicitly, firstly, we propose learning aided hard and soft detection for CS-MIM. Secondly, two novel neural network aided methods are 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 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 Channel Estimation (CE), which results in an improved transmission rate. Explicitly, we develop an advanced DL architecture for blind-detection-aided MIM for the first time in the open literature, where the HD and SD CS algorithms are implemented by learning without the need for CE. 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.**

***Index Terms*—Index Modulation, Compressed Sensing-Aided Multi-Dimensional Index Modulation (CS-MIM), Soft Detection, Machine learning, Neural Network, Iterative Detection, Blind Detection**

I. INTRODUCTION

DURING the past two decades, there has been a tremendous increase in the number of connected users and devices, which results in escalating data rate requirements. Against this background, Multiple-In-Multiple-Output (MIMO) systems have been proposed to provide beneficial throughput gains [1]. As a further advance, the concept of massive MIMO systems has been conceived, which also leads to the new challenge of acquiring Channel State Information (CSI) for an excessive number of antennas [2]. As the next evolutionary stage, the concept of Spatial Modulation (SM) was proposed as a low-complexity MIMO design alternative that can provide an improved trade-off between performance and complexity by relying on a single RF chain [3]–[5]. In recent years, the SM philosophy has been extended to the frequency and time dimensions as well as to multiple domains under the umbrella of Index Modulation (IM) [6], [7]. Indeed, IM schemes are considered as energy-efficient candidates for next-generation wireless systems as a benefit of their flexible resource activation [5].

As a member of the IM family, Space-Time Shift Keying (STSK) was proposed as a multi-functional MIMO technique, which is capable of striking a flexible diversity versus multiplexing gain trade-off without encountering MIMO's inter-antenna interference [8]. Briefly, in STSK, Q Dispersion Matrices (DM) are invoked, which spread the signal over T Time Slots (TSs) and M Transmit Antenna (TA) elements in the spatial and time domains. The IM design activates one of the Q DMs for transmission, where the activation pattern conveys extra IM bits. By appropriately adjusting these parameters, improved Bit Error Ratio (BER), throughput and complexity trade-offs can be optimized [9].

On the other hand, Subcarrier-IM combined with Orthogonal Frequency Division Multiplexing (SIM-OFDM) exploits the IM concept in the Frequency Domain (FD). In this scheme, a subset of subcarriers is activated for modulated signal transmission, where the IM-enabled subcarrier activation pattern conveys extra information bits [17]. As a benefit, the effective signal power of the activated subcarriers in the FD is increased, so that the same time domain signal power is retained after IFFT. This results in a higher SNR for the modulated symbols without requiring extra radiated power. Although Tsonev *et al.* [18] proposed enhanced SIM-OFDM and a novel IM-aided OFDM (OFDM-IM) scheme [19] for increasing the spectral efficiency, SIM-OFDM has been shown to suffer from significant throughput reduction compared to the classic OFDM. Hence, Zhang *et al.* [10] proposed the improved SIM-OFDM concept relying on Compressed Sensing (CS) [20], which beneficially exploits the sparsity of symbols in the FD to improve the throughput [21].

In order to improve the degree of freedom in IM, an advanced Multi-dimensional IM (MIM) scheme was proposed by Shamasundar *et al.* [22], which is capable of simultaneously

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@soton.ac.uk

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/X01228X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

exploiting the benefits of the IM concept in several domains without imposing any extra hardware costs, such as extra RF chains or transmission power. Furthermore, Lu *et al.* [23] proposed to combine CS techniques with STSK and OFDM-IM. This MIM system inherits the benefits of both STSK and OFDM-IM. As a further advance, SM was also integrated into this MIM scheme for TA selection [12]. In [24], the concept of multi-functional layered SM was proposed by Hemadeh *et al.*, which offers flexible trade-offs in terms of performance, hardware cost and power dissipation. *However, the increased number of IM dimensions leads to substantially increased signal detection complexity at the receiver side.*

The classic Maximum Likelihood (ML) detection has been widely employed for IM systems, which however suffers from an escalating complexity upon increasing the degrees of freedom or dimensions [25]. 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 [26]. Moreover, both the pilot overhead as well as the associated Channel Estimation (CE) complexity increases with the number of antennas [27] [28]. As a result, the excessive complexity of CE-aided ML detection of MIM schemes becomes unrealistic in practice. Furthermore, the Shannonian capacity can only be approached, when Soft-Decision (SD) detection is used together with the powerful tool of channel coding using Iterative Soft Detection (ISD) [29]. With the guidelines of ISD design in [30], Xu *et al.* discuss the trade-off between complexity and performance in the context of the model-based MIM solutions of [31] and [12]. In [23], 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.

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 [32], namely in channel decoding [33], networking [34], [35], and millimeter wave communications (mmWave) [36]. Deep Learning (DL) has also been shown to be efficient in signal detection [37], especially in MIMO detection [38]. For instance, a Deep Neural Network (DNN) was used for detecting MIMO signals in [39], while Samuel and Diskin [40] 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 [41].

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

Contribution	proposed*	[10]	[11]	[12]	[13]	[14]	[15]	[16]
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	✓							

TABLE II: NOMENCLATURE

Index Modulation	IM	Multi-dimensional IM	MIM
Compressed Sensing	CS	Deep Learning	DL
Compressed Sensing-aided MIM	CS-MIM	Hard-Decision	HD
Soft-Decision	SD	Iterative Soft Detection	ISD
Channel State Information	CSI	Multiple-In-Multiple-Output	MIMO
Spatial Modulation	SM	Space-Time Shift Keying	STSK
Dispersion Matrices	DM	Time Slot	TS
Channel Estimation	CE	Bit Error Ratio	BER
Subcarrier-IM combined with Orthogonal Frequency Division Multiplexing			SIM-OFDM
IM-aided OFDM	OFDM-IM	Maximum Likelihood	ML
Deep Neural Network	DNN	Convolutional Neural Networks	CNN
Transmit Antennas	TA	Receive Antennas	RA
Virtual Domain	VD	Frequency Domain	FD
Subcarrier Index	SI	Space-Time	ST
Cyclic Prefix	CP	Inverse Fast Fourier Transform	IFFT
Radio Frequency	RF	Fast Fourier Transform	FFT
Additive white Gaussian Noise	AWGN	Log Likelihood Ratios	LLR
Simultaneous Matching Pursuit	S-MP	Least Squared CE	LSCE
Mean Square Error	MSE	Minimum MSE	MMSE
Reduced-Complexity	RC	Long Short-Term Memory	LSTM
Maximum A Posteriori	MAP	Approximate Log-MAP	Approx-Log-MAP
Maximum Log-MAP	Max-Log-MAP	Long Short-Term Memory	LSTM
Discrete-Input Continuous-Output Memoryless Channels			DCMC
Root MSE	RMSE	MMSE aided CE	MMSE-CE

