

# Deep Learning Assisted Detection for Index Modulation Aided mmWave Systems

K. SATYANARAYANA<sup>1</sup>, (Student Member, IEEE), LUPING XIANG<sup>1</sup>, (Member, IEEE),  
YANQING ZHANG, MOHAMMED EL-HAJJAR<sup>1</sup>, (SENIOR MEMBER, IEEE), ALAIN MOURAD<sup>2</sup>,  
AND LAJOS HANZO<sup>1</sup>, (FELLOW, IEEE)

<sup>1</sup>Department of ECS, University of Southampton, Southampton SO17 1BJ, U.K

<sup>2</sup>InterDigital Inc., London EC2A 3QR, U.K.

This work was supported by InterDigital as well as in part by the Engineering and Physical Sciences Research Council under Project EP/N004558/1 and Project EP/P034284/1, in part by the Royal Society's Global Research Challenges Grant, in part by the European Research Council's Advanced Fellow Grant QuantCom, and in part by the Royal Academy of Engineering Industrial fellowship.

**ABSTRACT** In this paper, we propose deep learning assisted detection for index modulation millimeter wave (mmWave) systems, where we train a neural network (NN) to jointly detect the transmitted data and index information without relying on explicit channel state information (CSI). As a design example, we first employ multi-set space-time shift keying (MS-STSK) combined with beamforming for transmission over the mmWave channel, where the information is conveyed implicitly using the index of the antennas, the dispersion matrix and the M-ary constellation. Then, we analyze our design when MS-STSK transmission is considered in conjunction with beam index modulation (BIM), where the information is also conveyed by the beam index in addition to the MS-STSK information. In contrast to the MS-STSK's conventional maximum likelihood (ML) detector, our learning-assisted detection dispenses with the channel estimation stage. We demonstrate by simulations that the learning assisted detection outperforms the ML-aided detection in the face of channel impairments with low complexity. Furthermore, we show by simulations that ML-aided detection produces an error floor, when the MS-STSK transmission is coupled with BIM, when realistic channel estimation errors are considered. Additionally, we present qualitative discussions on the receiver complexity in terms of its search space as well as the number of computations required.

**INDEX TERMS** Index Modulation, Millimeter Wave, MIMO, Beamforming, Machine Learning, Detection.

## NOMENCLATURE

AA	Antenna Array
AC	Antenna Combination
AF	Activating Function
ANN	Artificial Neural Network
AoA	Angle-of-Arrival
AoD	Angle-of-Departure
BER	Bit Error Rate
BF	Beamforming
BIM	Beam Index Modulation
BLAST	Bell-Labs Layered Space-Time
BS	Base Station
CSI	Channel State Information

DCMC	Discrete-input Continuous-output Memoryless Channel
DM	Dispersion Matrix
DNN	Deep Neural Network
FDD	Frequency Division Duplex
IM	Index Modulation
LSSTC	Layered Steered Space-Time Coding
HBF	HBF
MF	Multi-Functional
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Squared Error
mmWave	Millimeter Wave
MS	Multi-Set
NN	Neural Network
OSTBC	Orthogonal Space-Time Block Coding

QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
SM	Spatial Modulation
STSK	Space-Time Shift Keying
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
TPC	Transmit Precoder
ULA	Uniform Linear Array
V-BLAST	Vertical-Bell Laboratories Layered Space-Time

## I. INTRODUCTION

THE challenge of providing massive connectivity to mobile users motivated wireless communication researchers to migrate from sub-6 GHz to the frequencies spanning between 30 GHz and 300 GHz, which is referred to as the millimeter wave (mmWave) band [1, 2]. Owing to the large available bandwidth, mmWave frequencies have the potential of accommodating a large number of users while simultaneously providing high data rates for each. However, harnessing mmWave frequencies faces several technical challenges, since they suffer from high propagation losses compared to that of the sub-6 GHz spectrum. To mitigate the substantial losses due to atmospheric absorption, rain-induced fading and foliage, beamforming-aided directional transmission is envisaged [3, 4] by relying on large antenna arrays (AA) to derive beamforming gains that can compensate for the loss [1].

Furthermore, the employment of spatial-multiplexing based multiple-input multiple-output (MIMO) transmission has been proved beneficial for the enhancement of the data rates. In the literature, there is a vast body of MIMO-aided transmission schemes, but Bell-Labs' Layered Space-Time (BLAST) [5] stands as the seminal technique of achieving high multiplexing gains. By contrast, when aiming for diversity gains, the family of diversity-oriented schemes exemplified by space time block coding (STBC) [7] and orthogonal space-time block coding (OSTBC) [6] may be invoked. A further extension of space diversity design for multiple access is known as space time spreading [20, 21]. Later, a new MIMO scheme has been born by the amalgamation of diversity-, multiplexing- and beamforming-oriented arrangements, which is referred to as a multi-functional MIMO (MF MIMO) design [11]. In this design, both diversity and multiplexing as well as beamforming gains can be obtained for enhancing the capacity. Explicitly, the MF MIMO concept relies on the amalgamation of two or more MIMO schemes. For example, Satyanarayana et al. [4] designed a MF MIMO that plays a dual role by providing both diversity and beamforming (BF) gains. Another example is the layered steered space-time coding (LSSTC) of [10], where both diversity and multiplexing gains as well as the beamforming gains are achieved by the amalgamation of V-BLAST, STBC and BF. Amongst other MF MIMO techniques, space-time shift keying (STSK) [12, 15] is popular as a benefit of striking a design trade-off

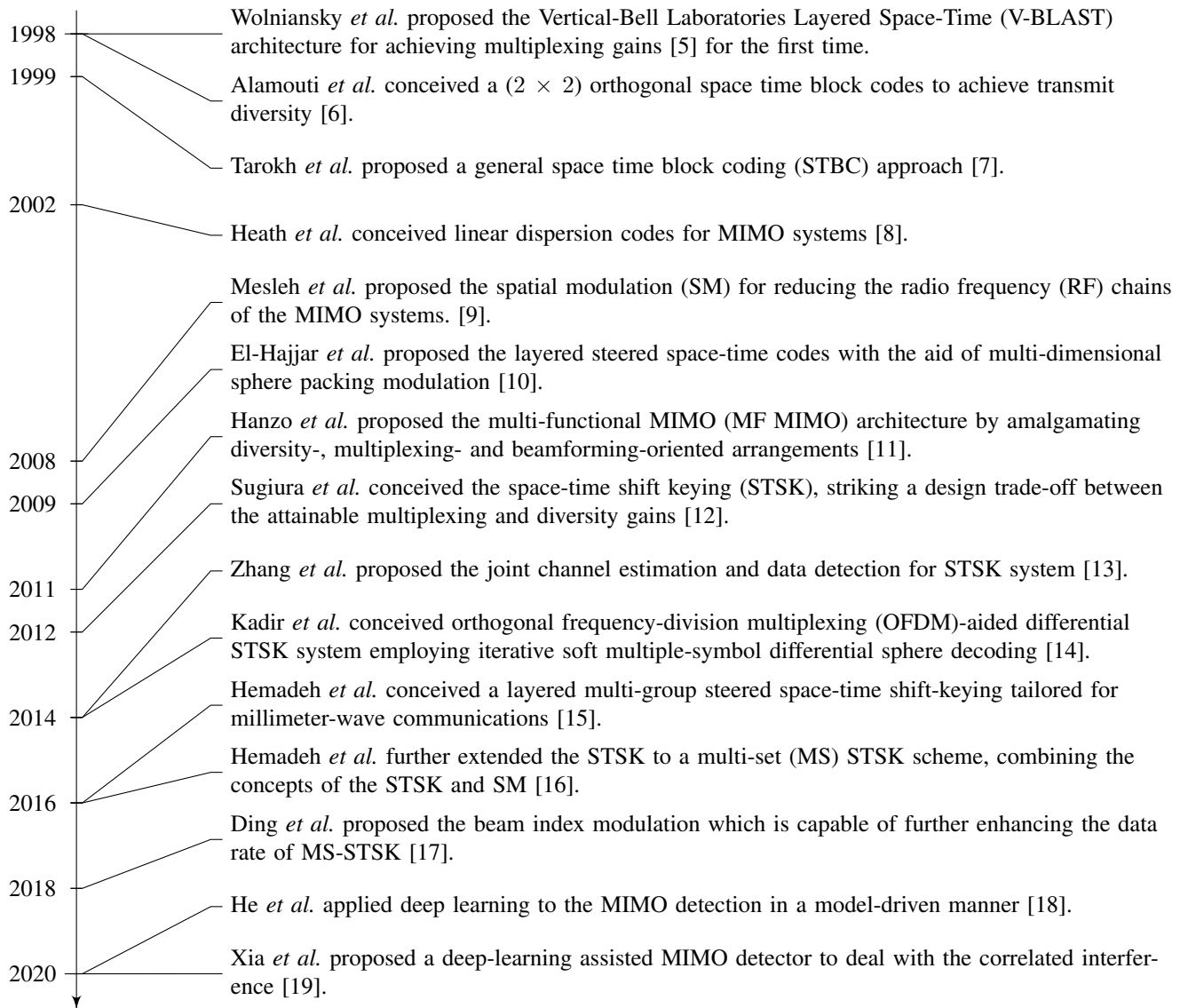
between the attainable multiplexing and diversity gains. The STSK design is conceived as an extension to the concept of spatial modulation (SM) [9, 22], where a single antenna is activated at any time. To elaborate a little further, in the STSK design, a single dispersion matrix (DM) [8] is activated from a set of DMs at any symbol instant. In other words, information is conveyed implicitly by the index of the DM in addition to the complex-valued signal drawn from an M-ary constellation. As an extension of the STSK concept, a multi-set (MS) STSK scheme was proposed in [16], which is formed by combining the concepts of the STSK and SM. This design is capable of increasing the data rate, since the information is carried by both the classic M-ary alphabet and by the DM index as well as by the antenna index combination. In mmWave communications, where the channel supports a few clusters of rays, the data rate of the MS-STSK design can be further enhanced by coupling it with the concept of beam index modulation (BIM) [17]. In the BIM aided transmission, information is implicitly conveyed by the index of the beams in addition to the classic M-ary constellation.

However, a growing concern in the index modulation transmission schemes, such as the MS-STSK is the search complexity imposed on the receiver [16]. Furthermore, like any communication system, MS-STSK also required accurate channel state information (CSI) at the receiver for achieving a low bit error ratio (BER) [23]. In frequency division duplex (FDD) systems CSI estimations relies on pilots, which reduces the effective data rate in addition to imposing extra complexity for channel estimation.

To circumvent this problem, a machine learning based detection approach may be employed, where symbol detection is carried out without explicit CSI knowledge. *This philosophy makes the design more spectrally efficient.*

There is a vast body of literature on utilizing machine learning aided physical-layer communications [24–26]. Samuel et al. [27] conceived a deep neural network (DNN) assisted architecture for data detection, while Aoudia et al. [28] demonstrated the feasibility of deep learning for end-to-end communications. By contrast, He et al. [18] employed deep learning for MIMO detection in a model-based scenario. Additionally, Xia et al. [19] designed a DNN-assisted MIMO detector operating in the presence of correlated interference, while Xiang et al. [29] employs DNN for joint channel estimation and data detection of SM. In the context of channel coding, Huang et al. et al. [30] employed reinforcement learning for the construction of polar codes, while Kurka et al. et al. [31] demonstrated the power of deep learning in improving the performance of joint source-channel coding. Furthermore, the so-called autoencoder-aided model was proposed by Van Luong et al. [32] for non-coherent index modulation (IM) detection. Table 1 contrasts the novelty of our design to the family of state-of-the-art MIMO transceiver designs. Against this backdrop, our contributions are as follows.

