Unity-Rate Coding Improves the Iterative Detection Convergence of Autoencoder-Aided Communication Systems

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

***Abstract*—A forward error correction (FEC) and unity-rate coded (URC) autoencoder (AE)-assisted communication system is proposed for the first time, which relies on soft iterative decoding for attaining a vanishingly low error probability. The AE-demapper is specifically designed for directly calculating the *extrinsic* logarithmic likelihood ratios (LLRs), which can be directly entered into the URC decoder for soft iterative decoding. This avoids the potential degradation due to the conversion of symbol probabilities to bit LLRs. A comprehensive capacity analysis of the AE is performed, which demonstrates the capacity advantage of the AE-aided constellation design over its conventional quadrature amplitude modulation (QAM)/ phase shift keying (PSK) counterpart. Furthermore, we carry out its EXtrinsic Information Transfer (EXIT) chart analysis, which indicates that as a benefit of our URC, the EXIT curve always reaches the [1,1] point of perfect convergence, leading to a vanishingly low error probability. More explicitly, our bit error ratio (BER) and block error ratio (BLER) results demonstrate that the proposed FEC-URC-AE system achieves significant iterative gains both in additive white Gaussian noise (AWGN) and Rayleigh channels, outperforming both its model-based FEC-AE and its conventional coded QAM/QPSK counterparts.**

***Index Terms*—Unity-rate code (URC), autoencoder (AE), iterative detection and decoding, EXtrinsic Information Transfer (EXIT) chart.**

I. Introduction

Although learning techniques have been used in wireless communications for three decades, advances in deep learning (DL) have led to their renaissance in the context of physical-layer communications [1–6]. The applications of deep neural networks (DNNs) to communication networks can be classified into two main categories, model-inspired and data-driven classes. The model-inspired solutions approximately adapt existing algorithms by invoking a DNN-like topology [7]. Different model-driven DNNs have been proposed for replacing either a single or multiple components of the conventional model-based communication systems, such as the multiple-input multiple-output (MIMO) detector [8, 9], channel estimator [10, 11], data detector [12, 13] or channel decoder [6]. By contrast, the family of data-driven approaches provide model-free solutions for these physical-layer communication challenges [14–16], which do not require a statistical input-output model, which is achieved by intrinsically integrating a conventional receiver into the DNN structure. Compared to the class of model-inspired DNN solutions, data-driven DL methods are capable of performing detection or decoding without the need for mathematical models, but this is achieved at the cost of requiring a large training dataset and a high training complexity.

Furthermore, joint optimization of the transmitter and receiver has been proposed for achieving end-to-end communications by using a so-called autoencoder (AE) [17], where the transmitter and receiver are implemented by a pair of DNNs. Additionally, the implementation of an AE-based communication system relying on software-defined radios was reported in [18]. The applications of the AE have penetrated into various scenarios of next-generation wireless communications, for instance, orthogonal frequency division multiplexing (OFDM) detection [19], random deployment of nonorthogonal multiple access (NOMA) [20], activity identification in grant-free random access [21], multicarrier communications [22], beamforming [23], just to name a few.

However, the aforementioned designs consider the optimization of symbol-based cross entropy, by maximizing the mutual information (MI) between the transmitter outputs and the receiver inputs. However, a holistic communication system always comprises a channel coding block, which typically processes bit-based information. In this case, the conversion of the softmax function outputs [24, 25], which converts the detected symbol probabilities to the bit-based logarithmic likelihood ratios (LLRs) imposes significant performance degradation. This prevents the potential implementation of holistic AE-based communication systems operating on the basis of bit LLRs. Hence, Cammerer *et al.* [26] proposed a novel bit-based AE mapper design, which maximized the MI between the transmitter outputs and the receiver inputs at the bit level, facilitating the bit interleaved coded modulation (BICM) for the family of AE-modulated communication systems.

L. Xiang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China. E-mail: luping.xiang@uestc.edu.cn. (*Corresponding author: Luping Xiang*)

C. Xu, T.V. Luong, R. G. Maunder, L.-L. Yang and L. Hanzo are with School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK. (E-mail: cx1g08, thien.luong, rm, lly, l-h@soton.ac.uk).

X. Zhang is with both the School of Electronics and Computer Science and the Optoelectronics Research Centre, University of Southampton, Southampton, SO17 1BJ, U.K. (E-mail: xy.zhang@soton.ac.uk).

L. Hanzo 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).

TABLE I: Contrasting the contributions of this work to the literature.

Contributions	**This work**	[8]	[14–16]	[17]	[18]	[20–23]	[26]
Autoencoder (AE)-based system	✓			✓	✓	✓	✓
Model-based communication system	✓	✓	✓			✓	
Soft iterative decoding	✓						✓
Unity-rate code (URC)	✓						
Capacity Analysis	✓						✓
EXtrinsic Information Transfer (EXIT) chart analysis	✓						✓

While the built-in iterative decoding structure proposed in [26] achieves iterative gains, it fails to retain the classic model-based communication architecture. Given its data-driven approach, the iterative decoder of [26] requires training for each and every legitimate block length and for each forward error correction (FEC) coding scheme. For instance, in the context of the 5G New Radio [27, 28], exhaustive training of every low-density parity-check (LDPC) and polar code block length [29, 30], which requires extensive training sets and imposes an extremely high training overhead. However, there is a paucity of literature on iterative decoding of model-based AE-aided communication systems.

Therefore, we fill this knowledge gap by exploiting the attractive properties of unity-rate codes (URC) [31], which can be harnessed for improving the iterative decoding performance of turbo-style receivers, while incurring only a moderate decoding complexity [32], as it will be detailed in Section II. The family of URCs has been proposed for diverse applications, such as improving quantum codes [33], for visible-light communications [34, 35], as well as for simultaneous wireless information and power transfer (SWIPT) [36]. Hence, by incorporating a carefully designed URC, we conceive an iterative decoder for AE-based systems, which circumvents the inflexibility of iterative decoding designed for AE-based systems, albeit at the cost of slightly increased complexity. The main contributions of this paper are boldly and explicitly contrasted to the literature in Table I and are summarized as follows.

- A FEC-URC-AE system is proposed for soft iterative decoding, which is capable of retaining the model-based structure of classic communication systems. By employing the URC, bit-correlations are created between the FEC-coded and AE-mapped sequences, hence increasing the error-correction capability. Additionally, the AE-mapper of our system is designed for directly generating the *extrinsic* LLRs, which may then be entered into the URC decoder for soft iterative decoding. Hence, the optimization and training of the AE-mapper and demapper will directly deal with the LLRs. This completely avoids the degradation due to the conversion of symbol probabilities to bit LLRs. For demonstrating the theoretical advantage of the AE-mapper, both the continuous-input continuous-output memoryless channel (CCMC) and the discrete-input continuous-output memoryless channel (DCMC) capacity of the AE-modulation is analysed. This shows the capacity advantage of the AE-aided constellation design over its conventional quadrature amplitude modulation (QAM)/ phase shift keying (PSK) counterpart, for the first time.
- We carry out the EXtrinsic Information Transfer (EXIT) chart analysis of the proposed FEC-URC-AE system for demonstrating the benefits of incorporating a URC, which always results in an EXIT curve reaching at the unity MI at the [1,1] point, regardless of the signal-to-noise ratio (SNR). This indicates the potential of attaining a vanishingly low bit error ratio (BER) by using iterative decoding.
- Our BER and block error ratio (BLER) simulations demonstrate that the proposed FEC-URC-AE achieves significant iterative gains for both additive white Gaussian noise (AWGN) and Rayleigh fading channels, and outperforms both the model-based FEC-AE and the conventional FEC-QAM/QPSK system.

The rest of this paper is structured as follows. Section II describes the transceiver structure of the proposed FEC-URC-AE system. Then Section III proposes our iterative detection and decoding scheme. Section IV and V analyse the FEC-URC-AE system from the perspective of its capacity and EXIT characteristics, respectively. Following this, Section VI demonstrates the error-correction capability of the proposed system in terms of both its BER and BLER. Finally, our main conclusions and some promising future research directions are outlined in Section VII.

Notations: In this paper, the uppercase and lowercase boldface letters, $\boldsymbol{X}$ and $\boldsymbol{x}$, denote matrices and vectors, respectively, while the subscript z of the lowercase letter x represents the z-th element in the vector $\boldsymbol{x}$. Furthermore, $(\cdot)^{-1}$, $(\cdot)^{T}$, and $(\cdot)^{H}$ represent the matrix inversion, transpose and Hermitian transpose operations, respectively. Finally, $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts, respectively, of a complex vector or matrix.

II. System Model

In this section, we assume an end-to-end single-input single-output (SISO) communication scenario, where the modulator and the demodulator are represented by the DNN-based AE structure, so that they can be jointly designed and optimized. We introduce our proposed FEC-URC-AE transmitter and receiver structure in Sections II-A and II-B, respectively.

