Turbo Detection Aided Autoencoder for Multi-Carrier Wireless Systems: Integrating Deep Learning into Channel Coded Systems

Chao Xu, *Senior Member, IEEE*, Thien Van Luong, Luping Xiang, *Member, IEEE*, Shinya Sugiura, *Senior Member, IEEE*, Robert G. Maunder, *Senior Member, IEEE*, Lie-Liang Yang, *Fellow, IEEE* and Lajos Hanzo* *Fellow, IEEE*

***Abstract*—A variety of deep learning schemes have endeavoured to integrate deep neural networks (DNNs) into channel coded systems by jointly designing DNN and the channel coding scheme in specific channels. However, this leads to limitations concerning the choice of both the channel coding scheme and the channel paramters. We circumvent these impediments and conceive a turbo-style multi-carrier auto-encoder (MC-AE) for orthogonal frequency-division multiplexing (OFDM) systems, which is the first one that achieves the flexible integration of DNN into any given channel coded systems while achieving an iteration gain. More explicitly, first of all, we design the MC-AE independently of both the channel coding arrangement and of the channel model, where the output layer of the MC-AE decoder is designed for both accepting and producing reliable soft-bit decisions. Owing to the fact that bit-dependency is imposed by the MC-AE mapping, our bespoke MC-AE decoder becomes capable of achieving a beneficial iteration gain, when the extrinsic information is exchanged between the soft-decision MC-AE decoder and the soft-decision channel decoder. Secondly, in order to be able to interpret the performance advantages of our MC-AE over the conventional OFDM, we map the MC-AE's input-output relationship to an equivalent model-based representation. The associated theoretical analysis verifies the fact that during the process of data-driven signal reconstruction across OFDM's subcarriers, a beneficial frequency diversity gain is achieved by the proposed MC-AE design. Finally, our simulation results demonstrate that the MC-AE is capable of achieving substantial performance advantages over both conventional OFDM and OFDM based index modulation (OFDM-IM) in channel coded systems.**

***Index Terms*—Orthogonal frequency-division multiplexing, deep learning, deep neural network, auto-encoder, turbo detection, low density parity check, binary convolutional coding, mutual information.**

C. Xu, T. V. Luong, R. G. Maunder L-L. Yang and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK (e-mail: {cx1g08,tvl1r20,rm,lly,lh}@soton.ac.uk). L. Xiang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: lxlg15@soton.ac.uk). S. Sugiura is with the Institute of Industrial Science, University of Tokyo, Meguro-ku, Tokyo 153-8505, Japan (e-mail: sugiura@ieee.org).

We would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028) The work of Luping Xiang is supported by International Postdoctoral Exchange Fellowship Program (YJ20210244). The work of S. Sugiura was supported in part by the Japan Science and Technology Agency (JST) Precursory Research for Embryonic Science and Technology (PRESTO) under Grant JPMJPR1933.

I. Introduction

The development of wireless communication systems has been centred on the Shannonian model-based block-structure, where each block is optimized based on its mathematical model for specific tasks, such as the channel encoder/decoder, bit-to-symbol mapper/demapper, channel estimator, equalizer as well as the multiple-input multiple-output (MIMO) signal processing schemes and orthogonal frequency-division multiplexing (OFDM) operations [1]–[6]. As a design alternative, inspired by their remarkable success in computer science [7], [8], the data-driven deep learning methods, especially deep neural networks (DNNs) have also attracted substantial attention [9]–[12], thanks to the following attractive benefits. Firstly, the data-driven DNN is trained to reconstruct the input at the output, especially when no tractable mathematical models are available. This facilitates a high degree of adaptivity and flexibility for DNN applications. Secondly, during the process of data-driven reconstruction, the underlying stochastic model can be inferred, so that the semantic information carried in extremely complex environments can be extracted and disentangled. Thirdly, the input-output relationship of DNN constitutes a "black box", which may incorporate multiple traditional signal processing blocks. This approach aims for finding the entire Pareto-front of a joint optimization problem, which cannot be directly solved by the conventional model-based methods. Finally, the inherent concurrent nature of DNN computation and memory access is capable of efficiently exploiting the parallel processing architectures of graphic processing units (GPUs), which have a significant faster computational capability than the traditional processor chips. However, the channels encounted in wireless communication are highly dynamic in nature, while their hardware resources available for training large DNNs are typically limited. Therefore, the training complexity, reliability, scalability, and the unexplainable nature of DNNs constitute major issues.

More explicitly, the DNN applications in the physcial layer are generally classified into two main categories, where the DNNs are invoked either to replace the entire end-to-end system or to be integrated into the classic model-based block-structure. On one hand, it was proposed in [13] that the end-to-end communication system can be interpreted as a so-called autoencoder (AE) in DNN, which maps the input data at the transmitter into a code, so that the output data at the

receiver may be optimized to match the input. Furthermore, it was demonstrated in [14] that such designs can be readily realized using off-the-shelf software-defined radios and open-source DNN libraries. Moreover, an end-to-end AE-based scheme capable of handling unknown channel models was conceived in [15], while a noncoherent AE design that relies on energy detection was proposed in [16]. However, the large DNNs designed for end-to-end systems tend to require a large number of parameters, especially when the important channel coding scheme is incorporated. On the other hand, the model-based DNN philosophy aims for replacing the individual signal processing blocks, such as equalization [17], channel coding [18], [19], MIMO detection [20], by DNNs. As a result, by exploiting the classic block-structure and injecting certain expert domain knowledge, the DNN computing resources and training time may be substantially reduced.

The wireless communication environment suffers both from interference and fading, which contaminate the signal transmission. In order to mitigate this problem, the essential wireless system component of channel coding artificially imposes correlation on its input data before its transmission, so that the errors caused by the propagation channel may then be corrected at the receiver by beneficially exploiting the intentionally imposed bit-dependency in the received data. Looking back over the past seven decades of wireless developments, major breakthroughs are marked by advances in channel coding, such as turbo coding (TC) and low density parity check (LDPC) coding. Therefore, in this work, we focus our attention on the integration of DNN into channel coded systems. More explicitly, a beneficial iteration gain can be achieved, when the extrinsic information is exchanged between the soft-decision MC-AE decoder and the soft-decision channel decoder. This facilitates performance improvements without the need for training large DNNs. However, the flexible integration of DNN with channel coded systems has not been demonstrated in the existing state-of-the-art. Hence, we boldly and explicitly contrast our novel contributions to the state-of-the-art in Table I. Specifically, the DNN's iteration gain is either unexploited or dependent on specific channel coding parameters in the existing literature. More explicitly, it is feasible to replace the channel encoder/decoder by DNNs, but the major challenges are the generalization of implementing different channel coding schemes such as TC and LDPC, the generalization of using different block lengths and the generalization of transmitting over different types of channels. At a glance, inspired by the concept of deep unfolding, where iterative decoding is expanded by serial concatenated DNN layers, the AE-based solution of [13] was shown to be capable of reconstructing the encoder and decoder of a Hamming code for transmission over additive white Gaussian noise (AWGN) channels. The model-based ViterbiNet concept was proposed in [18], where DNN is integrated into the Viterbi algorithm in both AWGN and block fading channels. Furthermore, the deep learning methods of [21] were shown to be able to achieve the decoding capability of the maximum likelihood sequence estimation based Viterbi algorithm and that of the optimal BCJR algorithm in AWGN channels. In these schemes, the space of codeword design is enormous. For example, a rate 1/2 code having a frame length of 100 information bits involves designing 2^{100} codewords in a 200 dimensional space [23]. Therefore, it is preferable to replace the modulation-related blocks by the DNN, which can be integrated back into channel coded systems performing turbo detection by iteratively exchanging extrinsic information between a soft-decision data-driven DNN and a classic model-based Viterbi/BCJR decoder. Based on this, the DNN solution of [19] invokes an AE to incorporate both the modulator at the transmitter as well as the demodulator at the receiver, where a substantial iteration gain is achieved by exchanging soft extrinsic information between the soft-decision DNN and the LDPC decoder. EXtrinsic Information Transfer (EXIT) chart analysis [2], [24]–[27] may be employed for visualizing the exchange of mutual information throughout the decoding iterations. However, the iterative decoder of [19] requires individual optimization both for different block lengths as well as for the different choices of channel coding schemes. Therefore, as boldly and explicitly highlighted in Table I, our proposed scheme is the first one that achieves both flexible integration of DNN into channel coded system and an AE-induced iteration gain. In summary, the novel contributions of this work are as follows:

- We propose to integrate the multi-carrier AE (MC-AE) of [28] into channel coded systems. First of all, the AE is designed independently of both the channel coding arrangement and of the channel model. Secondly, the domain knowledge of a channel equalizer is streamlined for the input to the AE decoder, which helps to reduce both the number of DNN parameters and the training complexity. Thirdly, the DNN's softmax output layer is appropriately modified for accepting and producing reliable soft decisions. Finally, the soft-decision MC-AE design is extended to support multi-carrier MIMO systems of V-BLAST and space-time block coding (STBC) [4], [5], [27], [29], [30]. We note that following the specific design guidelines conceived in this paper, the MC-AE may be replaced by any DNN schemes that are trained based on the stochastic gradient descent (SGD) with the gradients of the loss function back-propagated from the output layer to the input layer. Hence our solution is eminently suitable for the broad family of DNNs designed for OFDM [31]–[33], OFDM relying on index modulation (OFDM-IM) [34], nonorthogonal multiple access (NOMA) [35], [36] and interference cancellation [37]–[39].
- In order to be able to interpret and explain the performance advantages of MC-AE over OFDM and OFDM-IM, we map the input-output relationship to an equivalent model-based representation, which is guided by the unconstrained capacity of continuous-input continuous-output memoryless channels (CCMC). Following this, both the pairwise error probability (PEP) and the constrained capacity of discrete-input continuous-output memoryless channels (DCMC) are invoked in order to verify that the performance advantage of the MC-AE design stems from its improved frequency diversity gain, which is implicitly optimized in line with the DNN

TABLE I: State-of-the-art DNN schemes designed in channel coded systems.

	DNN setup		Explainable DNN		Turbo DNN		
	AE-based	Model-based	Capacity analysis	EXIT charts analysis	DNN indepdent to channel coding parameters	Soft-decision DNN	DNN with iteration gain
O'Shea *et al.* [13]	√						
Dorner *et al.* [14]	√						
Samuel *et al.* [20]		√					
Shlezinger *et al.* [18]	√	√					
Kim *et al.* [21]	√					√	√
Cammerer *et al.* [19]	√	√		√		√	√
Shlezinger *et al.* [22]	√						
Proposed Scheme	√	√	√	√	√	√	√

training loss function. Finally, the powerful analytical tool of EXIT charts is employed for analyzing the evolution of mutual information throughout the turbo detection iterating between the soft-decision DNN and the channel decoder.

- Finally, the proposed MC-AE scheme is examined in conjunction with both Binary Convolutional Coding (BCC) employed in the IEEE 802.11a/n/ac/ah standard family as well as the classic LDPC used in the operational 4G and 5G cellular networks [40], [41]. Our simulation results demonstrate that our improved soft-decision DNN is capable of beneficially exploiting the inter-bit correlation imposed by the AE, which leads to a precious iteration gain. As a result, the proposed MC-AE scheme is capable of achieving substantial performance advantages over its conventional OFDM and OFDM-IM counterparts in channel coded systems. As a further advance, in MIMO systems, the MC-AE aided STBC is capable of achieving a further improved performance over the MC-AE aided V-BLAST, thanks to both the frequency diversity gain attained by MC-AE and the space-time diversity gain achieved by STBC.

The rest of this paper is organized as follows. The system overview is provided in Sec. II, where the MC-AE design is integrated with the BCC and LDPC channel coding schemes. Sec. III offers the model-based performance analysis of MC-AE, where both the CCMC capacity, as well as the DCMC capacity and the PEP are analysed. The soft-decision DNN is derived in Sec. IV, where the MC-AE decoder is revised to be able to accept and produce soft-bit decisions. The performance characterization of turbo MC-AE is presented in Sec. V, and then the MC-AE design is extended to the MIMO setups of V-BLAST and STBC in Sec. VI and Sec. VII, respectively. Finally, our conclusions are offered in Sec. VIII. The notations used throughout the paper are summarized in Table II.