TABLE III: Table for symbols used in this paper

Symbol	meaning	Symbol	meaning
N_r	number of RAs	N_t	number of TAs
G	number of group	N_c	number of subcarrier for each frame
N_v	number of subcarrier in VD	N_f	number of subcarrier in FD
K	number of activated index	$\boldsymbol{X}[i]$	STSK codewords
$\boldsymbol{X}_t^m[k]$	STSK slice	$\boldsymbol{S}$	ST matrix
$\boldsymbol{s_a}$	ST vector in one TS for single TS	$\boldsymbol{A}$	measurement matrix
$\boldsymbol{s_m^{FD}}$	compressed FD ST vector of single TS	$\boldsymbol{Y}[\alpha]$	Received signal of each subcarrier
$\boldsymbol{H}[\alpha]$	CSI of each subcarrier	$\boldsymbol{S^{FD}}[\alpha]$	compressed FD symbol
$\boldsymbol{I_{AC}}$	sub matrix of antenna index selection pattern		
$\boldsymbol{I_{SI}}$	sub matrix of suncarrier index selection pattern		
$\mathcal{I}_{AC}(\gamma)$	possible realizations of the active TA indices		
$\mathcal{I}_{SI}(\beta)$	possible realizations of the active subcarrier indices		
$\mathcal{X}_{q,l}(\varphi)$	possible realizations of STSK codes blocks		
$\boldsymbol{Y}_p$	Received signal of transmitted pilot symbol of each group		
$\boldsymbol{S}_p$	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
z_k	output of LSTM	x_k	input of LSTM
ϕ_k	parameter of LSTM	$\boldsymbol{W}_n$	weight of each neuron
$\boldsymbol{\theta}_n$	bias of each neuron	$L_e(u_i)$	LLR information if bits
$\mathcal{X}_{\gamma,\beta,\varphi}$	realization of STSK codewords	$\lambda_{\gamma,\beta,\varphi}$	intrinsic metrix of STSK realization

In [42], a DNN model was also harnessed for estimating an OFDM channel. Learning has also been used for channel decoding by Al-Baidhani and Fan [43]. Additionally, Xu *et al* [44] 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 [45], a DNN based model is proposed for detecting the OFDM-IM signal by Luong *et al.* The authors of [46] and [47] investigated Convolutional Neural Networks (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 [15] and for multi-set STSK in [16]. However, the authors of [16] only investigated the combination of basic SD 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 I boldly contrasts the novelty of this paper to the literature. Explicitly, the contributions of this paper are boldly and explicitly contrasted to the state-of-the-art as follows:

1) 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.

The rest of the paper is organized as follows. In Section II, the system model of CS-MIM is presented. In Section III, we characterize the conventional HD and SD detectors relying on the proposed learning-aided detector. Finally, our simulation results are provided in Section IV, while our conclusions are offered in Section V . Additionally, Table II shows all the abbreviation used in this treatise and Table III defines the symbols used for describing the system.

II. SYSTEM MODEL

Fig. 1 shows the block diagram of the CS-MIM transceiver employing N_t TAs and N_r Receive Antennas (RAs). As shown in Fig. 1, the input bit sequence is split into three parts: STSK in the Virtual Domain (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. 1, 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 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. 1. At the receiver, the signal is transformed back to the subcarrier symbols and each IM group is detected separately. In the following subsections, we present the details of the processing stages both the transmitter and the receiver.

[1]VD is the actual domain for user data mapping/demapping, where subcarrier IM is applied before the CS process as shown in Fig. 1. This concept is firstly introduced in [10] 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.1.

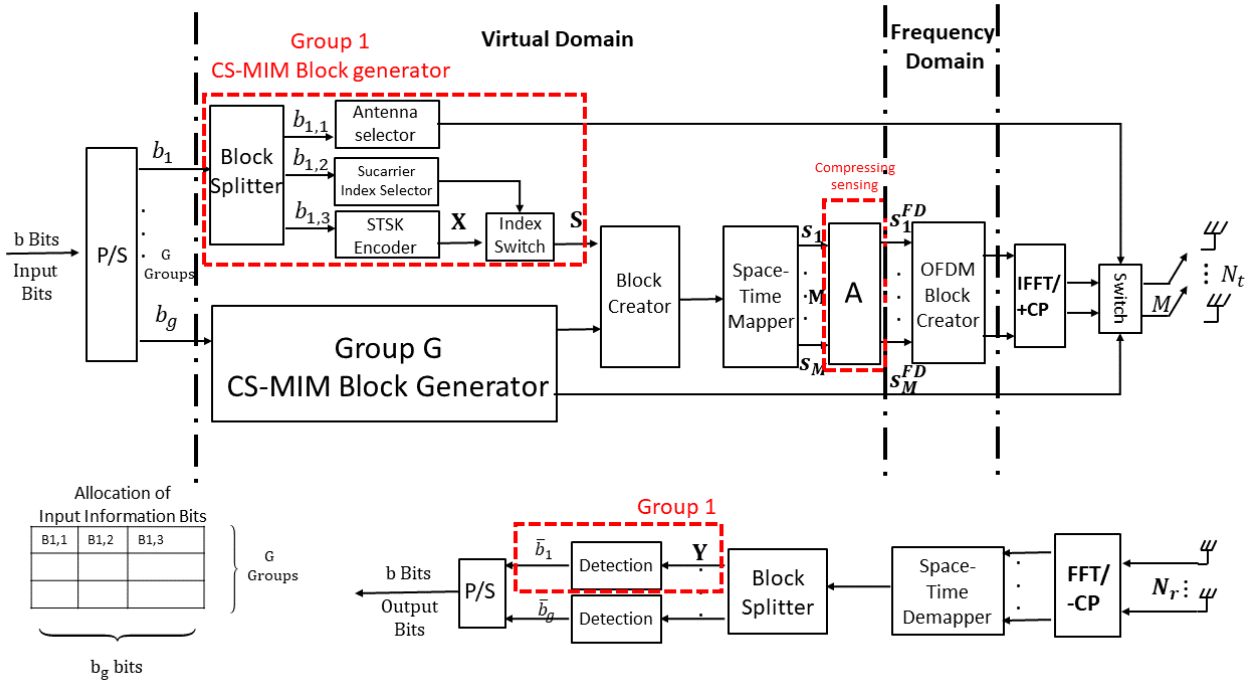


Fig. 1: CS-MIM system block diagram.

