Digital Object Identifier

Soft-Decoding for Multi-Set Space-Time Shift-Keying mmWave Systems: A Deep Learning Approach

K. SATYANARAYANA[1], (Student Member, IEEE), MOHAMMED EL-HAJJAR[1], (SENIOR MEMBER, IEEE), ALAIN MOURAD[2], PHILIP PIETRASKI[2], AND LAJOS HANZO[1], (FELLOW, IEEE)

[1]University of Southampton, Southampton SO17 1BJ, U.K
[2]InterDigital Inc., London EC2A 3QR, U.K.

This work was supported by InterDigital as well as in part by the EPSRC under Project EP/Noo4558/1 and Project EP/PO34284/1, in part by the Royal Society's Global Research Challenges Grant, and in part by the European Research Council's Advanced Fellow Grant QuantCom.

ABSTRACT In this paper, we propose a deep learning assisted soft-demodulator for multi-set space-time shift keying (MS-STSK) millimeter wave (mmWave) systems, where we train a neural network (NN) to provide the soft values of the MS-STSK symbol without relying on explicit channel state information (CSI). Thus, in contrast to the conventional MS-STSK soft-demodulator which relies on the knowledge of CSI at the receiver, the learning-assisted design circumvents the channel estimation while also improving the data rate by dispensing with pilot overhead. Furthermore, our proposed learning-aided soft-demodulation substantially reduces the number of cost-function evaluations at the output of the MS-STSK demodulator. We demonstrate by simulations that despite avoiding CSI-estimation and the pilot overhead, our learning-assisted design performs closely to the channel-estimation aided design assuming perfect CSI for BER $< 10^{-4}$, whilst imposing a low complexity. Furthermore, we show by simulations that upon using realistic imperfect CSI at the receiver employing conventional soft-demodulation, the learning-aided soft-demodulator outperforms the conventional scheme. Additionally, we present quantitative discussions on the receiver complexity in terms of the number of computations required to produce the soft values.

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
BLAST	Bell-Labs Layered Space-Time
BS	Base Station
CSI	Channel State Information
DCMC	Discrete-input Continuous-output Memoryless Channel
DM	Dispersion Matrix
EXIT	EXtrinsic Information Transfer
FDD	Frequency Division Duplex
LLR	Log-Likelihood Ratio
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
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 rapid proliferation of cellular devices has led to the requirement of providing massive wireless connectivity for a huge number of users, each having data rate demands [1]. However, given the dearth of bandwidth for accommodating a large number of users at sub-6GHz frequencies, wireless communication researchers have turned their attention to millimeter wave frequencies spanning from 28 GHz to 300 GHz [1]. Whilst utilizing mmWave frequencies looks promising, they impose a challenge in terms of coverage, since mmWave frequencies are highly susceptible to blockages. Furthermore, mmWave frequencies suffer from high propagation losses because of oxygen absorption, rain-induced fading and foliage attenuation. Therefore, to mitigate the propagation loss, high gain directional transmission, popularly referred to as beamforming, has to be harnessed [2]. Typically, beamforming gain is achieved by the employment of large antenna arrays (AA), where the antenna elements are separated by half a wavelength.

To further enhance the data rate, multiple-input multiple-output (MIMO) transmission in conjunction with beamforming may be introduced. There is a vast body of literature focusing on MIMO-aided transmission, as exemplified by Bell-Labs Layered Space-Time (BLAST) transmission, where a high spatial multiplexing gains may be achieved [3]. Another variant of MIMO-aided transmission which is the quintessential prerequisite for achieving diversity gains is space-time block coding (STBC) [4], [5]. With BLAST, STBC and beamforming as the building blocks, a novel MIMO transmission scheme has been conceived by the authors of [6], which is referred to as multi-functional MIMO (MF MIMO). The rationale of MF MIMO is to exploit the multiple features of MIMO transmission, such as multiplexing, diversity as well as beamforming for improving the throughput of the communication link. More particularly, El-Hajjar *et al.* conceived layered steered space-time coding (LSSTC) [7] by amalgamation of V-BLAST, STBC and BF in order to achieve the aforementioned gains. Recently, Satyanarayana *et al.* [2] proposed a MF MIMO for mmWave systems that exhibits dual functionality by providing both diversity and beamforming gains relying on the concept of sub-arrays.

A new arrival amongst the other MF MIMO techniques is space-time shift keying (STSK) [8], [9], which strikes a trade-off between multiplexing and diversity gains. More explicitly, the STSK design relies on the generalization of a technique referred to as spatial modulation (SM) [10], where at any given time a single radio-frequency (RF) chain is activated. If the information in SM transmission is implicitly conveyed by the index of the RF chain and the complex-valued signal, in STSK the information is conveyed by the classic complex-valued modulated signal as well as by the index of the dispersion matrix (DM), which spans several RF chains. Hemadeh *et al.* [9] extends the STSK philosophy by amalgamating SM and STSK to conceive multi-set (MS) STSK transmission. This design, as a descendant of the SM and STSK schemes, conveys information implicitly by the indices of the complex-valued signal, the dispersion matrix and the RF chain. Thus, this design is capable of further improving the data rate.

However, an impediment of index modulation systems such MS-STSK is their high search complexity at the receiver [11], since they have to perform an exhaustive search among all the possible combinations of the MS-STSK symbol. Furthermore, like in any typical communication system, its receiver requires accurate channel state information (CSI) in order to be able to decode the signal at a high integrity [12]. In frequency division duplex (FDD) systems, pilots are appended to the data frame for estimating the CSI at the receiver. Consequently, this design results in a reduced effective data rate and additionally imposes extra complexity by the channel estimation prior to the detection process.

To circumvent the problem of both the pilot overhead and channel estimation, a deep learning approach may be conceived, where the MS-STSK information bits can be decoded without explicit knowledge of the CSI. This philosophy makes the design spectral-efficient while also significantly reducing the complexity involved both in channel estimation and detection. This design can be interpreted as blind detection, as it turns a 'blind eye' to the CSI while providing soft information about the MS-STSK codeword.

At the time of writing, deep learning aided wireless solutions have gained significant attention as a benefit of their ability to learn patterns and detect trends. Samuel et al. [13], [14] advocated neural network (NN) assisted MIMO detection for a time-invariant channel. Dorner et al. [15] conceived a point-to-point communication system relying on NNs using asynchronous off-the-shelf software-defined radios. Furthermore, Jin [16] et al. proposed sub-optimal deep learning based detection for MIMO relay channels. In the context of channel decoding, Nachmani et al. [17] demonstrated that NN-aided belief propagation improves the performance at a reduced complexity, while Liang et al. [18] analyzed the performance of NN based iterative belief propagation. The theoretical model proposed by Yan [19] et al. for symbol detection using autoencoders may be expected to find further beneficial applications.