1) *We propose deep learning assisted detection for in-*



**FIGURE 1:** Timeline of spatial-multiplexing based multiple-input multiple-output (MIMO) systems.

**TABLE 1:** Summary of our design contrasting with state-of-the-art, where ✓ and the blank correspond to with and without contributions, respectively.

Themes of Contributions	This Paper	[27]	[28]	[18]	[19]	[32]
mmWave Frequencies	✓					
Diversity	✓					✓
Multiplexing	✓	✓		✓	✓	
Beamforming	✓					
Beam Index Modulation	✓					
Non-Coherent Detection	✓					✓
Deep Learning	✓	✓	✓	✓	✓	✓

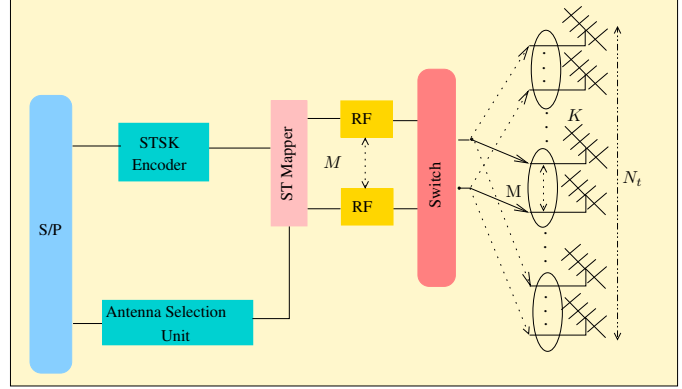
dex modulation based mmWave MIMO systems. More specifically, by employing multi-layer NNs, the detection of MS-STSK is performed without relying on the explicit knowledge of the CSI at the receiver. We also extend our design to a transmission scheme, where MS-STSK is intrinsically amalgamated with BIM, but requires limited CSI knowledge for detection. This philosophy makes our design spectrally efficient, since it eliminates the need for pilot-assisted channel estimation.

- 2) We demonstrate by simulations that our proposed design detects the MS-STSK information with reliability, despite circumventing CSI estimation.
- 3) We show by simulations that our design outperforms the ML-aided detector in the face of unavoidable channel impairments introduced during the CSI estimation. Furthermore, we demonstrate that the net Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of the ML-aided detector is lower than that of our proposed learning-assisted detection dispensing with any pilot overhead. More explicitly, this is because the DCMC capacity is significantly affected by the pilot overhead required by the classic ML-aided detection, since it requires accurate CSI in every frame of the transmission.
- 4) To improve the fidelity of our learning-assisted detection, we propose recalibration of the NN weights after a certain number of frames, which is contingent on the Doppler frequency. The recalibration of the NN weights is carried out using a modest amount of side-information or training information, which is known at both the transmitter and receiver. We show that the overhead needed for this training is negligible compared to the pilot overhead required for channel estimation.
- 5) We show by simulations that our proposed design outperforms ML-aided detection in the face of channel estimation errors whose variance is as low as 0.15.
- 6) We then extend our design to link-adaptation, where the transmitter can be adapted between the beamforming-aided MS-STSK and MS-STSK intrinsically amalgamated with BIM. The rationale of this design is to increase the data rate at a given target BER.
- 7) A qualitative complexity discussion is presented both for learning-assisted detection and for ML-aided detection in terms of the search space volume as well as the number of computations.

The rest of the paper is organized as follows. In Section II, we detail the system models of MS-STSK amalgamated with beamforming and MS-STSK with BIM in mmWave communication, while in Section III we present our proposed learning-assisted detector design. Section IV presents the complexity of the design in terms of the computations as well as the search space, while Section V and Section VI

discuss our results and conclusions, respectively.

**Notations:** We use upper case boldface,  $\mathbf{A}$ , for matrices and lower case boldface,  $\mathbf{a}$ , for vectors. We use  $(\cdot)^T$ ,  $(\cdot)^H$ ,  $\|\cdot\|_F$ ,  $\text{Tr}(\cdot)$   $\mathbb{E}(\cdot)$  for the transpose, Hermitian transpose, Frobenius norm, trace and expectation operator, respectively. We adopt  $\mathbf{A}(m, n)$  to denote the  $m^{\text{th}}$  row and  $n^{\text{th}}$  column of the  $\mathbf{A}$ ,  $\mathbf{I}_N$  is the identity matrix of size  $N \times N$ , and  $\mathbf{A} \succ 0$  indicate that  $\mathbf{A}$  is a positive definite matrix. Finally, we use  $\mathcal{CN}$ ,  $\mathcal{U}$ , and i.i.d. to represent complex-valued normal distribution, uniform distribution, and independent and identical distribution, respectively.



**FIGURE 2:** Block diagram of the MS-STSK transmitter having  $N_t$  AAs with  $K$  AEs in each AA. In our design, the inputs bit-sequence is mapped to one of  $M_Q$  dispersion matrices, one of class  $M_c$ -QAM symbols and also to  $M$  consecutive AA out of  $N_t$  AAs.

## II. SYSTEM MODEL

Let us consider the system model shown in Fig. 2, where the transmitter is equipped with  $N_t$  antenna arrays (AA)<sup>1</sup> of  $K$  antenna elements (AE) each. In Fig. 2, the transmitter employs an MS-STSK scheme, where the information is conveyed by both the STSK symbols and antenna combination (AC) information. In our system model of Fig. 2, the MS-STSK scheme relies on using  $M$  AAs (RF chains), where the AC selection is performed by selecting  $M$  AA out of  $N_t$  AA. More explicitly, the MS-STSK codeword is comprised of two parts, where the first part conveys  $\log_2(M_c M_Q)$  bits, with  $M_c$  being the constellation size and  $M_Q$  is the number of dispersion matrices [8]. The remaining  $\log_2(N_t/M)$  bits are mapped to a specific AC in the antenna selection unit of Fig. 2. It is important to emphasize that during the MS-STSK transmission only  $M$  AA are activated at any symbol interval, while the other antennas remain silent<sup>2</sup>

<sup>1</sup>The transmitter is equipped with a beamforming antenna array to derive for compensating the propagation loss at mmWave frequencies.

<sup>2</sup>The rationale of having  $K$  AE is to provide beamforming gain for compensating the propagation loss, while the time and spatial diversity is obtained by  $M$  AAs (RF chains) and dispersion matrices

The output of a typical STSK encoder is given by

$$\mathbf{X} = \mathbf{A}_q x_l, \quad (1)$$

where  $x_l$  is the M-QAM/PSK symbol, and  $\mathbf{A}_q$  is the  $q^{\text{th}}$  dispersion matrix of size  $M \times T$  from the set  $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_q, \dots, \mathbf{A}_{M_Q}\}$ , where  $\|\mathbf{A}_q\|^2 = T$ . The physical significance of the matrix  $\mathbf{A}_q$  is that it disperses the symbol  $x_l$  over  $M$  AA during  $T$  time slots. For example, for a 4-bit-sequence '0110', where the first two bits, '10', are mapped to one of the classic 4-QAM symbols, while the remaining two bits, '01', are mapped to one of the four dispersion matrices from the set  $\mathcal{A}$  having the cardinality of 4, i.e.  $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$ . It is also possible that the first three bits are mapped to 8-QAM while the last bit is used for the selection of one of two dispersion matrices, depending on the specific design requirements.

Having expounded on the MS-STSK design, in the next subsections, we focus our attention on the system model of the MS-STSK design combined with beamforming followed by the description of amalgamating MS-STSK with the BIM concept.

#### A. MS-STSK COMBINED WITH BEAMFORMING

It is important to emphasize that (1) represents the STSK symbol associated with a specific combination of active AA. However, since the transmitter is equipped with  $N_t$  AA, the total number of ACs ( $N_c$ ) for the STSK symbol transmission is

$$N_c = 2^k, k = \lfloor \log_2 N_t/M \rfloor. \quad (2)$$

The  $N_c$  value is rounded down using the floor function to allow only an integer number of bits. Note that the number of ACs ( $N_c$ ) is calculated by combining the set of  $M$  consecutive antennas together.

During transmission, an MS-STSK symbol is formed when an STSK symbol is fed to the ST mapper of Fig. 2, where a specific AC is selected depending upon the input bit-sequence. In other words, a part of the input bit-sequence determines the specific AC to be selected for transmission. To expound a little further, let us again consider the bit-sequence where two additional bits are appended to the left of the aforementioned bits, i.e., '110110'. In this scenario, the bits '11' convey the index of one of four ACs<sup>3</sup>. Thus, in this design we have a total of  $\log_2(N_c M_c M_Q)$  bits/channel use (bpcu), where the additional  $\log_2(N_c)$  bits pass the information of the AC. Therefore, an MS-STSK symbol formed at the output of the ST mapper of Fig. 3 from the  $c^{\text{th}}$  AC can be expressed as [15]

$$\tilde{\mathbf{X}} = \tilde{\mathbf{A}}_{q,c} s_l, \quad (3)$$

where  $\tilde{\mathbf{A}}_{q,c}$  is the MS-STSK dispersion matrix whose entries are constituted by the selected AC dispersion matrix and it

is given by

$$\tilde{\mathbf{A}}_{q,c} = [\mathbf{0} \dots \mathbf{A}_q \dots \mathbf{0}]^T, \quad (4)$$

where  $q$  denotes the index of the dispersion matrix, while  $c$  denotes the antenna combination. Then the MS-STSK symbol is steered in the desired direction for transmission over the mmWave channel using the RF analog BF matrix  $\mathbf{F}_{\text{RF}}$  to the desired user, as shown in Fig. 3(a), where the block-based received signal  $\mathbf{Y}$  of size  $N_r \times T$  after RF analog combining using  $\mathbf{W}_{\text{RF}}$  matrix is given by<sup>4</sup>

$$\mathbf{Y} = \mathbf{W}_{\text{RF}} \mathbf{H} \mathbf{F}_{\text{RF}} \tilde{\mathbf{X}} + \mathbf{V}, \quad (5)$$

where  $\mathbf{V}$  is the Gaussian noise having the distribution of  $\mathcal{CN}(0, \sigma^2)$ , while  $\mathbf{F}_{\text{RF}}$  is expressed as

$$\mathbf{F}_{\text{RF}} = [\mathbf{O} \dots \mathbf{F}_{\text{RF}_q} \dots \mathbf{O}] \in \mathbb{C}^{K N_t \times N_t}, \quad (6)$$

$$\mathbf{F}_{\text{RF}_q} = \text{diag}(\mathbf{F}_{\text{RF}}^1, \mathbf{F}_{\text{RF}}^2, \dots, \mathbf{F}_{\text{RF}}^M), \quad (7)$$

where  $\mathbf{F}_{\text{RF}}^i$  is the BF vector of the  $i^{\text{th}}$  AA of size  $K \times 1$ . Similarly,  $\mathbf{W}_{\text{RF}}$  is the analog RF combining matrix of size  $M \times N_r$ . Furthermore,  $\mathbf{H}$  represents the statistical mmWave channel model expressed as

$$\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_{N_t}], \quad (8)$$

where  $\mathbf{H}_i$  is a statistical channel matrix of size  $N_r \times K$ , which is expressed as

$$\mathbf{H}_i = \sqrt{\frac{N_r N_t}{N_c N_{\text{ray}}}} \sum_{n_c=1}^{N_c} \sum_{n_{\text{ray}}=1}^{N_{\text{ray}}} \alpha_{n_c}^{n_{\text{ray}}} \mathbf{a}_r(\phi_{n_c}^{n_{\text{ray}}}) \mathbf{a}_t^T(\phi_{n_c}^{n_{\text{ray}}}). \quad (9)$$