A. Transmitter

As shown in Fig. 1, 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. 1. 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-OFDM-STSK system of [12]. In the following we will detail the different blocks in the CS-MIM transmitter of Fig. 1.

1) Subcarrier Index 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. 1. 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 IV 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 IV.

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. 1 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 dispersion matrix spreads the information both over M TAs and over T TSs in each subcarrier. Each Space-Time (ST) codeword $\boldsymbol{X}[i] \in \mathbb{C}^{M \times T}$ of Fig. 1 is generated by spreading a conventional $\mathcal{L}$-ary constellation symbol using a specific dispersion matrix 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. 1.

TABLE IV: 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) 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. 1 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. 2 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 $\boldsymbol{S} = \{[\boldsymbol{0}\ \boldsymbol{X}[1]\ \boldsymbol{0}\ \boldsymbol{0}\}$ 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. 2, 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

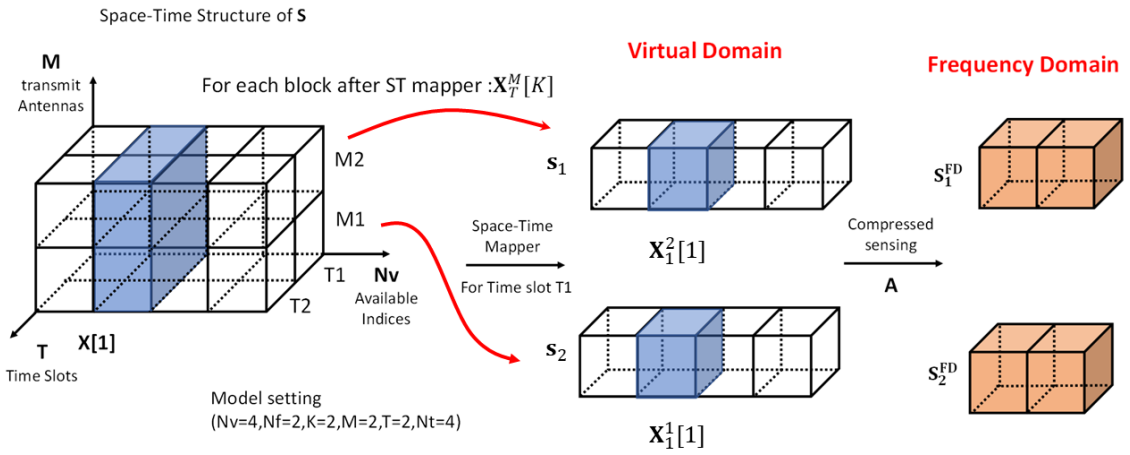


Fig. 2: The process of single subcarrier group STSK symbol $\boldsymbol{S}$ with ST mapper and CS techniques employing $M = 2, T = 2, K = 1, N_v = 4, N_f = 2, N_t = 4$.

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. 2, 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. 2. 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. 2.

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 Cyclic Prefix (CP) is added.

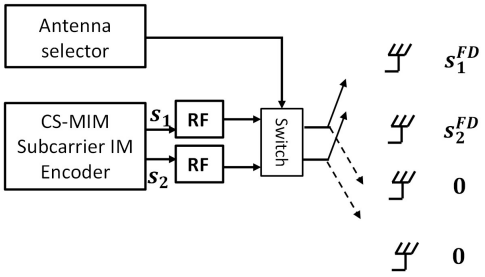


Fig. 3: The process of selecting active TAs for $M = 2$ and $N_t = 4$.

4) Antenna Selection: After Inverse Fast Fourier Transform (IFFT) and CP, the time domain ST symbol is transmitted using M TAs during T TS using a limited number of Radio Frequency (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. 1. 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 V, 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. Similarly, if the incoming bit is $b_{g,1} = [1]$, then the third and fourth antennas are selected to transmit the symbols.

TABLE V: 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)	$[TA^1TA^2\ 0\ 0]$
[1]	(3,4)	$[0\ 0\ TA^3TA^4]$

B. 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 [48]. 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 Fast Fourier Transform (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. 1 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. 1. 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 [12]

$$\boldsymbol{Y}[\alpha] = \boldsymbol{H}_\alpha \boldsymbol{I}_{AC} \boldsymbol{S}^{FD}[\alpha] + \boldsymbol{W}[\alpha], \tag{1}$$

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 Additive white Gaussian Noise (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 submatrix, 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.

III. DETECTION

As shown in Fig. 1, 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 (1) in the following compact form [12]:

$$\boldsymbol{Y} = \boldsymbol{H} \bar{\boldsymbol{I}}_{AC} \bar{\boldsymbol{A}} \boldsymbol{I}_{SI} \boldsymbol{X} + \boldsymbol{W}. \tag{2}$$

where $\bar{\boldsymbol{A}} \in \mathbb{C}^{MN_f \times MN_v}$ is the equivalent measurement matrix $\boldsymbol{A}$ shown in (1) used for compressing the M VD vectors. Similarly, $\bar{\boldsymbol{I}}_{\boldsymbol{AC}} \in \mathbb{C}^{N_r N_f \times N_t N_f}$ denotes the TA selection of N_f subcarriers in each subcarrier group. Observe from 1 and 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 [6]. Furthermore, in case of SD, the received signal is converted into probability values, which are referred to as Log Likelihood Ratios (LLR) [49], 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 reduced-complexity 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 Section III-B, where we first present the conventional SD detectors followed by our proposed non-iterative NN-aided SD and ISD receiver.

A. Hard-Decision Detection

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

1) Maximum Likelihood 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 (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 [12]

$$\langle \hat{\gamma}, \hat{\beta}, \hat{\varphi} \rangle = \arg \min_{\gamma, \beta, \varphi} \|\boldsymbol{Y} - \boldsymbol{H} \bar{\mathcal{I}}_{AC}(\gamma) \bar{\boldsymbol{A}} \bar{\mathcal{I}}_{SI}(\beta) \mathcal{X}_{q,l}(\varphi)\|^2, \tag{3}$$

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 [12]. 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.

2) Channel Estimation: Conventional pilot based CE, which inserts pilots in each symbol can be ineffective due randomly activating both the subcarriers and TAs [50]. 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. 4, 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 [26].

Let us model the received ST pilot symbol based on (1) as

$$\boldsymbol{Y}_p = \boldsymbol{H} \bar{\boldsymbol{I}}_{AC} \boldsymbol{S}_p + \boldsymbol{W}, \tag{4}$$

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.

Under the assumption that the fading envelope remains unchanged over a block of signal transmission, the Least Squared CE (LSCE) and the Minimum Mean Square Error (MMSE) based CE can be applied [51].