Against this backdrop, our contributions are as follows.

1) We propose a deep learning assisted soft-demodulator for multi-set space-time shift keying (MS-STSK) assisted millimeter wave (mmWave) systems, where we train a NN for providing the soft values of the MS-STSK symbol without relying on explicit CSI.
2) Our simulations demonstrate that despite dispensing with the CSI, our learning-assisted design performs closely to the conventional design using perfect CSI for BER $< 10^{-4}$. This is achieved at a low complexity.

Furthermore, our simulations demonstrated that upon using realistic imperfect CSI at the receiver for conventional soft-demodulation, the learning-aided soft-demodulator can outperform the latter.

3) Additionally, we quantify the receiver complexity in terms of the number of computations required to produce the soft values.

The rest of the paper is organized as follows. In Section II, we detail the system model of MS-STSK amalgamated beamforming in mmWave communication, while in Section III we present our proposed learning-assisted detector followed by quantifying the complexity of the design. Finally, Section IV and Section V 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 $(.)^T$, $(.)^H$, $\|.\|_{\mathrm{F}}$, $\mathrm{Tr}(.)$ $\mathbb{E}(.)$ for the transpose, Hermitian transpose, Frobenius norm, trace and expectation operator, respectively. We adopt $A(m,n)$ to denote the m^{th} row and n^{th} column of the $\mathbf{A}$, 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.

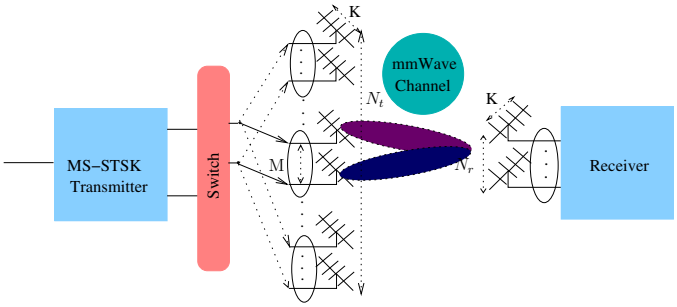


FIGURE 1: Block diagram of a typical MS-STSK point-to-point single-user link.

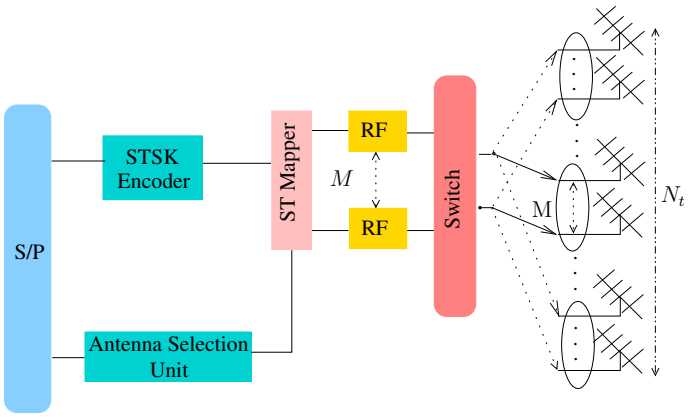


FIGURE 2: Block diagram of the MS-STSK modulator.

II. SYSTEM MODEL

Let us consider the single-user point-to-point link of Fig. 1, where the transmitter and receiver are equipped with N_t and N_r antenna arrays (AAs) of K antenna elements (AEs), respectively. Furthermore, the transmitter has M RF chains, where at any given time, only M AAs out of the total of N_t transmit antennas are activated, as illustrated in Fig. 1. The transmitter of Fig. 1 maps the input stream of bits to an MS-STSK codeword, where the information is implicitly conveyed by the indices of the activated transmit antennas, of the dispersion matrix and of the complex-valued QAM symbol. A typical MS-STSK encoder is shown in Fig. 2, where the input stream of N_b bits is mapped to a MS-STSK codeword comprising the index of one of the $N_c = \binom{N_t/M}{1}$ transmit antenna combinations (ACs), the index of the M_Q dispersion matrices as well as the index of the M_c complex-valued QAM symbols. The resultant codeword X transmitted by the MS-STSK encoder is expressed as:

$$X = A_{q,c} x_l, \tag{1}$$

where $A_{q,c}$ is the q^{th} dispersion matrix selected for the c^{th} AC and x_l is the l^{th} QAM symbol. Thus, by employing MS-STSK transmission, a total of $\log_2(N_c M_Q M_c)$ bits are used for conveying the information.

To elaborate further, let us consider a 'toy' example, where we assume $N_c = 2$, $M_Q = 4$ and $M_c = 4$. In other words, the transmitter is equipped with 2 ACs and transmitting 4-QAM symbols with the aid of 4 dispersion matrices. Thus, given the above MS-STSK parameters, a total of $\log_2(2 \times 4 \times 4) = 5$ bits can be transmitted. As a design example, let us assume having the input bit sequence of '01101'. Then, the mapping of the bit sequence to MS-STSK is carried out as follows. By considering the least significant bit first, the bit '1' is mapped to the transmit antenna index, while the ensuing bits '10' are mapped to one of the four dispersion matrices. And finally, the bits '01' are mapped to a 4-QAM symbol.

Thereafter, the MS-STSK codeword of (1) is transmitted over mmWave channel by employing beamforming. Then, the block-based received signal $\mathbf{Y}$ at the receiver after analog RF combining is given by

$$\mathbf{Y} = \mathbf{W}_{\mathrm{RF}} \mathbf{H} \mathbf{F}_{\mathrm{RF}} \tilde{\mathbf{X}} + \mathbf{V}, \tag{2}$$

where $\mathbf{V}$ is the Gaussian noise, and $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ are the analog RF beamforming and combining matrices of sizes $KN_t \times N_t$ and $N_r \times KN_r$, respectively. They are expressed as

$$\mathbf{F}_{\mathrm{RF}} = [\mathbf{0} \dots \mathbf{F}_{\mathrm{RF}_q} \dots \mathbf{0}] \in \mathcal{C}^{KN_t \times N_t}, \tag{3}$$

$$\mathbf{F}_{\mathrm{RF}_q} = \mathrm{diag}(\mathbf{F}^1_{\mathrm{RF}} \mathbf{F}^2_{\mathrm{RF}} \dots \mathbf{F}^M_{\mathrm{RF}}), \tag{4}$$

where $\mathbf{F}^i_{\mathrm{RF}}$ is the BF vector of the i^{th} AA of size $K \times 1$. Similarly,

$$\mathbf{W}_{\mathrm{RF}} = \mathrm{diag}\left(\mathbf{W}_{\mathrm{RF}_1} \dots \mathbf{W}_{\mathrm{RF}_i} \dots \mathbf{W}_{\mathrm{RF}_{N_r}}\right)^T \in \mathcal{C}^{N_r \times KN_r}, \tag{5}$$

where $\mathbf{W}_{\text{RF}}^{i}$ is the BF vector of the i^{th} AA of size $K \times 1$. Furthermore, $\mathbf{H}$ is the channel matrix of size $KN_r \times KN_t$ expressed as

$$\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_{N_c}], \tag{6}$$

where $\mathbf{H}_i$ is the sub-channel matrix of size $N_r K \times MK$, which is expressed as