A. Transmitter

The proposed FEC-URC-AE transmitter relies on the serially-concatenated scheme, as depicted in Fig. 1, where the A-bit information bit sequence $\boldsymbol{u}_1$ is first input to the FEC encoder, with the output being the N-bit FEC coded sequence $\boldsymbol{c}_1$. This is then interleaved by the interleaver π_1 and entered into the URC encoder giving the identical-length N-bit

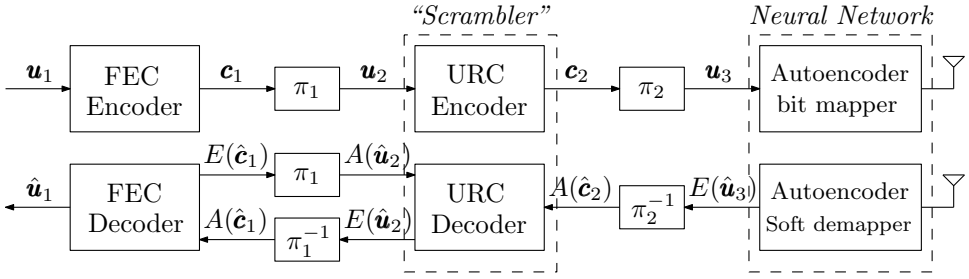

Fig. 1: The proposed FEC-URC-AE system with an iterative decoding scheme.

coded sequence $\boldsymbol{c}_2$. A URC encoder is a convolutional encoder having a coding rate of $R = 1$ [31]. Here, the URC encoder can be viewed as a 'scrambler' of the encoded bits, which beneficially 'smears and disperses' the *extrinsic* information within the code block without reducing the code rate R by convolving its input bit with the URC's generator polynomial. In this way, the URC encoding process may be viewed as a block-based *mapping* technique, which beneficially randomizes the positions of the *extrinsic* information bits in support of iterative detection. This is in contrast to the FEC-AE scheme of [26], where the *extrinsic* information propagation across symbols is not possible.

Following the concatenated encoding process of Fig. 1, the N-bit FEC-URC coded bit sequence $\boldsymbol{c}_2$ is rearranged by the interleaver π_2 and the resultant scrambled sequence $\boldsymbol{u}_3 = \pi_2(\boldsymbol{c}_2)$ is split into $Z = N/m$ groups, giving a $(m \times Z)$-dimensional coded bit matrix $\boldsymbol{U}_3$, where $m = \log_2 M$ represents the number of bits to be mapped into a single M-ary amplitude-phase modulation (APM) symbol and M is the modulation order. In this work, we adopt the one-hot mapping approach proposed in [8]. The z-th column $\boldsymbol{u}_{3,z}$ $(z = 1, 2, \cdots, Z)$ of $\boldsymbol{U}_3$ will be input to the AE-mapper of Fig. 1 for bit-to-symbol mapping during the z-th symbol duration.

A simple example illustrating the FEC-URC-AE transmitter processing is shown in Fig. 2, where the information bits $\boldsymbol{u}_1 = [1101]$ are encoded by the $R = 1/2$ recursive systematic convolutional (RSC) encoder relying on the generator polynomial of $[1011, 1101]$. After interleaving, the resultant interleaved sequence $\boldsymbol{u}_2 = [10110011]$ is entered into the URC encoder having the generator polynomial of $[10]$ and the feedback polynomial of $[11]$. After interleaving, one-hot mapping is applied to the interleaved bits $\boldsymbol{u}_3 = [11111001]$.

More specifically, the AE-mapper replaces the conventional QAM/PSK bit-to-symbol mapper by a fully-connected multi-layer DNN [8, 26]. Within the z-th symbol duration, the coded bit sequence $\boldsymbol{u}_{3,z}$ comprising m bits will be input to the AE-mapper at the transmitter after the one-hot preprocessin [8]. Specifically, the m-bit $\boldsymbol{u}_{3,z}$ is converted to the $M = 2^m$-bit one-hot vector $\boldsymbol{e}_{3,z}$, which uses the non-zero element position to indicate the index of the 'learned' constellation point. Considering the example of Fig. 2, the one-hot mapping of a 4-bit sequence $\boldsymbol{u}_{3,z}$ can be defined as

$$\begin{aligned}\boldsymbol{u}_{3,z} =& [0,0,0,0] \\ &\Rightarrow \boldsymbol{e}_{3,z} = \quad [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1], \end{aligned} \tag{1a}$$

$$\begin{aligned}\boldsymbol{u}_{3,z} =& [0,0,0,1] \\ &\Rightarrow \boldsymbol{e}_{3,z} = \quad [0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0], \end{aligned} \tag{1b}$$

$$\vdots$$

$$\begin{aligned}\boldsymbol{u}_{3,z} =& [1,1,1,1] \\ &\Rightarrow \boldsymbol{e}_{3,z} = \quad [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]. \end{aligned} \tag{1c}$$

Then, the M one-hot bits comprise the M input neurons of the AE-mapper, whereas the hidden layer of the AE-mapper is comprised of $2M$ neurons. The output layer is constituted by the real and imaginary parts of the "learned" constellation point, as exemplified in Fig. 2. Given the optimization parameters of the AE-mapper $\boldsymbol{\theta}_{\text{en}} = \left\{\boldsymbol{W}_1^{(1)}, \boldsymbol{b}_1^{(1)}, \boldsymbol{W}_2^{(1)}, \boldsymbol{b}_2^{(1)}\right\}$, where $\boldsymbol{W}_l^{(1)}$ and $\boldsymbol{b}_l^{(1)}$ represent the weights and biases to be optimized in the l-th $(l = 1, 2)$ layer, the 2×1 output $\boldsymbol{t}_z$ of the AE-mapper gives the real and imaginary parts of the 'learned' constellation. The input-output relationship of the AE-mapper can be expressed as

$$\boldsymbol{t}_z = \boldsymbol{W}_2^{(1)} \left[f_{\text{ReLU}} \left(\boldsymbol{W}_1^{(1)} \boldsymbol{e}_{3,z} + \boldsymbol{b}_1^{(1)} \right) \right] + \boldsymbol{b}_2^{(1)}, \tag{2}$$

where f_{ReLU} is the rectified linear unit (ReLU) function [24], which is defined as $f_{\text{ReLU}}(x) = \max(x, 0)$, and the optimization parameters $\boldsymbol{W}_1^{(1)}$, $\boldsymbol{W}_2^{(1)}$, $\boldsymbol{b}_1^{(1)}$ and $\boldsymbol{b}_2^{(1)}$ have dimensions of $2M \times M$, $2 \times 2M$, $2M \times 1$ and 2×1, respectively.

After obtaining the output $\boldsymbol{t}_z = [\Re(\bar{x}_z), \Im(\bar{x}_z)]^T$, which is comprised of the real and imaginary parts of the original 'learned' complex-valued constellation point $\bar{x}_z \in \bar{\mathcal{M}}$, respectively, where $\bar{\mathcal{M}}$ is the original 'learned' symbol set, normalization is performed to give the normalized complex-valued constellation point $x_z \in \mathcal{M}$, where $\mathcal{M}$ is the normalized symbol set. The normalization process of $\boldsymbol{t}$ to x_z can be formulated as,

$$\begin{aligned} x_z &= \Re(x_z) + j\Im(x_z) \\ &= \frac{\sqrt{ME_s}(\Re(x_z) + j\Im(x_z))}{\sqrt{\sum_{\bar{x}_z \in \bar{\mathcal{M}}} [\Re(\bar{x}_z)^2 + \Im(\bar{x}_z)^2]}}, \end{aligned} \tag{3}$$

where E_s is the transmit power. In this work, we assume the normalised transmit power of $E_s = 1$ without loss of

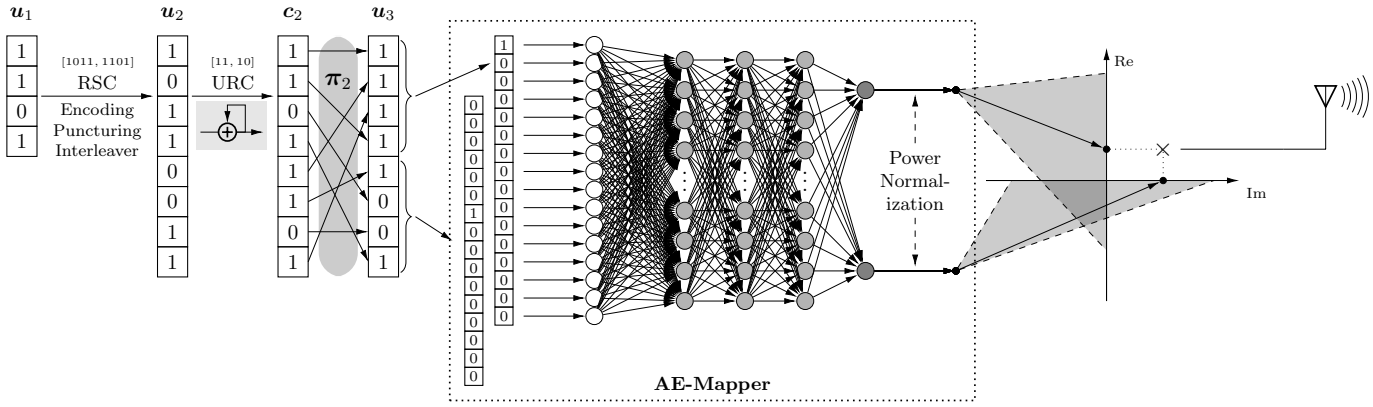

Fig. 2: Example of the RSC-URC-AE transmitter processing the information bit sequence of $\boldsymbol{u}_1 = [1101]$.

generality. Note that the learning process of the AE-mapper has been detailed in [26] and will be reviewed in Section .