II. System Overview

A. Turbo MC-AE

The serially concatenated code (SCC) constituted by the BCC/LDPC channel coding schemes and the data-driven MC-AE is portrayed in Fig. 1, where iterative demapping and decoding is invoked by the two-stage turbo receiver. More explicitly, the source bits b_u are encoded by the BCC/LDPC encoder into b_c, which is interleaved into b_m and then mapped by the MC-AE encoder at the transmitter of Fig. 1. At the turbo receiver, both the MC-AE's decoder and the BCC/LDPC decoder accept and then produce soft-bit decisions in the form of Log Likelihood Ratios (LLRs). The *extrinsic* LLRs

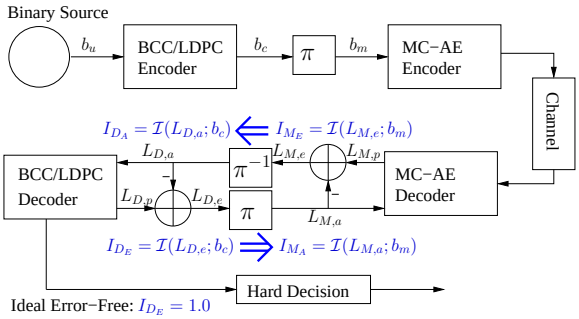


Fig. 1: Schematics of binary convolutional code (BCC) and low-density parity-check (LDPC) code aided multi-carrier autoencoder (MC-AE) and their two-stage turbo receiver.

produced by the MC-AE's decoder are given by $L_{M,e} = L_{M,p} - L_{M,a}$, where $L_{M,p}$ and $L_{M,a}$ refer to the *a posteriori* LLRs and the *a priori* LLRs of the MC-AE, respectively. The *extrinsic* LLRs $L_{M,e}$ produced by the MC-AE's decoder are de-interleaved and fed into the BCC/LDPC decoder as the *a priori* LLRs $L_{D,a}$. Hence the *extrinsic* mutual information $\mathcal{I}_{M_E} = \mathcal{I}(L_{M,e}; b_m)$ at the output of the MC-AE's decoder becomes equal to the *a priori* mutual information $\mathcal{I}_{D_A} = \mathcal{I}(L_{D,a}; b_c)$ at the input of the BCC/LDPC decoder. The calculation of mutual information $\mathcal{I}(L; b)$ between the LLRs L and binary bits b is given by (1). Symmetrically, the *extrinsic* LLRs produced by the BCC/LDPC decoder are given by $L_{D,e} = L_{D,p} - L_{D,a}$, where $L_{D,p}$ and $L_{D,a}$ refer to the *a posteriori* LLRs and the *a priori* LLRs of the BCC/LDPC, respectively. The *extrinsic* LLRs $L_{D,e}$ produced by the BCC/LDPC decoder are interleaved and fed into the MC-AE's decoder as the *a priori* LLRs $L_{M,a}$, which indicates that the *extrinsic* information $\mathcal{I}_{D_E} = \mathcal{I}(L_{D,e}; b_c)$ at the output of the BCC/LDPC decoder becomes the *a priori* information $\mathcal{I}_{M_A} = \mathcal{I}(L_{M,a}; b_m)$ at the input of the AE's decoder, as seen in Fig. 1. Decoding convergence to vanishingly low BER may become achievable, when the full $\mathcal{I}_{D_E} = 1.0$ is recovered by the BCC/LDPC decoder.

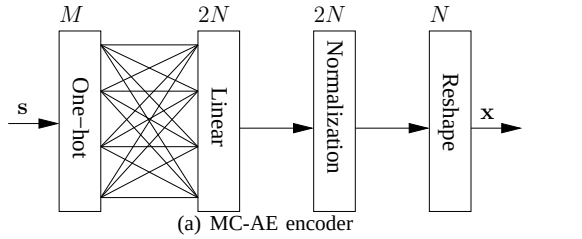
More explicitly, the mutual information attained is evaluated from:

$$\mathcal{I}(b; L) = \frac{1}{2} \sum_{b=1,0} \int_{-\infty}^{\infty} p(L|b) \log_2 \frac{2p(L|b)}{\sum_{b'=1,0} p(L|b')} dL, \quad (1)$$

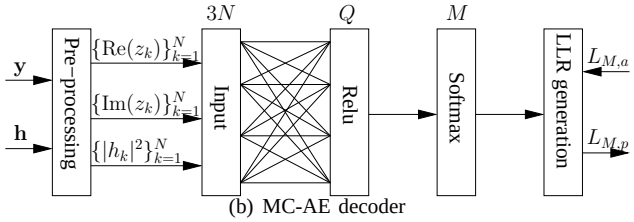
where an equiprobable source having $p(b = 1) = p(b = 0) = \frac{1}{2}$ is assumed, while the PDFs $p(L|b = 1)$ and $p(L|b = 0)$ are evaluated by the histograms of L [4], [27]. Based on this, the powerful tool of EXIT charts [2], [24]–[27] is capable of analysing the convergence behaviour of the turbo receiver, where both $\mathcal{I}_{M_E} = \mathcal{I}(L_{M,e}; b_m)$ and $\mathcal{I}_{D_E} = \mathcal{I}(L_{D,e}; b_c)$ of each iteration are evaluated, so that both the total number of iterations and the SNR required for reaching $\mathcal{I}_{D_E} = 1.0$ can

TABLE II: Notations used throughout the paper.

$(\cdot)^*$	the conjugate of a complex number	$(\cdot)^H$	the Hermitian transpose of a complex matrix
$\mathrm{Re}(\cdot)$	the real part of a complex number	$\mathrm{Im}(\cdot)$	the imaginary part of a complex number
$\ln(\cdot)$	natural logarithm function	$\exp(\cdot)$	natural exponential function
$p(\cdot)$	the probability	$\mathrm{E}(\cdot)$	the expectation
$[\mathbf{A}]_{a,b}$	the element on the a-th row and b-th column of matrix	$\mathbf{A} \in \mathbb{C}^{c\times d}$	denotes $\mathbf{A}$ is a complex-valued matrix of size $(c \times d)$
$\|\mathbf{A}\|$	the norm of a matrix	$\mathcal{I}(L;b)$	mutual information between LLRs L and binary bits b
$L_{M,e}$	the *extrinsic* LLRs produced by MC-AE decoder	$\mathcal{I}_{M_E}$	the *extrinsic* mutual information of MC-AE
$L_{M,a}$	the *a priori* LLRs produced by MC-AE decoder	$\mathcal{I}_{M_A}$	the *a priori* mutual information of MC-AE
$L_{D,e}$	the *extrinsic* LLRs produced by BCC/LDPC decoder	$\mathcal{I}_{D_E}$	the *extrinsic* mutual information of BCC/LDPC
$L_{D,a}$	the *a priori* LLRs produced by BCC/LDPC decoder	$\mathcal{I}_{D_A}$	the *a priori* mutual information of BCC/LDPC
$L_{M,p}$	the *a posteriori* LLRs produced by MC-AE decoder	$L_{D,p}$	the *a posteriori* LLRs produced by BCC/LDPC decoder
$\mathcal{CN}(\mu,\sigma^2)$	complex normal distribution with meanμ and variance σ^2		



(a) MC-AE encoder



(b) MC-AE decoder

Fig. 2: Schematic of multi-carrier autoencoder (MC-AE) encoder at the transmitter and decoder at the receiver for N subcarriers in the frequency-domain.

be predicted before running time-consuming BER simulations. Moreover, without having to access the source bits at the transmitter, the turbo receiver may evaluate the mutual information in an "on-line" fashion [2], [24]–[27] as:

$$\mathcal{I}(L;b)=1+E\left[\frac{e^L}{1+e^L}\log_2\left(\frac{e^L}{1+e^L}\right)+\frac{1}{1+e^L}\log_2\left(\frac{1}{1+e^L}\right)\right], \tag{2}$$

so that the iterations may be terminated, when no extra information can be gleaned. More details on EXIT charts will be presented in Sec. V.

B. MC-AE Structure

The MC-AE partitions a total number of N_c subcarriers (SCs) into G groups of N SCs $(N_c = GN)$. The MC-AE encoder of N SCs in each group is portrayed by Fig. 2(a). For each group, $(\mathrm{BPB} = \log_2 M)$ bits are transmitted, where the message is represented by the set $\mathcal{S} = \{\mathcal{S}_1, \cdots, \mathcal{S}_M\}$. The input of the MC-AE encoder is in the form of the one-hot vector of:

$$\mathbf{s} = [\underbrace{0 \cdots 0\, 1\, 0 \cdots 0}_{M}]^T, \quad (m\text{-th position}) \tag{3}$$

where we have $\mathbf{s} \in \mathbb{C}^{M\times 1}$ and the activation index m carries source information. After the input layer, the AE encoder has a linear fully-connected layer with a weight matrix $\mathbf{W} \in \mathbb{C}^{2N\times M}$ and a bias vector $\mathbf{b} \in \mathbb{C}^{2N\times 1}$. The output of this layer is given by $(\mathbf{v} = \mathbf{W}\mathbf{s} + \mathbf{b})$. Following this, a normalization layer is formulated as $(\overline{\mathbf{v}} = \sqrt{N}\mathbf{v}/\|\mathbf{v}\|)$, where the $2N$-element real-valued vector $\overline{\mathbf{v}}$ is reshapded into am N-element complex-valued signal vector $\mathbf{x}$ based on $(\overline{\mathbf{v}} = [\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})]^T)$, and then $\mathbf{x}$ is transmitted over N SCs in the classic OFDM system. In a nutshell, the AE encoder may be simply formuated as:

$$\mathbf{x} = f_{\boldsymbol{\theta}_{\text{enc}}}(\mathbf{s}), \tag{4}$$

where $\boldsymbol{\theta}_{\text{enc}} = \{\mathbf{W}, \mathbf{b}\}$ refers to the trainable parameters.

Following the frequency-domain operation of (4), the time-domain samples are generated by the N_c-point inverse fast fourier transform (IFFT) at the transmitter and then transmitted over a frequency-selective fading channel. Upon invoking the N_c-point FFT at the receiver, the frequency-domain received signal of each group is modelled by:

$$y_k = h_k x_k + n_k, \tag{5}$$

for $(k = 1, \cdots, N)$. The transmitted signal is contaminated by the Rayleigh fading of $h_k \in \mathcal{CN}(0,1)$ and additive white Gaussian noise (AWGN) of $n_k \in \mathcal{CN}(0, N_0)$, where N_0 refers to the noise power. However, we emphasize once again that in stark contrast to model-based detectors, the wireless channel's distribution is not known to the DNN, where the DNN is fed and trained by data.

In the frequency-domain, the MC-AE decoder for N SCs is portrayed by Fig. 2(b). The input of the AE's decoder is a $3N$-element vector of $\mathbf{u} = [\{\mathrm{Re}(z_k)\}_{k=1}^N, \{\mathrm{Im}(z_k)\}_{k=1}^N, \{|h_k|^2\}_{k=1}^N]^T$, where the decision variables are given by $(z_k = h_k^* y_k/|h_k|^2)$. This pre-processing beneficially reduces the AE input size from $4N$ to $3N$, where the conventional MC-AE design of [28] requires the input of $[\{\mathrm{Re}(y_k)\}_{k=1}^N, \{\mathrm{Im}(y_k)\}_{k=1}^N, \{\mathrm{Re}(h_k)\}_{k=1}^N, \{\mathrm{Im}(h_k)\}_{k=1}^N]^T)$. After the input layer, the AE's decoder is consituted by a linear fully connected hiden layer having Q nodes using the popular Relu activation function, followed by an output layer having M nodes employing the Softmax activation function. In summary, the AE's decoder is formulated as:

$$\hat{\mathbf{s}} = f_{\boldsymbol{\theta}_{\text{dec}}}(\mathbf{u}) = \sigma_{\text{Softmax}}[\mathbf{W}_2\sigma_{\text{Relu}}(\mathbf{W}_1\mathbf{u} + \mathbf{b}_1) + \mathbf{b}_2], \tag{6}$$

where $\boldsymbol{\theta}_{\text{dec}} = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2\}$ refers to the trainable parameters, while we have $\mathbf{W}_1 \in \mathbb{C}^{Q\times 3N}$, $\mathbf{b}_1 \in \mathbb{C}^{Q\times 1}$, $\mathbf{W}_2 \in \mathbb{C}^{M\times Q}$ and $\mathbf{b}_2 \in \mathbb{C}^{M\times 1}$.The activation functions are given by $\sigma_{\text{Relu}}(u_i) = \max(0, u_i)$ and $\sigma_{\text{Softmax}}(u_i) = \frac{e^{u_i}}{\sum_j e^{u_j}}$.