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{H}_1^1 & \dots & \mathbf{H}_1^m & \dots \mathbf{H}_1^M \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}_j^1 & \dots & \vdots & \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}_{N_r}^1 & \vdots & \vdots & \mathbf{H}_{N_r}^M \end{bmatrix},$$

while $\mathbf{H}_j^m$ is the mmWave channel matrix of size $K \times K$ spanning between the j^{th} AA at the receiver and the m^{th} AA at the transmitter. In this paper, we consider a statistical channel model having N_c clusters with N_{ray} each, which is given by

$$\mathbf{H}_j^m = \sqrt{\frac{1}{N_c N_{\text{ray}}}} \sum_{n_c=1}^{N_c^{\{m,j\}}} \sum_{n_{\text{ray}}=1}^{N_{\text{ray}}^{\{m,j\}}} \alpha_{n_c}^{n_{\text{ray}}} \mathbf{a}_r(\phi_{n_c}^{n_{\text{ray}}}) \mathbf{a}_t^T(\phi_{n_c}^{n_{\text{ray}}}). \tag{7}$$

For a uniform linear AAs, the response vectors $\mathbf{a}_r$ and $\mathbf{a}_t$ are given by

$$\mathbf{a}_r(\phi_r) = [1 \; e^{j\frac{2\pi}{\lambda} d\cos(\phi_r)} \dots e^{j\frac{2\pi}{\lambda}(K-1)d\cos(\phi_r)}]^T, \tag{8}$$

$$\mathbf{a}_t(\phi_t) = [1 \; e^{j\frac{2\pi}{\lambda} d\cos(\phi_t)} \dots e^{j\frac{2\pi}{\lambda}(K-1)d\cos(\phi_t)}]^T, \tag{9}$$

where ϕ_r and ϕ_t are the angles of arrival and departure, respectively.

Remark: Note that in our design learning-aided fingerprint based beam-alignment may be invoked for AoA-AoD information. Once the beam-alignment is carried out, we assume that the channel impulse response (α) 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

$$\zeta = J_\circ(2\pi f_d \tau), \tag{10}$$

where f_d is the maximum Doppler frequency and τ is the sample time.

At the receiver, Eq. (2) is vectorized during the detection stage, which is given by

$$\mathbf{y} = \tilde{\mathbf{H}} \mathcal{X} \mathcal{I}_c \mathbf{K}_{q,l} + \tilde{\mathbf{V}}, \tag{11}$$

where each constituent vectorized matrix is expressed as

$$\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{N_r T \times 1}; \tag{12}$$

$$\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}_{\text{eff}} \in \mathbb{C}^{N_r T \times N_t T}, \mathbf{H}_{\text{eff}} = \mathbf{W}_{\text{RF}} \mathbf{H} \mathbf{F}_{\text{RF}}; \tag{13}$$

$$\tilde{\mathbf{V}} = \text{vec}(\mathbf{V}) \in \mathbb{C}^{N_r T \times 1}; \tag{14}$$

$$\mathcal{X} = [\text{vec}(\tilde{\mathbf{A}}_{1,1}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,c}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,N_c})]; \tag{15}$$

$$\in \mathbb{C}^{N_t T \times N_c M_Q}; \tag{16}$$

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

Furthermore, in the conventional MS-STSK aided transceiver design, soft-decision detection is carried out by outputting the log-likelihood ratio (LLR) of the MS-STSK demodulator. The LLR of a bit is defined as the ratio of the probabilities associated with the logical bit '0' and logical bit '1', which is formulated as:

$$L(b) = \log \frac{p(b=1)}{p(b=0)}, \tag{18}$$

where $p(b=1)$ and $p(b=0)$ are the probabilities associated with the logical bit '1' and logical bit '0', respectively. The sign of $L(b)$ indicates the logical bit '1' or '0', while the magnitude indicates the confidence in that specific bit.

Then, the probability of receiving the signal $\mathbf{y}$ given that the symbol $\mathbf{K}_{q,l}$ is transmitted from the c^{th} AC is given by

$$p(\mathbf{y}|\mathbf{K}_{q,l,c}) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right). \tag{19}$$

On the other hand, the received signal $\mathbf{y}$ conveys the bit sequence $B = [b_1, \dots, b_{N_b}]$, where $N_b = \log(N_c M_Q M_c)$. Then the LLR of the bit b_i is given by

$$L(b_i) = \log \frac{p(\mathbf{y}|b_i=1)}{p(\mathbf{y}|b_i=0)}, \tag{20}$$

where

$$p(\mathbf{y}|b_i=1) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \times \tag{21}$$

$$\sum_{k,q,c \in b_i=1} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right) \tag{22}$$

and

$$p(\mathbf{y}|b_i=0) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \times \tag{23}$$

$$\sum_{k,q,c \in b_i=0} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right). \tag{24}$$

It is important to emphasize that (19) relies on the knowledge of CSI at the receiver. This requires pilots for channel estimation, hence reducing the effective data rate [20]. The effective Discrete-input Continuous-output Memoryless Channel (DCMC) capacity accounting for the pilot density

f_p, which is the ratio of the number of pilots to the number of data symbols, is given by [21], [22]

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

while

$$C_{\text{DCMC}}^{\text{MS-STSK}} = \log_2(N_c M_c M_Q) - \frac{1}{(N_c M_c M_Q)} \times \sum_{q,l,c} \mathbb{E}\left[\log_2 \sum_{q',l',c'} \exp(\psi_{\text{MS-STSK}} \Big| \mathbf{K}\right], \tag{25}$$

where

$$\psi_{\text{MS-STSK}} = -\frac{\left\| \mathbf{H}\mathcal{X}\left(\mathcal{I}_{c'}\mathbf{K}_{q,l} - \mathcal{I}_{c'}\mathbf{K}_{q',l'}\right) + \tilde{\mathbf{V}} \right\|^2 - \|\tilde{\mathbf{V}}\|^2}{\sigma^2}. \tag{26}$$

For N_d number of symbols and N_p number of pilots in a frame, f_p is expressed as

$$f_p = \frac{N_p}{N_d + N_p}. \tag{27}$$

Furthermore, when the minimum mean squared error (MMSE) channel estimate is considered, the channel estimate error variance (σ_h^2) for a total transmission signal power ρ_t is given by [23]

$$\sigma_h^2 = \frac{1}{1 + 100\rho_t f_p}. \tag{28}$$

Therefore, in the next section, we propose the deep learning aided MS-STSK demodulator, which provides the soft LLR values by circumventing the requirement of having any CSI knowledge.