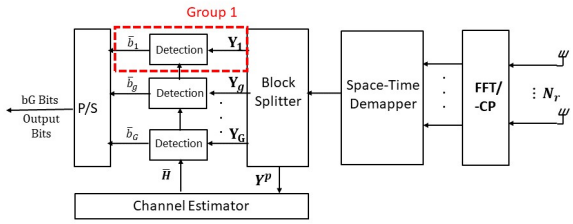


Fig. 4: CS-MIM system receiver block diagram

3) Reduced-Complexity Detection: To reduce the search-space of the ML detector, a Reduced-Complexity (RC) detector was proposed in [12] based on the S-MP scheme. Here we rewrite (2) as:

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W} = \boldsymbol{H}\bar{\boldsymbol{S}} + \boldsymbol{W} \\ &= \bar{\boldsymbol{H}}\boldsymbol{S}^{VD} + \boldsymbol{W} \\ &= \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W},\end{aligned} \tag{5}$$

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 $\mathcal{O}(N_{AC}N_{SI} + Q\mathcal{L}K)$ [12].

4) Approximate Message Passing Algorithm Aided Detection: It has been shown in [52] that the AMP algorithm benefits from low complexity and rapid convergence. Inspired by the application of AMP in OFDM-IM [53] and GSM [54], we apply the AMP detector for the CS-MIM system. We rewrite (2) as

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}\boldsymbol{X} + \boldsymbol{W} \\ &= \boldsymbol{H}\bar{\boldsymbol{I}}_{AC}\boldsymbol{S}^{\boldsymbol{FD}} + \boldsymbol{W}.\end{aligned} \tag{6}$$

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

$$\boldsymbol{Y}^{FD} = \boldsymbol{H}^{FD}\boldsymbol{S}^{\boldsymbol{FD}} + \boldsymbol{W}^{FD}, \tag{7}$$

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

$$\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)} = \boldsymbol{s}_{a,\tau}^{[t](FD)} + \bar{\boldsymbol{H}}_{a,\tau}^{H(FD)}\boldsymbol{z}_{a,\tau}^{[t]}, \tag{8}$$

$$\boldsymbol{s}_{a,\tau}^{[t+1](FD)} = \eta(\check{\boldsymbol{s}}_{\boldsymbol{a,\tau}}^{\boldsymbol{[t](FD)}}; \boldsymbol{\sigma_{a,\tau,t}}), \tag{9}$$

$$\begin{aligned}\boldsymbol{z}_{a,\tau}^{[t+1]} &= \boldsymbol{y}_{a,\tau}^{FD} - \bar{\boldsymbol{H}}_{a,\tau}^{FD}\boldsymbol{s}_{a,\tau}^{[t+1](FD)} \\ &\quad + \frac{\boldsymbol{z}_{a,\tau}^{[t]}}{2N_f}\nabla\eta_{MMSE}(\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)}; \sigma_{a,\tau,t}),\end{aligned} \tag{10}$$

where $\check{\boldsymbol{s}}_{a,\tau}^{[t](FD)}$ represents the updated FD 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 [53], where σ characterizes the Root MSE (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 $\mathcal{O}[N_{AC} + N_{SI}(Q\mathcal{L})^K]$.

5) Proposed Learning-Aided Hard-Decision 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.

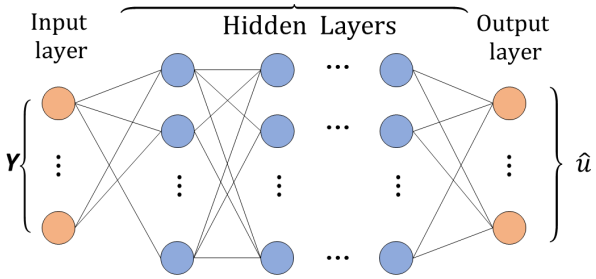


Fig. 5: Fully-connected DNN model for CS-MIM detection system

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. 5. 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 $\hat{u}$. To construct the associated model, we apply an Long Short-Term Memory (LSTM) layer as the first layer to capture the non-linear relationship between the transmitted signal and its observations. This technique is eminently suitable for processing sequential data in DL and can be characterized by

$$\{C_k, z_k\} = LSTM(C_{k-1}, z_{k-1}, x_k; \phi_{k-1}), \tag{11}$$

where C_k is also often termed as the cell state [55] 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 classicsoftmax 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

$$\begin{aligned}\hat{u} =& f_{softmax}(\boldsymbol{W_n}...f_{Relu}\{\boldsymbol{W_2}(f_{Relu_1}[\boldsymbol{W_1}f_{LSTM}(\boldsymbol{y})+\boldsymbol{\theta_1}]) \\ &+\boldsymbol{\theta_2}\}+...+\boldsymbol{\theta}_n).\end{aligned} \tag{12}$$

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

$$\mathcal{L}(\boldsymbol{u},\hat{\boldsymbol{u}};\boldsymbol{W_n},\boldsymbol{\theta_n})=\frac{1}{B}\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|^2, \tag{13}$$

where $\boldsymbol{u}$ represents the target labels, $\hat{\boldsymbol{u}}$ denotes the 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 $\mathcal{O}(n_in_h)+\mathcal{O}(n_h^2)+\mathcal{O}(n_hn_o)$ [16], 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.

B. 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 [56] and has been extensively used in wireless systems [29]. 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 [57]. This motivates the design of sub-optimal soft detectors. In light of this motivation, we propose learning aided SD detection, which is capable of providing near-optimal performance at a reduced complexity.

1) Conventional SD Detection: A CS-MIM ISD scheme is shown in Fig. 6. At the transmitter side of Fig. 6, the information bit sequence b is encoded by a Recursive Systematic Convolutional (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. 1.

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 [58]

$$\begin{aligned}&p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi})\\ &=\frac{1}{(\pi\boldsymbol{N}_0)^{NT}}exp(-\frac{||\boldsymbol{Y}-\boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2}{N_0}),\end{aligned} \tag{14}$$

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 (15) [12]. In (15), $\mathcal{X}_1^l$ 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 $X_1^l\equiv\{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}:u_l=1\}$ and $X_0^l\equiv\{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}:u_l=0\}$. Therefore, the extrinsic LLR of bit $u_l(i=1,2...,B)$ can be expressed by (15). The variable $L_a()$ in (15) represents the *a priori* LLR fed back from the RSC decoder to the demodulator.