The variables  $N_c, N_{\text{ray}}$  in (9) are the number of clusters and number of rays, respectively, while  $\alpha$  obeys the distribution of  $\mathcal{N}(0, 1)$ , and  $\mathbf{a}_i$  as well as  $\mathbf{a}_r$  represent the array response vectors at the  $i^{\text{th}}$  AA of the transmitter and the array at the receiver, respectively, expressed as follows:

$$\mathbf{a}_r(\phi_r) = [1 e^{j \frac{2\pi}{\lambda} d \cos(\phi_r)} \dots e^{j \frac{2\pi}{\lambda} (N_r-1) d \cos(\phi_r)}]^T, \quad (10)$$

$$\mathbf{a}_t(\phi_t) = [1 e^{j \frac{2\pi}{\lambda} d \cos(\phi_t)} \dots e^{j \frac{2\pi}{\lambda} (K-1) d \cos(\phi_t)}]^T. \quad (11)$$

Note that our system model of (5) corresponds to Fig. 3(a), where the MS-STSK symbol is transmitted over the channel matrix  $\mathbf{H}$  to its intended receiver with the aid of beamforming, where all the beams (one or many) supported by the channel are utilized.

Additionally, by letting  $\mathbf{W}_{\text{RF}} \mathbf{H} \mathbf{F}_{\text{RF}} = \mathbf{H}_{\text{eff}}$  and invoking the vectorial stacking operation, Eq. (5) becomes equivalent to an SM system, which is detailed in [15]. In other words, Eq. (5) can be re-written in the vector form of [15, 16]:

$$\mathbf{y} = \tilde{\mathbf{H}} \mathcal{X} \mathcal{I}_c \mathbf{K}_{q,l} + \tilde{\mathbf{V}}, \quad (12)$$

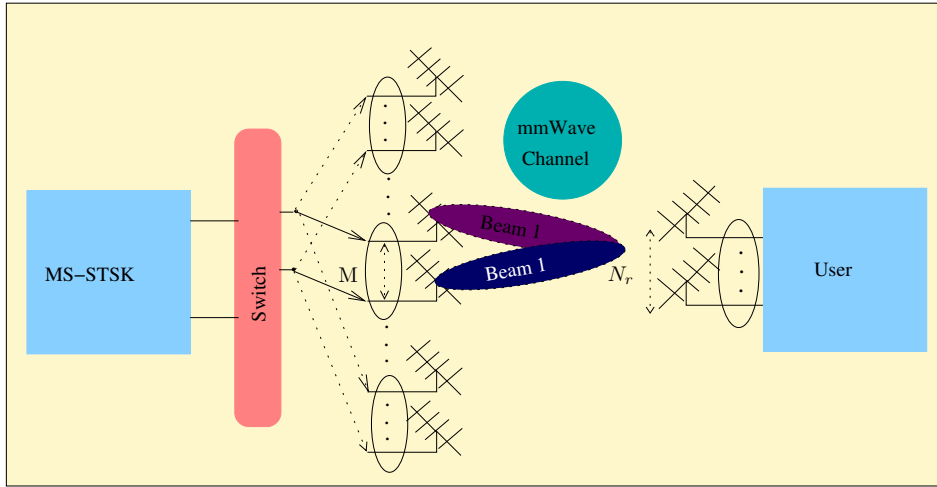
where the vectorized constituent matrices are expressed as

$$\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{N_r T \times 1}, \quad (13)$$

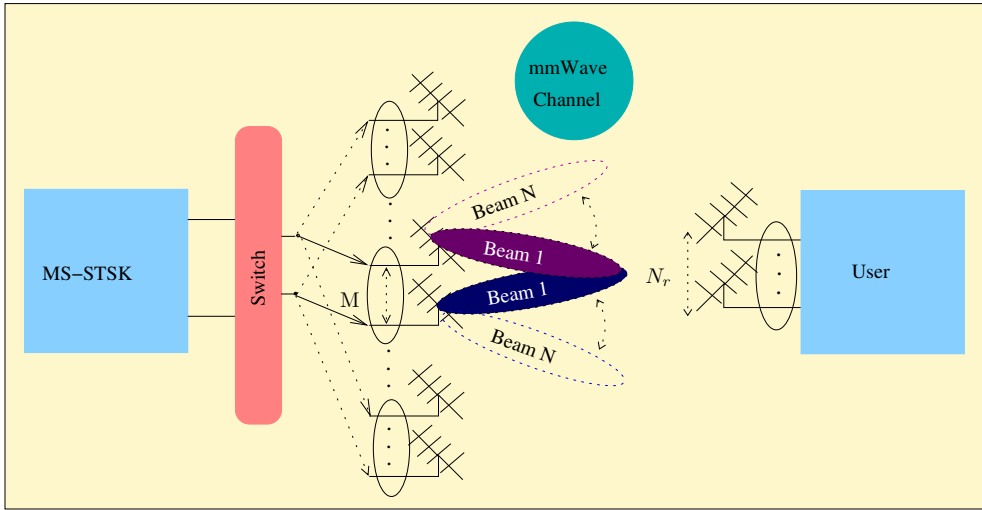
<sup>4</sup>Note that  $\mathbf{Y}$  is a matrix, where each column represent the received signal vector at time  $t$ . Therefore, for a time interval  $T$ , the received signal  $\mathbf{Y}$  will be of size  $N_r \times T$ .

<sup>3</sup>In this setting, the total number of AC assumed is 4.





(a) Beamformed MS-STSK symbol in the direction of the desired user. In this design, a total of  $\log_2(N_c M_c M_Q)$  bits are mapped to one of the classic  $M_c$ -QAM symbols, one of the  $M_Q$  dispersion matrices and one of the  $N_c$  ACs.



(b) Beamformed MS-STSK symbol coupled with the beam index in the direction of the desired user. In this design, a total of  $\log_2(N_b N_c M_c M_Q)$  bits are mapped to one of the classic  $M_c$ -QAM symbols, one of the  $M_Q$  dispersion matrices, one of the  $N_c$  ACs and one of the  $N_b$  beams.

**FIGURE 3:** MS-STSK symbol transmission - (a) when the channel supports only one beam (b) amalgamated with beam index when the channel supports a plurality of beams.

$$\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}_{\text{eff}} \in \mathbb{C}^{N_r T \times N_t T}, \quad (14)$$

$$\tilde{\mathbf{V}} = \text{vec}(\mathbf{V}) \in \mathbb{C}^{N_r T \times 1}, \quad (15)$$

$$\mathcal{X} = [\text{vec}(\tilde{\mathbf{A}}_{1,1}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,c}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,N_c})] \in \mathbb{C}^{K N_t T \times N_c M_Q}, \quad (16)$$

$$\mathbf{K} = [\underbrace{0 \dots 0}_{q-1} \underbrace{s_l 0 \dots 0}_{M_Q - q}]^T. \quad (17)$$

expressed as

$$\langle \hat{q}, \hat{l}, \hat{c} \rangle = \arg \min_{q,l,c} \|\mathbf{y} - \tilde{\mathbf{H}} \mathcal{X} \mathcal{I}_c \mathbf{K}_{q,l}\|^2. \quad (18)$$

Additionally, the DCMC capacity of the MS-STSK scheme is given by [33]

$$C_{\text{DCMC}}^{\text{MS-STSK}} = \log_2(N_c M_c M_Q) - \frac{1}{(N_c M_c M_Q)} \quad (19)$$

$$\sum_{q,l,c} \mathbb{E} \left[ \log_2 \sum_{q',l',c'} \exp(\psi_{\text{MS-STSK}}) \middle| \mathbf{K} \right], \quad (20)$$

At the receiver, the vectorized received signal  $\mathbf{y}$  is used during the detection process. Conventionally, the detection of the MS-STSK symbol, where the estimates  $\hat{q}, \hat{l}, \hat{c}$  of (5) are obtained, is carried out by employing ML detection relying on the CSI estimated at the receiver and it is

where

$$\psi_{\text{MS-STSK}} = -\frac{\|\mathbf{H}\mathcal{X}(\mathcal{I}_{c'}\mathbf{K}_{q,l} - \mathcal{I}_{c'}\mathbf{K}_{q',l'}) + \tilde{\mathbf{V}}\|^2 - \|\tilde{\mathbf{V}}\|^2}{\sigma^2}. \quad (21)$$

By taking into account the pilot overhead  $f_p$ , which is the ratio of the number of pilots to the number of data symbols, the effective DCMC capacity becomes:

$$C_{\text{eDCMC}}^{\text{MS-STSK}} = (1 - f_p)C_{\text{DCMC}}^{\text{MS-STSK}}.$$

We note that the index modulation, such as spatial modulation and MS-STSK, suffer from a potentially higher performance loss than the classic MIMO configurations, because, the CSI accuracy affects the antenna and beam indices of our system, both of which form an integral part of the total information symbol stream. Hence, index modulation systems are more susceptible to CSI impairments. In order to overcome this difficulty, typically a higher pilot overhead is imposed with the objective of obtaining a more accurate CSI. Furthermore, the channel estimation error variance using LMMSE, which is typically employed in practice, is given by [34]

$$\sigma^2 = \frac{1}{1 + \tau_p \rho_t}, \quad (22)$$

where  $\rho_t$  is the pilot transmission power and  $\tau_p = \tau - \tau_d$  is the pilot symbols' transmission time duration, while  $\tau$  is the total transmission time and  $\tau_d$  is the data duration. Equation (23) can also be equivalently written in terms of pilot ratio  $f_p$ <sup>5</sup> and total number of symbols  $\eta$  as

$$\sigma^2 = \frac{1}{1 + \eta f_p \rho_t}. \quad (23)$$

### B. MS-STSK AMALGAMATED WITH BIM

In contrast to Fig. 3(a), Fig. 3(b) shows the MS-STSK symbol coupled with the beam index before the final transmission. In this model of Fig. 3(b) information is also conveyed by the index of the beam in addition to the information conveyed by MS-STSK symbol. More explicitly, when the channel of (8) supports a plurality of beams, say  $N_b$  beams, instead of transmitting in all beams at once, the transmitter selects a specific beam for transmission depending upon the input bit-sequence. Thus, this design is capable of achieving an additional bit rate of  $\log_2(N_b)$  bits per channel use than that of its counterpart dispensing with BIM.