III. THE PROPOSED LEARNING-AIDED SOFT-DECODING

In this section, we propose our learning assisted MS-STSK soft-demodulator, which calculates the soft LLR values by applying a neural network, thereby circumventing the requirement of having CSI knowledge. We commence this section by presenting some preliminaries on deep learning, followed by our proposed design.

A. PRELIMINARIES ON NEURAL NETWORKS

Inspired by the biological nervous system, the deep learning relies on layers of artificial neurons, which process the information in a similar fashion to that of the human brain. Furthermore, akin to the structure of the human neural networks, deep learning is composed of an interconnected network of neurons, which led to the terminology of artificial neural networks (ANNs). Given its capability of learning patterns and detecting trends, ANNs became one of the important constituents of machine learning. A typical ANN is illustrated in Fig. 3, where each layer of neurons is interconnected with the layers succeeding and preceding it. Fig. 3 presents a 4-layered-network, where the first and last layers are the input and output layers, respectively, while the layers sandwiched between them are referred to as hidden layers. Note that except for the input layer of Fig. 3, every other layer has an activation function (AF) $f(.)$, weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$. The output values of the input layers

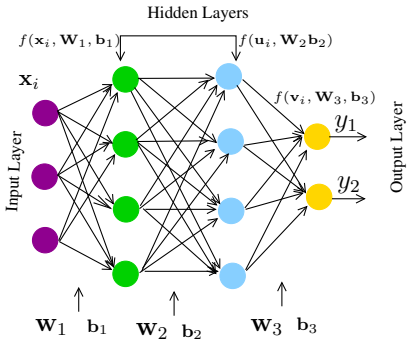


FIGURE 3: The structure of a typical feed-forward ANN.

are then forwarded to the next layer by invoking the AF. The resultant output value then serves again as the input of the next layer, where the AF of that layer is applied to its input. The resultant value can be interpreted as the stimulus it created for activating the corresponding neuron of that layer. This process is then carried on until the output layer is reached for each neuron. It is instructive to note that the best choice of the AF is problem-specific and each layer of the network may have different AF. The only constraint concerning on the choice of the AF is its differentiability. Some popularly used AFs are sigmoid, piecewise-linear and threshold functions [24].

Machine learning is broadly classified as: unsupervised and supervised learning. In unsupervised learning, the network aims for learning the probability distribution of the dataset by observing samples of it, while in supervised learning, the network aims for learning the mapping of inputs to labels and then predicts the label from the new input. In this paper, we invoke supervised learning for predicting the soft LLRs from the received signal vector.

To elaborate a little further, supervised learning is a twofold process that includes the training phase and testing phase. During the training phase, the known input and output data samples, which are labeled, are used for inferring the mapping function between the input and output. This is carried out by learning the network parameters such as the weight matrix $\mathbf{W}$ and the bias vector $\mathbf{b}$ for each layer. The network parameters are designed for minimizing an error, or loss, between the predicted output and the actual output. Having learned the network parameters, the testing phase commences, where the learned weights obtained during the training phase are invoked for predicting the output labels and for quantifying the performance of the ANN using data outside the training set.

B. LEARNING-AIDED SOFT-DECODING

Having presented the rudimentary philosophy of ANN, let us now discuss the proposed learning-aided soft-decoding. Note

that the rationale of using learning learning over conventional soft-decoding is that by employing the former pilot-assisted channel estimation can be eliminated. In contrast to (19), where an exhaustive search is carried out over all the legitimate combinations of the bit being either logical '0' or '1', the learning-aided design provides the soft-LLRs by employing the ANN weights designed during the training phase. In this design, the input training samples of the ANN are the received signal vectors $\mathbf{y}$, while the output labels are the LLRs, as shown in Fig.4. Then, the ANN is trained to infer the functional mapping between the input and output samples. However, since the received signal vectors are affected by the noise, the ANN may fail to accurately infer the function. Therefore, the choice of the SNR during the training is crucial, which can only be obtained empirically by varying the SNR[1].

For a given SNR, the ANN predicts the LLR value by employing the AF $f(v)$ in each layer of the network as shown in Fig. 5, where the input of each AF is the output of the preceding layer. In this paper, we opted for the sigmoid function as the AF, as a benefit of its smoothness which is formulated as

$$f(v) = \frac{1}{1+e^{-v}}. \tag{29}$$

For example, in Fig. 5, the input of the first neuron in the second layer is $v_1 = w_{11}y_1 + w_{21}y_2 + w_{31}y_3 + b_1$. Then the AF of (29) is applied to v_1 to obtain its mapping, which serves as one of the inputs to the next layer. In this way, the output of the AF of each layer is then fed to its subsequent layer as shown in Fig. 4. This process is carried on until the output layer is reached, where the final predicted values are obtained. Note that initially the weights of each layer are assigned to random values obeying the distribution $\mathcal{N}(0,1)$. These weights are then updated for ensuring that they minimize the error between the predicted LLR and the actual LLR. Mathematically, it is formulated by a loss function (LF) given by

$$\text{LF} = \frac{1}{\mathcal{S}}\sum_{i=1}^{\mathcal{S}} \left\| \hat{\mathbf{L}}_i - \mathbf{L}_i \right\| + \rho_1 \|\mathbf{W}_1\|_2^2 + \rho_2 \|\mathbf{W}_2\|_2^2 + \rho_3 \|\mathbf{W}_3\|_2^2, \tag{30}$$

where $\mathcal{S}$ is the cardinality of the training set, $\hat{\mathbf{L}}_i$ and $\mathbf{L}_i$ are the predicted and the known LLR value, respectively, of the i^{th} training sample, while ρ_1, ρ_2, ρ_3 are the regularization factors used for avoiding over-fitting [25].

To minimize the loss function of (30), the gradient of the loss with respect to the weights is computed and used for updating the weight values in a gradient descent procedure known as back-propagation. While only a local —rather than global— minimum is ensured, the procedure has been shown to be practical. A more detailed discourse on back-propagation is presented by Chauvin et al. in [26]. These weights, which are learned during the training phase are then stored in memory and are invoked during the testing phase. In other words, the ANN predicts the LLR value from the received signal vector $\mathbf{y}$ by employing the pre-determined weights. Thus, this design does not depend on the channel knowledge to obtain the LLR values. It is instructive to note that during the training of ANN weights, we assume the AoAs and AoDs of the channel matrix of (7) to be time-invarying, while the small-scale fading coefficient is assumed to evolve in time according to Jakes' model.