The expression in (15) can be further simplified by the Approximate Log Maximum A Posteriori (Approx-Log-MAP) algorithm based on the Jacobian Maximum operation [59] [60] as

$$\boldsymbol{L}_e(u_l)=\text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_1^l}(\lambda_{\gamma,\beta,\varphi})-\text{jac}_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_0^l}(\lambda_{\gamma,\beta,\varphi}), \tag{16}$$

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

$$\begin{aligned}\lambda_{\gamma,\beta,\varphi}=&(-||\boldsymbol{Y}-\boldsymbol{H}\bar{\boldsymbol{I}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{I}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2)/N_0\\ &+\sum_{j\neq l}u_jL_a(u_j).\end{aligned} \tag{17}$$

$$\begin{aligned}\boldsymbol{L}_e(u_l) &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) exp[\sum_{j\neq l} u_j L_a(u_j)]}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_0^l}} p(\boldsymbol{Y}|\mathcal{X}_{\gamma,\beta,\varphi}) exp[\sum_{j\neq l} b_j L_a(b_j)]} \\ &= \ln \frac{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}} exp[-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)\|^2/N_0 + \sum_{j\neq l} u_j L_a(u_j)]}{\sum_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}} exp[-\|\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2/N_0 + \sum_{j\neq l} u_j L_a(u_j)]}.\end{aligned} \tag{15}$$

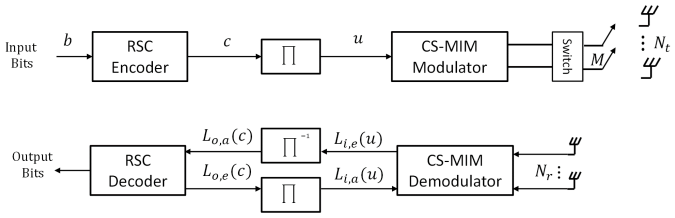


Fig. 6: The transceiver architecture of SD aided CS-MIM employing iterative extrinsic information exchange between the CS-MIM detector and the RSC decoder.

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.

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 III-B1), the extrinsic LLRs $L_{i,e}(u)$ are updated in every iteration by utilizing the *a priori* LLRs $L_{i,a}(u)$. By modifying (16) and (17), the extrinsic LLR of every iteration can be expressed as [12]

$$\boldsymbol{L}_e^n(u_l) = jac_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}}(\lambda^n_{\gamma,\beta,\varphi}) - jac_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_0^l}}(\lambda^n_{\gamma,\beta,\varphi}), \tag{18}$$

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

$$\begin{aligned}\lambda^n_{\gamma,\beta,\varphi} &= (-||\boldsymbol{Y} - \boldsymbol{H}\bar{\boldsymbol{\mathcal{I}}}_{AC}(\gamma)\bar{\boldsymbol{A}}\boldsymbol{\mathcal{I}}_{SI}(\beta)\mathcal{X}_{q,l}(\varphi)||^2)/N_0 \\ &\quad + \sum_{j\neq l} u_j L_a^n(u_j).\end{aligned} \tag{19}$$

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. 7(a), 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 III-A3) 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 [61] formulated as:

$$\begin{aligned}\hat{L}_e(u) =& \boldsymbol{W_n}...f_{Relu}(\boldsymbol{W_2}(f_{Relu}(\boldsymbol{W_1}f_{Relu}(\boldsymbol{y}) + \boldsymbol{\theta_1}) \\ &+ \boldsymbol{\theta_2}) + ... + \boldsymbol{\theta}_n),\end{aligned} \tag{20}$$

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 DNNs and $\hat{L}_e(u)$ represents the estimated extrinsic LLR information at the output of the demodulator.

As discussed in Section III-A3), 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:

$$\mathcal{L}(\boldsymbol{L}_i, \hat{\boldsymbol{L}}_{\boldsymbol{i}}; \boldsymbol{W}_i, \boldsymbol{\theta_i}) = \frac{1}{B}\sum_{i=1}^{S} \|\boldsymbol{L_i} - \hat{\boldsymbol{L}}_i\| + \sum_{i=1} \rho_n \|\boldsymbol{W}_n\|_2^2, \tag{21}$$

where B is the number of training samples, $\boldsymbol{L_i}$ and $\hat{\boldsymbol{L}}_{\boldsymbol{i}}$ represent the target and the predicted LLR, respectively of the i-th training sets and the regularization factors ρ_n are introduced for avoiding a over-fitting [58].

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.

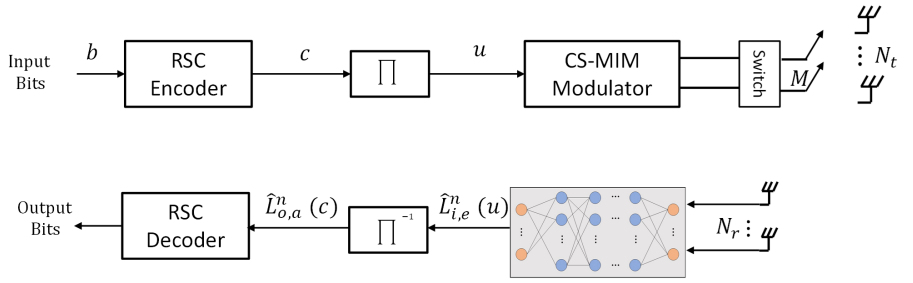
When using iterative extrinsic information exchange between the RSC decoder and our MIMO detector, the LLRs $L_e^n(u_l)(n = 1, 2, 3, \cdots)$ can be typically improved over iterations. We consider to use the updated extrinsic LLR $L_{i,e}^n$ after several iterations as our training target, with the received 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. 7(a) 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.

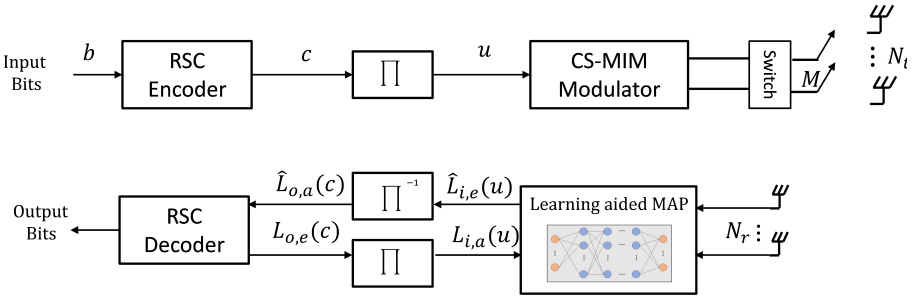
3) Iterative DNN-Aided MAP SD Detection: In this section, we activate and optimize the DNN for each iteration in the ISD assisted channel coded system.