To elaborate a little further, let us consider the 'toy' example shown in Fig. 4, where the channel seen from each AA supports a total of 4 beams for transmission. In order to increase the spectral efficiency, the BIM is invoked by relying on the index of the beam used for transmission. In other words, MS-STSK transmission can be carried out by one of the four beams from each AA by allowing additional

bits to be conveyed by the index of the beam. Naturally, this philosophy only can be exploited when there are more than one beams<sup>6</sup>. If there is only a single beam, then MS-STSK will be combined with conventional beamforming. As an example, in Fig. 4, beam 2 is selected for MS-STSK transmission according to the bit sequence '01' representing the beam index, while beam 3 is selected from the other AA for the bits '11'. Thus, the total number of bits per channel use when BIM is coupled with the MS-STSK example discussed in Section II-A is 10, i.e.,  $\underbrace{1101}_{\text{BI}} \underbrace{110110}_{\text{MS-STSK}}$ .

Now let us again consider Fig. 3(b), where the channel seen from each AA at the transmitter supports  $N_b$  beams and only one of the  $N_b$  beams is selected for transmission depending on the bit-sequence. Thus there are  $N_b$  possible combinations of beamforming for MS-STSK symbol transmission conveying a total of  $\log_2(N_b N_c M_c M_Q)$  information bits in contrast to the  $\log_2(N_c M_c M_Q)$  bits of the MS-STSK scheme of [15], transmission dispensing with the beam index mode. The block-based received signal  $\mathbf{Y}_{\text{BI}}$  of this scenario can be expressed as

$$\mathbf{Y}_{\text{BI}} = \mathbf{W}_{\text{RF}}^n \mathbf{H}_{\text{BI}}^n \mathbf{F}_{\text{RF}}^n \tilde{\mathbf{X}} + \mathbf{N}, \quad (24)$$

where  $\mathbf{H}_{\text{BI}}^n$  is the statistical channel model of (8) in the  $n^{\text{th}}$  beam, while the sizes of the matrices  $\mathbf{H}_{\text{BI}}^n$ ,  $\mathbf{W}_{\text{RF}}^n$ ,  $\mathbf{F}_{\text{RF}}^n$  and  $\tilde{\mathbf{X}}$  are still the same, as described in Sec. II-A. We observe that the only difference between (5) and (24) lies in the manner of exploiting the beams.

Similar to (13), Eq. (24) can be vectorized for the  $n^{\text{th}}$  beam. At the receiver, the vectorized received signal  $\mathbf{y}_{\text{BI}}$  is used during the detection process. In this setting, the detection of the estimates  $\hat{q}, \hat{l}, \hat{c}, \hat{n}$  of (5) is obtained by employing ML detection on the CSI of the  $n^{\text{th}}$  beam at the receiver and it is expressed as

$$\langle \hat{q}, \hat{l}, \hat{c}, \hat{n} \rangle = \arg \min_{q,l,c,n} \|\mathbf{y}_{\text{BI}} - \mathbf{H}_{\text{BI}}^n \mathcal{X} \mathbf{I} \mathbf{K}_{q,l}\|^2; \quad (25)$$

It is important to emphasize that both (18) and (25) are heavily reliant on the availability of accurate CSI for the successful detection of the symbols, thereby imposing both the usual pilot overhead required for channel estimation and an additional channel estimation complexity. Furthermore, we will show in the subsequent section that, both (18) and (25) produce an error floor when the CSI estimation error variance is set to 0.15.

Additionally, the DCMC capacity of the MS-STSK

<sup>5</sup>The typical values of  $f_p$  for normalized Doppler of 0.001 are 0.03 and 0.05 [35, 36].

<sup>6</sup>At this stage the question arises — how do we best configure our  $N_t K$ -element antenna for a specific diversity-, spatial multiplexing- and beamforming order? Naturally, this depends on the particular application in mind, as well as on its specifications. In the example of Fig. 4, we opted for invoking BIM for implicitly conveying two extra bits instead of the classic space-division multiple access principle, because BF relies on  $\lambda/2$ -spaced elements; but such a tight element-spacing would result in a modest STSK multiplexing gains, because the adjacent AEs receive correlated signals, which are hard to separate at the receiver. In a nutshell, the specific assignment of AEs to the baseband signal processing functions has to be carefully considered.

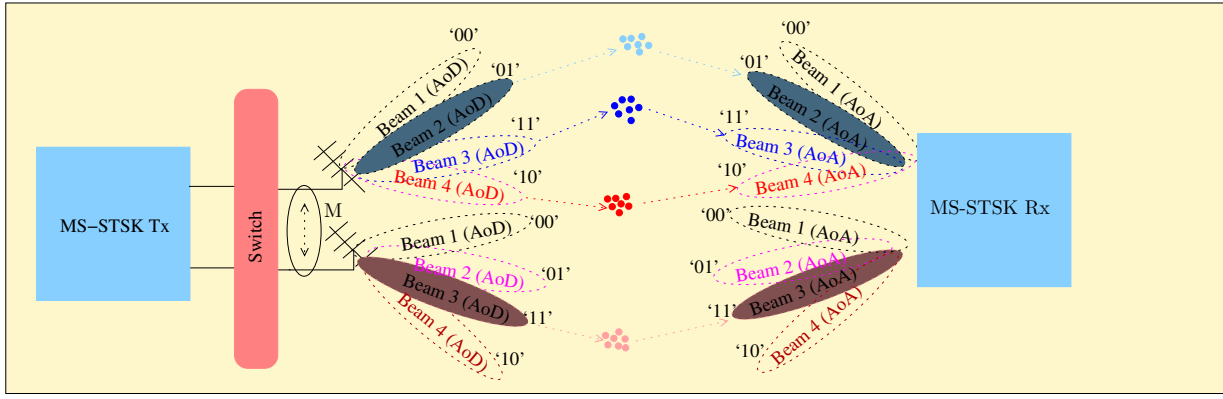


FIGURE 4: Illustration of beam index modulation coupled with MS-STSK.

scheme combined with BIM is expressed as

$$C_{\text{DCMC}}^{\text{MS-STSK-BIM}} = \log_2(N_b N_c M_c M_Q) - \frac{1}{(N_b N_c M_c M_Q)} \times \sum_{n,q,l,c} \mathbb{E} \left[ \log_2 \sum_{n',q',l',c'} \exp(\psi_{\text{MS-STSK-BIM}}) \middle| \mathbf{K} \right], \quad (26)$$

where we have:

$$\psi_{\text{MS-STSK-BIM}} = \quad (27)$$

$$- \frac{\left\| \mathbf{H}_n \mathcal{X} \mathcal{I}_c \mathbf{K}_{q,l} - \mathbf{H}_{n'} \mathcal{X} \mathcal{I}_{c'} \mathbf{K}_{q',l'} + \tilde{\mathbf{V}} \right\|^2 - \|\tilde{\mathbf{V}}\|^2}{\sigma^2}. \quad (28)$$

By taking into account the pilot percentage  $f_p$  the effective DCMC capacity becomes:

$$C_{\text{eDCMC}}^{\text{MS-STSK-BIM}} = (1 - f_p) C_{\text{DCMC}}^{\text{MS-STSK-BIM}}.$$

In the next section, we propose a design, where we employ deep detection relying on a neural network while dispensing with the requirement of having CSI knowledge at the receiver. The advantage of our design is that it avoids the reliance on CSI and consequently circumventing the pilot overhead and complexity of channel estimation. This philosophy makes our design spectral-efficient.

### III. PROPOSED LEARNING ASSISTED DETECTOR DESIGN

In this section we commence our discussion by presenting some preliminaries on deep learning, in order for the paper to be self-contained. Then we expound on the proposed design, where we invoke deep learning philosophy for the data detection.

#### A. DEEP LEARNING PRELIMINARIES

Artificial Neural Networks (ANN) rely on a computational model inspired by the structural and functional aspects of biological neural networks [37]. As a benefit of their ability to learn and generalize, they have become one of the most popular machine learning techniques. They mimic the functions of the human brain in terms of organizing neurons

to evaluate certain operations. The model of a typical neural network is shown in Fig. 5. A neural network as that of

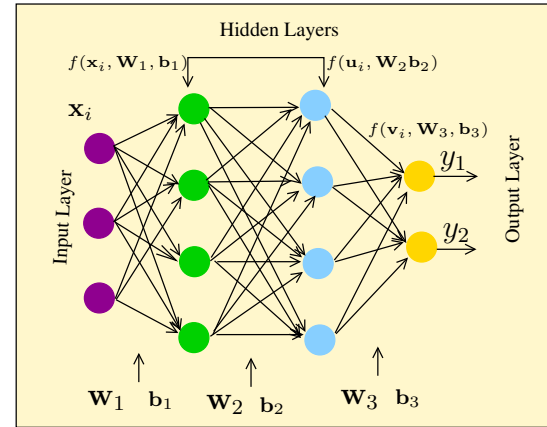


FIGURE 5: Model of a typical DNN architecture.

Fig. 5 consists of multiple layers, where the first and the last layers are the input and the output layers, while the layers between them are referred to as hidden layers. Note that a neural network having more than a single hidden layer is referred to as a deep neural network [38], which belongs to the family of deep learning techniques. To elaborate a little further, the input vector  $\mathbf{x}_i$  is fed to the input layer, which is connected to the left-most hidden layer of Fig. 5. The output from each neuron of the hidden layer is governed by a so-called activation function  $f(\cdot)$ . The activating function represents the rate of change neuron's potential from its resting state before it is stimulated. Typically, the activating function is a function of a weight vector and a bias. In our example of Fig. 5, the first column of the weight matrix  $\mathbf{W}_1$  and the first element of the bias vector  $\mathbf{b}_1$  belong to the first neuron of the first hidden layer. Similarly, the outputs from the other neurons are calculated and fed to the next layer whose outputs are again governed by the activating function of that layer. Finally, the predicted response of the neural network is obtained from the output layer. It is instructive to note that the only mathematical condition on



the choice of the activating function is differentiability [37, 38]. Some commonly used activating functions include a simple threshold function, ReLu function, piecewise-linear function, and sigmoid function.

Having discussed the structure of a typical neural network, we now focus our attention on the employment of an ANN. The employment of an ANN comprises two stages: the training phase and testing phase. In the training phase, the known input and output samples are used for computing the weight matrices and bias vectors. In other words, the weights and biases are specifically designed for minimizing the error between the known output and the predicted output. Then in the testing phase, the pre-designed weights and biases are applied to the new input data, outside the training set for predicting the output.

In the next subsection, we rely on this philosophy, where the detection is carried out relying on the weights trained during the training phase.

### B. LEARNING ASSISTED DETECTION

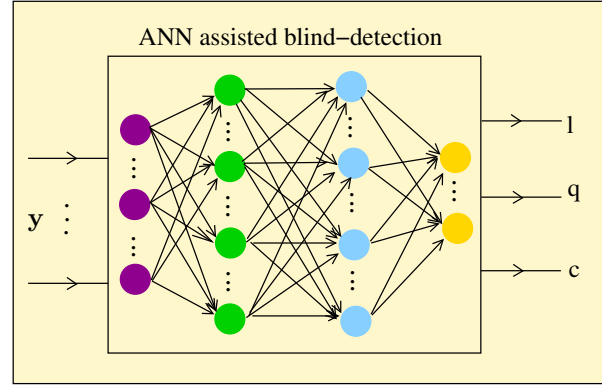
Let us now focus our attention on the learning-assisted detection for our system model. As discussed in the previous section, we first aim for designing the training weights and biases for our neural network. In our design, the number of hidden layers is set to 2, while the number of neurons is adjusted in such a way that it faithfully reproduces the output during the training stage. In our MS-STSK symbol transmission of Eq. (5), the vectorized matrix  $\mathbf{y}$  serves as the input to the neural network; while the detected dispersion matrix index, the antenna index and the complex-valued classic symbol drawn from the M-QAM constellation constitute the output vector, as shown in Fig. 6(a). Similarly, when the MS-STSK encoder is amalgamated with a beam-index, as in Eq. (24), the vectorized matrix  $\mathbf{y}_{BI}$  serves as the input of the ANN. In this scenario, we have an additional element, at the output of the ANN which is the beam index, as shown in Fig. 6(b).

Having defined the input and the output vectors of the neural network, training of the network is carried out using a set of known input and output samples. We note that before the training process, the weight matrices and the biases vectors of all layers are set to random values from the distribution  $\mathcal{N}(0, 1)$  [39]. Furthermore, in our design, a member of the hyperbolic tangent function is used as the activating function as a benefit of its smoothness and asymptotic properties, which is given as [37]

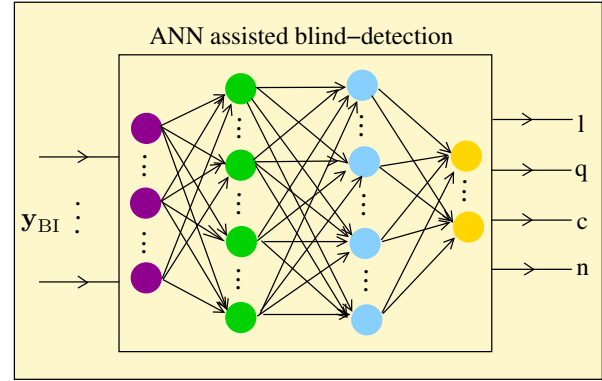
$$f(x, a, v) = \frac{2}{1 + e^{-2a(x-v)}} - 1, \quad (29)$$

where  $f(x, a, v)$  is mapping on  $x$ ,  $a$  is the slope parameter, and  $v$  affects the function position.

The activating function of (29) is applied at every neuron of the network whose output is fed as the input to the next layer of neurons. In other words, the activating function of (29) maps its input vector  $\mathbf{x}_i$  of the  $i^{th}$  training sample using the weight matrix  $\mathbf{W}$  and bias vector  $\mathbf{b}$  of that layer. Then



(a) Learning assisted detection for beamformed MS-STSK transmission.



(b) Learning assisted detection for MS-STSK with BIM transmission.

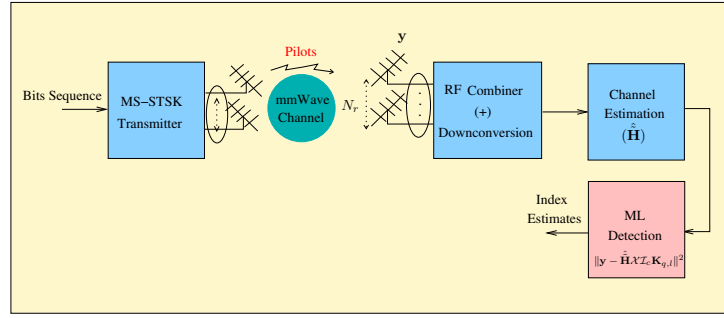
**FIGURE 6:** ANN assisted detection at the receiver. Since this design is capable of estimating the output while dispensing with the CSI, the pilot overhead is markedly reduced.

this mapping serves as the input vector to its succeeding layer and so forth. After the final mapping, which is at the output layer, the error between the known output and the predicted output is computed. The error is formulated for each layer as a loss function given by [40]

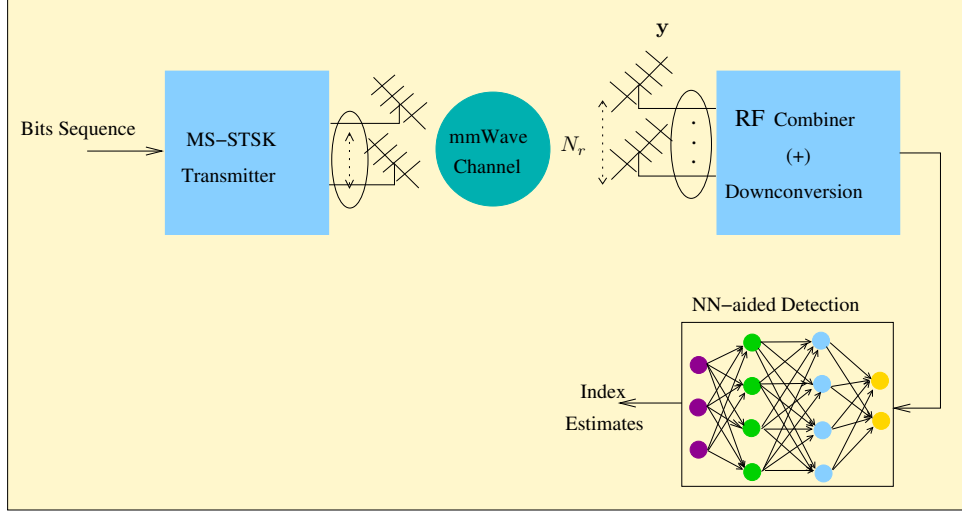
$$L = \frac{1}{S} \sum_{i=1}^S \|\hat{\mathbf{y}}_i - \mathbf{y}_i\| + \rho_1 \|\mathbf{W}_1\|_2^2 + \rho_2 \|\mathbf{W}_2\|_2^2 + \rho_3 \|\mathbf{W}_3\|_2^2, \quad (30)$$

where  $S$  is the cardinality of the training set,  $\hat{\mathbf{y}}_i$  and  $\mathbf{y}_i$  are the predicted and the known output vector of the  $i^{th}$  training sample, respectively, while  $\rho_1, \rho_2, \rho_3$  are the regularization factors used for avoiding over-fitting [40].

Then, the weight matrices and bias vectors are designed for minimizing the loss function of (30). Note that the solution obtained is not guaranteed to have global optimum solution. This is typically carried out using the technique of back-propagation. In back-propagation, the gradient of (30) is evaluated with respect to the weight matrices  $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$  and bias vectors  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ . A more detailed explanation of back-propagation is presented by Chauvin et al. [41].



(a) ML assisted detection for MS-STSK transmission. This design entails having CSI estimation for symbol detection.



(b) Learning assisted detection for MS-STSK transmission. This design dispenses with the CSI estimation for symbol detection.

**FIGURE 7:** ANN assisted detection at the receiver. Since this design is capable of estimating the output while dispensing with the CSI, the pilot overhead is markedly reduced.

Note that in our design we assumed that the channel evolves in time according to Jakes' model, where the channel's correlation coefficient in time is defined by the zero-order Bessel-function of the first kind as [42]

$$\zeta = J_0(2\pi f_d \tau), \quad (31)$$

where  $f_d$  is the maximum Doppler frequency and  $\tau$  is the sample time. Therefore, the number of neurons and the time required for designing the weights and biases during the training phase depend on the Doppler spread<sup>7</sup>  $f_d \tau$ . Furthermore, the Doppler spread also plays a key role in deciding how often the training of the weights and biases is required for estimating the indices with a high integrity.

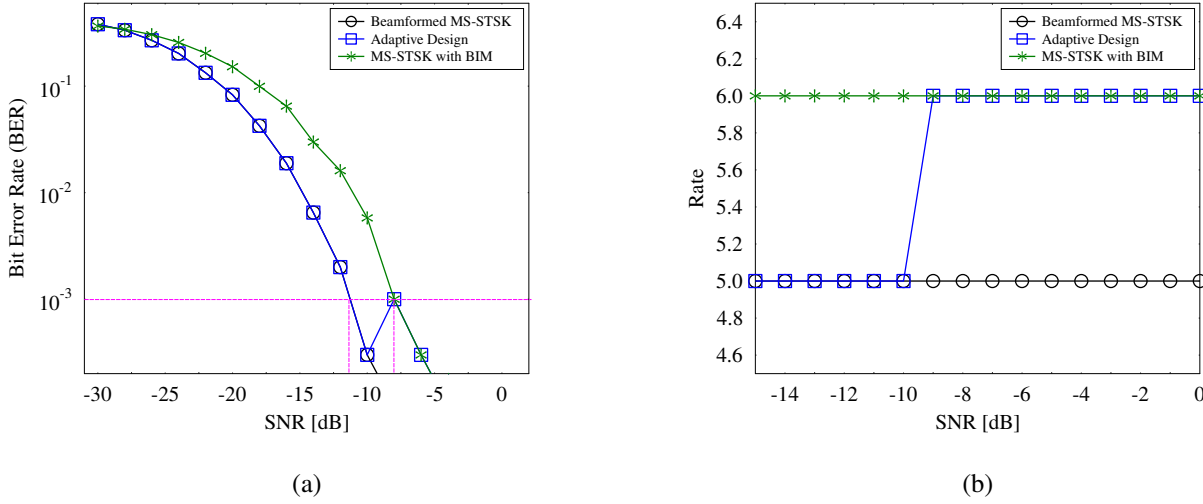
After designing the neural network parameters, the testing phase ensues, where the vector  $\mathbf{y}$  (or  $\mathbf{y}_{BI}$ ) from the receive AA is fed to the input of the neural network. Here the weights and biases computed during the training phase are applied to the input vector for estimating the indices at the output of the network.

<sup>7</sup>At high Doppler spread, more number of neurons may be required for training.

*Remark 2:* The input of the neural network takes only real values; therefore, we have split the received vector  $\mathbf{y}$  into real part  $\mathcal{R}(\mathbf{y})$  and imaginary part  $\mathcal{I}(\mathbf{y})$  before feeding it to the neural network.

### C. TRANSCEIVER ADAPTATION

Having discussed the learning assisted detection for the MS-STSK mmWave systems, we now extend our design to link-adaptation, where the transmitter adapts between beamforming-aided MS-STSK and MS-STSK intrinsically amalgamated with BIM. The rationale of this design is to increase the data rate at a given target BER. As a design example, we consider the target BER of  $10^{-3}$ . Then it can be seen in Fig. 8(a) that at the SNR of -10 dB, the beamforming aided MS-STSK design switches to MS-STSK combined with BIM for increasing the data rate while still maintaining the target BER. Fig. 8(b) shows the rate achieved by the adaptive design. More explicitly, it can be seen that the adaptive design supports higher rate transmission than that of the beamforming aided MS-STSK while also satisfying the target BER of  $10^{-3}$ . It is instructive to note that the MS-STSK with BIM provides higher rate than that of the



**FIGURE 8:** The transmitter adapts between beamforming-aided MS-STSK and MS-STSK combined with BIM modes of transmission (a) BER of the adaptive design (b) Rate of the adaptive design.

adaptive design for SNRs  $< -7.5$  dB, but it fails to satisfy our target BER criterion.

#### IV. RECEIVER COMPLEXITY

In this section, we focus our attention on the receiver design of the proposed model. Fig. 7 illustrates the block diagram of a typical receiver employing ML detection as well as the learning-assisted detection. To expound further, Fig. 7(a) shows the schematic of ML detection. In this design, the receiver first combines the signal in the RF stage and then performs down-conversion for further digital processing in the baseband. Given the necessity of having the CSI, channel estimation is carried out with the aid of pilots prior to the detection. The ML detection is invoked for estimating the MS-STSK symbol after the CSI estimate is obtained.

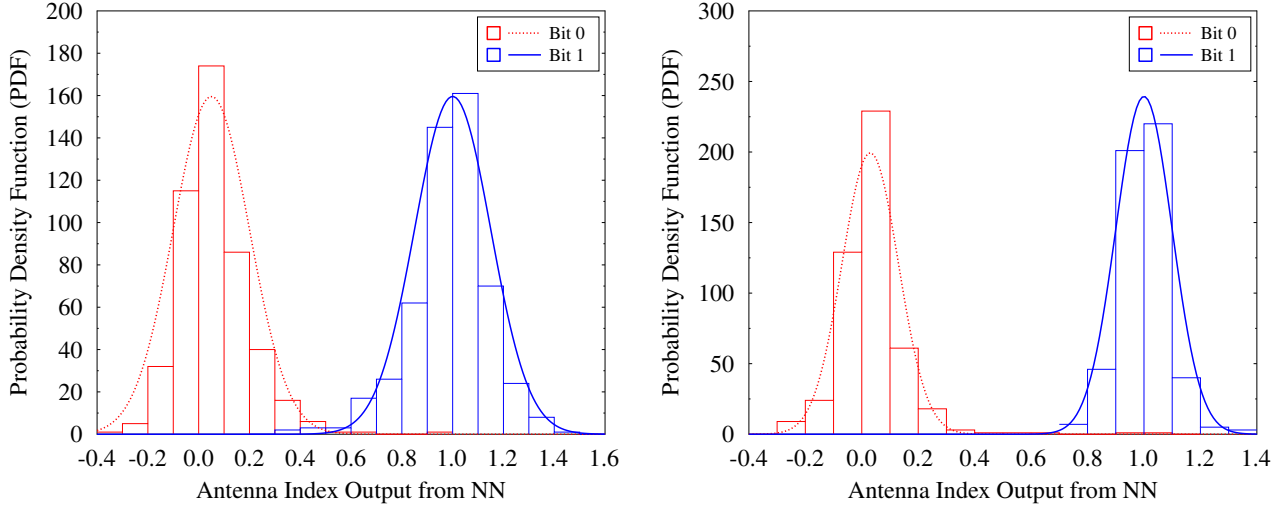
By contrast, Fig. 7(b) shows the receiver design relying on a neural network, where the training weights are designed offline. It can be seen in Fig. 7(b) that, in contrast to Fig. 7(a), this design dispenses with the channel estimation stage. More explicitly, the signal received after down-conversion is fed to an ANN, where the NN parameters learned during the training phase are applied for estimating the MS-STSK indices.

Having briefly discussed the receiver structure of both designs, let us now focus our attention on the complexity quantified in terms of search space volume and the number of computations. Let us consider again the MS-STSK symbol of (5) as a ‘toy’ example, where there are  $M_Q$  dispersion matrices,  $M_c$  complex-valued symbols and  $N_c$  ACs. In this scenario, the ML detection of (18) has to estimate both the index of the dispersion matrix and of the M-QAM symbol, as well as of the AC. Thus, the

run-time complexity relying on ML detection would be on the order of  $\mathcal{O}(M_c M_Q N_c)$ . Furthermore, the ML detection requires CSI knowledge which relies on pilots and imposes additional complexity during the channel estimation stage of Fig. 7 (a), while also significantly reducing the data rate because of the pilot overhead. In contrast to the ML detection, our learning assisted detection improves the data rate by eliminating the pilot overhead. In other words, once trained the neural network dispenses with the CSI. This philosophy makes our design spectral-efficient. Furthermore, the pre-determined parameters of the network learned during the training process allows us to estimate the indices of (5) with a high integrity as we will show later in Section V.

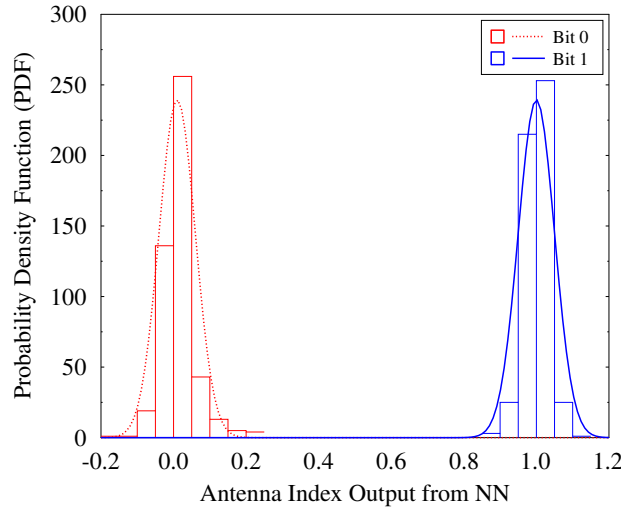
On the other hand, the complexity of the proposed design depends on the number of neurons in each hidden layer. More explicitly, the complexity of a typical NN is jointly determined by the forward propagation, and backward propagation. To elaborate a little further, let us assume that there are  $n$  neurons in a hidden layer. Let us also assume that the input and output vectors are of sizes  $n_i$  and  $n_o$ , respectively. We know that for each layer’s activating function of (29) is computed using the network parameters of the respective layer. In other words, the pre-determined weight matrix and bias vector values are substituted in the activating function of (29) relying on the input vector  $n_i$  for computing the intermediate output  $n_o$ , which serves as the input of the next layer. In (29),  $x$  is the input,  $a$  is the weight and  $v$  is the bias. Contrasting it to the ML receiver’s search complexity by considering each search operation as a node of Fig. 6, the complexity of the proposed design would be on the order of  $\mathcal{O}(n_i n_{h_1} n_{h_2} n_o)$ .<sup>8</sup> It is important to emphasize

<sup>8</sup>Note that this is only to check how it compares to the ML detection, albeit there is no search complexity associated with the NN aided detection.



(a) The PDF of antenna index estimate at SNR 0 dB.

(b) The PDF of antenna index estimate at SNR 5 dB.



(c) The PDF of antenna index estimate at SNR 15 dB.

**FIGURE 9:** Probability distribution of the antenna index estimated at the output of the NN for (a) SNR 0 dB (b) SNR 5 dB (c) SNR 15 dB.

that in contrast to the ML of Fig. 7(a), this design does not require additional computations for channel estimation and also avoids the pilot overhead.

Let us now delve into the complexity in terms of the number of complex multiplications for both designs. The total number of complex multiplications required by ML detection for the transmission parameters of (5) is  $\mathcal{O}(N_t N_r N_c M_Q T^2) + \mathcal{O}(N_c M_Q^3)$ , while it is  $\mathcal{O}(N_b N_t N_r N_c M_Q T^2) + \mathcal{O}(N_b N_c M_Q^3)$  for (24). By contrast, for the NN associated with the aforementioned parameters, the number of multiplications is  $\mathcal{O}(n_i n_{h_1}) + \mathcal{O}(n_{h_1} n_{h_2}) + \mathcal{O}(n_{h_2} n_o)$ .

Table 2 illustrates the number of complex multiplications required for (18) with the simulation parameters listed in Table 3.

**TABLE 2:** Computational complexity in terms of the number of complex-valued multiplications.

Design	Multiplications
Proposed ANN Detection for (5)	330
ML-Aided Detection for (5)	384
Proposed ANN Detection for (24)	1146
ML-Aided Detection for (24)	768

**TABLE 3:** Simulation Parameters.

Parameters	Values
Number of AEs in each AA at Tx ( $K$ )	64
Number of AEs in each AA at Rx ( $L$ )	16
Number of AAs at Tx ( $N_t$ )	4
Number of AAs at Rx ( $N_r$ )	2
Size of QAM symbol ( $M_c$ )	4
Number of dispersion matrices $M_Q$	4
Size of consecutive AA selected ( $M$ )	2
Time slots ( $T$ )	2
Number of clusters ( $N_c$ )	1 & 2
Number of rays ( $N_{ray}$ )	1
Number of beams ( $N_b$ )	2
AoA ( $\phi_r$ )	variable
$\phi_t$	variable

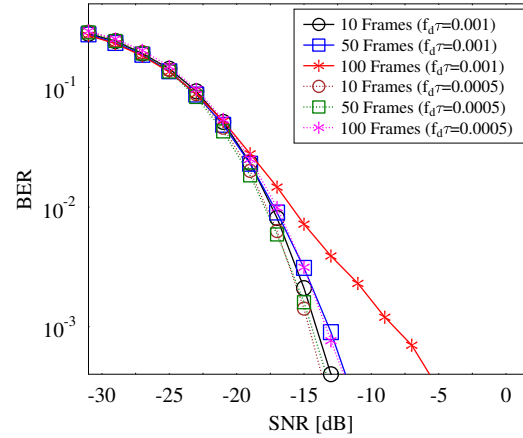
## V. SIMULATIONS

In this section we present our simulations characterizing the performance of the proposed design and of the ML detection. More particularly, we performed Monte Carlo simulations for comparing the performance of the learning-assisted detection and of the ML-aided detection. Our simulation parameters are listed in Table 3. Furthermore, in our simulations, we have configured the NN to have 2 hidden layers with 30 neurons each, where we trained the NN using 1000 samples and the classic stochastic gradient descent technique having a step size of 0.005 for 92 epochs.

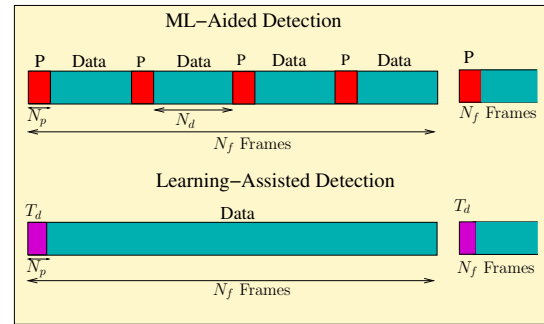
Fig. 9 shows the probability density function (PDF) of the antenna index estimate generated from the output of the NN for the system model of Fig. 3(a). The plots shown in Fig. 9 are generated during the testing phase of the NN for SNRs of 0, 5 and 15 dB. To obtain this plot, we have set the number of hidden layers to 2, where the NN underwent training using 2000 samples. Additionally, we note that the input vector of the NN of Fig. 6 takes only real values as discussed in Section III, hence the received vector  $y$  is split into its real and imaginary parts. In this setting, we empirically observed that the NN makes an accurate inference between the output and the input vectors, when the number of neurons is set to 12 and 13 in hidden layers 1 and 2, respectively for both the real and imaginary parts of the network.

In this setting, Fig. 9(a) shows the PDF of the antenna index output from the NN for an SNR of 0 dB. Ideally, the estimate of the output representing the antenna index is expected to be either '0' or '1', but it can be seen from the figure that the output of the NN is not exactly binary but a set of continuous values spanning from -0.4 to 1.5. More explicitly, for the bit '0' the output ranges from -0.4 to 0.5; while for the bit '1', the output of the NN is a real value between 0.3 to 1.6. It is observed empirically that the output for the bit '0' follows a near-Gaussian distribution with a mean of 0 and variance of 0.15. Similar trends are valid for the bit '1'. However, as the SNR increases from 0 to 5 to 15, as shown in Figures 9(b) and 9(c), it may be observed that the variances of the distributions reduce gradually. In other words, the range of values seen for an SNR of 0 dB

becomes narrower as the SNR increases, which results in a confident estimate of the antenna index. To understand the behavior of the NN in terms of index estimation, we have analyzed the output of the antenna index, as an example. The PDFs for the other indices can also be readily obtained.