The soft-LLR values predicted from the MS-STSK's ANN demodulator are then passed to the turbo channel decoder. In the next section, we examine the complexity of both the traditional MS-STSK receiver and the learning-aided MS-STSK receiver.

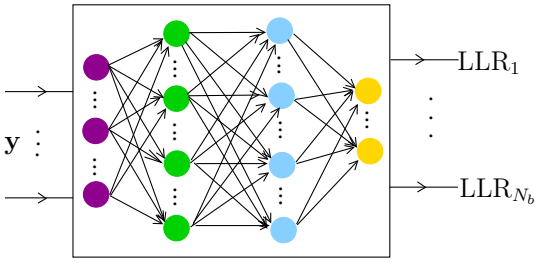


FIGURE 4: The input and output of the ANN proposed.

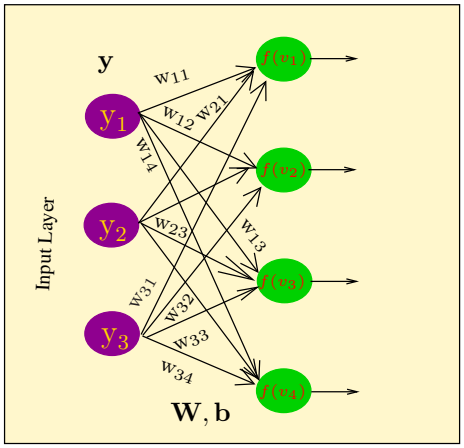


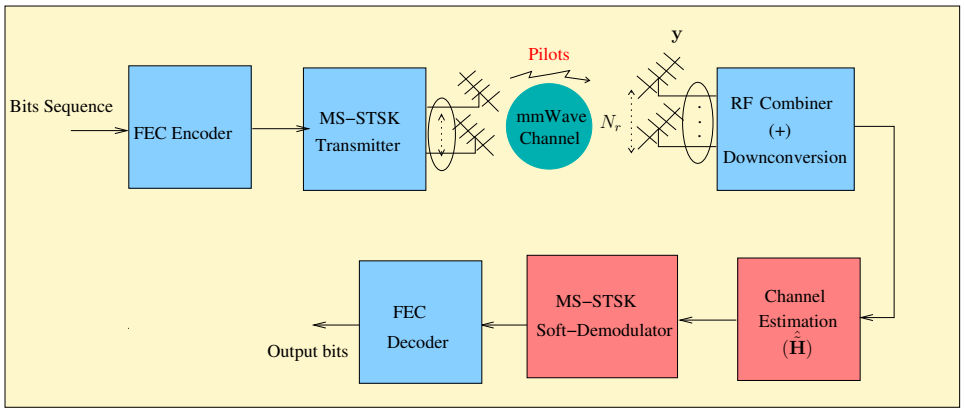
FIGURE 5: Illustration of the input to the AF of each neuron.

C. RECEIVER COMPLEXITY

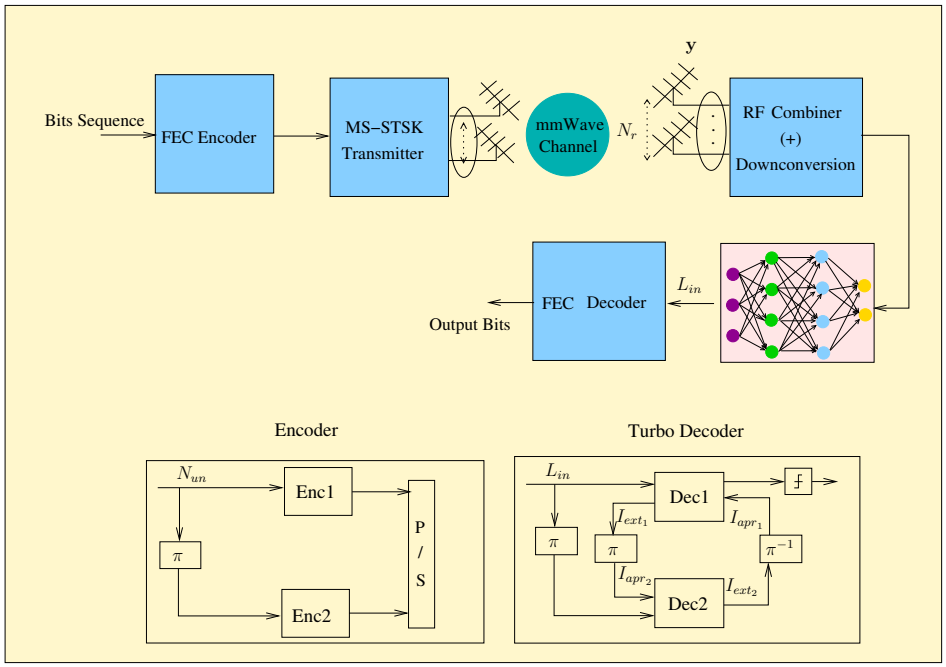
In this section, we commence by discussing the overhead associated with the MS-STSK soft-demodulator, followed by quantifying the receiver's complexity in terms of the number of computations for both the conventional and for the learning-aided soft-demodulator.

Fig. 6 shows the block diagram of both of the conventional MS-STSK and of the ANN assisted MS-STSK soft-demodulators. To expound further, Fig. 6(a) employs the conventional MS-STSK soft-demodulator, where first the

[1]There is a temptation to train the network at high SNRs, where the input samples are noise-free. However, using noiseless training data may result in performance degradation under realistic conditions.



(a) Conventional MS-STSK soft-demodulator.



(b) Learning MS-STSK soft-demodulator.

FIGURE 6: Block diagram of MS-STSK soft-demodulators: Conventional design and proposed ANN-assisted design.

signal is processed by analog RF combining and down-converted to the baseband. Then, the receiver estimates the channel matrix $\mathbf{H}$ with the aid of pilots. After the estimation of the channel the receiver employs the soft-demodulator of (20) to obtain the LLR values, which are then passed to the channel decoder. It is important to emphasize that in this design, the receiver has to estimate the channel using a pilot overhead of f_p, which has to be high enough for sampling the channel's complex-valued envelop at multiples of the Nyquist frequency for mitigating the effects of channel noise.