The AE-mapped symbols of all Z groups to be transmitted are expressed as

$$\boldsymbol{x} = [x_1, x_2, \cdots, x_Z]^T. \tag{4}$$

In this way, the bit-to-symbol mapping is achieved by the employment of the AE-mapper, rather than by the conventional QAM/PSK bit-to-symbol mapper. Additionally, since the AE-mapper is considered as a component of the classic model-based communication system, it can be readily implemented as an independent component of our proposed RSC-URC-AE system, with each column of the coded bit matrix $\boldsymbol{U}_3$ serially entered into the AE-mapper of Fig. 1.

B. Receiver

Given the channel impulse response (CIR) vector $\boldsymbol{h} = [h_1, h_2, \cdots, h_Z]^T$ spreading over Z symbol durations at the receiver, the $Z \times 1$ received observations $\boldsymbol{y} = [y_1, y_2, \cdots, y_Z]^T$ can be expressed as

$$\boldsymbol{y} = \text{diag}\{\boldsymbol{h}\}\boldsymbol{x} + \boldsymbol{n}, \tag{5}$$

where $\text{diag}\{\boldsymbol{h}\}$ represents the $Z \times Z$ diagonal matrix constructed from $\boldsymbol{h}$, and $Z \times 1$ vector $\boldsymbol{n}$ is the AWGN, which obeys the complex Gaussian distribution with a zero mean and a covariance matrix of $\boldsymbol{I}_Z \sigma_N^2$.

The receiver performs the inverse operations of the transmitter, as shown in 1 and will be detailed in the next section.

III. AE-aided Iterative Detection and Decoding

In contrast to [26], which proposed a 'built-in' iterative decoding architecture, in this section, we introduce our AE-aided block-based iterative detection and decoding in Section III-A, followed by the training of the AE in Section III-B.

A. RSC-URC-AE Decoding Process

To begin with, as shown in Fig. 3, zero-forcing (ZF) preprocessing is applied to the z-th received observation y_z, which can be expressed as $\bar{y}_z = y_z h_z^* / |h_z|^2$. The real $\Re(\bar{\boldsymbol{y}})$ and imaginary $\Im(\bar{\boldsymbol{y}})$ parts of the resultant $\bar{\boldsymbol{y}} = [\bar{y}_1, \bar{y}_1, \cdots, \bar{y}_Z]^T$ are then entered into the fully-connected DNN at the receiver, together with the magnitudes of the channel $\bar{\boldsymbol{h}} = [|h_1|^2, |h_1|^2, \cdots, |h_Z|^2]^T$, rather than both the real and imaginary parts of $\boldsymbol{h}$, as exemplified in Fig. 3. In this way, the length of the input-tuple is reduced from 4 to 3 at each symbol duration, hence reducing the training and testing complexity.

More specifically, the real $\Re(\bar{y}_z)$ and imaginary $\Im(\bar{y}_z)$ parts of the ZF-preprocessed received signal $\bar{y}_z$ as well as the magnitude $\bar{h}_z$ are the 3 inputs of the AE-demapper, as illustrated in Fig. 3, giving a 3×1 input vector $\boldsymbol{y}_{0,z}$ expressed as

$$\boldsymbol{y}_{0,z} = [\Re(\bar{y}_z), \Im(\bar{y}_z), \bar{h}_z]^T. \tag{6}$$

Furthermore, the activation function employed at each hidden layer of the AE-demapper is also the ReLU function f_{ReLU} as in the AE-mapper. After Z symbol durations, the output layer applies a linear activation function to directly create N *extrinsic* bit LLRs of the demapped bits. The L-layer DNN employed in the AE-demapper in our work is comprised of $(L-1) = 2$ hidden layers and an output layer, with the optimization weights and biases being $\boldsymbol{\theta}_{\text{de}} = \left\{\boldsymbol{W}_1^{(2)}, \boldsymbol{b}_1^{(2)}, \boldsymbol{W}_2^{(2)}, \boldsymbol{b}_2^{(2)}, \boldsymbol{W}_3^{(2)}, \boldsymbol{b}_3^{(2)}\right\}$. Then the N output *extrinsic* LLRs $E(\hat{\boldsymbol{u}}_3)$ by the AE-demapper are expressed as

$$\begin{aligned} E(\hat{\boldsymbol{u}}_3) =& \boldsymbol{W}_3^{(2)} \\ & \cdot f_{\text{ReLU}}(\boldsymbol{W}_2^{(2)} \cdot f_{\text{ReLU}}(\boldsymbol{W}_1^{(2)} \boldsymbol{y}_0 + \boldsymbol{b}_1^{(2)}) + \boldsymbol{b}_2^{(2)}) + \boldsymbol{b}_3^{(2)}. \end{aligned} \tag{7}$$

Note that each of the $(L-1)$ hidden layers is comprised of Q neurons, whereas the output layer has m neurons, which output m *extrinsic* bit LLRs of the demapped symbol $\hat{\boldsymbol{x}}_z$.

Then the output LLRs $E(\hat{\boldsymbol{u}}_3)$ are entered into the URC decoder after the de-interleaver π_2^{-1} of Fig. 1. The iterative decoding is performed by exchanging soft *extrinsic* information between the URC and FEC decoder, namely $E(\hat{\boldsymbol{u}}_2)$ obtained by the URC decoder and the *extrinsic* LLRs $E(\hat{\boldsymbol{c}}_1)$ after FEC decoding, respectively. This leads to significant iterative gains, as it will be demonstrated by our EXIT chart analysis in Section V as well as by the BER and BLER results of Section VI.

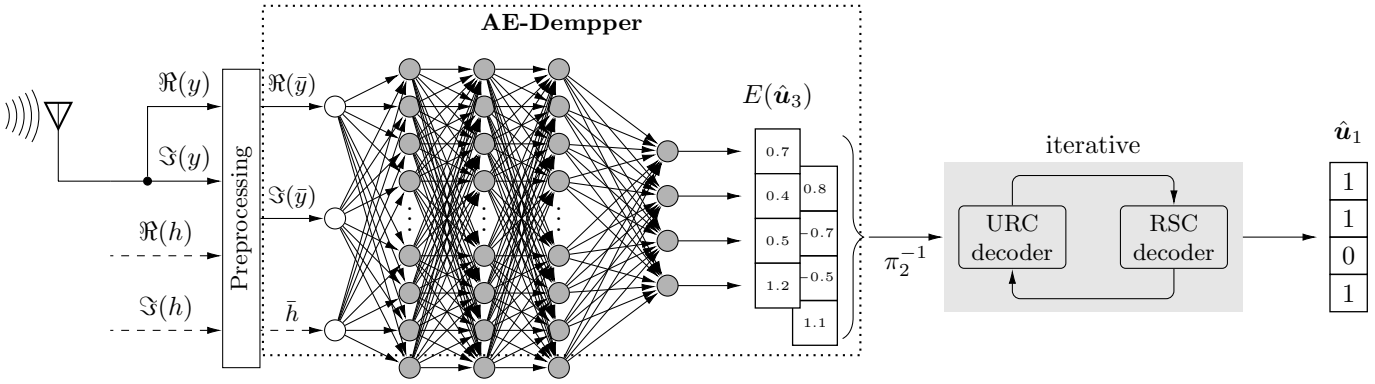

Fig. 3: Example of the RSC-URC-AE iterative receiver.