**FIGURE 10:** BER of the proposed design for different number of frames as channel evolves in time according to Jakes' correlation coefficient (31). The simulation parameters are listed in Table 3.



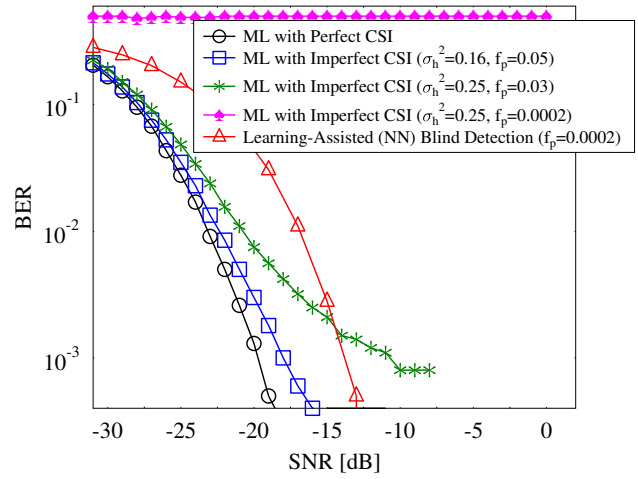
**FIGURE 11:** Pilots transmission for the proposed learning-assisted design and conventional design. In ML-aided detection, pilots  $P$  are transmitted in every frame which significantly affects the spectral-efficiency, while training data  $T_d$  is requested by the user only after  $N_f$  frames which is contingent on the Doppler spread. Note that the channel is not time-invariant over  $N_f$  frames. The channel is different for every frame according to the Doppler frequency. However, for NN assisted designs, the same training weights can be employed for the duration of  $N_f$  frames for successful symbol detection — thanks to the DNN.

Fig. 10 analyzes the BER of the proposed design for different number of frames as channel evolves in time according to Jakes' correlation coefficient of (31). It is



evident from Fig. 10 that as the number of frames increases from 10 to 100, the BER of the proposed design degrades. This phenomenon is observed because the training weights designed during the first few frames become outdated after a certain number of frames; hence retraining the NN parameters becomes necessary. We note that as expected, the number of frames transmitted before the NN weights become outdated directly depends on the Doppler spread. For example, in Fig. 10, the number of frames transmitted before the BER starts to degrade for the normalized Doppler spread of 0.0005 is higher than for 0.001. Therefore, in this scenario, the receiver requests the BS to transmit pilots to recalibrate its weights depending on the BER observed. Fig. 11 shows the schematic of the pilot transmission for both the learning-assisted design and ML detection. It is important to emphasize that the pilots are transmitted for every single frame in the case of ML detection. By contrast, our learning-assisted detection requires the training data for recalibrating the NN weights only after every  $N_f$  frames, as shown in Fig. 11, while performing detection without explicit CSI in the rest of the frames. This can also be interpreted as online learning. More explicitly, using the loss function of (30), the weights may be computed. Note that at the commencement of the process, random weights are used. By contrast, during retraining, initial weights are set to the previously trained weights. In this way, the NN adjusts its weights. The length of the known sequence depends on the Doppler frequency — and it is determined empirically. It is of salient importance to note that the amount of training data required for recalibration decreases with the reduction of the Doppler spread. The overhead involved in retraining the NN network of our design is proportional to  $N_p / [N_f(N_d + N_p)]$ , where  $N_f$  is the number of frames,  $N_d$  is the number of data streams, and  $N_p$  is the number of pilots; while the pilot overhead involved in the channel estimation for ML detection is  $N_p N_f / [N_f(N_d + N_p)]$ .

Fig. 12 shows the BER of both the learning-assisted detection and of the ML-aided detection with perfect CSI, as well as of the ML detection with imperfect CSI for the MS-STSK transmission dispensing with the BIM. Since no BF index is considered, it can be assumed that the channel supports only a single beam, or all potential beams are utilized for the transmission. It can be seen in Fig. 12 that for the aforementioned NN parameters, there is around 6 dB gap at the BER of  $10^{-3}$  between the learning-assisted detection and the ML-aided detection relying on perfect CSI. Although the ML-aided detection relying on perfect CSI outperforms the learning-assisted detection by 6 dB SNR gain, this is achieved under the idealized simplifying assumption of having perfect CSI. On the other hand, for the CSI estimate having an error variance of 0.16, which is obtained by substituting  $f_p = 0.05$  in (23), the SNR gap reduces to around 3 dB. Furthermore, the ML-



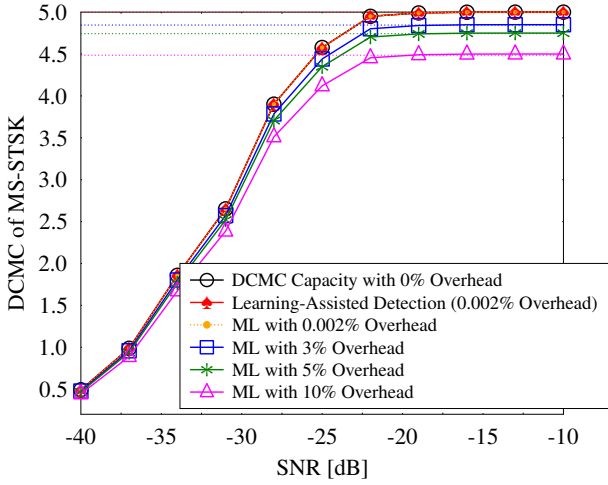
**FIGURE 12:** BER of the MS-STSK transmission. In this design the NN estimates indices of antenna, dispersion matrix and the symbol. The simulation parameters are listed in Table 3.

aided detection produces an error floor<sup>9</sup> for the CSI error variance of 0.25, when  $f_p = 0.03$ . By contrast, the learning-assisted detection remains capable of accurately estimating the indices of the MS-STSK transmission regardless of the nature of the CSI, while also circumventing pilot-assisted channel estimation all together.<sup>10</sup>

While Fig. 12 shows an SNR gap between the proposed scheme and ML detection, it is also pertinent to study the effective throughput of both designs for the sake of fairness, since the SNR gain observed for ML detection in Fig. 12 is critically contingent on the CSI estimation accuracy, which increases proportionally to the pilot density. However, increasing the pilot density would reduce the effective throughput of the design. Therefore, Fig. 13 characterizes the Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of the learning-assisted detection and of the ML detection for different pilot overheads. It can be seen in the figure that the capacity of the design is strictly governed by the pilot overhead. More explicitly, for the simulation parameters summarized in Table 3, the capacity of the ML detection for 3% pilot-overhead is limited to a maximum value of 4.85 [bpcu], while it is 4.75 [bpcu] for 5% pilot overhead. Furthermore, when the pilot

<sup>9</sup>Note that we have used LMMSE for channel estimation, since it is employed LTE. It is instructive to note that employing other sophisticated channel estimation techniques may also increase the complexity of the design, which defeats the purpose of our work.

<sup>10</sup>We emphasize here that the performance of DNN method cannot be better than the conventional ML method. It is also instructive to note that the loss function computed for designing the DNN weights is itself deduced from the ML method. The purpose of using DNN method is to reduce the detection complexity. Note that the DNN method can beat only in the face of channel impairments.

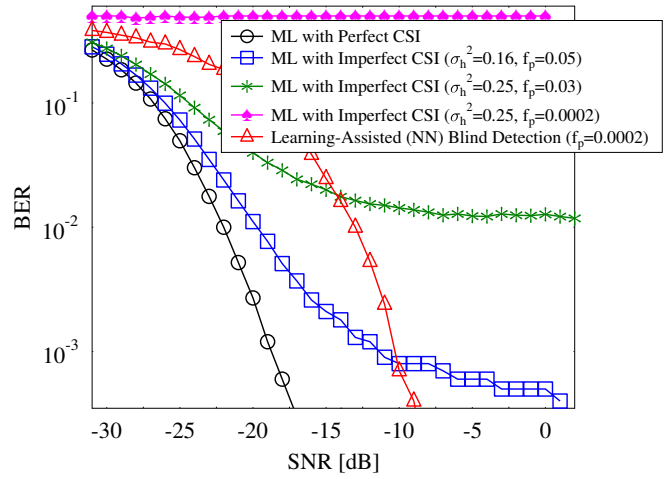


**FIGURE 13:** Discrete-input Continuous-output Memoryless Channel capacity of the learning-assisted detection and of the ML-aided detection with different pilot overheads for channel estimation when MS-STSK transmission with beamforming is considered. The simulation parameters are listed in Table 3.

overhead is increased to 10%, the DCMC capacity of the ML detection is reduced to a maximum value of 4.5 [bpcu]. By contrast, the DCMC capacity of the learning-assisted detection reaches the maximum value of 4.99 [bpcu], which is close to the DCMC capacity of 5 [bpcu], since the overhead involved in recalibrating the weights is marginally lower<sup>11</sup>. Additionally, we note that there is a SNR gain of 0.5 dB and 1 dB for our learning-assisted detection over the ML-detection with 5% overhead when aiming for achieving a rate of 3 [bpcu] and 4 [bpcu], respectively. This gain is around 1 dB and 3 dB for the learning-assisted detection over the ML-detection with 10% overhead at the rate of 3 [bpcu] and 4 [bpcu], respectively.

Fig. 14 shows the BER of both the learning-assisted detection and of the ML-aided detection with perfect CSI as well as of the ML-aided detection with imperfect CSI for MS-STSK transmission in conjunction with BIM. In this setting, we assumed that the channel supports two beams for each AA of Fig. 3(b). In other words, in Fig. 3(b) there is an extra index to be estimated, which is the beam index. In this scenario, it is empirically observed during the training phase that the number of neurons has to be set to 30 for

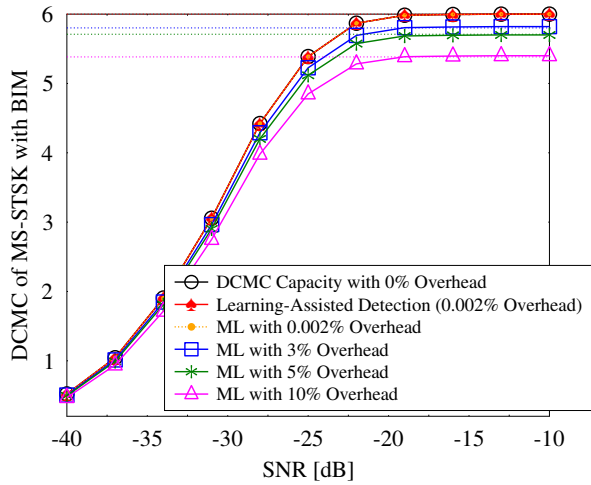
<sup>11</sup>For example, for 200 training samples, we can transmit 100 frames, where each frame has 100 symbols. In other words, we do have to retrain the neural network depending on how rapidly the propagation environment is fluctuating, but again, our method reduces the overall overhead. Boldly and explicitly, the main philosophy of our design is to reduce both the overhead and detection complexity. However, we note that the number of training symbols required for retraining for the transmission of say N frames, remains lower, since we only update the weights based on our already designed weights.



**FIGURE 14:** BER of the MS-STSK transmission amalgamated with the BIM. The simulation parameters are listed in Table 3.

both the real and imaginary constituents of the NN. It can be seen from Fig. 14 that adding an additional index for estimation increased the SNR gap between the learning-assisted detection and the ML detection using perfect CSI to 8 dB. Again, the superior performance of the ML detection is because of the unrealistic assumption of having perfect CSI. However, when a CSI estimate associated with the error variance of 0.15 is introduced, the ML-aided detection starts to produce an error floor from around -10 dB, while the BER remains flat for the CSI error variance of 0.25. On the other hand, despite the absence of CSI, the learning-assisted detection estimates both the MS-STSK indices and the beam index with high integrity.

Fig. 15 characterizes the DCMC capacity of the proposed design and of the ML design when MS-STSK transmission is amalgamated with the BIM. In this simulation, it is assumed that the channel supports two beams and BIM is employed for exploiting these two beams, where only one beam is activated depending on the input bit-sequence. It is evident from the figure that the DCMC capacity of the ML detection is inferior to that of the learning-assisted detection. This is because of the overhead imposed by the pilots used for channel estimation to aid the ML detection process. This becomes especially pronounced, when the pilot overhead is set to 10% as seen in the figure, where the DCMC capacity is less than 5.5 [bpcu], while that of learning-aided detection is 6 [bpcu]. In other words, the necessity of having pilots for estimating the CSI partly consumes the physical resources, thereby reducing the effective capacity of the system. However, this behavior is avoided by the learning-assisted system, since it achieves accurate symbol detection at a retraining overhead as low as 0.002%. Furthermore, we



**FIGURE 15:** Discrete-input Continuous-output Memoryless Channel capacity of the learning-assisted detection and of the ML-aided detection with different pilot overheads for channel estimation when MS-STSK transmission with BIM is considered. The simulation parameters are listed in Table 3.

observe an SNR gain of 3 dB at the rate of 5 [bpcu] for our learning-assisted detection over the ML detection at 10% overhead, while it is around 2 dB at the rate of 4 [bpcu].

## VI. CONCLUSIONS

In this paper, we proposed deep learning assisted detection for index modulation aided mmWave systems, where we trained a NN for estimating the symbol indices without relying on explicit CSI knowledge. As a design example, we first employed MS-STSK and then we analyzed our design when MS-STSK transmission is considered in conjunction with BIM. In contrast to the MS-STSK's conventional ML-detector, our learning-assisted detection dispenses with the channel estimation stage, which makes it more spectral-efficient than ML detection. We demonstrated by simulations that the learning-assisted detection outperforms ML detection in the face of realistic channel impairments. Furthermore, we show by simulations that ML detection produces an error floor, when the MS-STSK transmission is coupled with BIM. Additionally, we presented qualitative discussion on the receiver complexity in terms of both its search space as well as the number of computations required.

## REFERENCES

- [1] I. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions and link-budget," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, 2018.
- [2] A. V. Lopez, A. Chervyakov, G. Chance, S. Verma, and Y. Tang, "Oppor-

- tunities and challenges of mmwave NR," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 4–6, 2019.
- [3] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov 2005.
- [4] K. Satyanarayana, M. El-Hajjar, P. H. Kuo, A. Mourad, and L. Hanzo, "Dual-function hybrid beamforming and transmit diversity aided millimeter wave architecture," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2798–2803, March 2018.
- [5] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct 1998, pp. 295–300.
- [6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct 1998.
- [7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
- [8] R. W. Heath and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2429–2441, Oct 2002.
- [9] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, July 2008.
- [10] M. El-Hajjar, O. Alamri, J. Wang, S. Zummo, and L. Hanzo, "Layered steered space-time codes using multi-dimensional sphere packing modulation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3335–3340, July 2009.
- [11] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the mimo era: Evolution of standards, waveform design, and future perspectives," *Proc. of the IEEE*, vol. 99, no. 8, pp. 1343–1385, Aug 2011.
- [12] S. Sugiura, S. Chen, and L. Hanzo, "Space-time shift keying: A unified MIMO architecture," in *Proc. Globecom*, Dec 2010, pp. 1–5.
- [13] P. Zhang, S. Chen, and L. Hanzo, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE transactions on communications*, vol. 61, no. 5, pp. 1902–1913, 2013.
- [14] 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.
- [15] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Layered multi-group steered space-time shift-keying for millimeter-wave communications," *IEEE Access*, vol. 4, pp. 3708–3718, 2016.
- [16] —, "Multi-set space-time shift-keying with reduced detection complexity," *IEEE Access*, vol. 4, pp. 4234–4246, 2016.
- [17] Y. Ding, V. Fusco, A. Shitvov, Y. Xiao, and H. Li, "Beam index modulation wireless communication with analog beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6340–6354, July 2018.
- [18] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1702–1715, 2020.
- [19] J. Xia, K. He, W. Xu, S. Zhang, L. Fan, and G. K. Karagiannidis, "A MIMO detector with deep learning in the presence of correlated interference," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4492–4497, 2020.
- [20] L. Hanzo, L.-L. Yang, E. Kuan, and K. Yen, *Single-and multi-carrier DS-CDMA: multi-user detection, space-time spreading, synchronisation, standards and networking*. John Wiley & Sons, 2003.
- [21] Y. Liu, L.-L. Yang, and L. Hanzo, "Sparse space-time-frequency-domain spreading for large-scale non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, 2020.
- [22] M. Wen, B. Zheng, K. J. Kim, M. Di Renzo, T. A. Tsiftsis, K.-C. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging wireless systems: Research progresses and applications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 1949–1972, 2019.
- [23] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.
- [24] 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.



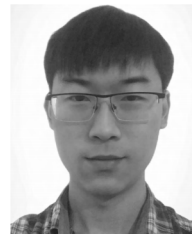
- [25] M.-S. Yee, B. L. Yeap, and L. Hanzo, "Radial basis function-assisted turbo equalization," *IEEE Transactions on Communications*, vol. 51, no. 4, pp. 664–675, 2003.
- [26] S. X. Ng, M.-S. Yee, and L. Hanzo, "Coded modulation assisted radial basis function aided turbo equalization for dispersive rayleigh-fading channels," *IEEE transactions on wireless communications*, vol. 3, no. 6, pp. 2198–2206, 2004.
- [27] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.
- [28] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2503–2516, 2019.
- [29] 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*, pp. 1–1, 2020.
- [30] L. Huang, H. Zhang, R. Li, Y. Ge, and J. Wang, "AI coding: Learning to construct error correction codes," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 26–39, 2019.
- [31] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, 2020.
- [32] T. Van Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, "Deep energy autoencoder for noncoherent multicarrier MU-SIMO systems," *IEEE Transactions on Wireless Communications*, 2020.
- [33] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multidimensional signal sets," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 528–536, March 2006.
- [34] A. Vakili, M. Sharif, and B. Hassibi, "The effect of channel estimation error on the throughput of broadcast channels," in *Proc. ICASSP*, vol. 4, May 2006, pp. IV–IV.
- [35] C. Xu, J. Zhang, T. Bai, P. Botsinis, R. G. Maunder, R. Zhang, and L. Hanzo, "Adaptive coherent/non-coherent single/multiple-antenna aided channel coded ground-to-air aeronautical communication," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1099–1116, Feb 2019.
- [36] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1462–1470, March 2014.
- [37] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.
- [38] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *CoRR*, vol. abs/1710.02913, 2017. [Online]. Available: <http://arxiv.org/abs/1710.02913>
- [39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [40] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [41] Y. Chauvin and D. E. Rumelhart, Eds., *Backpropagation: Theory, Architectures, and Applications*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1995.
- [42] W. C. Jakes and D. C. Cox, Eds., *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.



**K. SATYANARAYANA** ([www.satyanarayana.xyz](http://www.satyanarayana.xyz)) received his B. Tech. degree in Electrical Engineering from Indian Institute of Technology Madras, India, in 2014. During Jul'14-Aug'15, he worked as a research assistant at Indian Institute of Science, Bangalore. Currently, Satya is a research scholar in Wireless Communications at the University of Southampton in liaison with InterDigital Europe, London, UK. His research interests include millimeter wave communications, HBF, with an emphasis on transceiver algorithms for wireless communication systems and multi-functional MIMO.



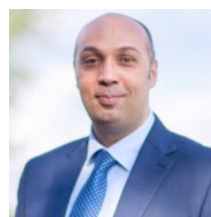
**LUPING XIANG** received the B.Eng. (Hons.) degree from Xiamen University, China, in 2015 and the Ph.D. degree from University of Southampton, in 2020. He is currently a research fellow with Next Generation Wireless Group, University of Southampton. His research interests include mMTC, URLLC, machine learning and channel coding.



**YANQING ZHANG** received the B.Eng. degree in telecommunications engineering with management from the Beijing University of Posts and Telecommunications, Beijing, China, the B.Sc. degree (Eng.) (Hons.) in telecommunications engineering with management from the Queen Mary University of London, London, U.K., in 2014, and the M.Sc. degree (Hons.) in wireless and optical communications from University College London, in 2015. He is currently pursuing the Ph.D. degree with the Next Generation Wireless Research Group, School of Electronics and Computer Science, University of Southampton. His research interests include joint source-channel coding, robust wireless video streaming, and multiview video coding.



**MOHAMMED EL-HAJJAR** is an Associate Professor in the department 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 80 journal and conference papers. Mohammed's research interests include the design of intelligent and energy-efficient transceivers, MIMO, millimeter wave communications, cross-layer optimization for large-scale networks and Radio over fiber network design.



**ALAIN A. M. MOURAD** holds a PhD degree in Telecommunications from ENST Bretagne in France. He has over 15 years' experience in the wireless networks industry. He is currently leading the research and development of Next Generation Radio Access Networks at InterDigital International Labs (London, Berlin, Seoul). Prior to joining InterDigital, Dr. Mourad was a Principal Engineer at Samsung Electronics R&D (UK) and previously a Senior Engineer at Mitsubishi Electric R&D Centre Europe (France). Throughout his career, Dr. Mourad has been active in the research and standardization of recent communication networks (5G/4G/3G) and broadcasting systems (ATSC 3.0/DVB-NGH/DVB-T2). He has held various leadership roles in the industry, invented over 35 granted patents and several other patent applications, and authored over 50 peer-reviewed publications. He received the Inventor of the Year Award from Samsung Electronics R&D (UK) twice in 2012 and 2013, and in 2016 InterDigital Innovation Award for the "idea, creation, and execution of InterDigital Europe".



**LAJOS HANZO** (<http://www-mobile.ecs.soton.ac.uk>, [https://en.wikipedia.org/wiki/Lajos\\_Hanzo](https://en.wikipedia.org/wiki/Lajos_Hanzo)) FIEEE, FREng, FIET, Fellow of EURASIP, DSc holds an honorary doctorate by the Technical University of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served as Governor of both IEEE ComSoc and of VTS. He has published 1900+ contributions at IEEE

Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 119 PhD students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry.

• • •