By contrast, Fig. 6(b) shows our proposed learning-assisted soft-demodulator. Like in the conventional design, the receiver first employs analog RF combining followed by the down-conversion of the received signal. However, in contrast to the conventional design, our proposed learning-aided soft-demodulator does not require the knowledge of the CSI. The down-converted signal vector is fed to the ANN, which then employs the learned weights to obtain the LLRs without explicit CSI. In other words, our design by passes the channel estimation stage, which is indispensable in the conventional design. The LLRs gleamed from the ANN are then passed to the turbo decoder, where the decoded bits are retrieved.

Having discussed the pilot overhead, we now focus our attention on the complexity in terms of the number of computations. To quantify the complexity, let us assume that the input and output vectors of the ANN shown in Fig. 4 are of sizes n_i and n_o, respectively. Let us also assume that the number of neurons in each hidden layer is n_h. The AF (29) is computed for each neuron of each layer by employing

the corresponding layer's weights and biases, as discussed in Sec. III-B. Then the total number of computations required for the ANN of Fig. 4 is on the order of $\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$.

On the other hand, the number of computations required for the conventional soft-demodulator is $2^{N_b}\left(\mathcal{O}(N_c M_Q^3)\right) + 2^{N_b}\left(\mathcal{O}(N_t N_r N_c M_Q T^2)\right)$. Table 1 shows the number of computations required for the simulation parameters summarized in Table 2.

TABLE 1: *Quantifying the complexity in terms of the number of complex-valued multiplications.*

Design	Computations
Learning assisted Blind Detection	330
ML-Aided Detection for	12288

IV. SIMULATIONS

In this section, we present our simulation results characterizing the performance of our proposed learning aided soft-demodulator and of the conventional soft-demodulator. We performed Monte Carlo simulations for analyzing the performance of both designs. Furthermore, in our simulations, we employed a half-rate turbo encoder using the LTE generator polynomials as that of the LTE in Fig. 6. The simulation parameters are summarized in Table 2.

TABLE 2: Simulation Parameters.

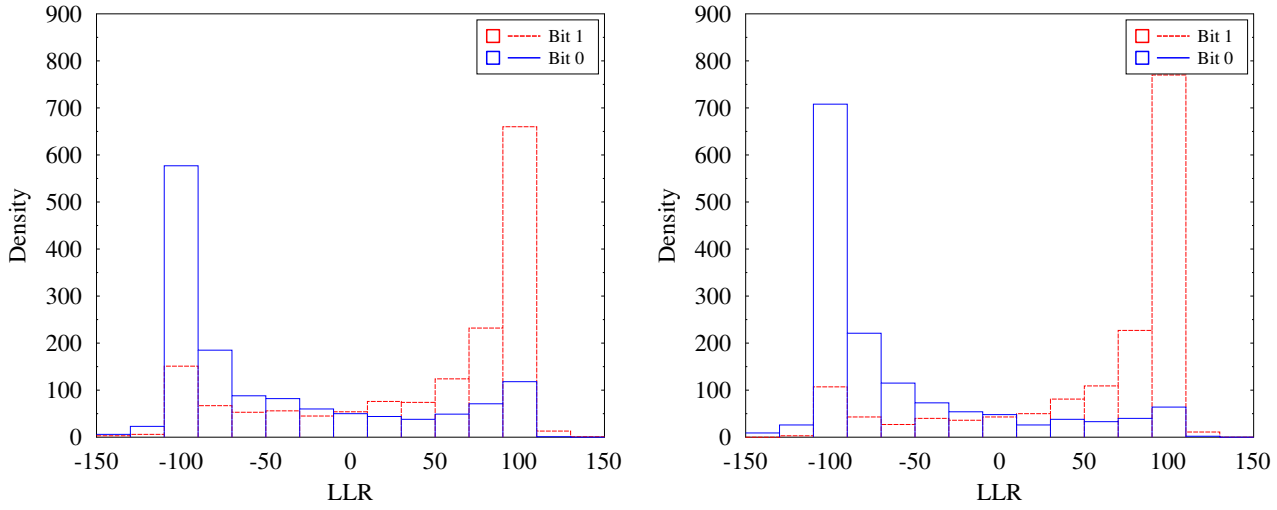
Parameters	Values
Number of AEs in each AA at Tx (K)	64
Number of AEs in each AA at Rx (L)	16
Number of AAs at Tx (N_t)	4
Number of AAs at Rx (N_r)	2
Size of QAM symbol (M_c)	4
Number of dispersion matrices M_Q	4
Size of consecutive AA selected (M)	2
Time slots (T)	2
Number of clusters (N_c)	1 & 2
Number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

Fig. 7 shows the histogram of the LLRs at the output of the ANN for different SNR values. In our simulation, we have limited the maximum and minimum values of the LLR to ± 100 for the conventional MS-STSK soft-demodulator. Therefore, we have trained the ANN using discrete output samples, i.e $+100$ or -100, according to the bits transmitted. Furthermore, the ANN adopted in our simulations is of continuous regression, since the values from the soft-demodulator are continuous-valued. Therefore, the maximum and minimum values of the ANN's LLRs hover around ± 100. It can be seen from Fig. 7 that for the SNR value of 0 dB, the densities of the ANN's LLRs for bit '1' and bit '0' overlap with a larger area. The physical significance of this is that the LLRs observed for bit '1' fall with a higher probability on the wrong side of the bit, i.e. bit '0', hence forwarding less reliable information to the turbo channel decoder. However, as the SNR is increased from 0-to-2-to-4 dB, observe in Fig. 7(a)-to-Fig. 7(c), that the overlapping area between the two histograms is gradually reduced. In other words, the ANN provides more reliable LLRs.

Fig. 8 characterizes the bit error rate of both the proposed learning-aided soft-demodulator and of the conventional MS-STSK soft-demodulator relying on the idealized simplifying assumption of having perfect CSI knowledge. It is evident from Fig. 8 that despite the lack of CSI knowledge, the learning-aided soft-demodulator performs closely to the conventional soft-demodulation. More particularly, at BER $< 10^{-4}$, the SNR gap between the two is 1.5 dB. Furthermore, when the realistic imperfect CSI is considered, the conventional soft-demodulation exhibits inferior performance. This becomes evident in Fig. 8, where the channel error variance is increased from as low as 0.16 to as high as 0.9. In other words, the conventional design is incapable of achieving a low BER for low pilot overheads. We note that the normalized pilot overhead required for training the learning-aided design and used in Fig. 8 is around 0.002. For the same pilot overhead, the conventional design produces an error floor. It is important to note that our learning-aided soft-demodulator relies on an extremely low pilot overhead as well as much lower search complexity than the conventional MS-STSK demodulator, as seen Table 1 and discussed in Sec. III-C.