First of all, the Approx-Log-MAP of (16) can be simplified by the Maximum Log-MAP (Max-Log-MAP) algorithm [56]

$$\boldsymbol{L}_e(u_l) = max_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_1^l}}(\lambda_{\gamma,\beta,\varphi}) - max_{\mathcal{X}_{\gamma,\beta,\varphi\in\mathcal{X}_0^l}}(\lambda_{\gamma,\beta,\varphi}). \tag{22}$$



(a) single-stage DNN-aided SD detection



(b) iterative DNN-aided MAP SD detection

Fig. 7: Structure of the learning-aided SD detector of CS-MIM. (a)one stage DNN-aided SD; (b)iterative DNN aided MAP SD.

In order to incorporate both (16) and (22), the extrinsic LLRs can be expressed as:

$$\boldsymbol{L}_e(u_l) = f_{map_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_1^l}}(\lambda_{\gamma,\beta,\varphi}) - f_{map_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_0^l}}(\lambda_{\gamma,\beta,\varphi}), \tag{23}$$

where $f_{map}(.)$ represents the Max-Log-MAP algorithm or the Approx-Log-MAP relying on the simplified Jacobian Maximum operation. Our proposed learning method aims for replacing the $f_{map}(.)$ function, where we use a pair of DNN models for replacing the functions of $f_{map_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_1^l}}$ and $f_{map_{\mathcal{X}_{\gamma,\beta,\varphi}\in\mathcal{X}_0^l}}$.

As shown in Fig. 7(b), the received signal Y and *a priori* LLR $L_{i,a}(u)$ feedback from the RSC decoder after interleaving will become the input data of the DNN models. 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_i, 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 *a priori* LLR $L_{i,a}$ of different iterations and the received signal.

IV. 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 VI. 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-based CE (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 in-

formation 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 VI: 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)
Transmit antennas, N_t	8
Receiver antennas, N_r	8
RSC code, (n, k, K)	(2,1,3)

TABLE VII: Configuration of the system presented in Section III

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	Perfect CSI at receiver	
	AMP aided detector	Perfect CSI at receiver	
3	DNN-aided	Blind	
SD		throughput,R_t	0.6667
4	ML	Perfect CSI at receiver	0
		Imperfect CSI at receiver	0
		Perfect CSI at receiver	1-5
		Imperfect CSI at receiver	1
5	Single-stage DNN-aided	Blind	0
		Blind	1-5
6	DNN-aided MAP detector	Blind	1-3

TABLE VIII: Training configuration for learning-aided method

Setting	Hard-decision	Soft-decision
layers	lstm+fc+fc+softmax	lstm+fc+fc+fc
Activation function	Relu	
output layer	softmax	fc
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	

Given the system parameters of Table VI, the achievable rate is $\frac{b_g G}{N_c+L_{CP}} = 1.333$ bits/sec/Hz. Fig. 8 characterizes the theoretical maximum rate of CS-MIM in Discrete-Input Continuous-Output Memoryless Channels (DCMC) 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. 9 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 VII.

As mentioned in Section II-B, the classic MMSE-CE is applied for estimating the CSI at the receiver. As shown in Fig. 9, 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. 9, 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. 9, 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.9. The DNN method is capable of achieving better performance without overhead compared to both RC methods, as shown in Fig. 9. Furthermore, we quantify the computational complexity order of each method, as shown in Table IX, 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 10, 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. Fig. 10 (a)-(c) 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. 10 (a)-(c). Nonetheless, Fig. 10 (c) demonstrates that when the fading channel varies rapidly, both the coherent and blind-detection-aided schemes suffer from a severe performance loss. Hence SD detection in combination with channel coding is needed to mitigate the channel impairments.

As discussed in Section III-A, 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 [12] 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 that of

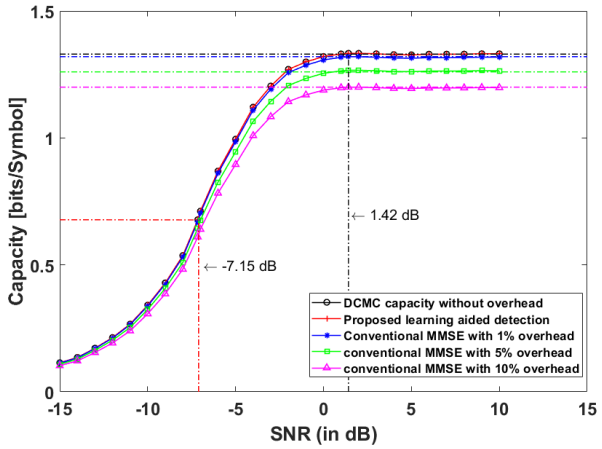


Fig. 8: The Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of the proposed design, and of the conventional design at 1%, 5% and 10% pilot overheads. Our simulation parameter are shown in Table VI and Table VIII.

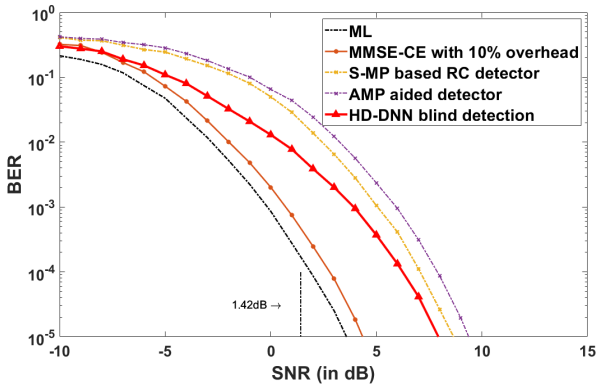


Fig. 9: BER performance comparison of CS-MIM using HD detection which Scheme 1(c), 2, 3 is considered. Our simulation parameter are shown in Table VI and Table VIII.

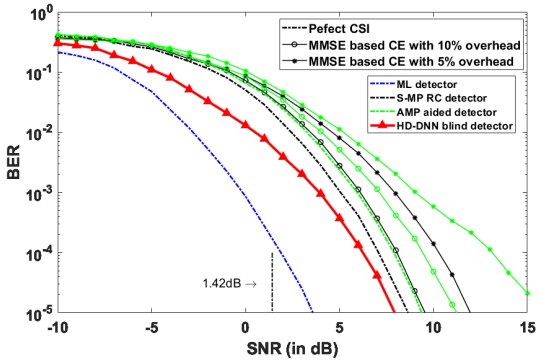
TABLE IX: Comparison between the computational complexities of different HD detection methods of CS-MIM

Detector	Complexity order	Computations
ML	$\mathcal{O}[N_{AC}N_{SI}(Q\mathcal{L})^K]$	8.51×10^6
SM-P	$\mathcal{O}(N_{AC}N_{SI} + Q\mathcal{L}K)$	1.64×10^5
AMP	$\mathcal{O}[N_{AC} + N_{SI}(Q\mathcal{L})^K]$	1.2×10^5
DNN	$\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$	9.2×10^3

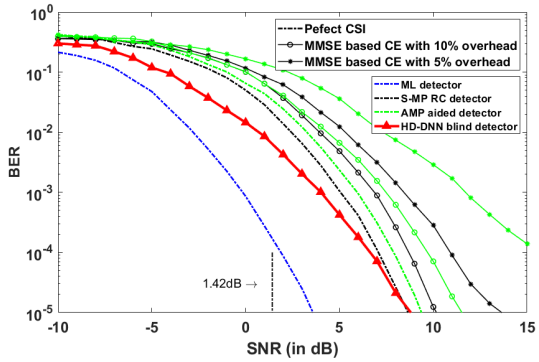
the forward and backward propagation, where we have the NN complexity order of $\mathcal{O}[n_i n_{h_1} n_{h_2} n_{h_3} n_o]$[3]. The computational complexity associated with the parameters of Table VI and Table VIII is shown in Table IX, which confirms that the proposed DNN method achieves the lowest detection complexity.

Let us now consider the performance of SD detection, where we employ a half-rate RSC encoder, as shown in Table VI. The corresponding maximum achievable rate is $R_t = 0.66667$

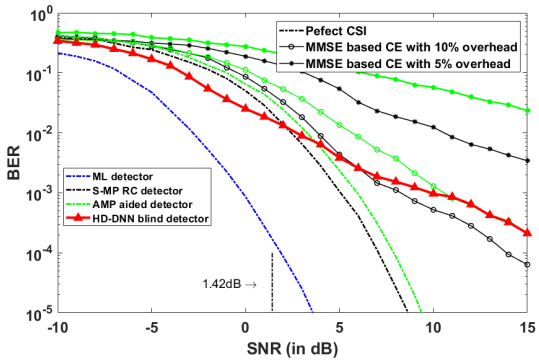
[3]Complexity order of NN only used to compare the ML detection, while there is no search complexity associated with the NN-aided detection. n_i and n_o denote the neuron size of input and output layer, $n_{h_i}(i = 1, 2, \cdots)$ denote the neuron size of hidden layer between input and output.



(a) Comparison of Scheme 1, 2, 3 under fading channel with normalized Doppler frequency of 10^-6.



(b) Comparison of Scheme 1, 2, 3 under fading channel with normalized Doppler frequency of 10^-5.



(c) Comparison of Scheme 1, 2, 3 under fading channel with normalized Doppler frequency of 10^-4.

Fig. 10: BER performance comparison of CS-MIM using HD detection which Scheme 1, 2, 3 is considered. Our simulation parameter are shown in Table VI and Table VIII.

bits/sec/Hz. In [12], it was shown that a half-rate RSC coded CS-MIM system with the same parameters as in Table VI achieves the theoretical maximum rate of 0.667 bits/sec/Hz at -7.25 dB in terms of its DCMC, which is also indicated in Fig. 8. We first examine the performance of Scheme 4(a) and 4(b) for the non-iterative SD detector. As shown in Fig. 11, 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 blind detection

DNN model is capable of achieving similar performance to the conventional SD detection, when the CE variance σ_h^2 is 0.16 and even better, when the CE variance is $\sigma_h^2 = 0.25$. Furthermore, note that our proposed learning aided method of Scheme 5 and Scheme 6 are blind and hence have no pilot overhead.

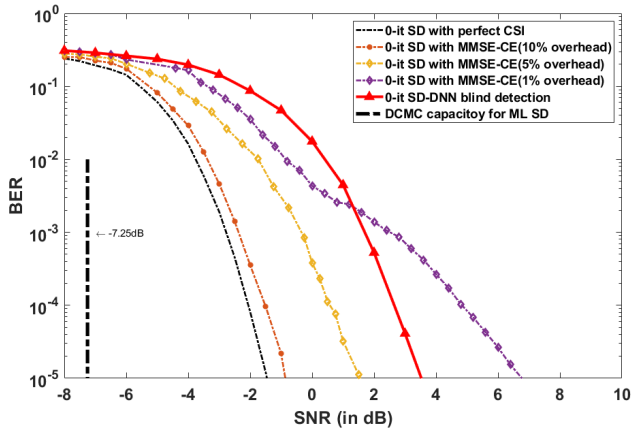
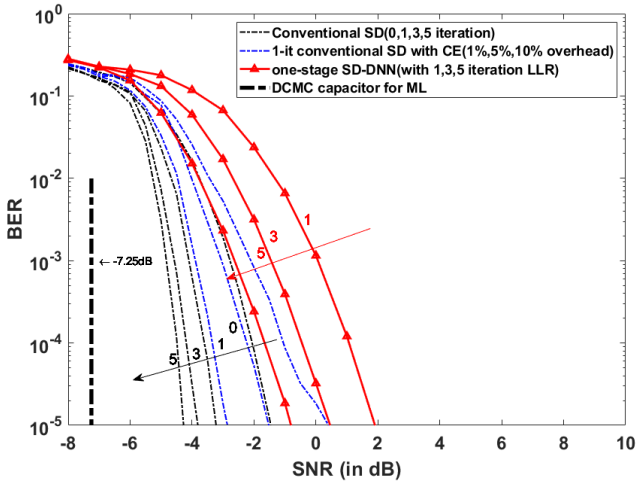


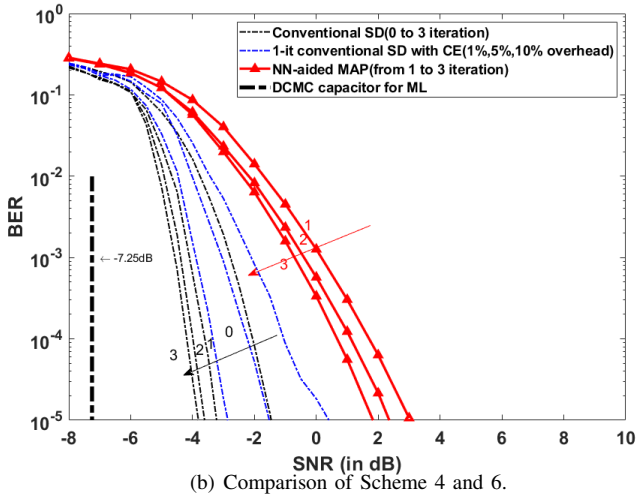
Fig. 11: BER performance comparison of the conventional SD detector with to proposed SD-DNN detector in non-iterative condition which are the comparison of Scheme 4(a), 4(b) and 5(a) as shown in Table VII. The simulation parameter are shown in Table VI and VIII.

The performance of SD can be further improved upon iteratively exchanging extrinsic information between the decoder and the demodulator. Fig. 12(a) 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. 12 characterizes 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. 12(b) 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 to train the learning model. When the number of ISD iterations is $n_{it} = 1-3$, the 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. 8.



(a) Comparison of Scheme 4 and 5.



(b) Comparison of Scheme 4 and 6.

Fig. 12: BER performance comparison of the ISD detector with NN detector of Scheme 4, 5 and 6 as shown in Table VII. The simulation parameter are shown in Table VI and Table VIII.

TABLE X: Comparison between the complexities of different SD detection methods of CS-MIM

Detector	Complexity order
Conventional	$O[n_{it} 2^{c_g} (N_{AC} N_{SI} (Q\mathcal{L})^K)]$
Single-Stage DNN-Aided	$\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$
Iterative DNN-Aided MAP	$2n_{it}(\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o))$

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

scheme	iteration	Runtime (in seconds)	iteration	Runtime (in seconds)
Conventional	0	0.735	1	3.43
	2	4.65	3	6.25
Single-Stage DNN	1	0.0064	3	0.0062
	5	0.0074		
Iterative DNN 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 X and the run time comparisons between conventional SD detection and the proposed methods are shown in Table XI. 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. 12(a), Fig. 12(b) and Table X, 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.

V. CONCLUSIONS

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.

REFERENCES

[1] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, feb 2016.

[2] D. A. Basnayaka, M. Di Renzo, and H. Haas, "Massive but few active MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 6861–6877, 9 2016.

[3] 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.

[4] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized mimo: Challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, 2014.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] J. Liu and H. Lu, "IMNet: A Learning Based Detector for Index Modulation Aided MIMO-OFDM Systems," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2020-May, 11 2019. [Online]. Available: http://arxiv.org/abs/1911.04133

[14] O. Shental and J. Hoydis, ""machine llrning": Learning to softly demodulate," pp. 1–7, 2019.

[15] 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.

[16] K. Satyanarayana, M. El-Hajjar, A. A. 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.

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

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

[19] 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.

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

[21] 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.

[22] B. Shamasundar, S. Bhat, S. Jacob, and A. Chockalingam, "Multidimensional Index Modulation in Wireless Communications," *IEEE Access*, vol. 6, pp. 589–604, 2018.

[23] 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.

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

[25] 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.

[26] 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.

[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] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. Murthy, and M. van der Schaar, "Machine learning in the air," 2019. [Online]. Available: https://arxiv.org/abs/1904.12385
[36] 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," 2019. [Online]. Available: https://arxiv.org/abs/1902.07678
[37] 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. Institute of Electrical and Electronics Engineers Inc., dec 2017, pp. 563–575.
[38] 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.
[39] 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.
[40] N. Samuel and T. Diskin, "Learning to Detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, May 2019.
[41] 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.
[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] 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.
[44] 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.
[45] T. V. Luong, Y. Ko, N. A. Vien, D. H. Nguyen, and M. Matthaiou, "Deep Learning-Based Detector for OFDM-IM," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1159–1162, aug 2019.
[46] 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.
[47] M. Turhan, E. Ozturk, and H. A. Cirpan, "Deep Convolutional Learning-Aided Detector for Generalized Frequency Division Multiplexing with Index Modulation," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2019-September. Institute of Electrical and Electronics Engineers Inc., sep 2019.
[48] L. Hanzo, B. Choi, T. Keller *et al.*, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2005.
[49] 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)," pp. 1679–1684, dec 2002.
[50] Y. Acar, H. Dogan, and E. Panayırcı, "Channel estimation for spatial modulation orthogonal frequency division multiplexing systems," pp. 382–385, 2015.
[51] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 805–808, 2012.
[52] 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.
[53] 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.
[54] 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.
[55] 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, mar 2015. [Online]. Available: http://arxiv.org/abs/1503.04069 http://dx.doi.org/10.1109/TNNLS.2016.2582924
[56] 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.
[57] 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.
[58] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning."
[59] 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.
[60] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *Electronics Letters*, vol. 32, no. 10, pp. 887–888, 1996.
[61] 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.

Xinyu Feng received a BSc(Eng) degree with First Class Honours from University of Nottingham Ningbo, china and a BSc(Eng) in 2014. He obtained a MSc degree with distinction in wireless communication from the University of Southampton, UK in 2019. He is currently a PhD student at Next Generation Wireless Research Group under the supervision of Prof. L. Hanzo and Dr. M. El-Hajjar, University of Southampton, UK. His research interests include index modulation, Channel estimation and Detection and machine learning.

Mohammed El-Hajjar Mohammed El-Hajjar is an Associate Professor in the School of Electronics and Computer Science in the University of Southampton. He received his PhD in Wireless Communications from the University of Southampton, UK in 2008. Following the PhD, he joined Imagination Technologies as a design engineer, where he worked on designing and developing Imagination's multi-standard communications platform, which resulted in three patents. He is the recipient of several academic awards and has published a Wiley-IEEE book and in excess of 140 journal and conference papers. His research is funded by the UK government and many industrial collaborators. Mohammed's research interests include the development of intelligent communications systems, energy-efficient transceiver design, MIMO, millimeter wave communications and Radio over fiber network design.

Chao Xu (S'09-M'14-SM'19) received a B.Eng. degree from Beijing University of Posts and Telecommunications, China, and a BSc(Eng) with First Class Honours from Queen Mary, University of London, UK, through a Sino-UK joint degree program in 2008, both in Telecommunications. He obtained a MSc degree with distinction in Radio Frequency Communication Systems and a Ph.D. degree in Wireless Communications from the University of Southampton, UK in 2009 and 2015, respectively. He is currently a senior research fellow working at Next Generation Wireless Research Group, University of Southampton, UK. His research interests include index modulation, reconfigurable intelligent surfaces, noncoherent detection and turbo detection. He was awarded the Best M.Sc. Student in Broadband and Mobile Communication Networks by the IEEE Communications Society (United Kingdom and Republic of Ireland Chapter) in 2009. He also received 2012 Chinese Government Award for Outstanding Self-Financed Student Abroad and 2017 Dean's Award, Faculty of Physical Sciences and Engineering, the University of Southampton.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo) (FIEEE'04) received Honorary Doctorates from the Technical University of Budapest and Edinburgh University. He is a Foreign Member of the Hungarian Science-Academy, Fellow of the Royal Academy of Engineering (FREng), of the IET, of EURASIP and holds the IEEE Eric Sumner Technical Field Award.