The training process of the MC-AE is independent of the channel coding schemes. More explicitly, the mean squared

TABLE III: Training hyperparameters.

Parameters	Value
Learning rate η	0.001
Batch size T	1000
Epochs	1000
Training data samples	10^5
Weight initialization	Xavier [42]
Optimizer	Adam [43]

error (MSE) loss function used for training the MC-AE is formulated as:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{s}^t - \hat{\mathbf{s}}^t\|^2, \tag{7}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\text{enc}}, \boldsymbol{\theta}_{\text{dec}}\}$ contains all trainable parameters, while T is the training batch size. Based on (7), the weights $\boldsymbol{\theta}$ are learned based on the stochastic gradient descent (SGD) with the gradient of the MSE loss function back-propagated from the output layer to the input layer as:

$$\boldsymbol{\theta} := \boldsymbol{\theta} - \eta \bigtriangledown \mathcal{L}(\boldsymbol{\theta}), \tag{8}$$

where η and $\bigtriangledown$ refer to the learning rate and the gradient, respectively. For improved training, the adaptive moment estimation (Adam) optimizer and the Xavier initialization are employed using the off-the-shelf Tensorflow of [28]. A summary of the hyperparameters selected for training the MC-AE is provided in Table III. Other important parameters such as the number of hidden nodes Q and the training SNR will be defined for each simulation in the following sections. Note that MC-AE is trained only at a single SNR, while tested at multiple SNRs.

Finally, we note that as the number of input bits (BPB $= \log_2 M$) increases, the dimension of the input layer $\mathbf{s} \in \mathbb{C}^{M\times 1}$ of (3) grows exponentially. In order to reduce the training and computational complexity, the embedded MC-AE (Emb-MC-AE) philosophy of [28] is harnessed. Instead of using the one-hot vector of (3), the reduced-size input embedding vector $\mathbf{s}$ is invoked, where all entries of $\mathbf{s}$ are trainable parameters. As a result, the MSE loss function of (7) is no longer applicable, since the input vector $\mathbf{s}$ has a smaller size. Therefore, the sparse categorical cross-entropy is used as our loss function formulated as $\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}\log([\hat{\mathbf{s}}^t]_m)$, where $[\hat{\mathbf{s}}^t]_m$ denotes the m-th element in $\hat{\mathbf{s}}^t$, and $m = \text{bin2dec}(b_1 \cdots b_{\text{BPB}})$ refers to the original data-carrying label modulated by the binary source bit-message. For example, when the one-hot vector has a large size such as $M = 256$, we may opt for a much smaller alternative embedding vector size of 16 [28].

III. Model-Based Analysis for MC-AE

The data-driven MC-AE is designed to be independent of the channel model. However, in order to be able to justify the performance advantage of MC-AE in physically tangible terms, we transform the MC-AE into an equivalent model-based representation, so that the model-based mathematical analysis metrics may be evaluated.

Definition 1 (CCMC capacity): Assuming that the noise is Gaussian-distributed as defined by $n_k \in \mathcal{CN}(0, N_0)$, the conditional probability based on (5) is given by:

$$p(y_k|x_k) = \frac{1}{\pi N_0}\exp\left(-\frac{|y_k - h_k x_k|^2}{N_0}\right). \tag{9}$$

Following this, the CCMC capacity is given by maximizing the mutual information $\mathcal{I}(x;y)$ between the channel's input and output signals over the distribution of $p(x)$ as $C^{\text{CCMC}}(SNR) = \max_{p(x)}\ \mathcal{I}(x;y) = \max_{p(x)}\ \mathcal{H}(y) - \mathcal{H}(y|x)$, where the entropies are maximized by the Gaussian input and output distributions, leading to $\mathcal{H}(y|x) = \log_2(\pi e N_0)$ and $\mathcal{H}(y) = \log_2\left[\pi e\left(|h|^2 + N_0\right)\right]$. Taking into account the SC index k, the unconstrained capacity is given by [4], [27], [44]:

$$C^{\text{CCMC}}(\text{SNR}) = \frac{1}{N}\sum_{k=1}^{N} E\left[\log_2\left(1 + \frac{|h_k|^2}{N_0}\right)\right]. \tag{10}$$

Alternatively to (5), the equivalent frequency-domain received signal model for N SCs can be expressed as:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}, \tag{11}$$

where we have $\mathbf{y} = [y_1, \cdots, y_N]^T \in \mathbb{C}^{N\times 1}$, $\mathbf{X} = \text{diag}([x_1, \cdots, x_N]) \in \mathbb{C}^{N\times N}$, $\mathbf{h} = [h_1, \cdots, h_N]^T \in \mathbb{C}^{N\times 1}$ and $\mathbf{n} = [n_1, \cdots, n_N]^T \in \mathbb{C}^{N\times 1}$. As a result, the received signal model of (11) becomes equivalent to that of a MIMO system associated with N transmit antennas (TAs), 1 receive antenna (RA) and N time slots for signal transmission, which allows us to use the classic MIMO rank and determinant criterion to interpret the MC-AE's diversity gain. The relationship between the two models is firstly demonstrated in terms of the CCMC capacity by the following proposition.

Proposition 1: The CCMC capacity of the frequency-domain representation of (10) based on the channel model of (5) is upper bounded by the MIMO's CCMC capacity based on the equivalent model of (11).

Proof. Based on the same AWGN assumption of $n_k \in \mathcal{CN}(0, N_0)$, the conditional probability based on (11) is given by:

$$p(\mathbf{y}|\mathbf{X}) = \frac{1}{(\pi N_0)^N}\exp\left(-\frac{\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2}{N_0}\right). \tag{12}$$

Following this, the CCMC capacity is given by $C^{\text{CCMC}}(SNR) = \max_{p(\mathbf{X})}\ \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{y}|\mathbf{X})$, where the entropies are also maximized by the Gaussian distributions, leading to $\mathcal{H}(\mathbf{y}|\mathbf{X}) = \log_2[\det(\pi e N_0 \mathbf{I}_N)]$ and $\mathcal{H}(\mathbf{y}) = \log_2\det\left\{\pi e\left[\frac{1}{N}\left(\mathbf{I}_N \otimes \mathbf{h}^H\right)\left(\mathbf{I}_N \otimes \mathbf{h}\right) + N_0\mathbf{I}_N\right]\right\}$. This is based on $\mathbf{y}^T = \text{rvec}(\mathbf{X})\left(\mathbf{I}_N \otimes \mathbf{h}\right) + \mathbf{n}^T$, where the operation $\text{rvec}(\cdot)$ forms a row vector by concatenating the rows in a matrix one-by-one. As a result, the equivalent MIMO's CCMC capacity is given by:

$$\begin{aligned} C^{\text{CCMC}}(\text{SNR}) &= \frac{1}{N}E\left\{\log_2\det(\mathbf{I}_N + \frac{1}{NN_0}\left[\mathbf{I}_N \otimes (\mathbf{h}^H\mathbf{h})\right]\right\} \\ &= E\left[\log_2\left(1 + \frac{\sum_{k=1}^{N}|h_k|^2}{NN_0}\right)\right]. \end{aligned} \tag{13}$$

Finally, based on Jensen's inequality, we have:

$$\frac{1}{N}\sum_{k=1}^{N} E\left[\log_2\left(1+\frac{|h_k|^2}{N_0}\right)\right] \leq E\left[\log_2\left(1+\frac{1}{N}\sum_{k=1}^{N}\frac{|h_k|^2}{N_0}\right)\right]. \tag{14}$$

□

Definition 2 (DCMC capacity): In practice, the finite-cardinality modulated signals constitute a non-Gaussian channel input. In this scenario, the mutual information is maximized, when the elements of the signal set $\{\mathbf{X}^m\}_{m=1}^{M}$ corresponding to the message set $\mathcal{S}=\{\mathcal{S}_1,\cdots,\mathcal{S}_M\}$ of (4) are equiprobable, i.e. we have $\{p(\mathbf{X}^m)=\frac{1}{M}\}_{m=1}^{M}$. As a result, the DCMC capacity [4], [27], [44] is formulated by (15).

Definition 3 (Error probability): Based on (11), the pairwise error probability (PEP) is evaluated by [4], [5], [27]:

$$\begin{aligned} E\left\{p(\mathbf{X}^m\rightarrow\mathbf{X}^{m'})\right\} &= E\left\{p\left(\|\mathbf{y}-\mathbf{X}^{m'}\mathbf{h}\|^2<\|\mathbf{n}\|^2\right)\right\} \\ &\leq E\left\{Q\left[\sqrt{\frac{\left\|(\mathbf{X}^m-\mathbf{X}^{m'})\mathbf{h}\right\|^2}{2N_0}}\right]\right\}, \end{aligned} \tag{16}$$

where $Q(\cdot)$ represents the integral form of the Q-function. The average PEP is given by:

$$\overline{P}_{e,bit}=E\left\{\sum_{m=1}^{M}\sum_{m'=1,m'\neq m}^{M}\frac{d_H(m,m')}{M\log_2 M}p(\hat{\mathbf{X}}=\mathbf{X}^{m'}|\mathbf{X}^m)\right\}, \tag{17}$$

where $d_H(m,m')$ refers to the Hamming distance between the bit-mappings of $\mathbf{X}^m$ and $\mathbf{X}^{m'}$, which may be directly obtained by mapping the indices m and m' back to $\log_2 M$ bits.

It is widely acknowledged that based on the PEP of (16), the minimum rank and determinant of the codeword difference $\mathbf{\Delta}=(\mathbf{X}^m-\mathbf{X}^{m'})^H(\mathbf{X}^m-\mathbf{X}^{m'})$ should be maximized for the diversity gain optimization. We observe here that the DCMC capacity of (15) also leads to the same conclusion, which is summarized by the following proposition.

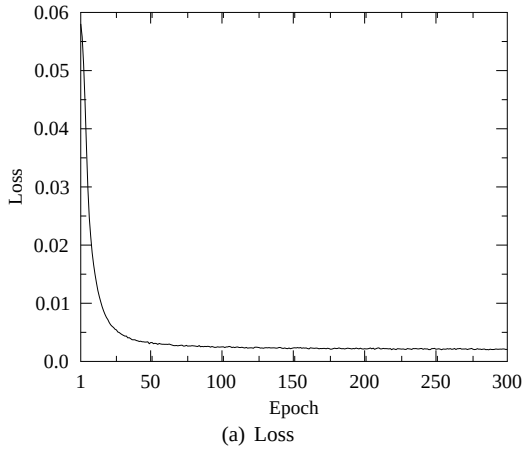
Proposition 2: Both the maximization of the DCMC capacity of (15) and the minimization of the PEP of (16) in the high-SNR region lead to the same rank and determinant criterion, which can be summarized as the maximization of the following diversity product [45], [46]:

$$\Lambda_p=\frac{1}{2}\min_{\forall m\neq m'}\ \det(\mathbf{\Delta})^{\frac{1}{2N}}. \tag{18}$$

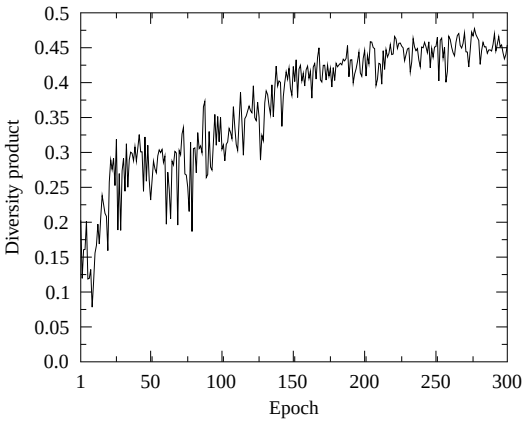
Proof. The maximization of the DCMC capacity in the high-SNR region is equivalent to minimizing the following metric:

$$\begin{aligned} E\left\{\exp\left(-\frac{\|(\mathbf{X}^m-\mathbf{X}^{m'})\mathbf{h}\|^2}{N_0}\right)\right\} &\leq \frac{1}{\det\left[\mathbf{I}_N+\frac{1}{N_0}\mathbf{\Delta}\right]} \\ &< \frac{1}{(\frac{1}{N_0})^{\text{rank}(\mathbf{\Delta})}\det(\mathbf{\Delta})}, \end{aligned} \tag{19}$$

where the first step is obtained by taking the expectation over the fading channel matrix $\mathbf{h}$, while the second step becomes approximately valid, when $N_0<<1$ in the high-SNR region.



(a) Loss



(b) Diversity product

Fig. 3: The loss and diversity product in training MC-AE associated with $N=4$, $M=16$ and $Q=128$, trained at SNR=7 dB, where epoch contains 20 batches of 1000 data samples.

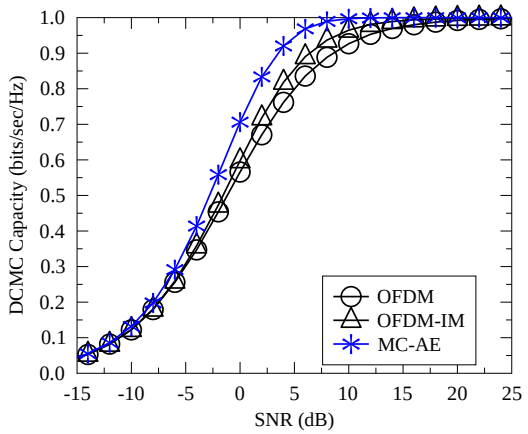
Similarly, upon applying the Chernoff bound, the PEP of (16) may be further extended as:

$$\begin{aligned} E\left\{p(\mathbf{X}^m\rightarrow\mathbf{X}^{m'})\right\} &\leq E\left\{\exp\left[-\frac{\left\|(\mathbf{X}^m-\mathbf{X}^{m'})\mathbf{h}\right\|^2}{4N_0}\right]\right\} \\ &\leq \left[\frac{1}{\det\left[\mathbf{I}_N+\frac{1}{4N_0}\mathbf{\Delta}\right]}\right] \\ &< \frac{1}{(\frac{1}{4N_0})^{\text{rank}(\mathbf{\Delta})}\det(\mathbf{\Delta})}. \end{aligned} \tag{20}$$

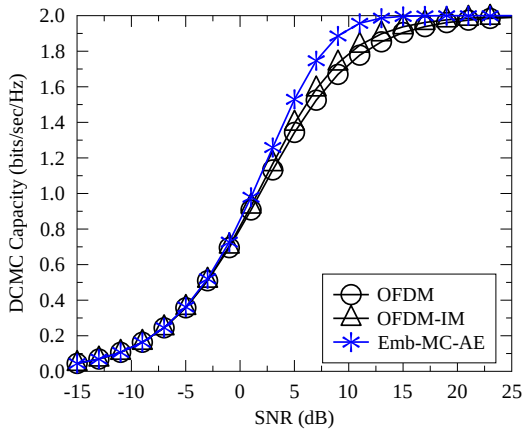
Therefore, the optimization of both the DCMC capacity and of the PEP in the high SNR region may be summarized as maximizing the diversity product of (18), which is the determinant metric normalized to be in the range of [0,1]. □

It may be readily seen that for both OFDM and OFDM-IM, the minimum rank and determiant are 1 and 0, respectively, where the worst case is encountered, when two codewords differ only in one SC. By contrast, Fig. 3 shows that as the MSE loss of (7) is reduced over the consecutive epoches, the diversity product of (18) is also implicitly maximized. This is due to the fact that as the difference between the input and output of the MC-AE is minimized during training, the model of (12) is also implicitly inferred by the DNN. As proven

$$
\begin{aligned}
C^{\text{DCMC}}(\text{SNR}) &= \max_{\{p(\mathbf{X}^m)\}_{m=1}^{M}} \sum_{m=1}^{M} \int p(\mathbf{y}|\mathbf{X}^m) p(\mathbf{X}^m) \log_2 \frac{p(\mathbf{y}|\mathbf{X}^m)}{\sum_{m'=1}^{M} p(\mathbf{y}|\mathbf{X}^{m'}) p(\mathbf{X}^{m'})} d\mathbf{y} \\
&= \frac{1}{NM} \sum_{m=1}^{M} E\left\{ \log_2 \left[\frac{M p(\mathbf{y}|\mathbf{X}^m)}{\sum_{m'=1}^{M} p(\mathbf{y}|\mathbf{X}^{m'})} \right] \right\} \\
&= \frac{\log_2 M}{N} - \frac{1}{NM} \sum_{m=1}^{M} E\left\{ \log_2 \left[\sum_{m'=1}^{M} \exp\left(-\frac{\|\mathbf{y}-\mathbf{X}^{m'}\mathbf{h}\|^2}{N_0} + \frac{\|\mathbf{y}-\mathbf{X}^{m}\mathbf{h}\|^2}{N_0} \right) \right] \right\} \\
&= \frac{\log_2 M}{N} - \frac{1}{NM} \sum_{m=1}^{M} E\left\{ \log_2 \left[\sum_{m'=1}^{M} \exp\left(-\frac{\|(\mathbf{X}^{m}-\mathbf{X}^{m'})\mathbf{h}+\mathbf{n}\|^2}{N_0} + \frac{\|\mathbf{n}\|^2}{N_0} \right) \right] \right\}.
\end{aligned} \tag{15}
$$

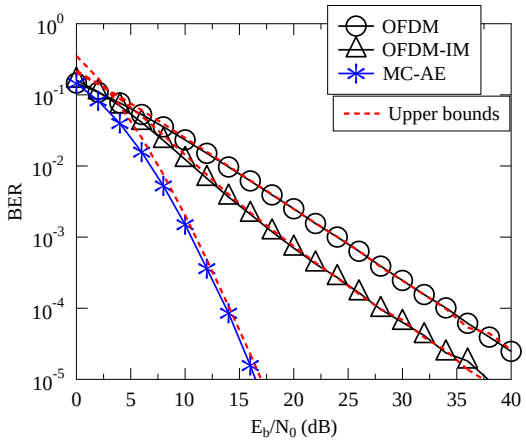


(a) $R_m = 1.0$

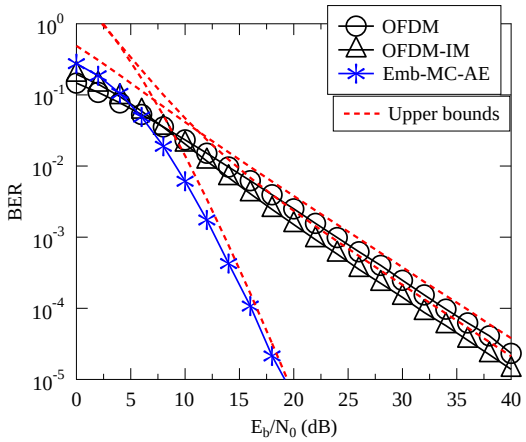


(b) $R_m = 2.0$

Fig. 4: DCMC capacity results of MC-AE and Emb-MC-AE in comparison to OFDM and OFDM-IM. The subfigures are (a) MC-AE associated with $N = 4$, $M = 16$ and $Q = 128$, trained at SNR=10 dB; (b) Emb-MC-AE associated with $N = 4$, $M = 256$ and $Q = 256$, trained at SNR=10.5 dB



(a) $R_m = 1.0$



(b) $R_m = 2.0$

Fig. 5: BER performance results of MC-AE and Emb-MC-AE in comparison to OFDM and OFDM-IM. The MC-AE decoder is temporarily replaced by the model-based ML detector. The subfigures are (a) MC-AE associated with $N = 4$, $M = 16$ and $Q = 128$, trained at SNR=10 dB; (b) Emb-MC-AE associated with $N = 4$, $M = 256$ and $Q = 256$, trained at SNR=10.5 dB

in Proposition 2, the optimization of this model results in codewords, which are as different as possible across the SCs. This can be successfully arranged by the DNN.

The DCMC capacity results of both the MC-AE and Emb-MC-AE in comparison to their OFDM and OFDM-IM counterparts are portrayed in Fig. 4, which confirms the performance advantages of the MC-AE design. Furthermore, the BER performance results are shown in Fig. 5, where the MC-AE decoder is temporarily replaced by the model-based maximum-likelihood (ML) detector, so that the simulation results can be compared to the theoretical ABER of (17). It is demonstrated by Fig. 5 that the ABER of (17) derived based on the equivalent model of (11) provides accurate upper bounds for the multi-carrier schemes of MC-AE, OFDM and OFDM-IM, where the MC-AE design is seen to be capable of achieving a substantial diversity gain.

IV. Soft-Input Soft-Out DNN

It is demonstrated in Sec. III that during the process of data-driven reconstruction, the trained MC-AE endeavours to recover the underlying model. This enables us to understand that the Softmax output probabilities $\{\hat{s}_m\}_{m=1}^{M}$ of (6) are the DNN's estimates on the probability distribution $p(\mathbf{y}|\mathcal{S}_m)$.

Based on the Log-MAP algortihm [4], [27], the MC-AE decoder may be revised to produce the *a posteriori* LLRs as:

$$L_{M,p}(b_k) = \ln \frac{\sum_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=1}} p(\mathbf{y}|\mathcal{S}_m)p(\mathcal{S}_m)}{\sum_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=0}} p(\mathbf{y}|\mathcal{S}_m)p(\mathcal{S}_m)}. \quad (21)$$

More explicitly, $\bar{b}_1 \cdots \bar{b}_{\text{BPB}}$ denotes the bit-mapping of the message $\mathcal{S}_m$, hence $\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=1}$ and $\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=0}$ refer to all codewords associated with $\bar{b}_k = 1$ and $\bar{b}_k = 0$, respectively. Then the *a priori* probabilities $\{p(\mathcal{S}_m)\}_{m=1}^{M}$ may be expressed as $p(\mathcal{S}_m) = \prod_{k'=1}^{\text{BPB}} \frac{\exp[\bar{b}_{k'} L_{M,a}(b_{k'})]}{1+\exp[L_{M,a}(b_{k'})]}$ based on the relationships of $p(\bar{b}_{k'} = 1) = \frac{\exp[L_{M,a}(b_{k'})]}{1+\exp[L_{M,a}(b_{k'})]}$ and $p(\bar{b}_{k'} = 0) = \frac{1}{1+\exp[L_{M,a}(b_{k'})]}$, where $\{L_{M,a}(b_k)\}_{k=1}^{\text{BPB}}$ denote the *a priori* LLRs gleaned from a channel decoder. Therefore, the Log-MAP algorithm of (21) may be rewritten as:

$$L_{M,p}(b_k) = \ln \frac{\sum_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=1}} d_m}{\sum_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=0}} d_m}, \quad (22)$$

where the probability metric d_m is given by:

$$d_m = \ln p(\mathbf{y}|\mathcal{S}_m) + \sum_{k'=1}^{\text{BPB}} \bar{b}_{k'} L_{M,a}(b_{k'}). \quad (23)$$

We note that the common constant of $\prod_{k'=1}^{\text{BPB}} \frac{1}{1+\exp[L_{M,a}(b_{k'})]}$ in all $\{p(\mathcal{S}_m)\}_{m=1}^{M}$ is eliminated by the division operation in (21). Moreover, we note that the MC-AE's Softmax outputs $\{\hat{s}_m\}_{m=1}^{M}$ are not the perfect estimates of $\{p(\mathbf{y}|\mathcal{S}_m)\}_{m=1}^{M}$, because the DNN does not have perfect knowledge of the channel model. As a result, directly using $\{\hat{s}_m \approx p(\mathbf{y}|\mathcal{S}_m)\}_{m=1}^{M}$ may leads to excess values of *a posteriori* LLRs of (22), which hence tends to mislead the channel decoder during iterations. As a remedy, we re-distribute the probabilities in the following way:

$$p(\mathbf{y}|\mathcal{S}_m) = \bar{p}\hat{s}_m + \frac{1-\bar{p}}{M}, \quad (24)$$

where $\bar{p}$ defines an artificially imposed probability upper bound, which is generally set to $\bar{p} = 0.8$ in the simulation results of this paper based on our empirical study.