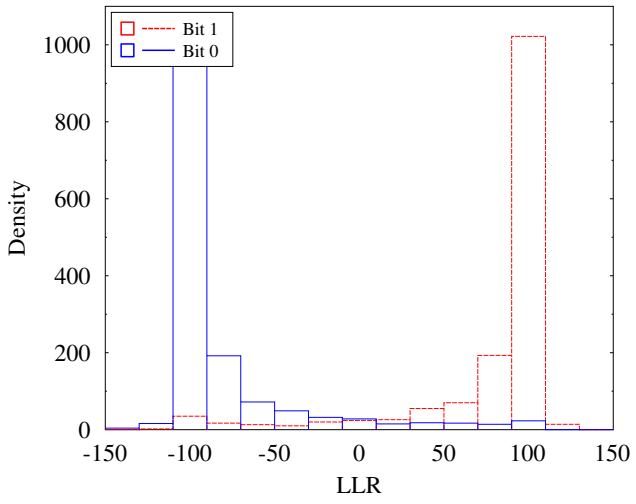
Our learning-aided soft demodulator requires a low pilot overhead for training the NN. We have found empirically that the pilot overhead required for normalized Doppler frequency of 10^{-3} is 0.002, while it is around 5% for conventional design, when MMSE based channel estimation is employed. We also note that by using (28) the channel estimation variance is around 0.16 for a pilot overhead of 5% [23]. Given these parameters, we observed in Fig. 8 that our proposed learning-aided soft-demodulation outperforms the conventional detection relying on practical MMSE channel estimation. However, in our design, retraining of the weights may be required depending on the environment.

Fig. 9 shows the Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of both the proposed design, and of the conventional design. Whilst our proposed learning-assisted design exhibits a 1.5 dB SNR loss at a BER of 10^{-4} as shown in Fig. 8, observe in Fig. 9 that it provides a pilot-overhead-dependent DCMC capacity reduction compared to the conventional proposed design. More explicitly, this is because of the pilot overhead required by the conventional design for channel estimation. For example, the pilot overhead of the conventional design may span from 3% to 10% of the data rate, depending on the Doppler spread. By contrast, the learning-assisted soft-demodulator does not rely on channel estimation, hence almost totally eliminating the pilot overhead, while providing a higher data rate. It is evident from Fig. 9 that the learning assisted design provides an SNR gain of 3 dB over the conventional design having a 10% pilot overhead at a throughput rate of 4 bps/Hz, while



(a)

(b)



(c)

FIGURE 7: Histogram depicting the densities of the LLRs at the output of the ANN assisted MS-STSK demodulator for (a) SNR= 0 dB (b) SNR= 2 dB (c) SNR= 4 dB.

it is around 1.5 and 0.8 dB, when having 5.7% and 3% pilot overhead, respectively.

Fig. 10 shows the EXtrinsic Information Transfer (EXIT) chart of our proposed design and of the conventional design. More particularly, Fig. 10 shows the EXIT chart of the conventional design at the SNR of -28 dB, where extrinsic soft-information is exchanged between the upper and lower recursive systematic convolutional turbo decoders, as shown in Fig. 6 , while the SNR is -26 dB for our learning-assisted design. More explicitly, the stair-case shaped curve of Fig. 10 portrays the extrinsic information exchange between I_{apri_1}/I_{ext_2} and I_{apri_2}/I_{ext_1}, when the input LLRs are provided by: a) conventional MS-STSK soft-demodulator and b) the learning-assisted soft-demodulator; where I_{apri_1}, I_{ext_1} denote the input and output mutual information of the Decoder 1[2], while I_{apri_2}, I_{ext_2} denote the input and output mutual information of the Decoder 2, as shown in Fig. 6. We observe from Fig. 10 that for the conventional design

[2]The decoder 1 and decoder 2 correspond to the constituent decoders in the turbo decoder, as shown in Fig. 6(b).

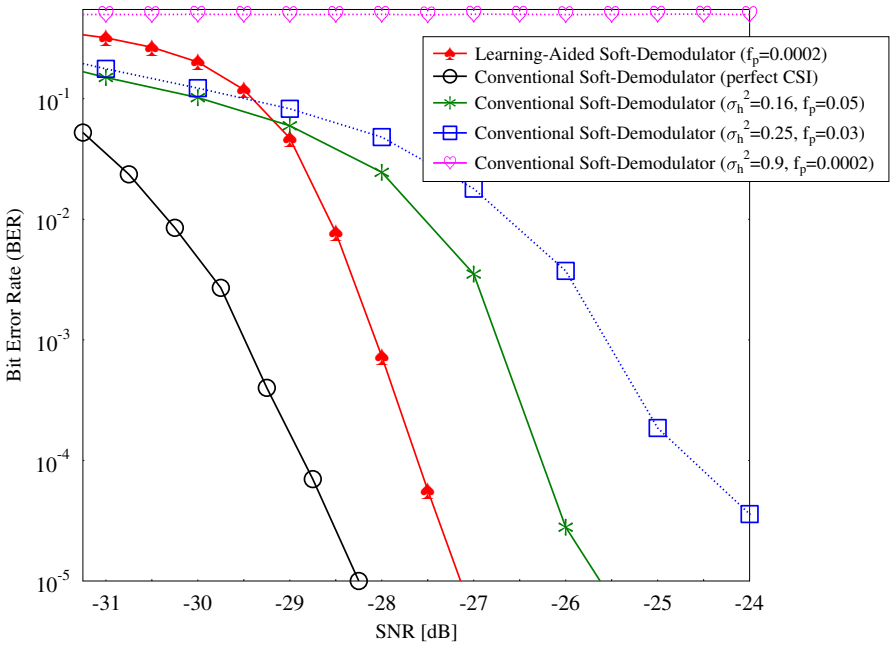
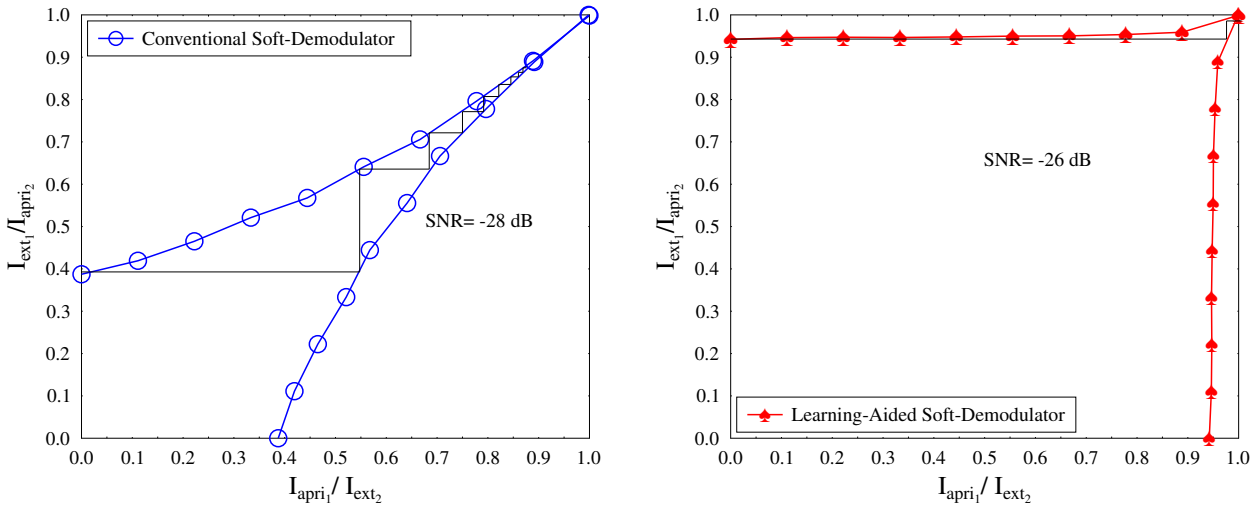