B. Training of the AE

The parameters of the AE that will be optimized by training are $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\text{en}}, \boldsymbol{\theta}_{\text{de}}\}$, where $\boldsymbol{\theta}_{\text{en}} = \left\{\boldsymbol{W}_1^{(1)}, \boldsymbol{b}_1^{(1)}, \boldsymbol{W}_2^{(1)}, \boldsymbol{b}_2^{(1)}\right\}$ is employed in (2) and $\boldsymbol{\theta}_{\text{de}} = \left\{\boldsymbol{W}_1^{(2)}, \boldsymbol{b}_1^{(2)}, \boldsymbol{W}_2^{(2)}, \boldsymbol{b}_2^{(2)}, \boldsymbol{W}_3^{(2)}, \boldsymbol{b}_3^{(2)}\right\}$ is employed in (7). Additionally, the training dataset of the AE is comprised of the FEC-URC encoded bit vector $\boldsymbol{u}_{3,b}$, the CIR $\bar{h}_b$ and the channel SNR γ, where $b = 1, 2, \cdots, B$ is the vector index of a training batch and B is the batch size. In this work, we choose $B = 1024$. Since our proposed FEC-URC-AE system is a model-based communication system, the AE-mapper is trained independently for each and every modulation order, and the trained AE-mapper can be used for the entire block, regardless of the block length.

In the training process, the AE aims for minimizing the difference of the FEC-URC encoded bit vector $\boldsymbol{u}_{3,b}$ and the corresponding detected bit vector at the receiver $\hat{\boldsymbol{u}}_{3,b}$. Hence, in this paper, the loss function $L(\boldsymbol{\theta})$ is expressed as

$$L(\boldsymbol{\theta}) = -\frac{1}{Bm}\sum_{b=1}^{B}\sum_{v=1}^{m} \left(u_{3,b}^{(v)}\log(\hat{u}_{3,b}^{(v)}) + (1-u_{3,b}^{(v)})\log(1-\hat{u}_{3,b}^{(v)})\right), \quad (8)$$

where $u_{3,b}^{(v)}$ is the v-th bit in $\boldsymbol{u}_{3,b}$, the *extrinsic* LLR $E(\hat{u}_{3,b}^{(v)})$ is the v-th element in the vector $E(\hat{\boldsymbol{u}}_{3,b})$ given by (7) and $\hat{u}_{3,b}^{(v)}$ is the output of the sigmoid function $f_{\text{sigmoid}}(\cdot)$, which may be expressed as

$$\begin{aligned}\hat{\boldsymbol{u}}_{3,b} &= f_{\text{sigmoid}}[E(\hat{\boldsymbol{u}}_{3,b})] \\ &= 1/[1+e^{-E(\hat{\boldsymbol{u}}_3)}]. \end{aligned} \quad (9)$$

By employing the loss function of (8), the parameters in $\boldsymbol{\theta}$ are updated for the batches and randomly picked from the dataset, using the classic stochastic gradient descent (SGD) algorithm [37] formulated as

$$\boldsymbol{\theta} := \boldsymbol{\theta} - \varepsilon\nabla L(\boldsymbol{\theta}), \quad (10)$$

where ε is the learning rate of the SGD and $\nabla L(\boldsymbol{\theta})$ represents the gradient of $L(\boldsymbol{\theta})$. In this paper, $\varepsilon = 10^{-4}$ is selected for the performance characterization of the proposed AE architecture. Note that during the training process, the widely-adopted adaptive moment estimation (Adam) optimizer is employed for the off-line learning [38].

IV. Capacity Analysis

In this section, we derive the continuous-input continuous-output memoryless channel (CCMC) and discrete-input continuous-output memoryless channel (DCMC) capacity of the DNN-based communication system in Sections IV-A and IV-B, respectively. Following this, the CCMC and DCMC capacity results are quantified and discussed in Section IV-C.

A. CCMC Capacity Analysis

The CCMC capacity models the channel by continuous inputs as well as continuous outputs associated with an infinite modulation order M [39], which is defined as

$$C_{\text{CCMC}} = \mathbb{E}_{h_z}\left[\log_2(1+|h_z|^2\gamma)\right] \text{ (bits/sec/Hz)}, \quad (11)$$

where the SNR γ under unity transmit power is defined as $\gamma = 1/\sigma_N^2$ and $\mathbb{E}_{h_z}[\cdot]$ is the expectation taken over the channel.

Observe from (11) that the constellation order or constellation design will not influence the CCMC capacity, as it will also be demonstrated in Section IV-C. However, in practical communications, discrete QAM/PSK symbols are input to the wireless channels. Hence, we will discuss the DCMC capacity in the next section, which quantifies the memoryless channel capacity relying on discrete inputs and gives more accurate and tight bound.

B. DCMC Capacity Analysis

We now derive the DCMC capacity of the system. The DCMC capacity (bits/sec/Hz) of an uncoded SISO system is given by [39] and is calculated as

$$C_{\text{DCMC}} = \max_{p(y_z|x_z), x_z\in\mathcal{M}} \sum_{x_z\in\mathcal{M}} \int_{-\infty}^{+\infty} p(y_z, x_z) \times \log_2\left(\frac{p(y_z|x_z)}{\sum_{v_z\in\mathcal{M}} p(y_z, v_z)}\right) \mathrm{d}y_z, \quad (12)$$

where $P(y_z|x_z)$ is the conditional probability density function (PDF) of the received signal y_z given the transmitted x_z, which

can be expressed as

$$P(y_z|x_z) = \frac{1}{2\pi\sigma_N^2} \exp\left(-\frac{|y_z - h_z x_z|^2}{2\sigma_N^2}\right). \quad (13)$$

Note that (13) is maximized when the transmitted symbols are equiprobable [39], i.e., $p(x_z = a_i) = 1/M$, $\forall a_i \in \mathcal{M}$. Hence, we arrive at

$$\begin{aligned}\log_2\left(\frac{p(y_z|x_z)}{\sum_{v_z\in\mathcal{M}} p(y_z, v_z)}\right) &= \log_2\left(\frac{p(y_z|x_z)}{\sum_{v_z\in\mathcal{M}} p(y_z|v_z)p(v_z)}\right) \\ &= -\log_2\left(\frac{1}{M}\sum_{v_z\in\mathcal{M}}\frac{p(y_z|x_z)}{p(y_z, v_z)}\right) \\ &= \log_2(M) - \log_2\sum_{v_z\in\mathcal{M}}\exp\left(\Psi\right), \end{aligned} \quad (14)$$

where by substituting (13) into (14), Ψ can be expressed as

$$\Psi = \frac{-|h_z\left(x_z - v_z\right) + n_z|^2 + n_z^2}{\sigma_N^2}. \quad (15)$$

Given a constellation set $\mathcal{M}$ having a modulation order M, (13) is maximized when all the constellation points have the same probability.

$$C_{\text{DCMC}} = \log_2(M) - \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}_{h_z,n_z}\left[\log_2\sum_{v=1}^{M}\exp(\Psi)\right]. \quad (16)$$

C. Simulation Results

Figs. 4 and 5 show the DCMC capacity of (16) for communication over both AWGN and Rayleigh fading channels, respectively, where an MPSK, MQAM or AE mapper is employed, using a modulation order of $M = 8$ or 16. Observe from both Figs. 4 and 5 that in the case of $M = 8$, the AE-mapper achieves a higher DCMC capacity in the low SNR region than 8QAM or 8PSK, which indicates that the 'learned' constellation achieves superior average Euclidean distance, compared to the 8QAM or 8PSK. By contrast, similar DCMC capacity can be observed when employing the conventional 16QAM and AE-mapper associated with $M = 16$, both of which show a higher DCMC capacity than 16PSK, before converging around SNR= 20 dB in Fig. 4. This lies in the perfect symmetry of the 16QAM. Additionally, the CCMC capacity is also included in Figs. 4 and 5 for upper-bounding the capacity, which quantifies the channel capacity associated with a continuous input.

V. EXIT Chart Analysis

In this section, we analyze the EXIT chart performance of the proposed FEC-URC-AE system.

To start with, following [40, 41], we assume that the *a priori* LLRs of the demapper $L^{(a)}(\boldsymbol{u}_3)$ obey the Gaussian distribution with a mean of μ_A and a variance of σ_A^2, where we have $\mu_A = \sigma_A^2/2$. Hence the conditional PDF associated with $L^{(a)}(\boldsymbol{u}_3)$ can be expressed as

$$p_a(\zeta|X = x) = \frac{1}{\sqrt{2\pi}\sigma_A}\exp\left(-\frac{(\zeta - \frac{\sigma_A^2}{2})^2}{2\sigma_A^2}\right). \quad (17)$$

Then, the MI between the *a priori* LLRs of the demapper $L^{(a)}(\boldsymbol{u}_3)$ and the transmitted signal sequence $\boldsymbol{x}$ of (4) can be

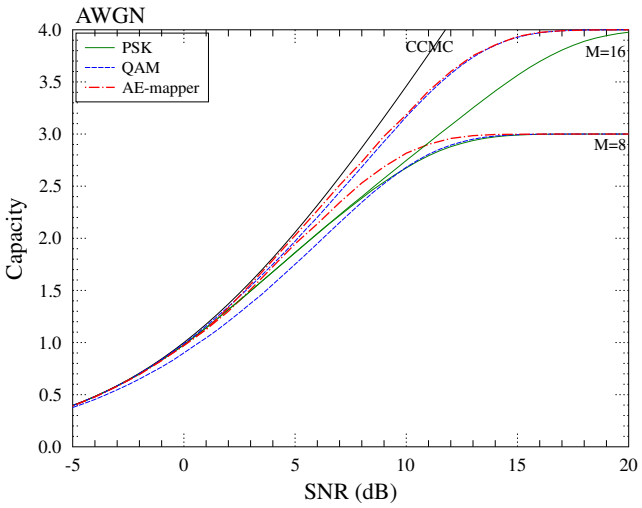

Fig. 4: DCMC and CCMC capacity of the conventional PSK, QAM, and AE-modulated communication system when communicating over AWGN channels, where the modulation order is $M = 8$ or 16.

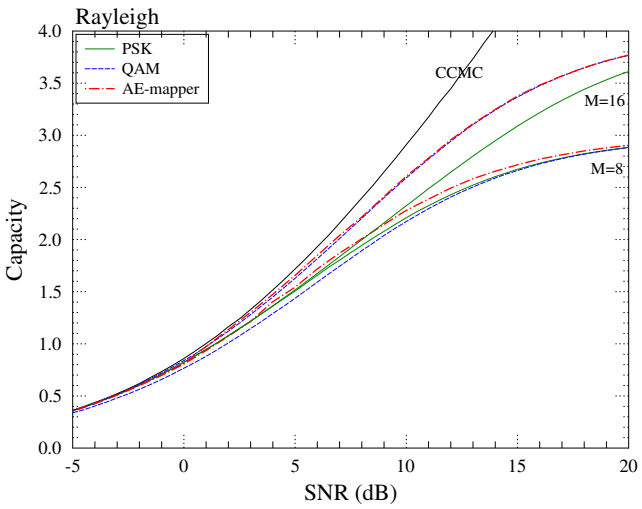

Fig. 5: DCMC and CCMC capacity of the conventional PSK, QAM, and AE-modulated communication system when communicating over uncorrelated Rayleigh fading channels, where the modulation order is $M = 8$ or 16.

expressed as

$$I_{a,3}(\boldsymbol{u}_3) = 1 - \frac{1}{\sqrt{2\pi}\sigma_A} \int_{-\infty}^{+\infty} \exp\left(-\frac{(\zeta - \frac{\sigma_A^2}{2})^2}{2\sigma_A^2}\right) \times \log_2\left[1 + e^{-\zeta}\right] \mathrm{d}\zeta. \quad (18)$$

By contrast, the MI between the *extrinsic* LLRs of the demapper $L^{(e)}(\boldsymbol{u}_3)$ and the transmitted signal sequence $\boldsymbol{x}$ can be expressed as

$$I_{e,3}(\boldsymbol{u}_3) = \frac{1}{2} \sum_{x=-1,+1} \int_{-\infty}^{+\infty} p_e\left(\zeta|Z=z\right) \times \log_2 \frac{2p_e\left(\zeta|Z=z\right)}{p_e\left(\zeta|Z=0\right) + p_e\left(\zeta|Z=1\right)} \mathrm{d}\zeta, \quad (19)$$

where $p_e\left(\zeta|Z=z\right)$ is obtained by the Monte-Carlo simulations. From (19), we can express $I_e(u)$ as a function of $I_{e,1}$ and the SNR γ. Hence the extrinsic information transfer characteristic $I_{e,3}(\boldsymbol{u}_3)$ can be expressed by an EXIT function $\mathcal{T}_3(\cdot)$ as

$$I_{e,3}(\boldsymbol{u}_3) = \mathcal{T}_3\left[I_{a,3}(\boldsymbol{u}_3), \gamma\right]. \quad (20)$$

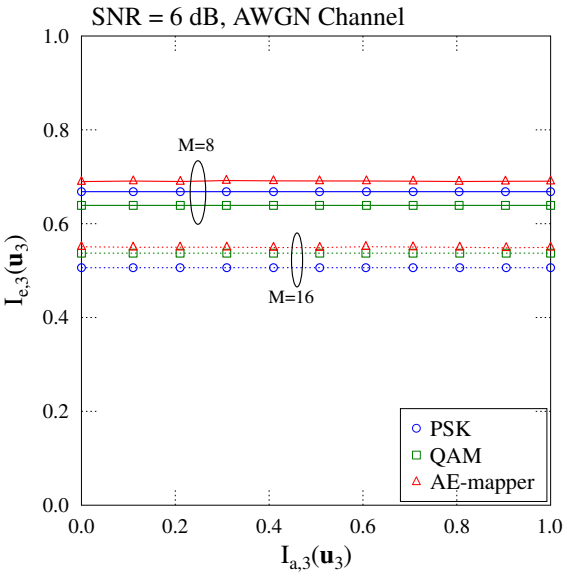

Fig. 6: The extrinsic information transfer characteristic of the QAM/PSK and AE-mapper having a modulation order of $M = 8$, 16 and 32 when communicating over AWGN channel.

The extrinsic information transfer characteristic of Gray-mapping based QAM/PSK and of the AE-mapper using a modulation order of $M = 8, 16$ and 32 is shown in Fig. 6, where 8PSK is employed for $M = 8$ for the conventional modulation schemes, while 16QAM and 32QAM are employed for $M = 16$ and 32, respectively. Observe from Fig. 6 that the AE-mapper has the advantage of faster convergence to $I_{e,3}(\boldsymbol{u}_3) = 1$ than QAM/PSK.

The relationship of the MI between the *a priori* LLRs $A(\hat{\boldsymbol{u}}_2)$ of the URC decoder and the URC encoded bits $\boldsymbol{u}_2$ as well as the MI between the *extrinsic* LLRs $E(\hat{\boldsymbol{u}}_2)$ of the URC decoder

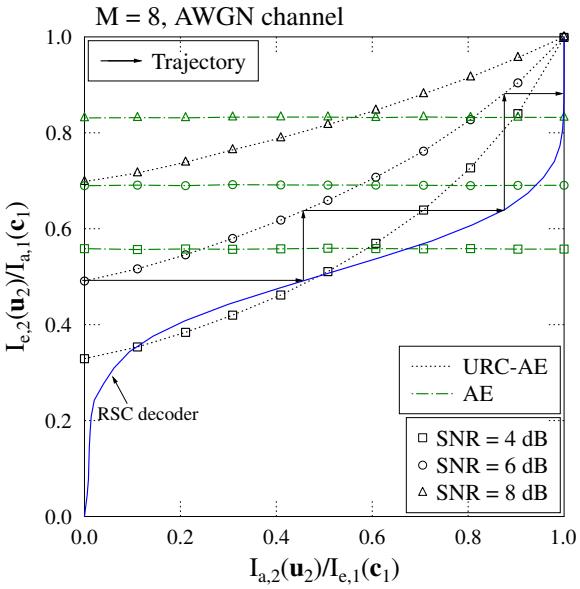

Fig. 7: The EXIT chart of the FEC-AE and of the proposed FEC-URC-AE system having a modulation order of $M = 8$ and a coded block length of $N = 960$ when communicating over AWGN channels at different SNRs, $\gamma = 4, 6$ and 8 dB.

and the URC encoded bits $\boldsymbol{u}_2$ can be also quantified by an EXIT function $\mathcal{T}_2(\cdot)$, expressed as

$$I_{e,2}(\boldsymbol{u}_2) = \mathcal{T}_2\left[I_{a,2}(\boldsymbol{u}_2), \gamma\right]. \quad (21)$$

By contrast, since the only input to the FEC decoder is the soft-interleaved *extrinsic* LLRs of the URC decoder, as shown in Figure 1, the EXIT function $\mathcal{T}_1(\cdot)$ of the FEC decoder is independent of γ. Therefore, the EXIT characteristic is defined as

$$I_{e,1}(\boldsymbol{c}_1) = \mathcal{T}_1\left[I_{a,1}(\boldsymbol{c}_1)\right], \quad (22)$$

where $I_{e,1}(\boldsymbol{c}_1)$ is the MI between the *extrinsic* LLRs $E(\hat{\boldsymbol{c}}_1)$ of the FEC decoder and the FEC encoded bits $\boldsymbol{c}_1$, while $I_{a,1}(\boldsymbol{c}_1)$ is the MI between the *a priori* LLRs $A(\hat{\boldsymbol{c}}_1)$ of the URC decoder and the FEC-encoded bits $\boldsymbol{c}_1$.

Figs. 7 to 9 show the EXIT chart of the FEC-AE and the proposed FEC-URC-AE system having a coded block length of $N = 960$, where $M = 8, 16$ or 32, when communicating over AWGN channels. The FEC employed in this paper is a RSC code having a code rate of $R = 1/2$ and memory of 3, as well as the generator polynomial of $[1011, 1101]$[1] and 8 trellis states. Indeed, this is the well-known the basic component of turbo codes. Note that turbo codes having two identical RSC encoders and an interleaver will always attain superior error-correction performance over an identical-complexity RSC decoder itself. However, when a URC is serially concatenated to the RSC, the convergence rate of its EXIT curve may be adjusted by assigning different number

[1]The first polynomial indicates the feedback parameter, while the second represents the feed-forward parameters. The coding rated could be adjusted by variable-rate puncturer.

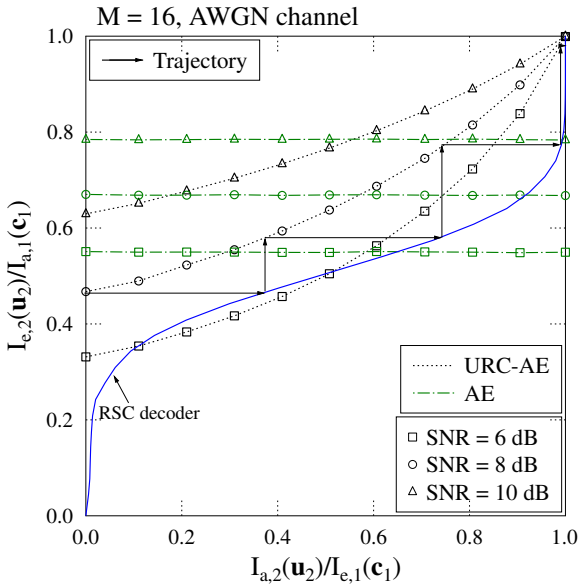

Fig. 8: The EXIT chart of the FEC-AE and of the proposed FEC-URC-AE system having a modulation order of $M = 16$ and a coded block length of $N = 960$ when communicating over AWGN channels at different SNRs, $\gamma = 6, 8$ and 10 dB.

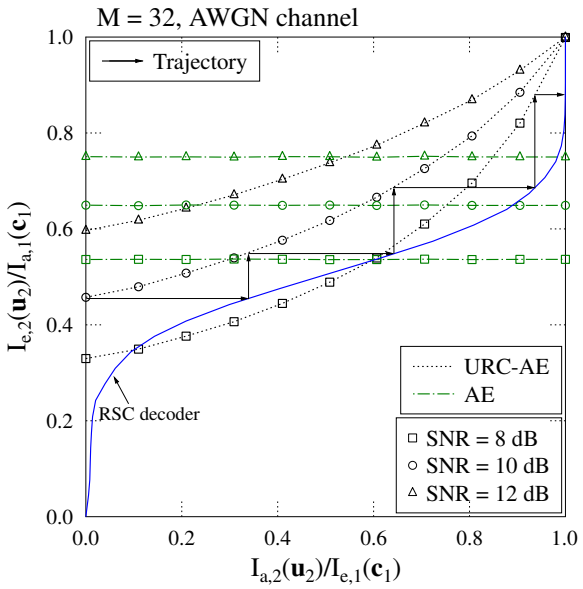

Fig. 9: The EXIT chart of the FEC-AE and of the proposed FEC-URC-AE system having a modulation order of $M = 32$ and a coded block length of $N = 960$ over AWGN channels at different SNRs, $\gamma = 8, 10$ and 12 dB.

of iterations between RSC and URC. In this way, there would always be an open tunnel between the EXIT curves of the URC-RSC decoder and of the QAM detector in the EXIT chart. By contrast, the turbo-coded system has a higher fixed convergence rate than that of the RSC-URC coded system [42, 43], indicating that the EXIT curve of turbo codes may cross that of the QAM detector before reaching the point (1,1). Hence, employing the turbo code may not guarantee better error-correction performance.

We can readily see from all the three EXIT charts that similar to the Gray mapping, the AE also generates a horizontal EXIT curve, indicating that no extra gains can be attained by employing iterative decoding for SISO FEC-AE systems. Hence, we employ a serially concatenated URC in the two-stage system, which is capable of scrambling the bit sequence. Observe from Figs. 7 to 9 that our proposed FEC-URC-AE system is capable of attaining a substantial iterative gain by exploiting the benefit of our serially concatenated URC, while retaining the model-based structure of the communication system, where the conventional modulator and demodulator are replaced by a single AE. In this way, the training only has to be performed once for a given modulation order, regardless of the block length, hence increasing the flexibility of the AE. This is a substantial benefit compared to the solution proposed by Cammerer *et al.* [26]. To elaborate a little further, their model-free structure also achieves iterative gains by integrating the FEC and mapping process, which requires independent training for each and every legitimate block length and for each and every FEC scheme. Additionally, we can see from the trajectory of Figs. 7 to 9 that the EXIT curve of the URC decoder always reaches the $I_{e,2}(\boldsymbol{u}_2) = I_{a,2}(\boldsymbol{u}_2) = 1$, regardless of the SNR. Hence, after a sufficiently high number of iterations between the URC decoder and RSC decoder, we can always reach the $I_{e,2}(\boldsymbol{u}_2) = I_{a,2}(\boldsymbol{u}_2) = 1$ point, as long as an open tunnel exists between URC decoder and the RSC decoder curves.

VI. Error-correction performance

In this section, the error-correction capability of the proposed FEC-URC-AE system is quantified in terms of both its BER and BLER for different system parameters, when communicating over AWGN or Rayleigh fading channels. The simulation parameters are summarized in Table II. Note that for consistence with the EXIT chart analysis of Section V, the same half-rate RSC is adopted and logarithmic Bahl-Cocke-Jelinek-Raviv (Log-BCJR) decoding is applied. Furthermore, each of the $L-1 = 2$ hidden layers is comprised of $Q = 4M$ neurons. The AE-mapper is implemented by employing the standard DL libraries TensorFlow and Keras, and we adopt the Adam optimizer.

Firstly, we consider 8QAM as an example to demonstrate the advantage of the AE-mapper. Specifically, we compare the constellation set of the AE-learned 8QAM to that of the conventional square-8QAM at an SNR of 9 dB for transmission over Rayleigh fading channels, as shown in Fig. 10. Our "learned" 8QAM constellation obtained by the AE-mapper has a peak-to-average power ratio (PAPR) of 1.79, whereas

TABLE II: SIMULATION PARAMETERS

Parameters	Values
Frame length (N)	$240, 480, 720, 960$
Modulation order (M)	$8, 16, 32$
Coding rate (R)	1/2
Number of iterations (T)	$16, 32, 64$
RSC polynomial	$[1011, 1101]$
URC polynomial	[11,10]
Number of hidden layers ($L-1$)	2
Learning rate	10^{-3} to 10^{-5}
Mini-batch size	1024

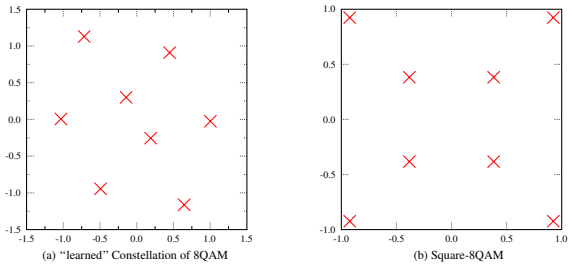

Fig. 10: (a) The constellation of 8QAM "learned" by the AE-mapper; and (b) the conventional square-8QAM constellation.

the conventional square-8QAM has a PAPR of 1.71, so the latter is marginally better. However, the PAPR of the "learned" constellation can be reduced if we include the PAPR into the training loss function and perform joint multiple-objective optimization. Furthermore, it can be observed from Fig. 10 that the minimum distance between the neighbouring constellation points of the AE-learned constellation is higher than that of the conventional square-8QAM. It is also notable that the position of the constellation points obtained by the AE-mapper may vary as a function of the SNR, since the SNR value is also an input of the training network, hence influencing the geometric distribution of the constellation points.

We now characterize the iterative gains in Figs. 11 and 12, which are in line with our EXIT chart analysis of Section V. More specifically, Figs. 11 and 12 show the BER performance of the proposed FEC-URC-AE system, which is compared to the two-stage FEC-AE system proposed in [26], using no URC scheme, where $M = 8, 16$ or 32. The number of iterations between the FEC decoder and URC decoder is $T = 16, 32$ or 64 and the half-rate RSC has $N = 960$ encoded bits. Note that since the AE-demapper does not have a priori feedback input, we assume no iterations between the decoder and AE-demapper. We can see that a steep 'turbo-cliff' occurs when using FEC-URC iterative decoding. This is consistent with our observations from Figs. 7 to 9, where the trajectory gradually increases to the [1,1] point. When $T = 64$ iterations are employed between the RSC decoder and URC decoder, over 2.4 dB gain can be observed from Fig. 11, when employing the proposed FEC-URC-AE system over the non-iterative FEC-AE system at a BER of 10^{-5} for transmission over AWGN channel. In all cases $M = 8, 16$ and 32 are used. Similar observations may be obtained in Fig. 12, where the same parameters are employed for communication over uncorrelated Rayleigh fading channels. Still, significant BER vs. SNR gains

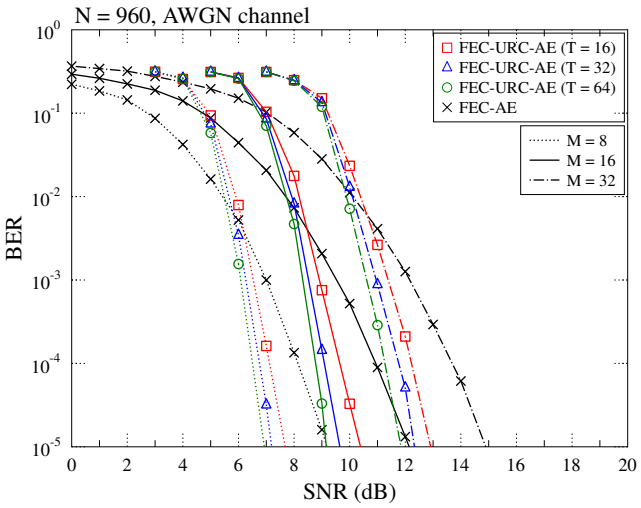

Fig. 11: BER performance of the proposed FEC-URC-AE and the FEC-AE system having a coded block length of $N = 960$ when communicating over AWGN channel, where $M = 8, 16$ and 32, $T = 16, 32$ and 64 are used.

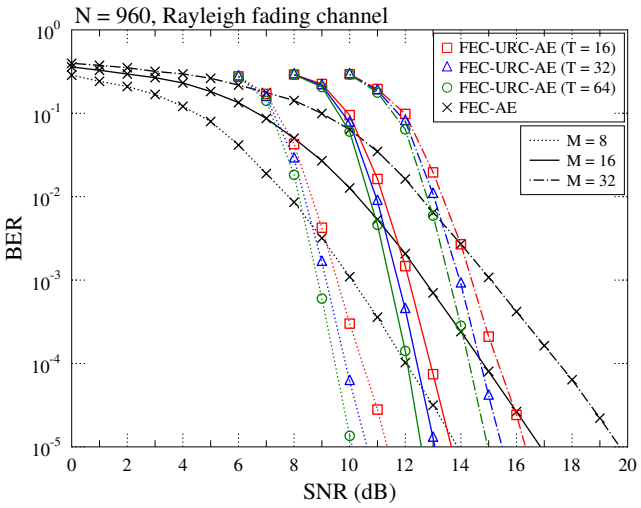

Fig. 12: BER performance of the proposed FEC-URC-AE and the FEC-AE system having a coded block length of $N = 960$ when communicating over uncorrelated Rayleigh fading channels, where $M = 8, 16$ and 32, $T = 16, 32$ or 64 are used.

are achieved with the aid of the proposed iterative decoding scheme. It can be observed from Fig. 11 that the results of FEC-AE exhibit a steeper slope than those of FEC-URC-AE, since the concatenation of the URC further improves the exploitation of the extrinsic information without increasing the overall delay in the iterative soft-information exchange between the URC decoder and the FEC decoder, hence also improving the BER performance. This may also be interpreted as increasing the time-diversity attained by our sophisticated coding scheme.

Furthermore, the BLER of the proposed FEC-URC-AE

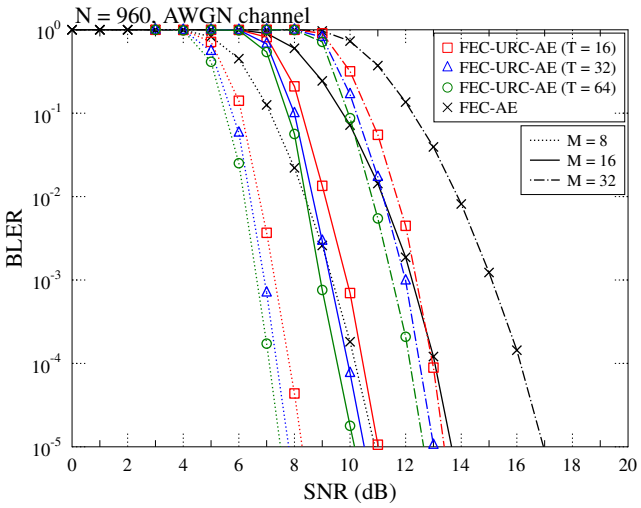

Fig. 13: BLER performance of the proposed FEC-URC-AE and the FEC-AE system having a coded block length of $N = 960$ when communicating over AWGN channel, where $M = 8, 16$ and 32, $T = 16, 32$ and 64 iterations are employed.

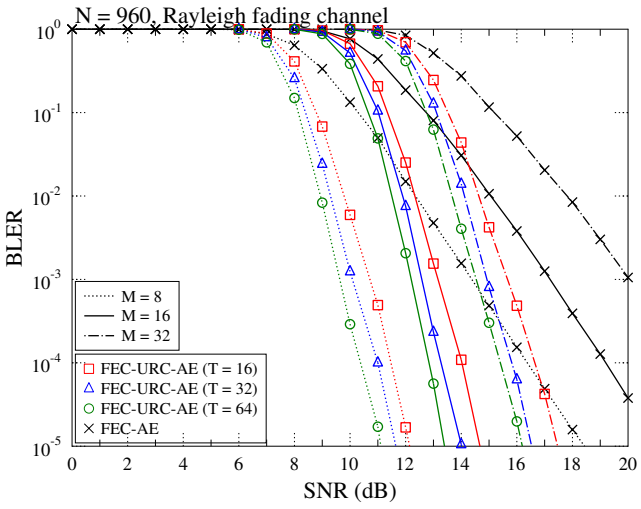

Fig. 14: BLER performance of the proposed FEC-URC-AE and the FEC-AE system having a coded block length of $N = 960$ when communicating over uncorrelated Rayleigh fading channels, where $M = 8, 16$ and 32, $T = 64$ iterations are employed.

systems is quantified in Figs. 13 and 14, using the same parameters as in Figs. 11 and 12, when communicating over AWGN and Rayleigh fading channels, respectively. When the BLER is considered, there are no intersections between the proposed FEC-URC-AE and the FEC-AE systems. Specifically, a 4.5 dB SNR gain is observed at a BER of 10^{-5} from Fig. 13 when employing a URC, in the case of $M = 32$ and $T = 64$, compared to that of the FEC-AE system of [26]. Furthermore, the results show that only as few as $T = 16$ iterations are required between the URC and FEC decoder for achieving significant SNR gains, compared to the FEC-AE decoder, as

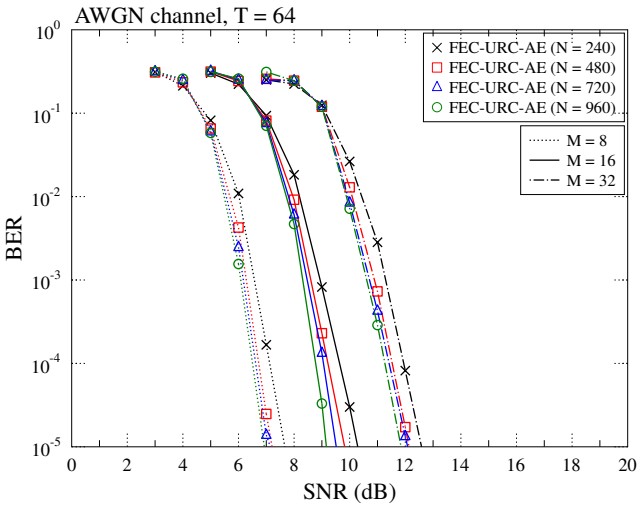

Fig. 15: BER performance of the proposed FEC-URC-AE and the FEC-URC-QAM/PSK system having different coded block lengths of $N = 240, 480, 720$ and 960 over AWGN channel, where $M = 8, 16$ and 32, and $T = 64$ iterations are employed.

shown in Fig. 14.

Additionally, the influence of the block lengths is characterized in Fig. 15, where the proposed FEC-URC-coded AE-modulated system has a code block length of $N = 240, 480, 720$ and 960 for communication over AWGN channels. It can be observed from Fig. 15 that a higher block length N always leads to an improved BER. Note that since we consider the AE-mapper as a concatenated part, simulations of all the block lengths are carried out for the same AE-mapper and the training is only performed once for each modulation order, regardless of the code block lengths N and FEC schemes used.

Finally, the BER comparison of the half-rate FEC-URC-coded QAM/PSK system and the proposed FEC-URC-coded AE-modulated system using the same number of $T = 64$ iterations for communicating over both AWGN and Rayleigh fading channels are shown in Figs. 16, where we have $N = 960$, $M = 8$ and 32. To avoid the overlap of the BER curves, we omit the case of $M = 16$, which shows similar trends to those of $M = 8$ or 32. Note furthermore that when $M = 8$, 8PSK, rather than 8QAM is employed, since 8PSK attains superior BER performance over 8QAM, as detailed in [44] and shows a higher DCMC capacity in the low SNR regions, as demonstrated in Figs. 4 and 5. However, as shown in Fig. 12, when $M = 8$ or 32 is employed, about 0.7 dB BER vs. SNR gain can be attained at a BER of 10^{-5}, which demonstrates the merits of our FEC-URC-coded AE-modulated system. Therefore, by exploiting the beneficial properties of URC, iterative decoding is conceived for our FEC-URC-AE system, while retaining the model-based processing structure and attaining superior BER performance over the conventional systems employing QAM/QPSK.

To summarize, for explicitly comparing FEC-URC-AE, FEC-URC-QAM/PSK and FEC-AE that we have considered in this paper, Table III quantifies the SNR required for different

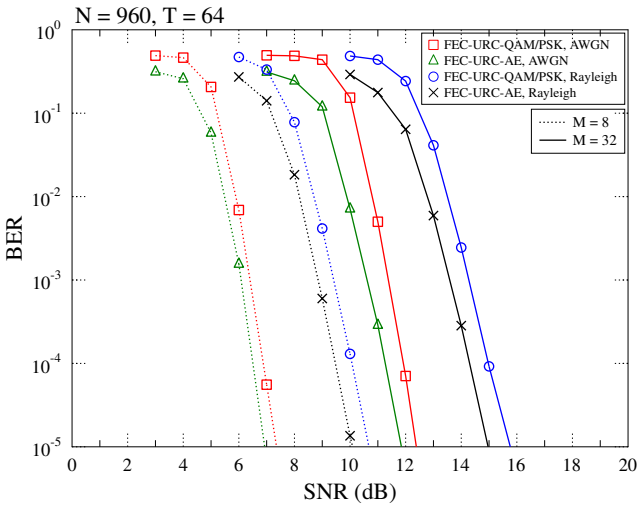

Fig. 16: BER performance of the proposed FEC-URC-AE and the conventional FEC-URC-QAM/PSK system having a coded block length of $N = 960$ over uncorrelated Rayleigh fading channels, where $M = 8$ and 32, and $T = 64$ iterations are employed.

TABLE III: SNR (dB) REQUIRED FOR FEC-URC-AE, FEC-URC-QAM/PSK AND FEC-AE TO ATTAIN A BER OF 10^{-5}, WHERE A CODED BLOCK LENGTH OF $N = 960$ IS EMPLOYED.

	AWGN		Rayleigh	
	M=8	M=32	M=8	M=32
FEC-URC-AE	6.94	11.86	10.11	14.96
FEC-URC-QAM/PSK	7.37	12.40	10.69	15.77
FEC-AE	9.19	14.90	13.89	19.73

schemes and parameters at a BER of 10^{-5}, where $N = 960$ is employed. We can see that our proposed FEC-URC-AE achieves the highest SNR gain among the three schemes for all modulation orders and channels, demonstrating the benefits of our scheme.

VII. Conclusions

A FEC-URC-AE system was proposed for soft iterative decoding, while retaining the model-based structure of communication systems. Additionally, the AE-mapper was specifically designed for directly producing *extrinsic* LLRs, which can be directly entered into the URC decoder for soft iterative decoding. This avoids the degradation due to the conversion of symbol probabilities to bit LLRs. Furthermore, the corresponding DCMC and CCMC analysis was carried out, explicitly demonstrating the advantages of the AE-aided constellation design over the conventional BICM, for the first time. The EXIT chart analysis of the proposed scheme also demonstrated the benefits of incorporating a URC scheme, showing that the MI of 1 was reached, regardless of the SNR, indicating the explicit benefits of iterative decoding. Our BER and BLER simulations demonstrated that the proposed FEC-URC-AE communication system is capable of achieving iterative gains for transmission over both AWGN and Rayleigh fading channels and that it outperforms its FEC-AE systems and its conventional FEC-URC-coded QAM/PSK counterparts. Our future work will focus on the implementation of our FEC-URC-AE system in diverse communication environments, including high-Doppler aeronautical scenarios.

References

[1] M.-S. Yee, B. L. Yeap, and L. Hanzo, "Radial basis function-assisted turbo equalization," *IEEE Transactions on Communications*, vol. 51, no. 4, pp. 664–675, 2003.

[2] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, 2019.

[3] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 214–222, 2019.

[4] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep learning for wireless communications: An emerging interdisciplinary paradigm," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 133–139, 2020.

[5] 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 Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1472–1514, 2020.

[6] H. Kim, S. Oh, and P. Viswanath, "Physical layer communication via deep learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 5–18, 2020.

[7] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.

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

[9] J. Xia, K. He, W. Xu, S. Zhang, L. Fan, and G. K. Karagiannidis, "A MIMO detector with deep learning in the presence of correlated interference," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4492–4497, 2020.

[10] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[11] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

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

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

[14] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Viterbinet: A deep learning based viterbi algorithm for symbol detection," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3319–3331, 2020.

[15] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Data-driven factor graphs for deep symbol detection," *arXiv preprint arXiv:2002.00758*, 2020.

[16] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.

[17] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[18] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2017.

[19] E. Balevi and J. G. Andrews, "One-bit OFDM receivers via deep learning," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4326–4336, 2019.

[20] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[21] W. Zhang, S. Li, and Y. Cui, "Jointly sparse support recovery via deep auto-encoder with applications in MIMO-based grant-free random access for mMTC," *arXiv preprint arXiv:2005.02389*, 2020.

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

[23] J. Tao, J. Chen, J. Xing, S. Fu, and J. Xie, "Autoencoder neural network based intelligent hybrid beamforming design for mmwave massive mimo systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 1019–1030, 2020.

[24] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016.

[25] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2503–2516, 2019.

[26] S. Cammerer, F. A. Aoudia, S. Dörner, M. Stark, J. Hoydis, and S. Ten Brink, "Trainable communication systems: Concepts and prototype," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5489–5503, 2020.

[27] 3GPP TS 38.212 V15.1.1, "NR Multiplexing and channel coding," *3rd Generation Partnership Project Std. 3GPP*, 2018.

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

[29] T. Richardson and S. Kudekar, "Design of low-density parity check codes for 5G new radio," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 28–34, 2018.

[30] Z. B. K. Egilmez, L. Xiang, R. G. Maunder, and L. Hanzo, "The development, operation and performance of the 5G polar codes," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 96–122, 2019.

[31] D. Divsalar, S. Dolinar, and F. Pollara, "Serial concatenated trellis coded modulation with rate-1 inner code," in *Globecom'00-IEEE. Global Telecommunications Conference*, vol. 2, pp. 777–782, IEEE, 2000.

[32] W. Zhang, M. F. Brejza, T. Wang, R. G. Maunder, and L. Hanzo, "Irregular trellis for the near-capacity unary error correction coding of symbol values from an infinite set," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 5073–5088, 2015.

[33] Z. Babar, H. V. Nguyen, P. Botsinis, D. Alanis, D. Chandra, S. X. Ng, and L. Hanzo, "Serially concatenated unity-rate codes improve quantum codes without coding-rate reduction," *IEEE Communications Letters*, vol. 20, no. 10, pp. 1916–1919, 2016.

[34] Z. Babar, H. V. Nguyen, P. Botsinis, D. Alanis, D. Chandra, S. X. Ng, and L. Hanzo, "Unity-rate codes maximize the normalized throughput of on–off keying visible light communication," *IEEE Photonics Technology Letters*, vol. 29, no. 3, pp. 291–294, 2016.

[35] Z. Babar, M. A. M. Izhar, H. V. Nguyen, P. Botsinis, D. Alanis, D. Chandra, S. X. Ng, R. G. Maunder, and L. Hanzo, "Unary-coded dimming control improves on-off keying visible light communication," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 255–264, 2017.

[36] J. Hu, M. Li, K. Yang, S. X. Ng, and K.-K. Wong, "Unary coding controlled simultaneous wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 637–649, 2019.

[37] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, pp. 177–186, Springer, 2010.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[39] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multidimensional signal sets," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 528–536, 2006.

[40] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 127–153, 2013.

[41] 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.

[42] W. T. Webb and L. Hanzo, *Modern Quadrature Amplitude Modulation: Principles and applications for fixed and wireless channels: one*. IEEE Press-John Wiley, 1994.

[43] H. Chen, R. G. Maunder, and L. Hanzo, "A survey and tutorial on low-complexity turbo coding techniques and a holistic hybrid arq design example," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1546–1566, 2013.

[44] Y. Liu, L.-L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1474–1487, 2017.