In practice, the Log-MAP algorithm of (22) is simplied by the low-complexity Max-Log-MAP [4], [27] as:

$$L_{M,p}(b_k) = \max_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=1}} d_m - \max_{\forall \mathcal{S}_m \in \mathcal{S}_{\bar{b}_k=0}} d_m. \quad (25)$$

Let us consider the example of $(M = 4)$. The bit-representations are given by $\bar{b}_1\bar{b}_2 = \{00, 01, 10, 11\}$ for the codewords $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$. Accordingly, the two *a posteriori* LLRs produced by the Max-Log-MAP of (25) are given by $L_{M,p}(b_1) = \max(d_3, d_4) - \max(d_1, d_2)$ and $L_{M,p}(b_2) = \max(d_2, d_4) - \max(d_1, d_3)$, where the probability metrics are given by $d_1 = \ln p(\mathbf{y}|\mathcal{S}_1)$, $d_2 = \ln p(\mathbf{y}|\mathcal{S}_2) + L_{M,a}(b_2)$, $d_3 = \ln p(\mathbf{y}|\mathcal{S}_3) + L_{M,a}(b_1)$ and $d_4 = \ln p(\mathbf{y}|\mathcal{S}_4) + L_{M,a}(b_1) + L_{M,a}(b_2)$.

V. Performance Results of Turbo MC-AE

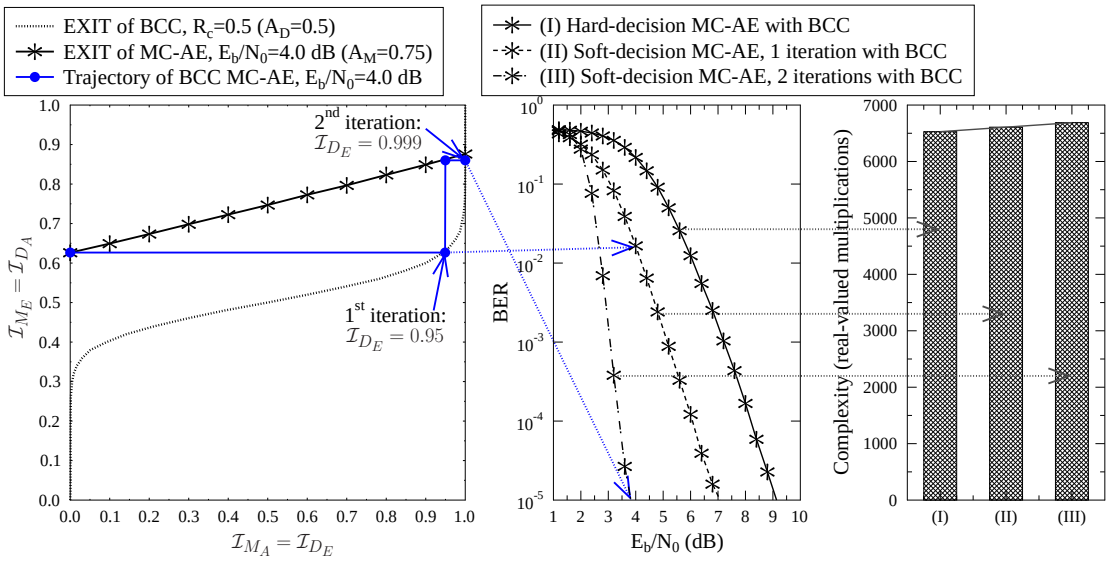
In this section, the performance results of the channel coded MC-AE schemes are compared to their multi-carrier counterparts of OFDM and OFDM-IM. The parameters are summarized in Table IV, where the modulation throughput and the channel coding rate are represented by R_m and R_c, respectively. The throughput of the MC-AE and OFDM are given by $(R_m = \frac{\text{BPB}}{N})$ and $(R_m = \text{BPS})$, respectively, where BPS denotes the number of bits per symbol for the modulated PSK/QAM scheme. The throughput of OFDM-IM [47], [48] is given by $(\log_2 \mathcal{C}_{N_G}^{N_A} + N_A\text{BPS})$, where $(\mathcal{C}_{N_G}^{N_A})$ represents all the possible combinations of activating N_A SCs out of N SCs.According to the IEEE 802.11a/n/ac/ah standard family, the BCC scheme invoked in this work is specified by the octal generator polynomials of $(g_0 = 133_8)$ and $(g_1 = 171_8)$ as well as the constrain length of $(K = 7)$. The coding frame length is set to 10 000 bits. The BCC rate is adjustable by puncturing, and we use $R_c = \{1/2, 2/3, 4/5\}$. Moreover, in the LDPC scheme, the number of neighbors of a variable node is set to $(d_v = 3)$ and that of a check node to $(d_c = 6)$, where the sum-product decoding algorithm is invoked.

The effect of integrating DNN into channel coded system is showcased by Fig. 6. First of all, the EXIT charts of Fig. 6(a) visualize the convergence behaviour of the turbo receiver of Fig. 1. In Fig. 6(a), the extrinsic information produced by the MC-AE decoder becomes the *a priori* information of the BCC decoder, i.e. $\mathcal{I}_{M_E} = \mathcal{I}_{D_A}$. Similarly, the extrinsic information produced by the BCC decoder becomes the *a priori* information of the MC-AE decoder, i.e. $\mathcal{I}_{D_E} = \mathcal{I}_{M_A}$. It can be seen in Fig. 6(a) that $\mathcal{I}_{D_E} = 0.95$ is produced after the first iteration at $E_b/N_0 = 4$ dB, which leads to $BER = 0.017$ in Fig. 6(b) as marked by the blue arrow with dashed line. Furthermore, Fig. 6(a) shows that $\mathcal{I}_{D_E} = 0.999$ is produced after the second iteration at $E_b/N_0 = 4$ dB, which leads to BER below 10^{-6} in Fig. 6(b) as marked by another blue arrow with dashed line. Therefore, the EXIT charts analysis is confirmed by the BER performance, where the beneficial iteration gain is indicated by both the improved $\mathcal{I}_{D_E}$ in Fig. 6(a) and the improved BER performance in Fig. 6(b). Secondly, Fig. 6(b) confirms that the performance of directly amalgamating hard-decision MC-AE with BCC performs the worst, which is improved by using the proposed soft-decision MC-AE having increased iterations with BCC. Thirdly, Fig. 6(c) demonstrates that the complexity differences between the MC-AE decoders are modest, when iterations are used between MC-AE and BCC. This is due to the fact that the proposed soft-decision MC-AE only needs to update the probabilities based on Sec. IV, where the DNN calculations in (6) only have to be invoked once before the ensuing iterations.

The EXIT charts of Fig. 7 compares the performance differences between MC-AE, OFDM and OFDM-IM, when they are integrated into the channel coded systems of LDPC and BCC. More explicitly, the area property of EXIT charts [2], [24]–[27] quantifies the relationship between the mutual information of (1) and the DCMC capacity as $A_M(\text{SNR}) = \int_0^1 \mathcal{I}_{M_E} d\mathcal{I}_{M_A} \approx \frac{\mathcal{C}^{\text{DCMC}}(\text{SNR})}{\text{BPB}}$. For the data-driven MC-AE decoder, the area $A_M(\text{SNR})$ indicates its decoding capabillity. Similarly, the area property for the channel coding results in $A_D = R_c$, which is independent of the SNR, as seen in Fig. 7. Hence, in order to achieve decoding convergence to a vanishingly low BER, the following two requirements have to

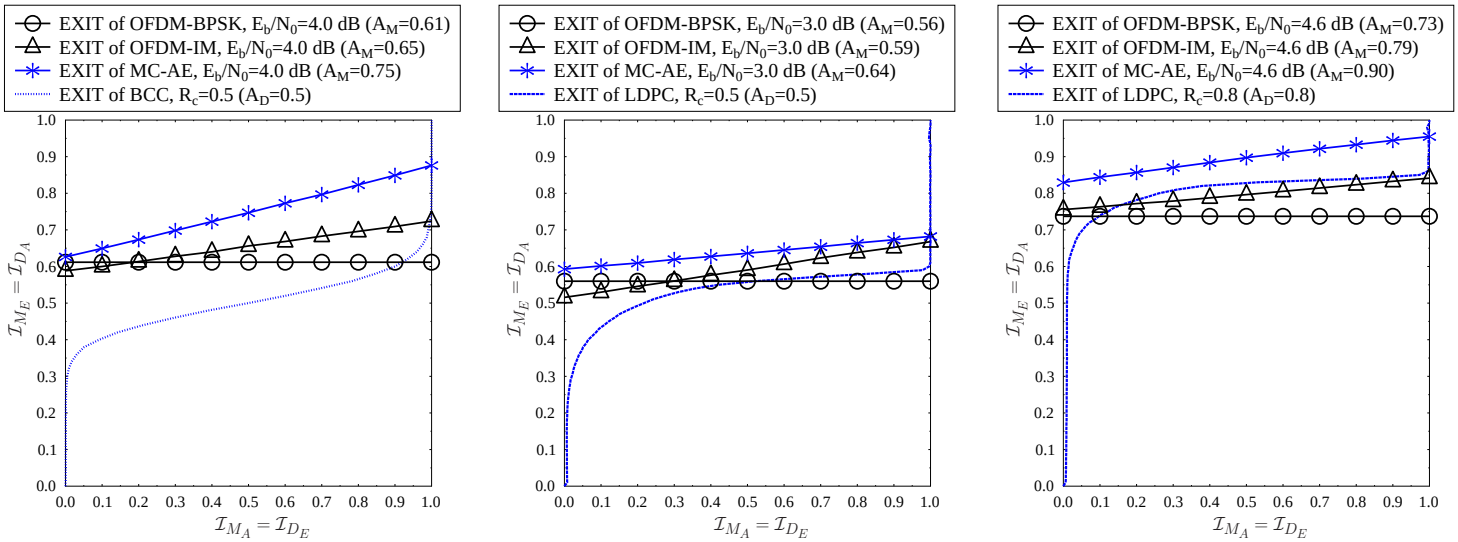
TABLE IV: Summary of parameters.

Scheme	Modulation Rate	Iterations	Coding Rate	Modulation Parameters
BCC MC-AE	$R_m = 1.0$	2	$R_c = 1/2$	$N = 4$, $M = 16$, trained at SNR=4dB
			$R_c = 2/3$	$N = 4$, $M = 16$, trained at SNR=0dB
			$R_c = 4/5$	$N = 4$, $M = 16$, trained at SNR=6dB
BCC OFDM-IM	$R_m = 1.0$	2	$R_c = \{1/2, 2/3, 4/5\}$	$N_G = 2$, $N_A = 1$, BPSK
BCC OFDM	$R_m = 1.0$	2	$R_c = \{1/2, 2/3, 4/5\}$	BPSK
LDPC MC-AE	$R_m = 1.0$	25 Inner 2 Outer	$R_c = 0.5$	$N = 2$, $M = 4$, trained at SNR=-1dB
			$R_c = 0.8$	$N = 4$, $M = 16$, trained at SNR=4dB
LDPC OFDM-IM	$R_m = 1.0$	25 Inner 2 Outer	$R_c = \{0.5, 0.8\}$	$N_G = 2$, $N_A = 1$, BPSK
LDPC OFDM	$R_m = 1.0$	50 Inner 1 Outer	$R_c = \{0.5, 0.8\}$	BPSK
BCC Emb-MC-AE	$R_m = 2.0$	2	$R_c = 1/2$	$N = 4$, $M = 256$, trained at SNR=3dB
			$R_c = 2/3$	$N = 4$, $M = 256$, trained at SNR=6dB
			$R_c = 4/5$	$N = 4$, $M = 256$, trained at SNR=6dB
BCC OFDM-IM	$R_m = 1.0$	2	$R_c = \{1/2, 2/3, 4/5\}$	$N_G = 4$, $N_A = 3$, QPSK
BCC OFDM	$R_m = 1.0$	2	$R_c = \{1/2, 2/3, 4/5\}$	QPSK



(a) EXIT charts (b) BER performance (c) Complexity

Fig. 6: The effect of integrating DNN into channel coded system. The MC-AE scheme is associated with $R_m = 1.0$. The BCC scheme is associated with $R_c = 0.5$. The complexity of the MC-AE decoder is quantified by the total number of real-valued multiplications.



(a) BCC, $R_c = 0.5$ (b) LDPC, $R_c = 0.5$ (c) LDPC, $R_c = 0.8$

Fig. 7: EXIT charts of LDPC and BCC aided MC-AE associated with $R_m = 1.0$.

be satisfied as seen in Fig. 7: (I) the area under the EXIT curve of the inner MC-AE decoder is higher than that of the outer BCC/LDCP decoder, i.e. $A_M > A_D$; (II) the only intersection between the EXIT curves of the BCC/LDPC decoder and the MC-AE decoder should occur at $\mathcal{I}_{D_E} = 1.0$.

Fig. 7 confirms that the MC-AE scheme exhibits an improved iteration gain, due to the fact that the MC-AE imposes inter-bit correlation in its mapping process. Specifically, Fig. 7(a) demonstrates that the MC-AE scheme substantially outperforms its OFDM and OFDM-IM counterparts, when the BCC is invoked. However, observe in Fig. 7(b) that the performance advantage of the MC-AE over OFDM and OFDM-IM becomes less significant in LDPC coded systems associated with $R_c = 0.5$. This is due to the fact that the LDPC associated with $R_c = 0.5$ represents a strong channel coding arrangement, which operates in the low SNR region at the cost of high channel decoding complexity and the throughput loss due to its low coding rate. By contrast, the beneficial diversity gain of MC-AE manifests itself better at higher SNRs. Therefore, Fig. 7(c) demonstrates that as expected, the performance advantage of MC-AE over OFDM and OFDM-IM becomes significant again, when the LDPC rate is increased to $R_c = 0.8$.

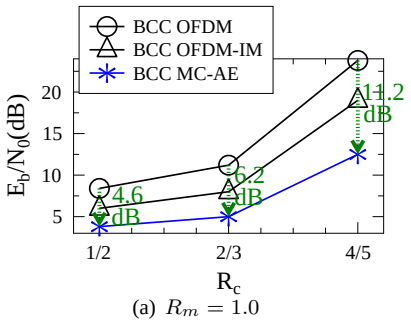
The EXIT charts analysis trends of Fig. 7 are further confirmed by the BER performance results of Fig. 8. More explicitly, Fig. 8(a) demonstrates that the proposed BCC MC-AE outperforms both BCC OFDM and BCC OFDM-IM at $R_c = 0.5$, where the power efficiency gain of BCC MC-AE over BCC OFDM-IM is a substantial 2.2 dB at BER=10^{-5}. The performance advantage of BCC MC-AE over BCC OFDM and BCC OFDM-IM is further increased, when a higher $R_c = 0.8$ is invoked in Fig. 8(b). In this scenario, the power efficiency gain of BCC MC-AE over BCC OFDM-IM is increased to an even more substantial 6.3 dB at BER=10^{-5}. However, the performance differences between the multi-carrier schemes become smaller in LDPC systems associated with $R_c = 0.5$, as seen in Fig. 8(c), where the power efficiency gain of LDPC MC-AE over LDPC OFDM-IM at BER=10^{-5} is reduced to 0.8 dB. Nonetheless, Fig. 8(d) once again demonstrates that as the LDPC rate is increased to $R_c = 0.8$, the performance advantages of LDPC MC-AE over LDPC OFDM and LPDC OFDM-IM become more significant, where the power efficiency gain of LDPC MC-AE over LDPC OFDM-IM at BER=10^{-5} is increased to 1.5 dB.

In summary, the performance advantage of the MC-AE design analysed in Sec. III manifests itself more explicitly for the BCC scheme. Fig. 9 summarizes our power efficiency comparison in terms of E_b/N_0 required for BER=10^{-5} between the MC-AE, OFDM and OFDM-IM in BCC coded scenarios. As expected, Fig. 9 demonstrates that for both the cases of $(R_m = 1.0)$ and $(R_m = 2.0)$, the performance advantage of BCC MC-AE over its OFDM and OFDM-IM counterparts becomes more substantial, as the BCC rate R_c increases.

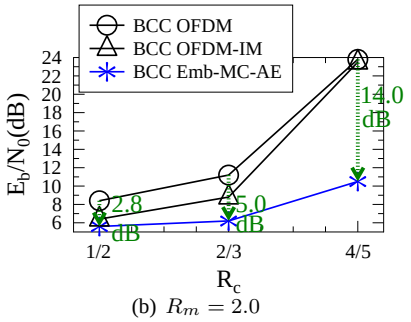
VI. Extension to V-BLAST MIMO

A. V-BLAST MC-AE Structure

In this section, the MC-AE design is extended to support V-BLAST MIMO systems associated with N_t TAs and N_r RAs.

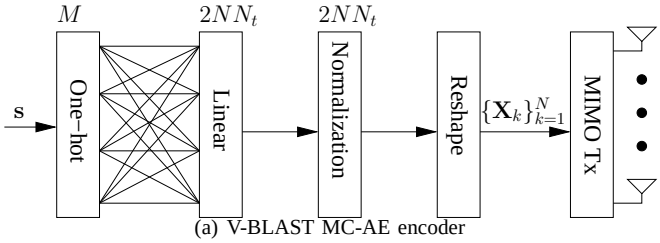


(a) $R_m = 1.0$

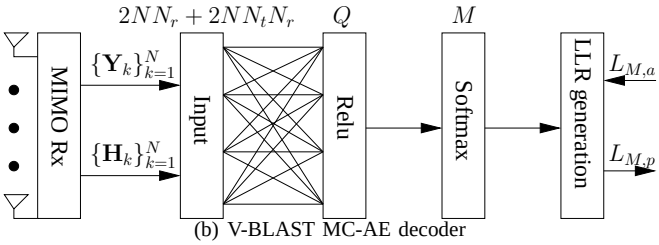


(b) $R_m = 2.0$

Fig. 9: Power efficiency comparison (E_b/N_0 required for BER=10^{-5}) between MC-AE, OFDM and OFDM-IM in BCC coded scenarios.



(a) V-BLAST MC-AE encoder

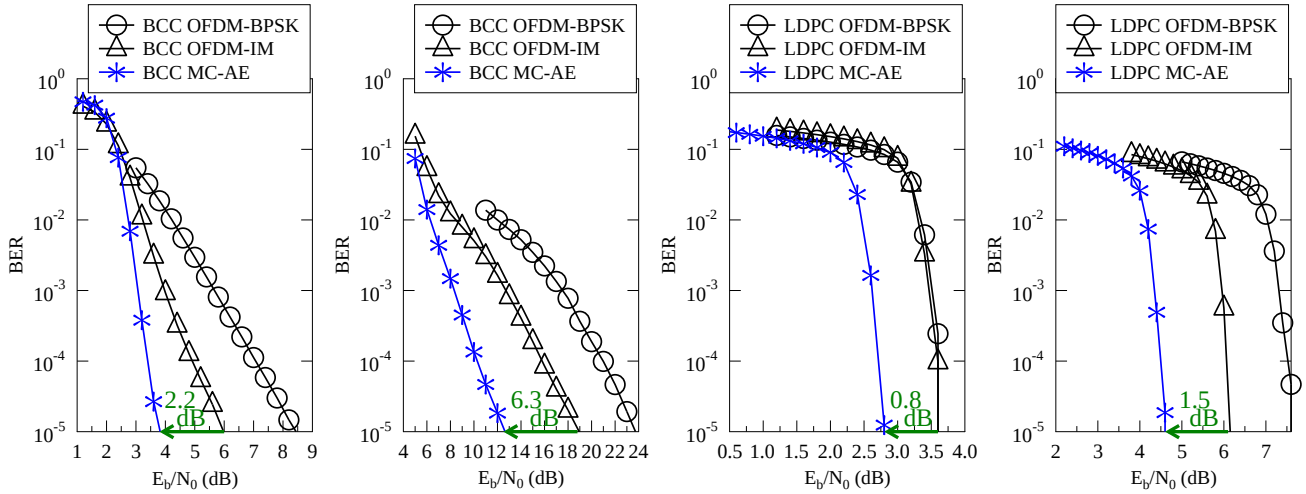


(b) V-BLAST MC-AE decoder

Fig. 10: Schematic of multi-carrier autoencoder (MC-AE) encoder at the V-BLAST MIMO transmitter and decoder at the V-BLAST MIMO receiver.

The V-BLAST MC-AE encoder is portrayed in Fig. 10(a). The input layer is also in the form of the one-hot vector $\mathbf{s} \in \mathbb{C}^{M\times 1}$ of (3). Following this, in order to prepare for the signal transmission over N_t TAs, the linear fully-connected layer is extended to have a larger weight matrix $\mathbf{W} \in \mathbb{C}^{2NN_t\times M}$ and bias vector $\mathbf{b} \in \mathbb{C}^{2NN_t\times 1}$. Upon obtaining the linear layer output $(\mathbf{v} = \mathbf{W}\mathbf{s} + \mathbf{b}) \in \mathbb{C}^{2NN_t\times 1}$, a normalization layer is invoked as $(\overline{\mathbf{v}} = \sqrt{N}\mathbf{v}/\|\mathbf{v}\|)$, and then the $2NN_t$-element real-valued vector $\overline{\mathbf{v}}$ is reshaped into the V-BLAST MIMO signal transmission over N_t TAs and N SCs, which is represented by $\{\mathbf{X}_k\}_{k=1}^{N}$.

Following the IFFT-based modulator at the transmitter and the FFT-based demodulator at the receiver, the frequency-



(a) BCC, $R_c = 0.5$ (b) BCC, $R_c = 0.8$ (c) LDPC, $R_c = 0.5$ (d) LDPC, $R_c = 0.8$

Fig. 8: Performance results of LDPC and BCC aided MC-AE associated with $R_m = 1.0$.

domain V-BLAST received signal model is given by:

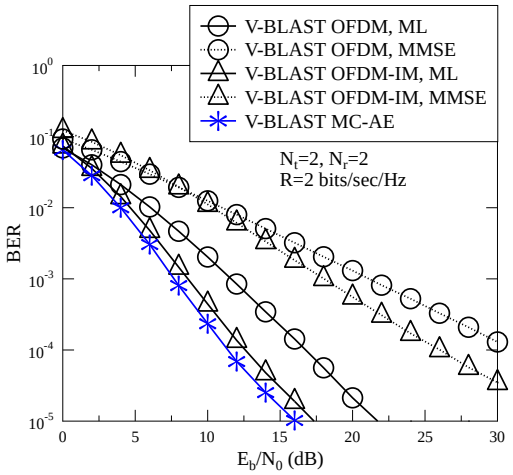
$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{N}_k, \tag{26}$$

for $(k = 1, \cdots, N)$, where we have $\mathbf{Y}_k \in \mathbb{C}^{N_r \times 1}$, $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{X}_k \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{N}_k \in \mathbb{C}^{N_r \times 1}$. Owing to the inherent inter-antenna interference in V-BLAST, the pre-processing of the MC-AE decoder seen in Fig. 2(b) can no longer be used. Instead, the input layer is constituted by $\mathbf{u} = [\{\mathrm{Re}(\mathbf{Y}_k^T)\}_{k=1}^N, \{\mathrm{Im}(\mathbf{Y}_k^T)\}_{k=1}^N, \{\mathrm{Re}\left(\mathrm{vec}(\mathbf{H}_k)^T\right)\}_{k=1}^N, \{\mathrm{Im}\left(\mathrm{vec}(\mathbf{H}_k)^T\right)\}_{k=1}^N]^T$, which has a total number of $(2NN_r + 2NN_tN_r)$ elements. The V-BLAST MC-AE decoder can be expressed in the same form as the MC-AE of (6), where the matrix and vector dimensions are adjusted to $\mathbf{W}_1 \in \mathbb{C}^{Q \times (2NN_r + 2NN_tN_r)}$, $\mathbf{b}_1 \in \mathbb{C}^{Q \times 1}$, $\mathbf{W}_2 \in \mathbb{C}^{M \times Q}$ and $\mathbf{b}_2 \in \mathbb{C}^{M \times 1}$.

The training procedure of the V-BLAST MC-AE is similar to that of the MC-AE presented in Sec. II-B. In particular, both the MSE loss function of (7) and the hyperparameters of Table III are employed for training the V-BLAST MC-AE, while the number of hidden nodes of the decoder and the training SNR will be detailed for each simulation in the next subsection.

B. Performance Results of V-BLAST MC-AE

The BER performance results of the V-BLAST MC-AE in comparison to its OFDM and OFDM-IM counterparts in uncoded scenarios are portrayed in Fig. 11, where the MC-AE is shown to be capable of achieving a beneficial diversity gain under the MIMO setup of $N_t = N_r = 2$. We note that first of all, the inherent concurrently nature of DNN relying on a parallel processing architecture of GPUs is generally considered to be capable of excuting the V-BLAST MC-AE decoder operations faster than the conventional CPU slowed down by the exponential complexity escalation of the ML detectors of V-BLAST OFDM and V-BLAST OFDM-IM [7], [8], [10], [11], [13], [49]. In light of this, the performance of the V-BLAST MC-AE is also compared to the MMSE detectors of V-BLAST OFDM and V-BLAST OFDM-IM, both of which have linearly increasing signal detection complexity. Observe in Fig. 11 that V-BLAST MC-AE substantially outperforms V-BLAST OFDM and V-BLAST OFDM-IM using



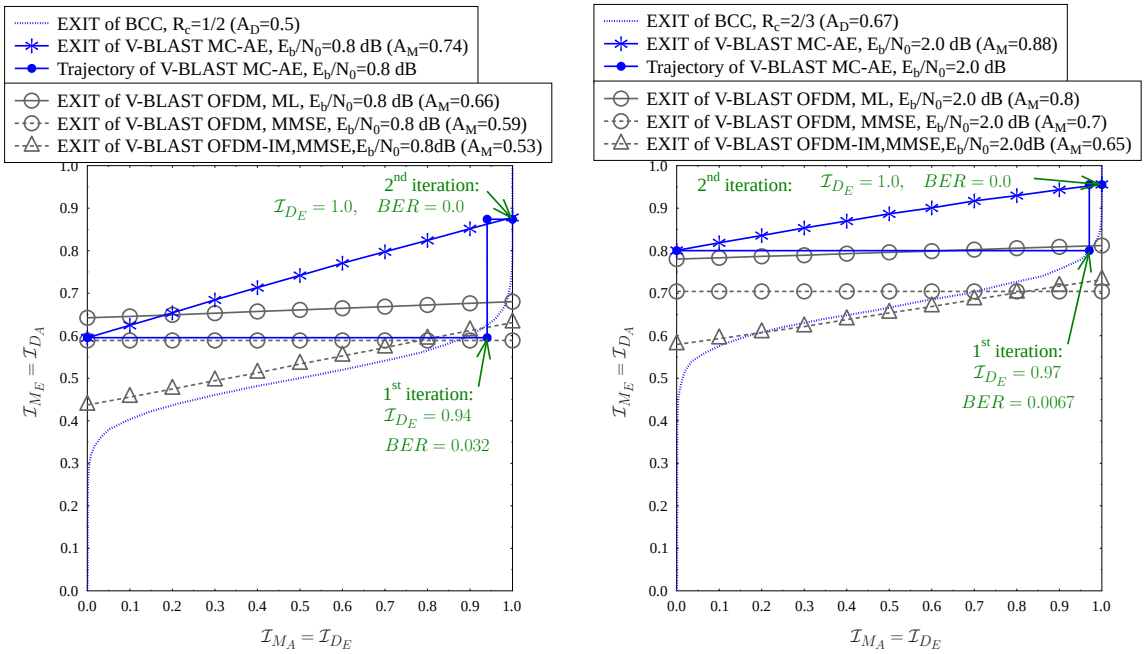
(a) $N_t = 2, N_r = 2$

Fig. 11: BER performance results of V-BLAST MC-AE in comparison to its OFDM and OFDM-IM counterparts in uncoded scenarios. The parameters of MC-AE are given by $N = 2$, $M = 16$ and $Q = 256$.

MMSE detectors. Furthermore, we note that V-BLAST MC-AE is expected to have an even further improved performance advantage over OFDM and OFDM-IM in channel coded scenarios, which is due to the fact that the MC-AE mapping procedure imposes correlation across the bits, hence leading to a beneficial turbo-iteration gain. This will be confirmed by the following performance results.

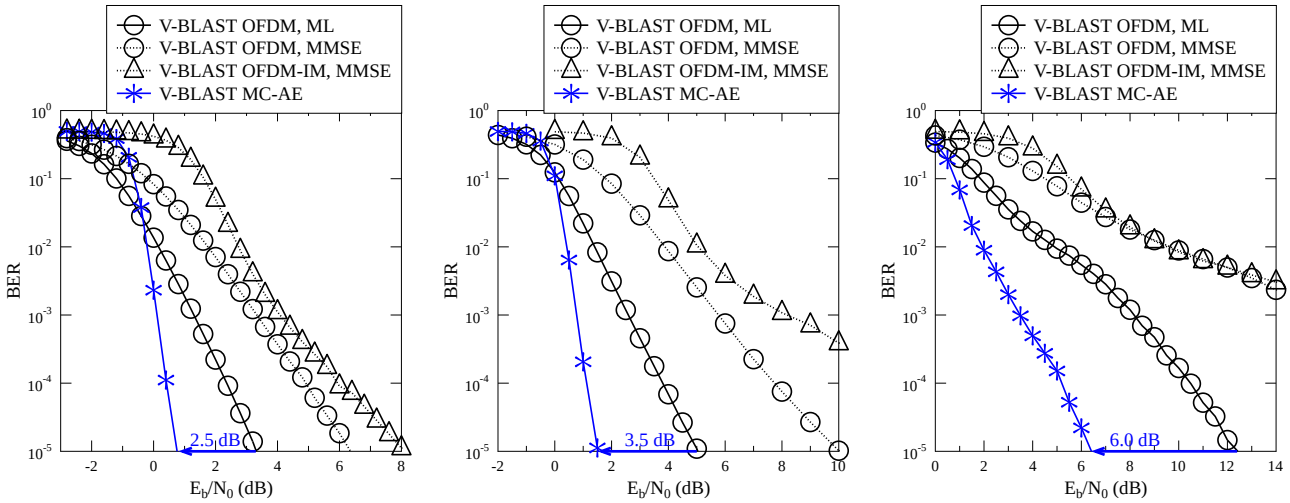
The EXIT charts of Fig. 12 confirm that first of all, based on the area property introduced in Sec. V, the V-BLAST MC-AE is capable of achieving a better performance than its OFDM and OFDM-IM counterparts in channel coded scenarios. Secondly, Fig. 12 further evidences that the V-BLAST MC-AE is capable of achieving a beneficial iteration gain, thanks to the bit-dependency imposed by the MC-AE encoder. Finally, the stair-case-shaped decoding trajectories portrayed in Fig. 12 confirm that the BCC V-BLAST MC-AE scheme is capable of achieving convergence at $E_b/N_0 = 0.8$ dB and $E_b/N_0 = 2$ dB for BCC rates of $R_c = 1/2$ and $R_c = 2/3$, respectively.

The EXIT charts analysis trends of Fig. 12 are further confirmed by the BER results of Fig. 13. Explicitly, it is



(a) BCC, $R_c = 1/2$ (b) BCC, $R_c = 2/3$

Fig. 12: EXIT charts and decoding trajectories of BCC aided V-BLAST MC-AE with $N_t = 2$, $N_r = 2$ and $R_m = 2.0$.



(a) $N_t = 2, N_r = 2, R_c = 1/2$ (b) $N_t = 2, N_r = 2, R_c = 2/3$ (c) $N_t = 2, N_r = 2, R_c = 4/5$

Fig. 13: BER performance results of V-BLAST MC-AE in comparison to its OFDM and OFDM-IM counterparts in BCC coded scenarios.

demonstrated by Fig. 13(a) that in terms of E_b/N_0 required for BER=10^{-5}, the BCC V-BLAST MC-AE scheme achieves a substantial 2.5 dB gain over its best-performing BCC V-BLAST OFDM counterpart using the ML detector, where we have the MIMO setup of $N_t = 2$ and $N_r = 2$, while the BCC rate is given by $R_c = 1/2$. Furthermore, this power efficiency gain is increased to even more substantial 3.5 dB and 6.0 dB, for the BCC rates of $R_c = 2/3$ and $R_c = 4/5$, as evidenced by Fig. 13(b) and Fig. 13(c), respectively.

VII. Extension to STBC MIMO

A. STBC MC-AE Structure

In this section, let us consider the orthogonal STBC design associated with N_t TAs and N_r RAs, which transmits N_Q symbols over N_P time slots. For the classic Alamouti's G2 STBC [29], we have $N_t = N_P = N_Q = 2$. For the amicable orthogoanl STBC [30] associated with $N_t = 2^\iota$ for any integer $\iota > 1$, we have $N_t = N_P$ and $N_Q = \iota + 1$. The STBC throughput is given by $R_m = \frac{N_Q}{N_P} \log_2 M$. The STBC MC-AE encoder is in the same form as the V-BLAST MC-AE encoder of Fig. 10(a), i.e. we have $(\mathbf{v} = \mathbf{W}\mathbf{s} + \mathbf{b}) \in \mathbb{C}^{2NN_Q \times 1}$, where the dimensions are updated for STBC as $\mathbf{W} \in \mathbb{C}^{2NN_Q \times M}$ and $\mathbf{b} \in \mathbb{C}^{2NN_Q \times 1}$. Upon obtaining the normalization layer $(\overline{\mathbf{v}} = \sqrt{N}\mathbf{v}/\|\mathbf{v}\|)$, the $2NN_Q$-element real-valued vector $\overline{\mathbf{v}}$ is reshaped into N_Q complex-valued symbols $\{x_{k,q}\}_{q=1}^{N_Q}$ for each subcarrier index $k = 1, \cdots, N$. Following this, the STBC signal matrix for each subcarrier may be formulated in the following dispersion matrix form [27]:

$$\begin{aligned} \mathbf{X}_k &= \sum_{q=1}^{N_Q} \left[\mathbf{A}_q \mathrm{Re}(x_{k,q}) + j\mathbf{B}_q \mathrm{Im}(x_{k,q})\right] \\ &= \sum_{q=1}^{N_Q} \left(\mathbf{D}_q^+ x_{k,q} + \mathbf{D}_q^- x_{k,q}^*\right). \end{aligned} \tag{27}$$

$$\mathbf{X}_k = \begin{bmatrix} x_{k,1} & -x_{k,2}^* \\ x_{k,2} & x_{k,1}^* \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{28}$$

For example, the Alamouti's G2 STBC signal matrix as well as its dispersion matrices are given by (28). Furthermore, the alternative dispersion matrices in (27) are given by $\left\{\mathbf{D}_q^+ = \frac{1}{2}(\mathbf{A}_q + \mathbf{B}_q)\right\}_{q=1}^{N_Q}$ and $\left\{\mathbf{D}_q^- = \frac{1}{2}(\mathbf{A}_q - \mathbf{B}_q)\right\}_{q=1}^{N_Q}$. Following this, the received signal at each subcarrier is given by:

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{N}_k, \tag{29}$$

where we have $\mathbf{Y}_k \in \mathbb{C}^{N_r \times N_P}$, $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{X}_k \in \mathbb{C}^{N_t \times N_P}$ and $\mathbf{N}_k \in \mathbb{C}^{N_r \times N_P}$. Owing to the STBC orthogonality that ensures $\mathbf{X}_k \mathbf{X}_k^H = \sum_{q=1}^{N_Q} |x_{k,q}|^2 \mathbf{I}_{N_t}$, the detection based on (29) may be decoupled as (30), where the decision variable is given by $\overline{z}_{k,q} = \frac{1}{\|\mathbf{H}_k\|^2} \mathrm{tr}\left[\mathbf{Y}_k (\mathbf{D}_q^+)^H \mathbf{H}_k^H + \mathbf{H}_k \mathbf{D}_q^- \mathbf{Y}_k^H\right]$ and the constant is given by $\varrho_k = \|\mathbf{Y}_k\|^2 - \sum_{q=1}^{N_Q} \|\mathbf{H}\|^2 |\overline{z}_{k,q}|^2$. Following this, the input layer of (10(b)) is streamlined for STBC as $\mathbf{u} = [\{\{\mathrm{Re}(\overline{z}_{k,q})\}_{k=1}^{N}\}_{q=1}^{N_Q}, \{\{\mathrm{Im}(\overline{z}_{k,q})\}_{k=1}^{N}\}_{q=1}^{N_Q}]^T$, which has a total number of $(2NN_Q)$ elements. The STBC MC-AE decoder can be expressed in the same form as the MC-AE of (6), where the matrix and vector dimensions are adjusted to $\mathbf{W}_1 \in \mathbb{C}^{Q \times 2NN_Q}$, $\mathbf{b}_1 \in \mathbb{C}^{Q \times 1}$, $\mathbf{W}_2 \in \mathbb{C}^{M \times Q}$ and $\mathbf{b}_2 \in \mathbb{C}^{M \times 1}$.

B. Performance Results of STBC MC-AE

The BER performance results of STBC MC-AE in comparison to its V-BLAST MC-AE counterpart in uncoded and BCC coded scenarios are portrayed in Fig. 14, where we have $N_t = 2$, $N_r = 2$ and $R_m = 2.0$. In the uncoded scenario, Fig. 14(a) demonstrates that STBC MC-AE achieves 9.8 dB power-efficiency gain over its V-BLAST MC-AE counterpart at BER=10^{-5}. In the BCC coded scenarios, Figs. 14(b)-14(d) show that STBC MC-AE achieves 3.6 dB, 4.1 dB and 5.1 dB power-efficiency gains at BCC rates of $R_c = 1/2$, $R_c = 2/3$ and $R_c = 4/5$, respectively, where two iterations are invoked between the soft-decision MC-AE decoder and BCC decoder. The performanc improvement achieved by STBC MC-AE over V-BLAST MC-AE is due to the space-time diversity gain of the STBC design, which is beneficially retained in the MC-AE system.

VIII. Conclusions

In this paper, we conceived a turbo MC-AE scheme for the family of OFDM system, where the DNN training is independent of both the choice of channel coding arrangement and of the channel model. Following this, we mapped the MC-AE's input-output relationship to an equivalent model-based representation, so that the performance advantage of MC-AE was intepreted in the form of theoretical capacity and PEP analysis. Based on this, we demonstrated that during the process of data-driven signal reconstruction, the MC-AE is capable of achieving a frequency diversity gain over both OFDM and OFDM-IM. In line with the model-based representation, we appropriately modified the MC-AE decoder for accepting and producing reliable soft-bit decisions. Furthermore, we extended the MC-AE scheme to support the V-BLAST and STBC MIMO setups. Our simulation results demonstrate that substantial performance advantages are achieved by the MC-AE design over its model-based counterparts of OFDM and OFDM-IM, when it is intrinsically amalgamated with the BCC and LDPC schemes considered.

References

[1] R. Steele and L. Hanzo, *Mobile Radio Communications*. John Wiley & Sons, May 1999.

[2] L. Hanzo, T. Liew, B. Yeap, R. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. John Wiley & Sons, 2011.

[3] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. John Wiley & Sons, 2004.

[4] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. John Wiley & Sons, May 2009.

[5] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

[6] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.

[9] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrows intelligent network traffic control systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2432–2455, 2017.

[10] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 2019.

[11] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 2018.

[12] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 2020.

[13] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognitive Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.

[14] S. Dorner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, 2018.

[15] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 298–303, 2018.

[16] T. Van Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, "Deep energy autoencoder for noncoherent multicarrier MU-SIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3952–3962, 2020.

[17] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, 2018.

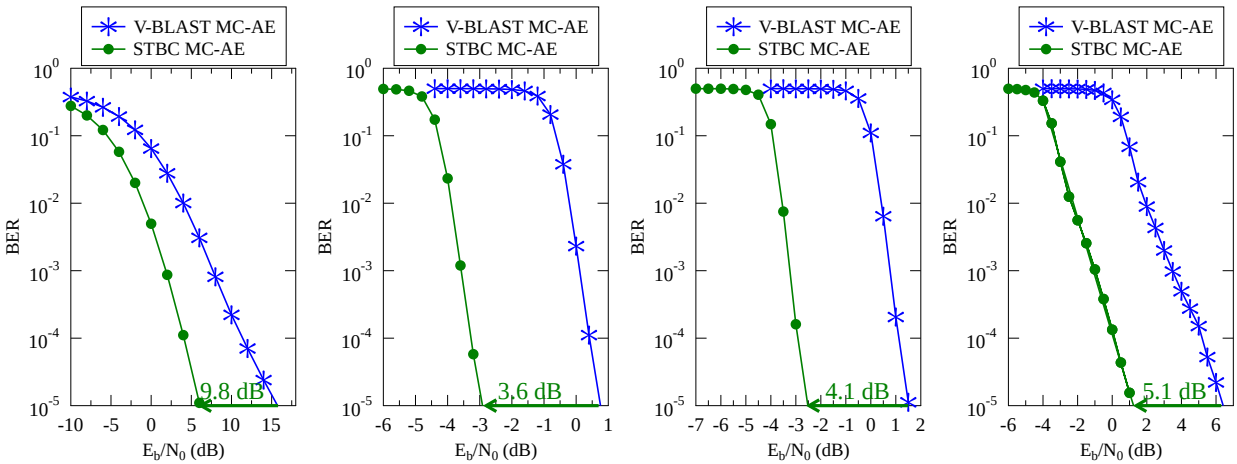
[18] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "ViterbiNet: A deep learning based Viterbi algorithm for symbol detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3319–3331, 2020.

[19] S. Cammerer, F. A. Aoudia, S. Drner, M. Stark, J. Hoydis, and S. ten Brink, "Trainable communication systems: Concepts and prototype," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5489–5503, 2020.

[20] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, 2019.

[21] H. Kim, S. Oh, and P. Viswanath, "Physical layer communication via deep learning," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 5–18, 2020.

$$\begin{aligned}\|\mathbf{Y}_k - \mathbf{H}_k\mathbf{X}_k\|^2 &= \|\mathbf{Y}_k\|^2 - \mathrm{tr}(\mathbf{Y}_k\mathbf{X}_k^H\mathbf{H}_k^H) - \mathrm{tr}(\mathbf{H}_k\mathbf{X}_k\mathbf{Y}_k^H) + \mathrm{tr}(\mathbf{H}_k\mathbf{X}_k\mathbf{X}_k^H\mathbf{H}_k^H)\\ &= \|\mathbf{Y}_k\|^2 + \textstyle\sum_{q=1}^{N_Q}\Big\{ -\mathrm{tr}\Big[\mathbf{Y}_k(\mathbf{D}_q^+)^H\mathbf{H}_k^H x_{k,q}^* + \mathbf{Y}_k(\mathbf{D}_q^-)^H\mathbf{H}_k^H x_{k,q} + \mathbf{H}_k\mathbf{D}_q^+\mathbf{Y}_k^H x_{k,q}\\ &\quad + \mathbf{H}_k\mathbf{D}_q^-\mathbf{Y}_k^H x_{k,q}^*\Big] + \|\mathbf{H}_k\|^2|x_{k,q}|^2\Big\}\\ &= \textstyle\sum_{q=1}^{N_Q}\left(\|\mathbf{H}_k\|^2|\overline{z}_{k,q} - x_{k,q}|^2\right) + \varrho_k,\end{aligned} \tag{30}$$



(a) Uncoded (b) BCC, $R_c = 1/2$ (c) BCC, $2/3$ (d) BCC, $R_c = 4/5$

Fig. 14: BER performance results of STBC MC-AE in comparison to its V-BLAST MC-AE counterpart in uncoded and BCC coded scenarios, where we have $N_t = 2$, $N_r = 2$ and $R_m = 2.0$.

[22] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Data-driven factor graphs for deep symbol detection," in *2020 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2682–2687, 2020.

[23] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication algorithms via deep learning," *arXiv preprint arXiv:1805.09317*.

[24] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, 2001.

[25] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, 2004.

[26] M. El-Hajjar and L. Hanzo, "Exit charts for system design and analysis," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 127–153, 2014.

[27] C. Xu, S. Sugiura, S. X. Ng, P. Zhang, L. Wang, and L. Hanzo, "Two decades of MIMO design tradeoffs and reduced-complexity mimo detection in near-capacity systems," *IEEE Access*, vol. 5, pp. 18564–18632, 2017.

[28] T. Van Luong, Y. Ko, M. Matthaiou, N. A. Vien, M. T. Le, and V. D. Ngo, "Deep learning-aided multicarrier systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.

[29] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[30] G. Ganesan and P. Stoica, "Space-time block codes: a maximum SNR approach," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1650–1656, 2001.

[31] A. Felix, S. Cammerer, S. Drner, J. Hoydis, and S. Ten Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 1–5, 2018.

[32] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, 2018.

[33] M. Kim, W. Lee, and D. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 510–513, 2018.

[34] T. V. Luong, Y. Ko, N. A. Vien, D. H. N. Nguyen, and M. Matthaiou, "Deep learning-based detector for OFDM-IM," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1159–1162, 2019.

[35] M. Kim, N. Kim, W. Lee, and D. Cho, "Deep learning-aided SCMA," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 720–723, 2018.

[36] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.

[37] 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 Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4492–4497, 2020.

[38] H. He, C. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[39] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.

[40] 3GPP Technical Report 21.915, "Summary of Rel-15 work items," *[Online]. Available: http://www.3gpp.org/release-15*.

[41] C. Xu, N. Ishikawa, R. Rajashekar, S. Sugiura, R. G. Maunder, Z. Wang, L. Yang, and L. Hanzo, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178246–178299, 2019.

[42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. AISTATS*, pp. 249–256, 2010.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ICLR*, 2014.

[44] 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, 2006.

[45] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, Z. Wang, and L. Hanzo, ""Near-perfect" finite-cardinality generalized space-time shift keying," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2146–2164, 2019.

[46] C. Xu, T. Bai, J. Zhang, R. G. Maunder, S. Sugiura, Z. Wang, and L. Hanzo, "Constant-envelope space-time shift keying," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1387–1402, 2019.

[47] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, pp. 5536–5549, Nov 2013.

[48] C. Xu, Y. Xiong, N. Ishikawa, R. Rajashekar, S. Sugiura, Z. Wang, S. X. Ng, L. L. Yang, and L. Hanzo, "Space-, time-and frequency-domain index modulation for next-generation wireless: A unified single-/multi-carrier and single-/multi-RF MIMO framework," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[49] Y. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 127–138, 2017.