FIGURE 8: Bit error ratio of the proposed learning-aided soft-demodulator and of the conventional soft-demodulator. The simulation parameters are listed in Table 2.



(a) (b)

FIGURE 10: Exit chart (a) of the conventional design; (b) of the learning-assisted design.

relying on perfect CSI, the tunnel begins to open at SNR of -28 dB allowing the stair-case-shaped decoding trajectory to reach the $(1, 1)$ point of perfect convergence to a vanishingly low BER. By contrast, the machine learning aided design

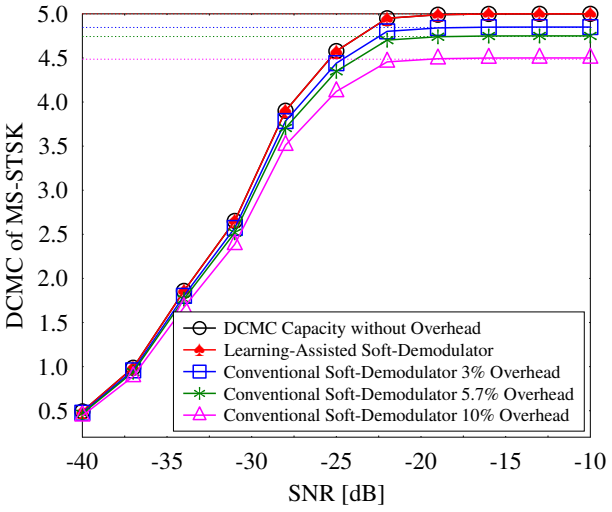


FIGURE 9: The Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of the proposed design, and of the conventional design at $3\%, 5\%$ and 10% pilot overheads. The simulation parameters are listed in Table 2.

achieves this at SNR of -26 dB with the aid of as few as two iterations.

V. CONCLUSIONS

We have proposed a deep learning assisted soft-demodulator for MS-STSK mmWave systems, where we trained a NN to provide the soft values of the MS-STSK symbol without relying on explicit CSI knowledge. We demonstrated that in contrast to the conventional MS-STSK soft-demodulator, which requires the knowledge of CSI at the receiver, the learning-assisted design circumvents channel estimation, while also improving the data rate, since the latter does not impose a pilot overhead. Furthermore, our proposed learning-aided soft-demodulator avoids the exhaustive search complexity of evaluating the soft values at the output of the MS-STSK demodulator. Our simulations demonstrated that despite dispensing with CSI knowledge, our learning-assisted design performs similarly to the conventional design relying on perfect CSI knowledge for BER $< 10^{-4}$, while upon using realistic CSI at the receiver for the conventional soft-demodulation, the learning-aided soft-demodulator outperforms the latter.

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] 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.

[3] 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 in Proc. URSI Int. Symp. Signals, Syst., Electron., Oct 1998, pp. 295–300.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] S. Sugiura, S. Chen, and L. Hanzo, "Space-time shift keying: A unified mimo architecture," in Proc. Globecom, Dec 2010, pp. 1–5.

[9] 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.

[10] 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.

[11] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Multi-set space-time shift-keying with reduced detection complexity," IEEE Access, vol. 4, pp. 4234–4246, 2016.

[12] S. U. H. Qureshi, "Adaptive equalization," Proc. IEEE, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.

[13] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in Proc. SPAWC, July 2017, pp. 1–5.

[14] ——, "Learning to detect," CoRR, vol. abs/1805.07631, 2018. [Online]. Available: http://arxiv.org/abs/1805.07631

[15] S. Dorner, S. Cammerer, J. Hoydis, and S. T. Brink, "Deep learning based communication over the air," IEEE J. Sel. Topics Sig. Proc., vol. 12, no. 1, pp. 132–143, Feb 2018.

[16] X. Jin and H.-N. Kim, "Deep learning detection networks in MIMO decode-forward relay channels," arXiv e-prints, p. arXiv:1807.09571, Jul. 2018.

[17] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," IEEE J. Sel. Topics Sig. Proc., vol. 12, no. 1, pp. 119–131, Feb 2018.

[18] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," IEEE J. Sel. Topics Sig. Proc., vol. 12, no. 1, pp. 144–159, Feb 2018.

[19] X. Yan, F. Long, J. Wang, N. Fu, W. Ou, and B. Liu, "Signal detection of MIMO-OFDM system based on auto encoder and extreme learning machine," in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 1602–1606.

[20] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, MIMO-OFDM Wireless Communications with MATLAB. Wiley Publishing, 2010.

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

[22] 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.

[23] 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.

[24] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[25] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag, 2006.

[26] Y. Chauvin and D. E. Rumelhart, Eds., Backpropagation: Theory, Architectures, and Applications. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1995.

K. SATYANARAYANA (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.

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".

PHILIP PIETRASKI received the PhD, EE degree from Polytechnic University, now NYU (2000). Phil's background is in signal processing and communication. He has a long history of radio, propagation, and PHY layer R&D spanning HSPA, LTE, 5G, mmW, WiFi, and WiGig, modem and beamforming development at InterDigital with over 100 patents. More recently, Phil is leading research on machine learning applied to the wireless domain with a focus on the lower layers of the communications stack. Prior to joining InterDigital, he worked as a Research Engineer at the DoE lab National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, working in several areas including pulsed-power RF systems, electron-beam control and diagnostics, and high-rate X-ray detectors; ultimately becoming Lead Engineer of the X-ray detector development project at the NSLS. He also does volunteer work by serving on the board of trustees until 2019 and currently on the advisory committee for DeVry University, NJ and coaching/mentoring for FIRST Robotics.

LAJOS HANZO (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 44-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 119 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1900+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing.