Data-Rate Driven Transmission Strategies for Deep Learning Based Communication Systems

SCHOLARONE™
Manuscripts

1

Data-Rate Driven Transmission Strategies for Deep Learning Based Communication Systems

Xiao Chen, Julian Cheng, *Senior Member, IEEE*, Zaichen Zhang, *Senior Member, IEEE*, Liang Wu, *Member, IEEE*, Jian Dang, *Member, IEEE*, Jiangzhou Wang, *Fellow, IEEE*

Abstract

Deep learning (DL) based autoencoder is a promising architecture to implement end-to-end communication systems. One fundamental problem of such systems is how to increase the transmission rate. Two new schemes are proposed to address the limited data rate issue: adaptive transmission scheme and generalized data representation (GDR) scheme. In the first scheme, an adaptive transmission is designed to select the transmission vectors for maximizing the data rate under different channel conditions. The block error rate (BLER) of the first scheme is 80% lower than that of the conventional one-hot vector scheme. This implies that higher data rate can be achieved by the adaptive transmission scheme. In the second scheme, the GDR replaces the conventional one-hot representation. The GDR scheme can achieve higher data rate than the conventional one-hot vector scheme with comparable BLER performance. For example, when the vector size is eight, the proposed GDR scheme can double the date rate of the one-hot vector scheme. Besides, the joint scheme of the two proposed schemes can create further benefits. The effect of signal-to-noise ratio (SNR) is analyzed for these DL-based communication systems. Numerical

X. Chen, Z. Zhang, L. Wu and J. Dang are with National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096, China (email: {chen_xiao, zczhang, wuliang, dangjian}@seu.edu.cn). Zaichen Zhang is the corresponding author.

J. Cheng is with School of Engineering, The University of British Columbia, Kelowna, V1V 1V7, BC, Canada (email: julian.cheng@ubc.ca).

J. Wang is with School of Engineering and Digital Arts, University of Kent, Canterbury, CT2 7NT, United Kingdom (email: J.Z.Wang@kent.ac.uk).

2

results show that training the autoencoder using data set with various SNR values can attain robust BLER performance under different channel conditions.

Index Terms

Autoencoder, communication systems, data rate, deep learning, transmission strategy.

I. INTRODUCTION

To satisfy growing demand for various communication applications and services, the next-generation network must deliver enhanced mobile broadband, ultra-reliable and low-latency communications, and massive Internet of Things (IoT) ecosystems [1]–[4]. One primary concern is to accommodate the exponential rise in the number of user equipments and the traffic capacity in future communication systems. Hence, several promising technologies have been proposed, and they include massive multi-input and multi-output (MIMO) transmissions, millimeter wave communications, ultra-dense networks, and non-orthogonal multiple access [5]–[10]. For these conventional communication systems, there exist a number of limitations, such as unavailable channel state information in complex transmission scenario, high complexity to process big data, and sub-optimal performance caused by conventional block structure. For these reasons, with the significant development of deep learning (DL) [11]–[13], researchers have applied the machine learning (ML), especially DL technologies, to design communication systems for benefits that cannot be obtained using the conventional approaches [14]–[18].

As a promising technique, deep learning implements communication systems using deep neural networks (NNs). Different from the conventional communication system that consists of multiple independent blocks (e.g., source/channel coding, modulation, channel estimation, equalization), the DL-based communication system can jointly optimize transmitter and receiver for end-to-end performance without a block structure [19], [20]. DL-based system design is promising for the following reasons: (i) A DL-based communication system can be optimized for end-to-end performance by using deep NNs, which is fundamentally different from the block-structure in conventional communication systems; (ii) A DL-based communication system can be optimized for a practical system over any type of channel without requiring a tractable mathematical model, and this includes the channel models that take into account of different transmission scenarios and non-linearities; (iii) DL algorithms can provide faster processing speed than conventional

3

communication algorithms, since the execution of NNs can be highly parallel on concurrent architectures and can be implemented using low-precision data types [21].

Attracted by these advantages, there have been a number of studies on DL-based communications and signal processing using state-of-the-art tools and hardware [19], [20], [22]–[37]. The DL method is used to deal with certain challenges in existing communication systems. For example, the DL-based belief propagation algorithm was originally used to improve the performances of channel decoding, where low-complexity and near optimal decoder performance were obtained [22]–[24]. Around the same time, autoencoder was developed to address the problem of learning an efficient physical layer [25]. In DL theory, an autoencoder describes a deep NN to find a low-dimensional representation of its input at certain intermediate layer that allows reconstruction at the output with minimal error [38, Ch. 14]. The DL-based communication system can be represented and implemented by an autoencoder that is trained using the dataset offline. Then, the trained autoencoder can be directly applied to practical systems online. A DL-based communication system, interpreted as an autoencoder, performs an end-to-end reconstruction task that jointly optimizes transmitter and receiver as well as learns signal encoding [19], [25], [26], [34]. To address the challenges of frame synchronization, an autoencoder was proposed to represented a complete communication system [20], [28], and comparable performance can be achieved even without extensive hyperparameter tuning. More recently, a DL-based algorithm has been used to solve the channel state information feedback and channel estimation problems in massive MIMO systems, and it outperforms the state-of-the-art compressive sensing based algorithms [29]–[31].

For future communication systems, there is a huge demand for data rate due to an increasing number of communication devices and equipment types, and improved quality of services (QoS). Consequently, high data-rate schemes should be developed in DL-based communication systems for future wireless networks. However, one-hot vector [39], being the most commonly used data representation in existing studies [16], [19], [20], [24], [26]–[28], [34], has a low data rate in DL-based communication systems. The reason is that an $M \times 1$ one-hot vector consists of 0s in all entries with the exception of a single 1, e.g., $[0, \ldots, 0, 1, 0, \ldots, 0]^T$, and there are only M possible transmitted messages, which leads to limited data rate. This becomes a barrier for developing future DL-based communication systems. Besides, the autoencoder with one-hot vector is typically trained using a fixed vector size M, which becomes a constraint when

designing communication systems having different data rate requirements. Also, the conventional autoencoder is trained under a fixed signal-to-noise ratio (SNR) value with an unrealistic expectation to operate well for a wide range of SNR values in practical transmission scenarios [2]. It was reported that training the autoencoder at different SNR values will affect autoencoder performances [19], but there is no detailed study on the effect on such a system. Therefore, our objective is to design a new transmission scheme and replace the conventional one-hot vector scheme to achieve higher data rate. As well, we will investigate the effect of training SNR on the performance of DL-based communication systems. Here, training SNR denotes the fixed SNR used for training the autoencoder offline, and it can be different from the practical SNR of a communication system when it is operating online.

In this paper, an adaptive transmission scheme is first designed for different communication scenarios to maximize the data rate in DL-based communication systems having a QoS constraint. Then, we propose a generalized data representation (GDR) scheme to improve the data rate of DL-based communication systems. Finally, we analyze the effect of SNR and mean squared error (MSE) performance in DL-based communication systems. Comparable block error rate (BLER) performance can be achieved by the proposed transmission schemes which has lower complexity and higher data rate than the conventional DL-based communication system[1].

The major contributions of this paper are summarized as follows:

1) In DL-based communication systems, we pointout the limited data rate problem of the conventional one-hot vector scheme. To address this issue, we design an adaptive transmission scheme having a QoS constraint for different channel conditions. In the proposed scheme, the optimal transmission vectors are adaptively selected for different SNR values, where the goal is to maximize the data rate with a constraint on MSE performance. It is shown that, when both two schemes have the same data rate, the proposed adaptive transmission scheme can reduce BLER of the conventional one-hot vector scheme by 80%.

2) Furthermore, we propose a generalized data representation scheme to improve the data rate in DL-based communication systems. The proposed scheme represents the message by using a probability vector having multiple non-zero elements, instead of the conventional

---

[1]Notably, throughout this paper, the conventional DL-based communication system refers to an autoencoder based communication system that adopts the one-hot vector data representation.

5

one-hot vector having only one non-zero element. As expected, higher data rate is obtained by the proposed GDR scheme with comparable BLER performance and low complexity. When the vector size is eight, as an example, the proposed GDR scheme can double the data rate of the conventional one-hot vector scheme. To the best of the authors' knowledge, this is the first time that the GDR scheme is proposed and its effectiveness is verified.

3) We investigate the effect of SNR on the system performances in DL-based communication systems. Simulation results show that the high training SNR can improve the convergence performance in training, but it can also degrade the BLER performance in practical transmission. As a tradeoff, we introduce a training SNR set strategy, which shows trade-off between convergence and BLER performance. Furthermore, it is shown that training the autoencoder at low SNR can achieve BLER and MSE performance gains when the trained autoencoder is applied to high SNR scenario. These results provide a reliable design guidance to select the suitable training SNR and achieve optimal system performance.

For potential applications, the DL-based autoencoder-represented communication system can be applied to complex channel conditions without a mathematically tractable model in, for examples, massive IoT ecosystems and high-speed Internet of Vehicles systems.

The remainder of this paper is organized as follows. In Section II, we describe the system model of a DL-based communication system. Section III presents an adaptive transmission scheme. Section IV proposes the generalized data representation scheme for DL-based communication systems. Section V investigates the effect of SNR and analyzes the MSE performance of the autoencoder. In Section VI, we present the numerical results of the proposed schemes and system performances. Section VII concludes this paper.

II. DEEP LEARNING BASED COMMUNICATION SYSTEMS

In this section, we describe the DL-based autoencoder for an end-to-end communication system, and then provide the research motivations of this paper.

A. Autoencoder for End-to-End Communication Systems

We consider a DL-based communication system represented as an autoencoder consisting of transmitter, channel, and receiver as shown in Fig. 1, where the corresponding NN structure is shown below. The autoencoder describes a deep NN that applies unsupervised learning in order to
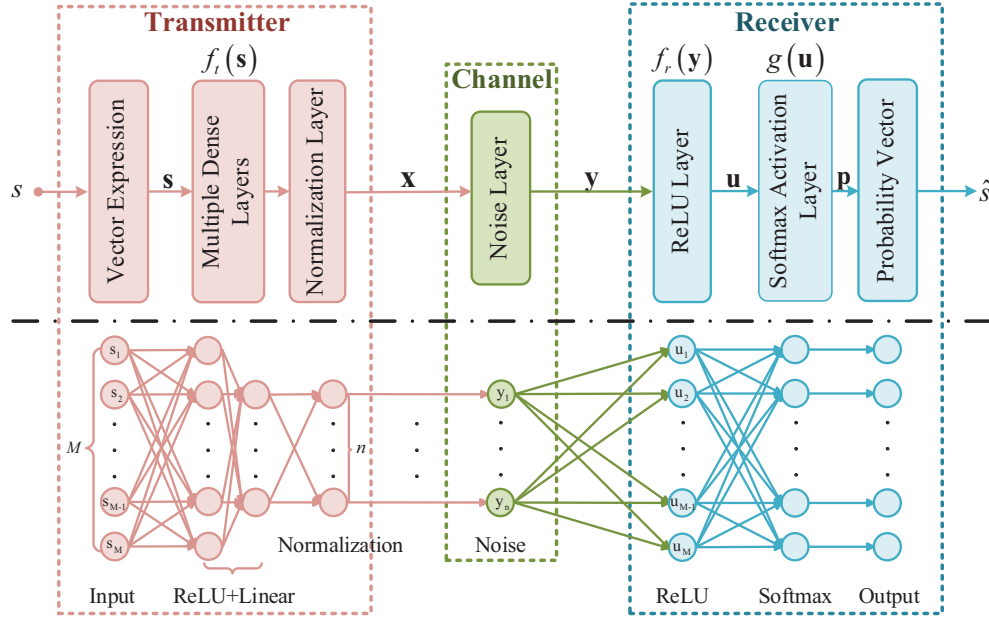
Fig. 1.   A DL-based communication system represented as an autoencoder with its NN structure [19].

TABLE I

ACTIVATION FUNCTIONS AND LOSS FUNCTIONS

Activation functions	Linear	s_i
	ReLU	$\max\{s_i, 0\}$
	Softmax	$\frac{e^{u_i}}{\sum_{j=1}^{M} e^{u_j}}$
	Sigmoid	$\frac{1}{1+e^{-u_i}}$
	tanh	$\tanh(u_i)$
Loss functions	MSE	$\|\mathbf{s} - \mathbf{p}\|_2^2$
	Categorical cross-entropy	$-\sum_{i=1}^{M} s_i \log(p_i)$

reconstruct the input at the output [38, Ch. 14]. At the transmitter, a message $s \in \{1, 2, \ldots, M\}$ is first transformed to a vector $\mathbf{s} \in \mathbb{R}^M$ after the vector expression processing, where, say, $M \in \{4, 8, 16, 32, 64\}$. For example, if the message $s = 2$ is transmitted, the corresponding vector expression is a one-hot vector $\mathbf{s} = [0, 1, 0, \ldots, 0]^T$ in a conventional DL-based communication system. Then, the multiple dense layers, including a rectified linear unit (ReLU) layer and a linear layer, apply the transformation $f_t : \mathbb{R}^M \mapsto \mathbb{R}^n$ to produce the transmitted signal for n discrete

channel uses [20]. The commonly used activation functions are shown in TABLE I. Finally, the normalization layer ensures the power constraint of the transmitted signal $\mathbf{x} = [x_1, \ldots, x_n]^T$ as $\mathbb{E}\{x_j^2\} \leq 1$ $(j = 1, \ldots, n)$, where $\mathbb{E}\{\cdot\}$ denotes expectation.

The transmit channel is implemented by a noise layer with its output being the received signal $\mathbf{y}$ given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ denotes zero-mean additive white Gaussian noise (AWGN) vector where each element has variance $\sigma^2 = (2RE_b/N_0)^{-1}$, and where R is the data rate, E_b is the energy per bit, and N_0 denotes the noise power spectral density. Notably, there is no complex operation in the existing NN architectures, and the complex number is represented by two real numbers [19]. Consequently, we assume that all the channel coefficients have real values. Furthermore, the autoencoder-represented communication system is suitable for any type of channel without a tractable mathematical model[2]. That is to say, the autoencoder can be applied to any type of channel model as long as real datasets are available for training and learning.

At the receiver, the received signal $\mathbf{y}$ is passed through the ReLU layer[3] to realize the transformation $f_r : \mathbb{R}^n \mapsto \mathbb{R}^M$. The last layer of the receiver has a softmax activation as shown in TABLE I, which is a generalization of the logistic function that compresses an M-dimensional vector of arbitrary real values to an M-dimensional probability vector $\mathbf{p} = [p_1, \ldots, p_M]^T$, where each element p_i $(i = 1, 2, \ldots, M)$ lies in the range (0, 1], and all the elements add up to one [38]. For the conventional autoencoder scheme, the estimated message $\hat{s}$ is obtained from the index of the element having the highest probability in $\mathbf{p}$. Here, the BLER of DL-based communication systems is defined as

$$\text{BLER} = \frac{1}{M} \sum_s \Pr(\hat{s} \neq s). \tag{2}$$

Notably, the BLER equals the symbol error rate (SER) of the DL-based communication system.

The autoencoder based communication system can be trained offline using a large training dataset, while the iterative training process depends on the value of loss function in each iteration. The most common loss functions are MSE and categorical cross-entropy as shown in TABLE

---

[2]We note that a real-world communication channel often does not have a tractable mathematical model.

[3]It can be shown by simulation, multiple ReLU layers do not improve the BLER performance for our problem.

I, and these loss functions are determined by the vector expression s and the probability vector p. The training parameters of the autoencoder are produced to minimize the loss function. Furthermore, the trained autoencoder with the fixed NN parameters is applied to practical communication scenarios online.

B. Motivations

The one-hot vector is the conventional data representation having only one non-zero element. Thus, the data rate of the conventional DL-based communication system with one-hot vector is limited to

$$R_C = \frac{\log_2 M}{n} \quad \text{bits/channel use.} \quad (3)$$

Over the last few years, the demand for high data rates has experienced unprecedented growth in communication systems [1], [2]. Therefore, providing a high data rate is essential for DL-based communication systems in future communications.

To improve the data rate, we propose two new autoencoder schemes:

1) *Adaptive transmission scheme.* For the conventional one-hot vector scheme, the DL-based autoencoder is trained over a fixed-size transmission vector with dimension M at fixed SNR value, which can introduce two limitations. On one hand, the trained autoencoder for a certain value of M cannot work in the scenarios with different values of M. On the other hand, the performance of DL-based communication systems is suboptimal when the trained autoencoder is applied to different SNR values. For these reasons, there is a need for a new transmission scheme for the autoencoder to improve the system performances, such as maximizing the data rate while satisfying the QoS constraint [40], [41]. Therefore, we propose an adaptive transmission scheme by adaptively selecting the optimal transmission vectors for different SNR values, where the optimization objective is to maximize the data rate with certain MSE constraint.

2) *Generalized data representation scheme.* From the definition of the data rate $R_{def} = \frac{\text{Number of bits}}{\text{Channel uses}}$, it is obvious that, for the same channel environment, the data rate is proportional to the number of bits being conveyed. However, the size of transmission vector M cannot be infinite due to the high complexity associated with deep NNs. Therefore, a new data representation scheme is required to meet the high data rate requirements in future
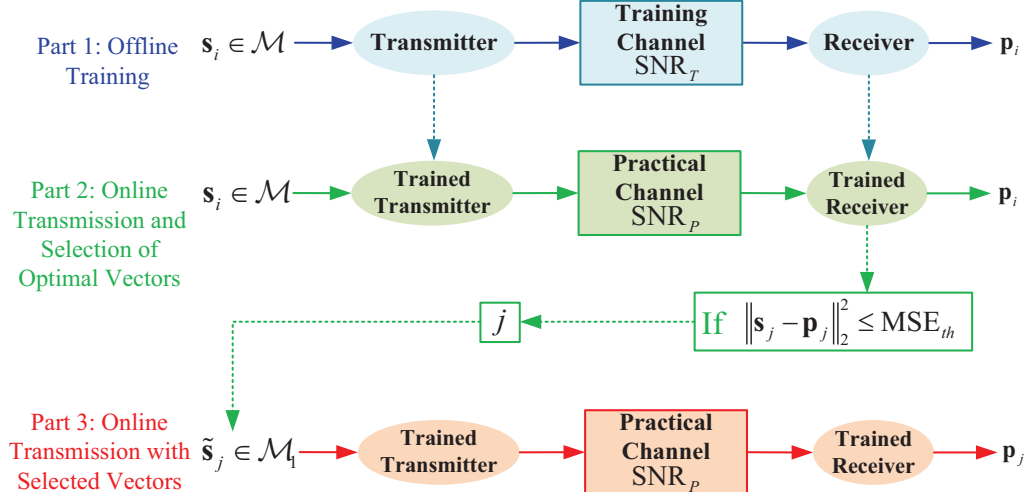
Fig. 2.   Adaptive transmission scheme applied to the DL-based communication system.

communication systems. To address this issue, we design a generalized data representation scheme that employs a new vector structure instead of the one-hot vector. The new vector structure can be generalized and used for communication scenarios having different data rate requirements.

Based on above discussions, we are motivated to develop data-rate driven transmission strategies for DL-based communication systems.

As for the system performances, the autoencoder that is trained offline using a fixed SNR value is expected to have robust performance for a wide SNR region online. In [19], it was found that an unaccommodated training SNR will result in performance degradation of DL-based communication systems, but there is little theoretical analysis. Consequently, the effect of the training SNR needs to be investigated and a reliable criterion needs to be developed for selecting training SNR values. Furthermore, current literature on DL-based autoencoder research do not analyze its performance. Therefore, we are motivated to develop an analytical framework to gain insights into the performance of DL-based communication systems.

III. ADAPTIVE TRANSMISSION SCHEME

In this section, an adaptive transmission scheme is employed in the DL-based communication system to maximize the data rate with the MSE constraint for different channel conditions.

Figure 2 shows the adaptive transmission scheme for the DL-based communication system, which consists of three parts.

The first part is offline training. The autoencoder including the transmitter and receiver is trained offline using one-hot vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$ over a fixed training SNR (SNR_T), while M should be suitably large[4], for example $M = 64$. After training, the trained transmitter and receiver, which will be used in the second part and the third part, are produced with fixed parameters.

The second part is online transmission and selection of optimal vectors. The second part includes three steps. First, each one-hot vector $\mathbf{s}_i$ in set $\mathcal{M} = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ is transmitted through the trained transmitter/receiver once over the practical channel using an operating SNR value (SNR_P). Here, the receiver can obtain the probability vector $\mathbf{p}_i$ corresponding to the transmitted vector $\mathbf{s}_i$. Second, the receiver calculates the MSE of each one-hot vector. If the MSE of the jth vector (MSE_j) is less than or equal to an MSE threshold, the receiver sends the label j back to the transmitter. In total, the receiver sends M_1 labels. Third, according to the feedback labels, the transmitter forms a new vector set $\mathcal{M}_1$, which is defined as $\mathcal{M}_1 = \{\tilde{\mathbf{s}}_j\}$, $j = 1, \ldots, M_1$, where $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \ldots, \tilde{\mathbf{s}}_{M_1}$ are the M_1 one-hot vectors selected from $\{\mathbf{s}_i\}$ with M_1 smallest MSE values. The selection goal is to maximize the data rate and satisfy the MSE requirement as

$$R_1 = \max \frac{\log_2 M_1}{n}$$
$$\text{s.t.} \quad \|\mathbf{s}_j - \mathbf{p}_j\|_2^2 \leq \mathrm{MSE}_{th}, \quad j = 1, \ldots, M_1 \tag{4}$$

where $M_1 \leq M$ satisfying $M_1 \in \{4, 8, 16, 32, 64\}$, and MSE_{th} is a preset MSE threshold.

The third part is online transmission with the selected vectors. The selected M_1 one-hot vectors are used for the autoencoder online over the current channel with SNR_P.

The main steps of the adaptive transmission scheme are summarized as follows:

Steps of the Adaptive Transmission Scheme

---

1) Train the autoencoder with a large training dataset consisting of all M possible one-hot vectors offline.

---

[4]If M is too large, the training complexity is prohibitive since the autoencoder must see every message at least once [19].

2) Each one-hot vector in $\mathcal{M}$ is transmitted through the trained autoencoder over the practical channel online.

3) Calculate the practical MSE of each vector and select $\mathbf{s}_j$ according to (4).

4) Feedback the label j and form $\mathcal{M}_1 = \{\tilde{\mathbf{s}}_j\}$.

5) Encode the message symbol using $\mathcal{M}_1$ and transmit.

IV. GENERALIZED DATA REPRESENTATION SCHEME

In this section, we propose a generalized data representation scheme to improve the data rate for DL-based communication systems.

Instead of the conventional one-hot vector containing one non-zero entry, we consider a bit vector containing m non-zero entries to improve the data rate for DL-based communication systems. An m-order bit vector $\mathbf{b} \in \mathbb{R}^M$ is defined as

$$\mathbf{b} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & \cdots & 1 & 0 \end{bmatrix}}_{m \text{ 1's}}^T \tag{5}$$

where $m = 1, 2, \cdots, \lfloor M/2 \rfloor$ denotes the number of non-zero entries in $\mathbf{b}$, and $\lfloor \cdot \rfloor$ is the floor operation. The bit vector provides $\binom{M}{m}$ possible messages for the transmission. In general, the number of possible symbols in the constellation diagram is a power of 2. For this reason, we only select $2^{\lfloor \log_2 \binom{M}{m} \rfloor}$ out of $\binom{M}{m}$ possible symbols for communications.

Furthermore, for the autoencoder shown in Fig. 1, the vector $\mathbf{s}$ at the transmitter can be viewed as a probability distribution, and the probability vector $\mathbf{p}$ at the receiver is the corresponding estimated probability distribution. The training goal of the autoencoder is to optimize $\mathbf{p}$ and reconstruct $\mathbf{s}$ while minimizing the loss function.

Thus, motivated by the above discussions, we propose a generalized data representation as a probability distribution

$$\mathbf{s} = \underbrace{\begin{bmatrix} \dfrac{1}{m} & 0 & \cdots & 0 & \dfrac{1}{m} & \cdots & \dfrac{1}{m} & 0 \end{bmatrix}}_{m \text{ non-zero entries}}^T \tag{6}$$

where the estimated message $\hat{s}$ can be obtained from the indices of elements with the m highest probabilities in $\mathbf{p}$. The conventional one-hot vector is a special case of the proposed GDR scheme when $m = 1$. Furthermore, the proposed GDR will be employed for the vector expression

TABLE II

RESULTS OF MESSAGES TRANSFORMED TO VECTORS

Message	16×1 **One-hot Vector**	8×1 **GDR Vector**
1	$[1, 0, 0, 0, 0, \ldots, 0]^T$	$[\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0]^T$
2	$[0, 1, 0, 0, 0, \ldots, 0]^T$	$[\frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0, 0]^T$
3	$[0, 0, 1, 0, 0, \ldots, 0]^T$	$[\frac{1}{2}, 0, 0, \frac{1}{2}, 0, 0, 0, 0]^T$
$\vdots$	$\vdots$	$\vdots$
14	$[0, \ldots, 0, 0, 1, 0, 0]^T$	$[0, 0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0]^T$
15	$[0, \ldots, 0, 0, 0, 1, 0]^T$	$[0, 0, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0]^T$
16	$[0, \ldots, 0, 0, 0, 0, 1]^T$	$[0, 0, \frac{1}{2}, 0, 0, \frac{1}{2}, 0, 0]^T$

processing of the transmitter in Fig. 1. As an example, when $M = 16$, there are 16 messages need to be transmitted. For the conventional one-hot scheme, the corresponding vectors are 16 different 16×1 one-hot vectors $\mathbf{s}_i$, which are shown in the first column of TABLE II. For the proposed GDR scheme, the corresponding vectors are also 16 different vectors, which can be 8×1 GDR vectors with $m = 2$. The GDR scheme provides $\binom{M}{m} = 28$ possible vectors for transmission, and we can randomly choose 16 vectors[5] as shown in the second column of TABLE II.

The data rate of the DL-based communication system can be improved by employing the proposed GDR as

$$R = \frac{\left\lfloor \log_2 \binom{M}{m} \right\rfloor}{n} \quad \text{bits/channel use.} \tag{7}$$

When $m = 1$, the data rate is obtained for the conventional one-hot vector scheme in (3). The data rate increases with m, while the value of M is suitably chosen and remains fixed. The performance gain of the proposed GDR scheme will increase with vector size M.

The maximum achievable rate of the proposed GDR scheme in the DL-based communication

[5]The vector selection is done here arbitrarily, and we leave the optimal vector selection as an open research problem.

system is derived as

$$C = \log_2(1 + \text{SNR}) = \log_2\left(1 + \frac{1}{\sigma^2}\right) \tag{8}$$

$$= \log_2\left(1 + \frac{2E_b \cdot \left\lfloor \log_2\binom{M}{m}\right\rfloor}{N_0 \cdot n}\right) \quad \text{bits/s/Hz.}$$

It can be shown that the achievable rate can be improved by using the proposed GDR scheme in the DL-based communication system. For example, when $M = 16$, the proposed GDR scheme with $m = 6$ has nearly 1.58 (bits/s/Hz) performance gain compared with the conventional one-hot vector scheme at $E_b/N_0 = 20$ dB for seven channel uses.

Furthermore, the proposed GDR can be directly applied to the proposed adaptive transmission scheme by using the generalized data representation. Combining the proposed two schemes, we obtain an adaptive GDR-based transmission scheme that can create further benefits for the DL-based communication system.

V. PERFORMANCE ANALYSIS OF THE AUTOENCODER

In this section, we provide a theoretical analysis of MSE performance for DL-based communication systems. Such an analysis can be applied to two proposed schemes and other autoencoder-represented schemes.

A. MSE Performance Analysis

In Fig. 1, the output of the ReLU layer at receiver can be written as

$$\mathbf{u} = f_r(\mathbf{y}) \triangleq f_{ReLU}\left(\mathbf{W}_r\mathbf{y} + \mathbf{b}_r\right) \tag{9}$$

where $f_{ReLU}(a) = \max\{a, 0\}$; $\mathbf{W}_r$ and $\mathbf{b}_r$ denote the trainable parameters of the ReLU layer, and they are defined as

$$\mathbf{W}_r = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{Mn} \end{pmatrix} \quad \text{and} \quad \mathbf{b}_r = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} \tag{10}$$

14

respectively, where w_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, n$, represents the symmetric interaction term between unit u_i and unit y_j in Fig. 1, and b_i is the bias term. Thus, from (1) and (9), the ith element of $\mathbf{u}$ is given by

$$u_i = \max\left\{[\mathbf{W}_r]_{i,:}(\mathbf{x} + \mathbf{n}) + b_i, 0\right\} \tag{11}$$

where $[\mathbf{W}_r]_{i,:}$ is the ith row of $\mathbf{W}_r$.

Next, a probability vector is derived from the softmax function at the receiver, and its ith element can be written as

$$p_i = \frac{e^{u_i}}{\sum_{k=1}^{M} e^{u_k}}. \tag{12}$$

From (11)-(12), in the offline training processing, different $\mathrm{SNR} = \frac{1}{\sigma^2}$ will lead to different trainable parameters $\mathbf{W}_r$ and $\mathbf{b}_r$, which will affect u_i in (11). As a result, p_i, the probability of the ith element is directly affected by the training SNR. Also, in the online practical transmission, the trainable parameters $\mathbf{W}_r$ and $\mathbf{b}_r$ are constant since the autoencoder has been trained. When the autoencoder is applied to a different SNR scenario online, it will lead to a different estimated probability vector $\mathbf{p}$ as well. The effect of SNR will also be studied through simulations.

In Appendix A, it is shown that, based on (12), the probability vector at the receiver in Fig. 1 can be approximated as

$$\mathbf{p} \approx \mathbf{Fu} \tag{13}$$

where $\mathbf{F} \in \mathbb{R}^{M \times M}$ is a diagonal matrix that is equivalent to the effect of softmax activation layer. It must be highlighted that, after training, the obtained $\mathbf{F}$ is constant when applying to online transmissions.

At the receiver, the output of the ReLU layer $\mathbf{u}$ consists of zero and non-zero elements as shown in (11). In this paper, we aim to analyze the effect of SNR on MSE performance. While the zero elements cannot reflect the characteristic of MSE, the non-zero output of the ReLU layer is considered and can be derived from (11) as

$$\mathbf{u}_+ = \mathbf{W}_r(\mathbf{x} + \mathbf{n}) + \mathbf{b}_r \tag{14}$$

if

$$[\mathbf{W}_r]_{i,:}(\mathbf{x} + \mathbf{n}) + b_i > 0. \tag{15}$$

Thus, the probability vector $\mathbf{p}$ under the assumption of (15) can be expressed as

$$\mathbf{p}_+ \approx \mathbf{F}_+ \mathbf{u}_+ \tag{16}$$

where $\mathbf{F}_+ \in \mathbb{R}^{M \times M}$ is the equivalent matrix of softmax activation layer in the non-zero case as (15), and entries of $\mathbf{F}_+$ are fixed after training.

Here, the average MSE of the DL-based communication system in the case of (15) can be given from (14) and (16) as

$$
\begin{aligned}
\mathrm{MSE} &= \mathbb{E}\left\{\|\mathbf{p}_+ - \mathbf{s}\|_2^2\right\} \\
&\approx \mathbb{E}\left\{\|\mathbf{F}_+(\mathbf{W}_r \mathbf{x} + \mathbf{b}_r) + \mathbf{F}_+ \mathbf{W}_r \mathbf{n} - \mathbf{s}\|_2^2\right\} \\
&= \mathbb{E}\left\{\|\mathbf{F}_+(\mathbf{W}_r \mathbf{x} + \mathbf{b}_r) - \mathbf{s}\|_2^2\right\} + \|\mathbf{F}_+ \mathbf{W}_r\|_2^2 \sigma^2.
\end{aligned}
\tag{17}
$$

After the autoencoder is trained over SNR_T, the transformation parameters $\mathbf{F}_+$, $\mathbf{W}_r$ and $\mathbf{b}_r$ in (17) are constant, where $\sigma_{n_T}^2$ is the noise variance at the training scenario. When the trained autoencoder is applied to the practical communication scenario with SNR_P, the noise variance of the current practical channel scenario is $\sigma_{n_P}^2$. For the non-zero case, it can be observed from (17) that, when $\sigma_{n_P}^2 < \sigma_{n_T}^2$, the practical MSE performance will be better than that of the training scenario; when $\sigma_{n_P}^2 > \sigma_{n_T}^2$, the converse is true. It indicates that the trained autoencoder can attain better system performance when it is applied to higher SNR scenario. For the zero case in (11), the variance of noise has no effect on the MSE performance. The MSE performance of the DL-based communication system will also be verified through simulations.

B. Training SNR Set Strategy

In conventional DL-based communication systems, the autoencoder is trained over a fixed SNR value offline, and it can suffer performance degradation when operating in environments having mismatched SNR values. Here, we propose a training SNR set strategy by employing multiple training SNRs, and it will improve the diversity of training dataset to obtain robust performance. For example, the training SNR set can be designed to $\mathcal{SNR}_{\mathcal{T}} = \{-20, -10, 0, 10, 20\}$ dB for offline training. Also, the system performance gain of the proposed training SNR set strategy will be shown by simulation results.

TABLE III

PARAMETERS FOR THE AUTOENCODER SETUP

Parameter	Value
Optimizer	Adam [42]
Loss function	MSE[6]
Epoch	150
Batch size	45
Trained samples	2×10^4
Test samples	1×10^6

TABLE IV

TRAINING PARAMETERS OF AUTOENCODER

	Vector size	Multiple dense layers	Normaliz-ation layer	ReLU layer	Softmax layer	Total
	$M = 4$	55	14	32	20	121
	$M = 8$	135	14	64	72	285
Simulated parameters	$M = 16$	391	14	128	272	805
	$M = 32$	1287	14	256	1056	2613
	$M = 64$	4615	14	512	4160	9301
Theoretical parameters	M	$(M+1)(M+n)$	$2n$	$M(n+1)$	$M(M+1)$	$(2M+3)(M+n)$

VI. NUMERICAL RESULTS

In this section, we evaluate the numerical results of the proposed adaptive transmission scheme, the GDR scheme, the adaptive GDR-based transmission scheme, and the system performances in the DL-based communication system via simulations on the TensorFlow framework. In all the simulations, the autoencoder is trained over the stochastic AWGN channel model with $n = 7$ channel uses without exhaustive hyperparameter tuning. Here, we use the same set of parameters for the autoencoder setup as described in TABLE III.

TABLE IV presents the simulated and theoretical number of training parameters in autoen-

---

[6]For convenience, the MSE loss function is used to show the effect of SNR on the MSE performance and to verify the analysis in Subsection V-A.
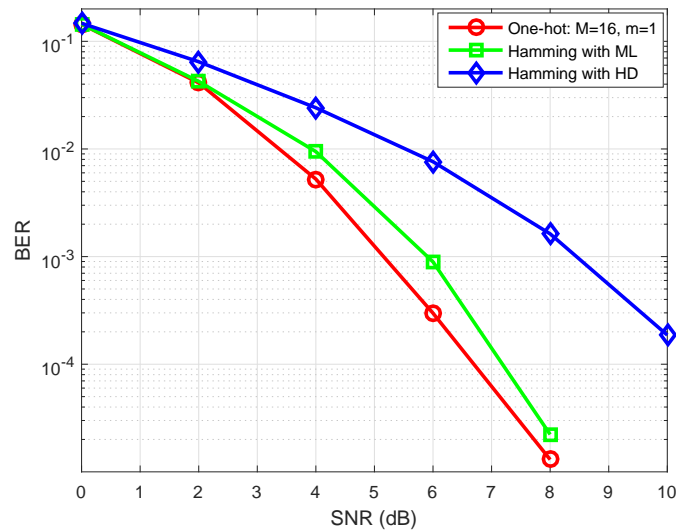
Fig. 3.   Simulated BER performance for the autoencoder and conventional communication schemes.

coder, where different size of the data representation M is employed. From TABLE IV, it is clear that the simulated number of trainable parameters increases with M from 4 to 64, including the total number (of trainable parameters) and the number (of trainable parameters) in each layer except for the normalization layer. The simulated results agree with the theoretical number of parameters as shown in the last row of TABLE IV. The increasing number of training parameters leads to an increased complexity for training. For the conventional one-hot vector, the data rate can be improved by increasing M as shown in (3) at the cost of high complexity. While the data rate of the proposed GDR scheme can be improved by controlling the number of non-zero elements m as well as the value of M as shown in (7).

A. Performance of the Autoencoder and Conventional Communication System

This subsection shows the simulated bit-error rate (BER) performance of the autoencoder scheme with one-hot vectors and the conventional communication scheme employing Hamming code, where the training SNR is 10 dB.

Figure 3 shows the simulated BER performance of the DL-based autoencoder scheme with $M = 16$ and $m = 1$ (one-hot vector) and the conventional communication scheme, where the conventional communication scheme employs binary phase-shift keying (BPSK) modulation and
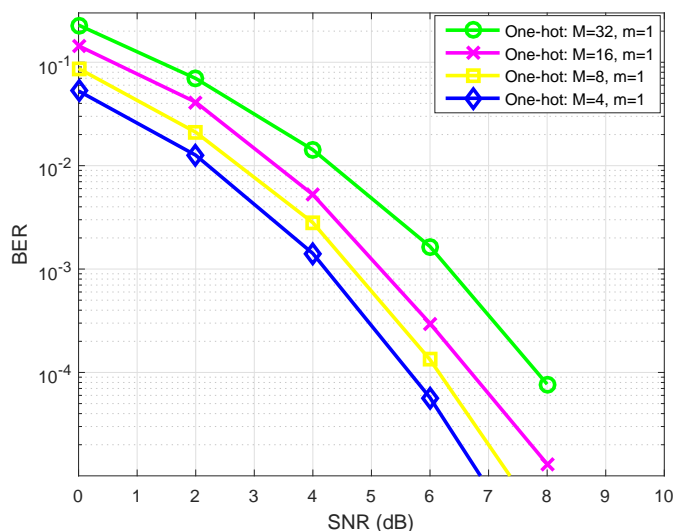
18

Fig. 4.   Simulated BER for the autoencoder employing the conventional one-hot vector scheme with different vector size M.

a $(7, 4)$ Hamming code with either binary hard-decision (HD) or maximum-likelihood (ML) decoding. Given the same information transmission rate (transmitting four information bits over seven channel uses), it can be seen that the BER performance of the autoencoder scheme is better than that of the conventional communication scheme employing Hamming code with ML decoding or HD decoding. It is worth pointing out that the autoencoder approach does not use any error control strategy for the noisy channel, and it still outperforms a classical scheme that employs error control strategy. It was reported in [19] that an autoencoder can achieve similar BLER performance compared to a conventional channel-coded scheme.

Figure 4 depicts the simulated BER performance of the DL-based autoencoder that employs gray coding and the conventional one-hot vector with the vector size $M = 4, 8, 16, 32$, where the training SNR is 10 dB. In Fig. 4, the BER of the conventional one-hot vector scheme increases when M is varied from 4 to 32.

B. Performance of the Proposed Adaptive Transmission Scheme

In this subsection, we show the simulated BLER and MSE performance of the proposed adaptive transmission scheme in the DL-based communication system. Here, the autoencoder is trained using $\mathrm{SNR}_T = 5$ dB.
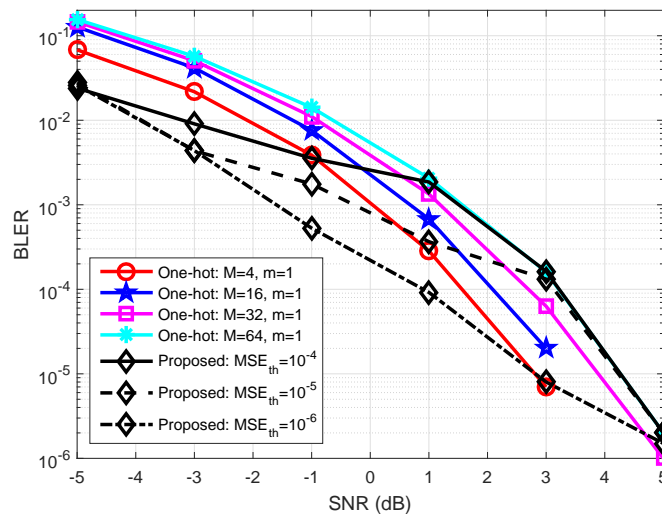
Fig. 5. Simulated BLER for the autoencoder with conventional one-hot vector and proposed adaptive transmission schemes, while the MSE thresholds are 10^{-4}, 10^{-5}, and 10^{-6}.

TABLE V

THE NUMBER OF ADAPTIVELY SELECTED VECTORS M_1 FOR DIFFERENT SNR VALUES AND MSE_{th}

M_1	SNR$=-5$ dB	SNR$=-3$ dB	SNR$=-1$ dB	SNR$=1$ dB	SNR$=3$ dB	SNR$=5$ dB
$\mathrm{MSE}_{th} = 10^{-4}$	4	16	32	64	64	64
$\mathrm{MSE}_{th} = 10^{-5}$	4	4	16	32	64	64
$\mathrm{MSE}_{th} = 10^{-6}$	4	4	4	16	32	64

Figure 5 depicts the simulated BLER performance of the DL-based autoencoder that employs the proposed adaptive transmission scheme and the conventional one-hot vector scheme, where the MSE thresholds are 10^{-4}, 10^{-5}, and 10^{-6}. First, it can be seen from Fig. 5 that the BLER of the conventional one-hot vector scheme increases when M is varied from 4 to 64, since smaller value of M requires less trainable parameters as shown in TABLE IV. With the same training dataset, the less trainable parameters contribute to better training accuracy. Second, for the proposed adaptive transmission scheme, the BLER increases when the MSE threshold is increased from 10^{-6} to 10^{-4} in Fig. 5. The reason is that, to maximize the data rate, a lower MSE threshold (means the tighter bound) requires smaller M_1 to satisfy the MSE constraint,
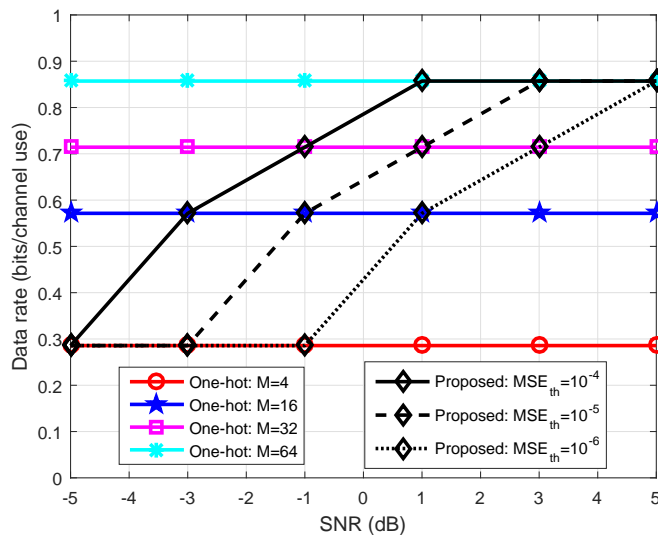
Fig. 6.   Data rate performance for the autoencoder with conventional and adaptive transmission schemes, while the MSE thresholds are 10^{-4}, 10^{-5}, and 10^{-6}.

which results in lower BLER. As shown in TABLE V, for each MSE threshold, the number of selected vectors M_1 adaptively increases from 4 to 64 with the increasing SNR value. For example, for $\text{MSE}_{th} = 10^{-5}$, when SNR is changing from -5 dB to 5 dB, the M_1 value changes accordingly as $4, 4, 16, 32, 64, 64$. For this reason, higher SNR value makes it easy to meet the MSE requirement and as a result, a larger value M_1 is obtained for maximizing the data rate. Fig. 5 shows that, when the data rates are the same, i.e. $M = M_1$, the adaptive transmission scheme can reduce the BLER of the one-hot vector scheme by 80%. The reason for the performance gain is that the proposed adaptive transmission scheme can select the optimal vectors that meet the MSE requirement as shown in (4).

Figure 6 illustrates the data rate performance of the autoencoder that employs the conventional one-hot vector scheme and the proposed adaptive transmission scheme with the MSE thresholds being 10^{-4}, 10^{-5}, and 10^{-6}. From Fig. 6, we observe that the data rates of the conventional one-hot vector scheme are constant for all SNR values. However, in Fig. 6, the data rate of the proposed adaptive transmission scheme increases with SNR as shown in TABLE V. From Fig. 5 and Fig. 6, it can be seen that the proposed adaptive transmission scheme can obtain better BLER performance than that of the conventional one-hot vector scheme when operating at the
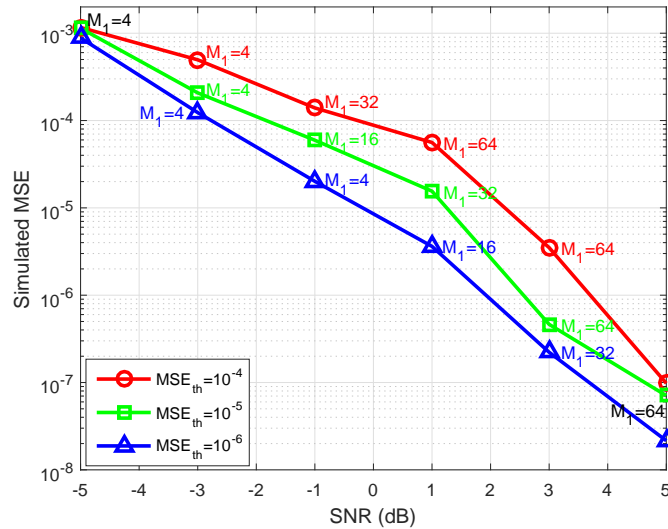
Fig. 7. Simulated MSE for the autoencoder employing the adaptive transmission scheme with the MSE thresholds being 10^{-4}, 10^{-5} and 10^{-6}.

same data rate.

Figure 7 presents the simulated MSE performance for a practical communication system that employs the proposed adaptive transmission scheme with MSE thresholds being 10^{-4}, 10^{-5} and 10^{-6}. It is seen from Fig. 7 that the simulated MSE of the proposed adaptive transmission scheme increases with MSE threshold. Furthermore, the simulated MSE of the proposed scheme decreases while the SNR increases, which is consistent with the prediction in (17). As expected, when the simulated MSE reaches the corresponding MSE threshold, the number of selected vectors M_1 is almost 64 which is the maximum value, and the maximum data rate is obtained.

C. Performance of the Proposed GDR Scheme

This subsection shows the BLER performance and the maximum achievable rate of the proposed GDR scheme in the DL-based communication system, where the training SNR is 5 dB.

Figure 8 shows the simulated BLER performance of the DL-based communication system that employs the proposed GDR and conventional one-hot vector schemes, while the schemes in (a) have the same vector size $M = 8$ and the schemes in (b) have the same data rate $R = 6/7$
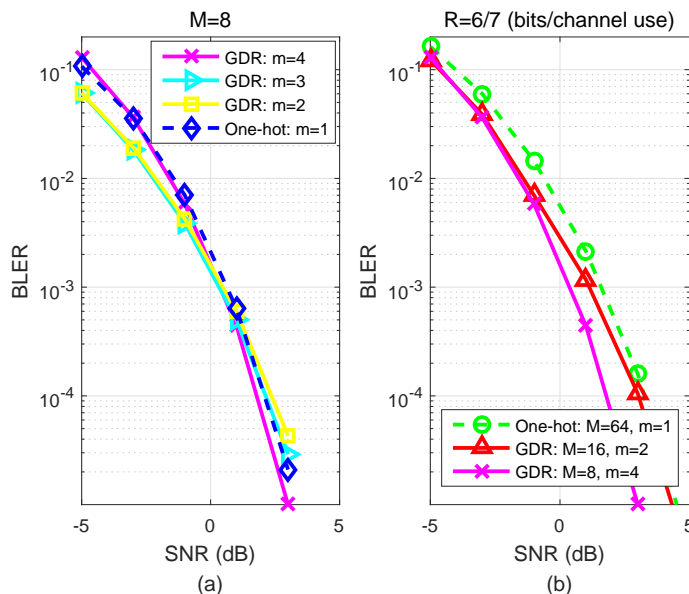
Fig. 8.   Simulated BLER for the autoencoder employing different data representations with (a) $M = 8$ and (b) $R = 6/7$ (bits/channel use), while the trained SNR is 5 dB.

TABLE VI

DATA RATE OF THE DL-BASED COMMUNICATION SYSTEM

	One-hot $M=8, m=1$	GDR $M=8, m=2$	GDR $M=8, m=3$	GDR $M=8, m=4$	GDR $M=16, m=2$	One-hot $M=64, m=1$
Data rate (bits/channel use)	3/7	4/7	5/7	6/7	6/7	6/7

(bits/channel use). In Fig. 8 (a), for the same vector size $M = 8$, the proposed GDR schemes ($m = 2, 3, 4$) obtain comparable BLER performances when compared to the conventional one-hot vector scheme ($m = 1$). It indicates that, with the same vector size, the number of non-zero elements in s has little effect on the BLER performance. Even the BLER performances are similar, the data rates of the GDR schemes and the one-hot vector scheme are different and they are shown in TABLE VI. It can be seen from TABLE VI that, with $M = 8$, the data rates of the proposed GDR schemes are $R = 6/7, 5/7, 4/7$ (bits/channel use) respectively with $m = 4, 3, 2$. The data rates of all GDR schemes are greater than that of the conventional one-hot
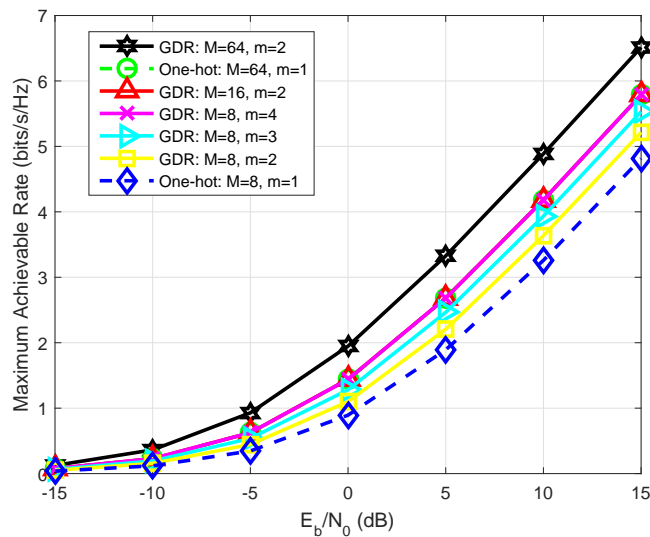
Fig. 9.   Maximum achievable rate for the autoencoder with different data representations.

vector scheme as $R = 3/7$ (bits/channel use), and the GDR scheme with $m = 4$ can double the data rate of the one-hot vector scheme. In Fig. 8 (b), with the same data rate $R = 6/7$ (bits/channel use) including the proposed schemes $M = 8$ with $m = 4$, $M = 16$ with $m = 2$, and the conventional scheme $M = 64$ with $m = 1$, the proposed GDR schemes have better BLER performance than that of the conventional one-hot vector scheme, and these performance gains are achieved with the GDR schemes with lower training complexity, i.e., less number of training parameters as shown in TABLE IV. Obviously, the BLER decreases with the vector size M for the same reason as that in Fig. 5. Furthermore, it can be found that the proposed GDR scheme can avoid the performance degradation by increasing m, when the transmission message size is large. For example, the proposed GDR scheme $M = 16$ with $m = 8$ can transmit $2^{\lfloor \log_2 \binom{16}{8} \rfloor} = 32768$ messages by using 16×1 vectors. However, to achieve the same data rate, the size of one-hot vector should be 32768×1 at least, which will lead to significant BLER performance degradation. In both Fig. 8 (a) and (b), the simulated BLER is less than 10^{-5} when the SNR is 5 dB, which demonstrates that the autoencoder attains a high accuracy with sufficient training over $\mathrm{SNR}_T = 5$ dB.

Figure 9 illustrates the maximum achievable rate of a DL-based communication system employing different data representations. It can be seen from Fig. 9 that, with $M = 8$, the
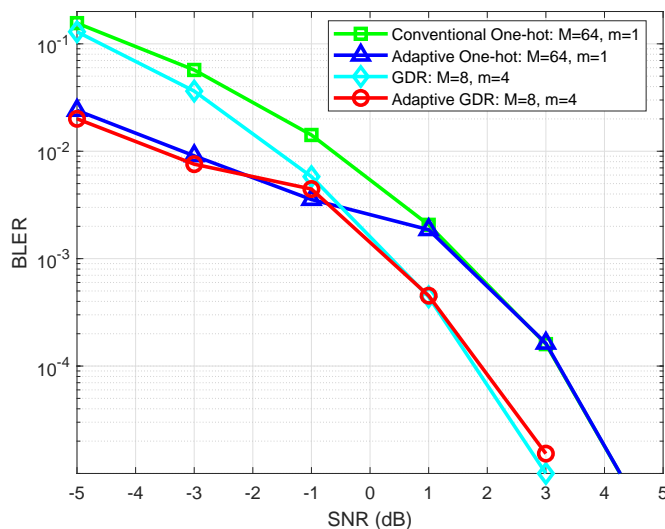
Fig. 10.   Simulated BLER for the autoencoder employing four different schemes, while the trained SNR is 5 dB.

maximum achievable rate increases when the order m increases from 1 to 4, which is consistent with the result in (8). This shows that the proposed GDR scheme can obtain a remarkable achievable rate improvement. Notably, the performance gain of the proposed GDR scheme is increased when the vector size M increases. As shown in Fig. 9, the GDR scheme employing $M = 64$ with $m = 2$ has a great performance gain when compared with the conventional scheme employing $M = 64$ with $m = 1$. Besides, the maximum achievable rate of the proposed GDR schemes ($M = 8$ with $m = 4$ and $M = 16$ with $m = 2$) is same as that of the conventional one-hot vector scheme ($M = 64$, $m = 1$) in Fig. 9. To obtain the same achievable rate with the GDR scheme, the conventional one-hot vector scheme needs to increase the vector size M, which has been shown in Fig. 8 to degrade the BLER performance.

Figure 10 presents the simulated BLER performance of the DL-based communication system that employs four schemes, including the conventional one-hot vector ($M = 64$, $m = 1$), the adaptive transmission (based on one-hot vector), the GDR ($M = 8$, $m = 4$), and the adaptive GDR-based transmission schemes[7]. The MSE threshold of the two adaptive schemes

[7]Except the adaptive GDR-based transmission scheme, the BLER performance results of the other three schemes have been shown in Fig. 5 and Fig. 8 (b).
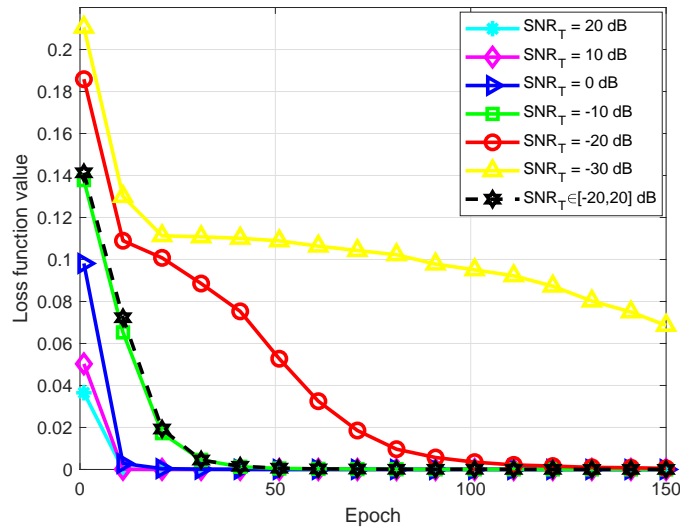
Fig. 11.   Simulated loss function performance of autoencoder in training process, while different fixed training SNRs and training SNR set are employed.

is $\text{MSE}_{th} = 10^{-4}$. The achievable maximal data rate of all four schemes is the same, $R = 6/7$ (bits/channel use). In Fig. 10, the adaptive GDR-based transmission scheme achieves the best BLER performance in all SNR regions, i.e., the adaptive GDR-based transmission scheme outperforms the GDR ($M = 8, m = 4$) scheme in low SNR region and the adaptive transmission (based on one-hot vector) scheme in high SNR region. With the $\text{MSE}_{th} = 10^{-4}$, the adaptive GDR-based transmission scheme selects M_1 vectors for transmission, when SNR is changing from -5 dB to 5 dB, the M_1 value changes accordingly as $8, 16, 32, 64, 64, 64$. It can be seen that, when $\text{SNR} = -5, -3, -1$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the adaptive transmission (based on one-hot vector) scheme since they have similar M_1; when $\text{SNR} = 1, 3, 5$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the GDR scheme ($M = 8, m = 4$), and the reason is that the adaptive GDR-based transmission scheme selects all 64 vectors for usage, in this case, the adaptive GDR-based transmission scheme is equal to the GDR scheme ($M = 8, m = 4$).

26

D. Performance Comparison of Different Training SNR

In this subsection, we investigate the effect of training SNR on system performance including the loss function performance in training process, the simulated BLER, and MSE performances in practical transmission process. Here, the data representation parameters are $M = 8$ and $m = 1$, and SNR_T denotes the training SNR.

Figure 11 shows the simulated loss function performance in training processing, when the autoencoder is trained over different SNRs and SNR set. The SNR set is designed as $\mathcal{SNR}_{\mathcal{T}} = \{-20, -10, 0, 10, 20\}$ dB, which includes all the fixed SNRs except for -30 dB. In Fig. 11, an epoch is the process that the entire training dataset is passed through the autoencoder once. As shown in Fig. 11, when the SNR_T is increased from -20 dB to 20 dB, the loss function value decreases and the convergence of loss function improves, which indicates that the good channel environment contributes to the improvement of the training performance. However, with $\mathrm{SNR}_T = -30$ dB, the loss value does not converge within 150 epoches. Furthermore, it can be seen from Fig. 11 that, the loss value of the autoencoder training with SNR set is similar to that of the autoencoder training with $\mathrm{SNR}_T = -10$ dB. The simulated results suggest that the training SNR has significant effect on the training performance of the autoencoder.

Figure 12 depicts the simulated BLER performance of the practical DL-based communication system employing the trained autoencoder with different fixed training SNRs and training SNR set. In Fig. 12, the BLER decreases with SNR_T ranging from 20 dB to -20 dB. The reason is that, with the lower training SNR (that is to say the worse channel environment), the autoencoder needs to learn more features to reconstruct the input at the output, which leads to a robust autoencoder and better BLER performance. However, the training SNR has a lower bound for the autoencoder. As shown in Fig. 12, when $\mathrm{SNR}_T = -30$ dB, the BLER is approximately 0.6, which demonstrates that the autoencoder trained over this channel environment cannot learn the features anymore. It is consistent with the non-convergence performance of the loss function with $\mathrm{SNR}_T = -30$ dB in Fig. 11. Besides, Fig. 12 shows that the BLER performance of the training SNR set scheme is similar to that of $\mathrm{SNR}_T = -10$ dB scheme, which is almost the best performance except for the $\mathrm{SNR}_T = -20$ dB scheme. It shows that training with SNR set can improve the generalization performance of the autoencoder. From Fig. 11 and Fig. 12, it can be found that, with a higher training SNR value, the autoencoder obtains better convergence

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
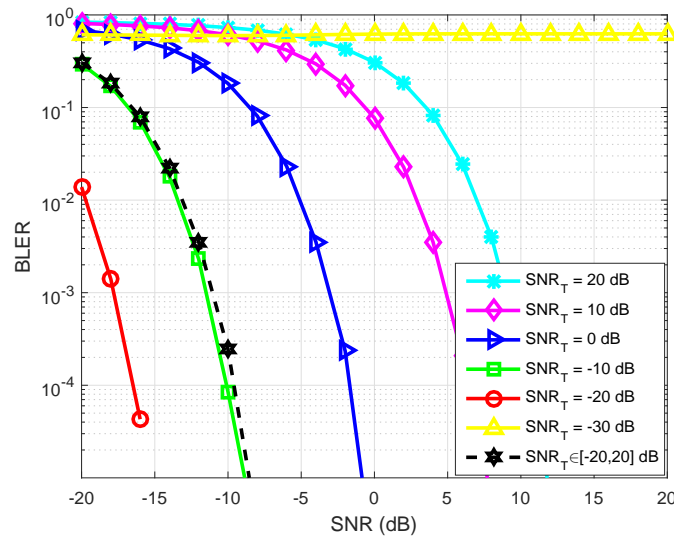51
52
53
54
55
56
57
58
59
60

Fig. 12. Simulated BLER for the DL-based communication system employing the trained autoencoder with different fixed training SNRs and training SNR set.

performance in training but worse BLER performance. The simulated results indicate that the training SNR will directly affect the system performance, which agrees with the analysis in Subsection V-A.

Figure 13 illustrates the simulated MSE performance of the practical DL-based communication system employing different trained autoencoders, while the training SNRs include different fixed SNRs and SNR set. In Fig. 13, it is clear that the MSE decreases when SNR is increased. It indicates that the MSE performance improves when the trained autoencoder is applied to a higher SNR scenario, which is consistent with the analysis in (17). Furthermore, the simulated MSE performance in Fig. 13 is similar to the BLER performance as shown in Fig. 12 for the same reasons.

VII. CONCLUSION AND FUTURE RESEARCH

In this paper, we proposed two new transmission schemes to address the problem of limited data rate in DL-based communication system using autoencoder. We designed an adaptive transmission scheme for different channel conditions to maximize the data rate with a mean square error constraint. Furthermore, we proposed the GDR scheme to obtain higher data rate
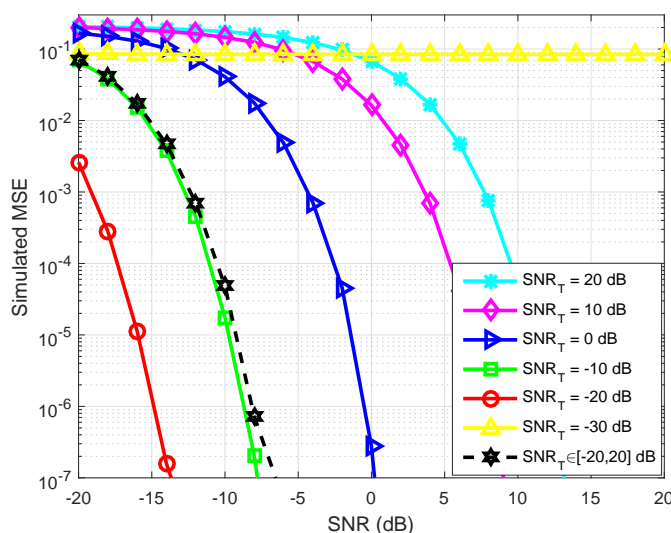
Fig. 13.   Simulated MSE for the DL-based communication system employing the trained autoencoder with different fixed training SNRs and training SNR set.

than the conventional one-hot vector scheme with a similar BLER performance. Besides, the effect of training SNR and MSE performance were analyzed and verified by simulations. We discovered that high training SNR can lead to good convergence in training process but worse BLER performance for practical transmission. We also introduced a training SNR set strategy to address the tradeoff between convergence and error rate. It was shown that the autoencoder trained over a low SNR can attain better BLER and MSE performances when operating in the high SNR region. As a result, it is concluded that training the autoencoder at a lower SNR value, in general, will lead to good system performance.

For a low SNR value, say, $\mathrm{SNR}_T = -30$ dB, numerical results indicate that the loss function value does not converge and the BLER degrades dramatically. This suggests that the studied autoencoder system is unable to learn from very noisy data set. It should be emphasized that the current system assumes neither knowledge about the noise nor about the system model. Therefore, one interesting research problem is to study the low SNR communication using DL techniques when partial knowledge about the noise and the system model is known.

To further improve the performance of the DL-based communication systems, we can possibly employ the ensemble method where the results of a set of individual NNs are combined to

estimate the transmitted message. In classifier problems, it has been shown that the ensemble method is effective in improving accuracy or decomposing a complex problem into easier subproblems [43].

APPENDIX A

DERIVATION OF (13)

Let

$$\mathbf{p} = \mathbf{F}\mathbf{u}. \tag{18}$$

According to (12), eq. (18) can be formulated as

$$\frac{1}{\sum_{k}^{M} e^{u_k}} \begin{pmatrix} e^{u_1} \\ e^{u_2} \\ \vdots \\ e^{u_M} \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1M} \\ f_{21} & f_{22} & \cdots & f_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{MM} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{pmatrix} \tag{19}$$

and we can obtain that

$$e^{u_i} = (f_{i1}u_1 + f_{i2}u_2 + \cdots + f_{ii}u_i + \cdots + f_{iM}u_M) \sum_{k=1}^{M} e^{u_k}. \tag{20}$$

Next, e^{u_i} can be approximated according to the Taylor's theorem as

$$e^{u_i} \approx 1 + u_i + \frac{u_i^2}{2!} + \cdots + \frac{u_i^N}{N!} \tag{21}$$

where N is a sufficiently large integer.

Finally, combining (20) and (21), we can derive the elements of matrix $\mathbf{F}$ as

$$f_{ij} \approx \begin{cases} \frac{1}{\sum_{k=1}^{M} e^{u_k}} \left(u_i^{-1} + 1 + \frac{u_i}{2!} + \cdots + \frac{u_i^{N-1}}{N!} \right) & i = j \\ 0 & i \neq j \end{cases}. \tag{22}$$

Thus, eq. (22) shows that the probability vector $\mathbf{p}$ at the receiver can be approximated as (13).

30

References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart. 2016.

[3] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. 1st Int. Conf. 5G Ubiq. Connect.*, Nov. 2014, pp. 1–6.

[4] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.

[5] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[7] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, June 2011.

[8] X. Ge, S. Tu, G. Mao, C. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[9] H. Zhang, Y. Dong, J. Cheng, M. J. Hossain, and V. C. M. Leung, "Fronthauling for 5G LTE-U ultra dense cloud small cell networks," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 48–53, Dec. 2016.

[10] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.

[11] X. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, May 2014.

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

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

[14] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing, exploratory DSP," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 145–154, Jan. 2011.

[15] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[16] T. Wang, C. K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, Nov. 2017.

[17] X. You, C. Zhang, X. Tan, S. Jin, and H. Wu, "AI for 5G: Research directions and paradigms," *Science China Information Sciences*, Sep. 2018.

[18] N. Huang, M. Chen, W. Xu, and J.-Y. Wang, "Incorporating importance sampling in EM learning for sequence detection in SPAD underwater OWC," *IEEE Access*, vol. 7, pp. 4529–4537, 2019.

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

[20] S. Dörner, 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, Feb. 2018.

31

[21] V. Vanhoucke, A. Senior, and M. Z. Mao, "Improving the speed of neural networks on CPUs," in *Deep Learn. Unsupervised Feature Learn. NIPS Workshop*, 2011.

[22] E. Nachmani, Y. Be'ery, and D. Burshtein, "Learning to decode linear codes using deep learning," in *IEEE Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep. 2016, pp. 341–346.

[23] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Beery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[24] T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding," in *IEEE Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2017, pp. 1–6.

[25] T. O'Shea, K. Karra, and T. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2016, pp. 1–6.

[26] T. Erpek, T. J. O'Shea, and T. C. Clancy, "Learning a physical layer scheme for the MIMO interference channel," in *IEEE Int. Conf. Commun (ICC)*, May 2018, pp. 1–5.

[27] T. J. OShea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[28] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. t. Brink, "OFDM-autoencoder for end-to-end learning of communications systems," *arXiv preprint arXiv:1803.05815*, 2018.

[29] C. K. Wen, W. T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, to be published, DOI 10.1109/LWC.2018.2832128.

[30] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[31] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, to be published, DOI 10.1109/LWC.2018.2874264.

[32] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," *arXiv preprint arXiv: 1706.01151*, 2017.

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

[34] S. Xue, Y. Ma, N. Yi, and R. Tafazolli, "Unsupervised deep learning for MU-SIMO joint transmitter and noncoherent receiver design," *IEEE Wireless Commun. Lett.*, to be published, DOI 10.1109/LWC.2018.2865563.

[35] K. Kim, J. Lee, and J. Choi, "Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 828–831, Apr. 2018.

[36] 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, Feb. 2018.

[37] H. Kim, Y. Jiang, S. Kannan, S. Oh, and P. Viswanath, "Deepcode: Feedback codes via deep learning," in *Advances in Neural Inf. Process. Syst.*, 2018, pp. 9436–9446.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[39] A. D. Stefano, O. Mirabella, G. D. Cataldo, and G. Palumbo, "On the use of neural networks for Hamming coding," in *IEEE Int. Symp. Circuits and Systems*, vol. 3, Jun. 1991, pp. 1601–1604.

[40] L. Wu, Z. Zhang, and H. Liu, "Adaptive modulation with finite rate feedback for QR decomposition-successive interference cancellationbased multiple-in multiple-out systems," *IET Commun.*, vol. 7, no. 5, pp. 456–462, Mar. 2013.

[41] X. Chen, L. Wu, Z. Zhang, J. Dang, and J. Wang, "Adaptive modulation and filter configuration in universal filtered multi-carrier systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1869–1881, Mar. 2018.

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

[43] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Information Fusion*, vol. 37, pp. 132–156, 2017.

1

Response on TCOM-TPS-19-0361.R1

Dear Editor and Reviewers:

Thank you very much for your kind handling of our manuscript entitled "Data-Rate Driven Transmission Strategies for Deep Learning Based Communication Systems". We would like to thank you all for your careful reading and constructive comments, which are very valuable and helpful for improving our manuscript. We have taken into account these comments carefully and revised the manuscript accordingly. Detailed responses to all the comments are provided in this point-to-point response. Also, for the convenience of the reviewers, we have marked the revisions in red color in the revised manuscript.

Again, we would like to thank you all for reviewing this manuscript and thank Prof. Vaneet Aggarwal for coordinating the review of this manuscript.

Sincerely yours,

Xiao Chen, Julian Cheng, Zaichen Zhang, Liang Wu, Jian Dang, and Jiangzhou Wang

2

RESPONSE TO EDITOR'S COMMENTS

We are grateful to the Editor for the valuable comments. Each comment has been taken into account in the revised manuscript. The reply to each of your comments is given as follows.

Comment 1:

"The paper has two schemes - one as adaptive transmission for one-hot vectors, and another as generalized data representation. However, the two are independent extensions and thus do not make a coherent story. It would be great to have a scheme that used both at the same time, and evaluate the joint scheme where benefits for both can be seen. In other words, adaptive transmission on GDR should be used."

Response:

Thanks very much for this comment.

The motivation of this manuscript is to design high data-rate schemes in DL-based communication systems for future wireless networks. However, the conventional one-hot vector has a low data rate in DL-based communication systems. The reason is that an $M \times 1$ one-hot vector consists of 0s in all entries with the exception of a single 1, e.g., $[0, \ldots, 0, 1, 0, \ldots, 0]^T$, and there are only M possible transmitted messages, which leads to limited data rate. This becomes a barrier for developing future DL-based communication systems. Besides, the conventional autoencoder is trained under a fixed signal-to-noise ratio (SNR) with unrealistic expectation to operate well for a wide range of SNR values in practical transmission scenarios [R1].

Thus, in order to achieve higher data rate, an adaptive transmission scheme is designed for different communication scenarios (SNR values) to maximize the data rate in DL-based communication systems having a QoS constraint. Then, we propose a generalized data representation (GDR) scheme to improve the data rate of DL-based communication systems as well, and the GDR scheme can replace the conventional one-hot vector scheme.

Furthermore, we agree with the Editor that the joint scheme is necessary to make a coherent story. Per the Editor's suggestion, an adaptive GDR-based transmission scheme has been added and evaluated in Section IV and Section VI-C of the revised manuscript. The simulation results show that the adaptive GDR-based transmission scheme achieves the best BLER performance in all SNR regions, i.e., the adaptive GDR-based transmission scheme outperforms the GDR
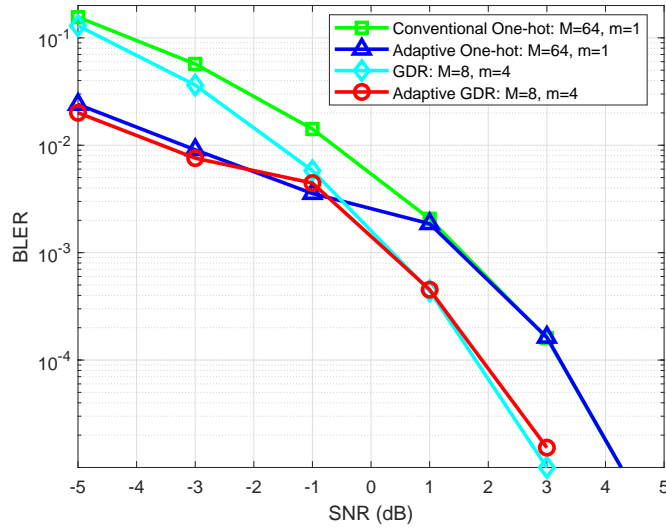
Fig. R1. Simulated BLER for the autoencoder employing four different schemes, while the trained SNR is 5 dB.

scheme in low SNR region and the adaptive transmission scheme in high SNR region.

For the Editor's convenience, we have quoted the related revisions in the revised manuscript as follows. (Fig. R1 corresponds to Fig. 10 in the revised manuscript.)

"*Furthermore, the proposed GDR can be directly applied to the proposed adaptive transmission scheme by using the generalized data representation. Combining the proposed two schemes, we obtain an adaptive GDR-based transmission scheme that can create further benefits for the DL-based communication system.*"

"*Figure 10 presents the simulated BLER performance of the DL-based communication system that employs four schemes, including the conventional one-hot vector ($M = 64$, $m = 1$), the adaptive transmission (based on one-hot vector), the GDR ($M = 8$, $m = 4$), and the adaptive GDR-based transmission schemes[7]. The MSE threshold of the two adaptive schemes is $\mathrm{MSE}_{th} = 10^{-4}$. The achievable maximal data rate of all four schemes is the same, $R = 6/7$ (bits/channel use). In Fig. 10, the adaptive GDR-based transmission scheme achieves the best BLER performance in all SNR regions, i.e., the adaptive GDR-based transmission scheme out-*

---

[7]Except the adaptive GDR-based transmission scheme, the BLER performance results of the other three schemes have been shown in Fig. 5 and Fig. 8 (b).

4

performs the GDR ($M = 8$, $m = 4$) scheme in low SNR region and the adaptive transmission (based on one-hot vector) scheme in high SNR region. With the $\mathrm{MSE}_{th} = 10^{-4}$, the adaptive GDR-based transmission scheme selects M_1 vectors for transmission, when SNR is changing from -5 dB to 5 dB, the M_1 value changes accordingly as $8, 16, 32, 64, 64, 64$. It can be seen that, when $\mathrm{SNR} = -5, -3, -1$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the adaptive transmission (based on one-hot vector) scheme since they have the same MSE threshold and similar M_1; when $\mathrm{SNR} = 1, 3, 5$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the GDR scheme ($M = 8$, $m = 4$), and the reason is that the adaptive GDR-based transmission scheme selects all 64 vectors for usage, in this case, the adaptive GDR-based transmission scheme is equal to the GDR scheme ($M = 8$, $m = 4$). "

We thank the editor again for suggesting us the adaptive GDR-based transmission scheme that unifies the two contributions so that the revised manuscript has now been strengthened with a more coherent story. To acknowlege this formally and openly, we have added an Acknowledgment Section in the revised manuscript as follows:

"*We thank all anonymous reviewers and the editor for their constructive comments that have significantly improve the original manuscript. In particular, we thank the editor Prof. Vaneet Aggarwal for suggesting us investigate the adaptive GDR-based transmission scheme.*"

Comment 2:

"Added Section VI-A does not seem relevant to this paper, since that is not using any of the two proposed schemes. In that sense, this section should be removed and results should be based on the proposed methodologies in this paper."

Response:

Thanks for this comment.

Per Comment 3 and Comment 5 of Reviewer 3 in the previous review, we added Section VI-A to show the improvement of DL-based communication systems over the traditional communication systems. The Comments of Reviewer 3 in the previous review are quoted as follows.

"*Comment 3: ...Also, while the authors mention that they don't plot the traditional codes, I think readers will still appreciate the plot for the performance of traditional codes.*"

5

"*Comment 5: BER or BLER. Related to the point above, I would suggest plotting BER (as well as BLER) because several curves include different values of block length M (e.g., 64, 16, and 8 in Figure 6). As of now, only with BLER, comparison of different schemes with different blocklengths is not immediate.*"

The purpose of the added Section VI-A is to reinforce the advantage of the autoencoder communication scheme over the conventional communication scheme. We are flexible to remove Section VI-A entirely if the Editor requires us to do so.

Comment 3:

"The concept of GDR is simple enough, with using m-choose-k vector rather than one-hot vectors. However, it is unclear why this waited till the main section. Intro could have the two schemes in words of what the methodology is."

Response:

Thank you very much for this comment.

To address the Editor's comment, we have revised the methodology of the proposed GDR scheme in the Introduction Section so that the introduction of the GDR occurs much earlier in the revised manuscript.

For the convenience of the Editor, we have quoted the revised text as follows.

1) "*In DL-based communication systems, we pointout the limited data rate problem of the conventional one-hot vector scheme. To address this problem, we design an adaptive transmission scheme with a QoS constraint for different channel conditions. In the proposed scheme, the optimal transmission vectors are adaptively selected for different SNR values, where the goal is to maximize the data rate while having a constraint on MSE performance. It is shown that, when both two schemes have the same data rate, the proposed adaptive transmission scheme can reduce BLER of the conventional one-hot vector scheme by 80%.*

2) *Furthermore, we propose a generalized data representation scheme to improve the data rate in DL-based communication systems. The proposed scheme represents the message by using a probability vector having multiple non-zero elements, instead of the conventional one-hot vector having only one non-zero element. As expected, higher data rate is obtained*

6

by the proposed GDR scheme with comparable BLER performance and low complexity. When the vector size is eight, as an example, the proposed GDR scheme can double the data rate of the conventional one-hot vector scheme. To the best of the authors' knowledge, this is the first time that the GDR scheme is proposed and its effectiveness is verified."

Comment 4:

"There are typos like "ont-hot", "there is yet a". A pass on the paper can be helpful."

Response:

Thank you very much for the detailed comment and suggestion.

We have carefully checked and revised the spelling and grammar throughout the revised manuscript.

We thank the Editor again for his significant comments that have helped us strengthen the contributions and presentations of this manuscript. We hope the Editor will be satisfied with our clarifications and modifications.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

7

RESPONSE TO REVIEWER 1'S COMMENTS

We are grateful to Reviewer 1 for the valuable comments. Each comment has been taken into account in the revised manuscript. The reply to your comments is given as follows.

Comment:

"Thank you very much for considering my suggestions in the comments. I don't have further questions about this work. Just one thing, on page 15 of the response letter, you thanked me (Reviewer 1) again in Reviewer 2's place."

Response:

We thank Reviewer 1 for his/her precious time in reviewing our resubmission and acknowledging this work. Also, we have checked the manuscript and the response carefully, and we believe the revised manuscript and response have been improved over the previous submission.

We thank Reviewer 1 again for his/her significant comments that have helped us improve the quality of this manuscript.

8

RESPONSE TO REVIEWER 3'S COMMENTS

We are grateful to Reviewer 3 for the valuable comments. Each comment has been taken into account in the revised manuscript. The replies to your comments are as follows.

Comment 1:

"I sincerely appreciate the authors' response and additional experimental results. It took me lots of thoughts to decide on my recommendation, but in the end, I can't quite recommend the acceptance of this manuscript."

Response:

We thank Reviewer 3 for his/her precious time in reviewing our manuscript and response. We have revised the manuscript according to the comments/suggestions raised by Reviewer 3, and we believe the revised manuscript has been improved significantly over the previous submission. In the following please kindly find the point-to-point responses.

Comment 2:

"While several results were interesting (e.g., BLER performance of high data rate GDR is similar to the BLER performance of one-hot encoding scheme), I think still there's one major missing piece - that is a large message size (M), which I believe shouldn't be ignored. One-hot vector itself has a problem of scaling to large message size M (as shown in Fig. R9 in the response), and GDR does not seem to improve in this aspect. Hence, it is hard to find myself convinced that GDR scheme will be practically useful."

Response:

Thanks very much for this detailed comment.

We are encouraged to learn that Reviewer 3 finds several results are interesting. The one-hot scheme is a special case of the GDR scheme, while $m = 1$. Therefore, as Reviewer 3 can see, the GDR scheme has the same performance degradation as the that of the one-hot scheme when the vector size M is too large. However, to obtain the same data rate, the GDR scheme can keep a small vector size M and large number of non-zero elements m, which will not lead to performance degradation. The simulation results are shown as follows. (Fig. R2 corresponds to
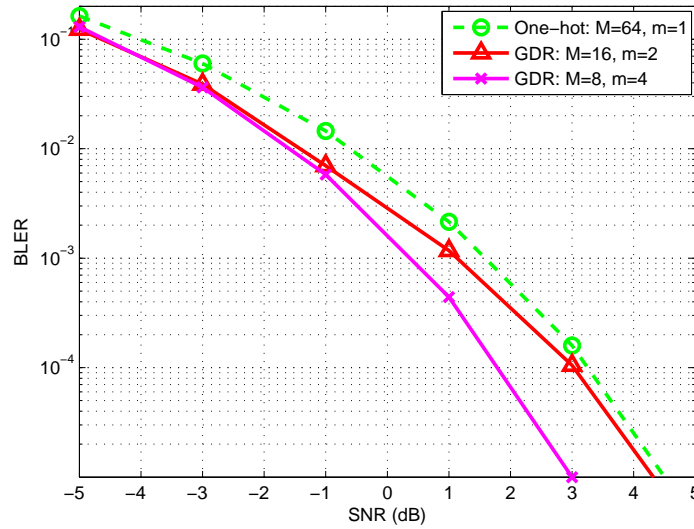
9

Fig. R2. Simulated BLER for the autoencoder employing different data representations with $R = 6/7$ (bits/channel use), while the trained SNR is 5 dB.

Fig. 8 (b) in the revised manuscript.)

Figure R2 shows the simulated BLER performance of the DL-based communication system that employs the proposed GDR and conventional one-hot vector schemes, while the schemes including the proposed GDR schemes $M = 8$ with $m = 4$, $M = 16$ with $m = 2$, and the conventional scheme $M = 64$ with $m = 1$. All three schemes have the same data rate $R = 6/7$ (bits/channel use). In Fig. R2, the proposed GDR schemes have better BLER performance than that of the conventional one-hot vector scheme. For example, when the SNR value is 3 dB, the BLER of proposed GDR scheme $M = 8$ with $m = 4$ is only 1×10^{-5}, however, the BLER of the one-hot scheme $M = 64$ with $m = 1$ is about 2×10^{-4}.

Thus, the proposed GDR scheme can avoid the performance degradation by increasing m, when the transmission message size is large. For example, the proposed GDR scheme $M = 16$ with $m = 8$ can transmit $2^{\lfloor \log_2 \binom{16}{8} \rfloor} = 32768$ messages by using 16×1 vectors. However, to achieve the same data rate, the size of one-hot vector should be 32768×1 at least, which will lead to significant BLER performance degradation due to the high complexity of NNs. Furthermore, when the message size is large enough, parallel transmission will be an effective strategy, which is an interesting and meaningful problem for our future research.

10

To address the comment of Reviewer 3, we have provided detailed descriptions in Section VI-C of the revised manuscript, and we have quoted the related revisions as follows.

"*...Obviously, the BLER decreases with the vector size M for the same reason as that in Fig. 5. Furthermore, it can be found that the proposed GDR scheme can avoid the performance degradation by increasing m, when the transmission message size is large. For example, the proposed GDR scheme $M = 16$ with $m = 8$ can transmit $2^{\lfloor \log_2 \binom{16}{8} \rfloor} = 32768$ messages by using 16×1 vectors. However, to achieve the same data rate, the size of one-hot vector should be 32768×1 at least, which will lead to significant BLER performance degradation.* "

Comment 3:

"On the other hand, from a technical perspective, whereas it is true that GDR has not been applied to channel coding, the idea itself is not novel, and I can't quite find novel techniques in neural network training or architecture, or theoretical development."

Response:

Thank you very much for the comment.

We are grateful to learn that Reviewer 3 confirms the proposed GDR has not been applied to channel coding.

To the best of the authors' knowledge, this is the first time that the GDR scheme is proposed and its effectiveness is verified. The proposed scheme represents the message by using a probability vector having multiple non-zero elements, instead of the conventional one-hot vector having only one non-zero element. As expected, higher data rate is obtained by the proposed GDR scheme with comparable BLER performance and low complexity. When the vector size is eight, as an example, the proposed GDR scheme can double the data rate of the conventional one-hot vector scheme. The proposed GDR scheme is novel for the reason that the GDR scheme is the first time to be proposed in the DL-based communication system and it obtains significant performance gain.

On the other hand, neural network is used as a tool to deal with certain challenges in existing and future communication systems.

This manuscript aims to increase the transmission rate, which is a totally fundamental problem of DL-based communication systems. The authors expect to improve the data rate and provide
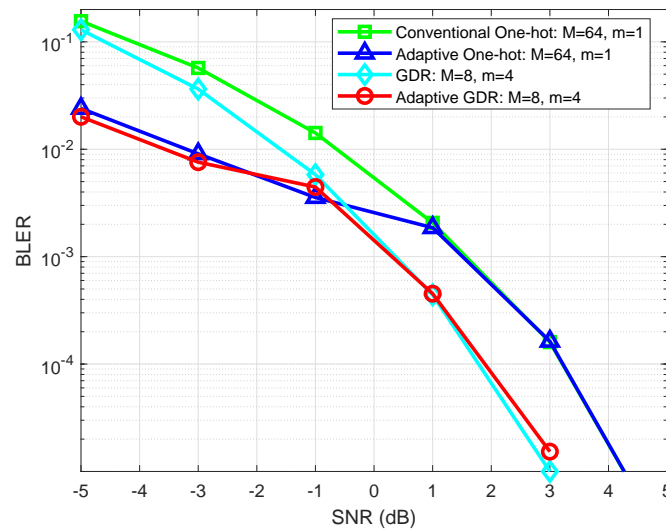
Fig. R3.   Simulated BLER for the autoencoder employing four different schemes, while the trained SNR is 5 dB.

an effective transmission basis for future researches of DL-based communication systems, and the proposed GDR scheme makes it possible. The proposed GDR scheme is simple but it plays an irreplaceable role in future DL-based communication systems.

Furthermore, the proposed GDR can be directly applied to the proposed adaptive transmission scheme by using the generalized data representation. Combining the proposed two schemes, we obtain an adaptive GDR-based transmission scheme that can create further benefits for the DL-based communication system.

For the convenience of Reviewer 3, we have quoted the related revisions in the revised manuscript as follows. (Fig. R3 corresponds to Fig. 10 in the revised manuscript.)

"*Figure 10 presents the simulated BLER performance of the DL-based communication system that employs four schemes, including the conventional one-hot vector ($M = 64$, $m = 1$), the adaptive transmission (based on one-hot vector), the GDR ($M = 8$, $m = 4$), and the adaptive GDR-based transmission schemes[7]. The MSE threshold of the two adaptive schemes is $\mathrm{MSE}_{th} = 10^{-4}$. The achievable maximal data rate of all four schemes is the same, $R = 6/7$*

[7]Except the adaptive GDR-based transmission scheme, the BLER performance results of the other three schemes have been shown in Fig. 5 and Fig. 8 (b).

12

(bits/channel use). In Fig. 10, the adaptive GDR-based transmission scheme achieves the best BLER performance in all SNR regions, i.e., the adaptive GDR-based transmission scheme out-performs the GDR ($M = 8$, $m = 4$) scheme in low SNR region and the adaptive transmission (based on one-hot vector) scheme in high SNR region. With the $\mathrm{MSE}_{th} = 10^{-4}$, the adaptive GDR-based transmission scheme selects M_1 vectors for transmission, when SNR is changing from -5 dB to 5 dB, the M_1 value changes accordingly as $8, 16, 32, 64, 64, 64$. It can be seen that, when $\mathrm{SNR} = -5, -3, -1$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the adaptive transmission (based on one-hot vector) scheme since they have the same MSE threshold and similar M_1; when $\mathrm{SNR} = 1, 3, 5$ dB, the BLER of the adaptive GDR-based transmission scheme is similar to that of the GDR scheme ($M = 8$, $m = 4$), and the reason is that the adaptive GDR-based transmission scheme selects all 64 vectors for usage, in this case, the adaptive GDR-based transmission scheme is equal to the GDR scheme ($M = 8$, $m = 4$). "

Comment 4:

"These two aspects together led me to the conclusion of recommendation of rejection at this point, but I look forward to see future work and advances in this direction and area."

Response:

Thanks very much for these two aspects comments. We have revised the manuscript according to the comments/suggestions raised by the reviewer, and we hope Reviewer 3 will be satisfied with the updated version. Furthermore, we are encouraged that Reviewer 3 looks forward to future work in this direction and area.

We thank Reviewer 3 again for his/her significant comments that have helped us improve the quality of this manuscript. We hope Reviewer 3 will be satisfied with our clarifications and modifications.

We believe this work studies an important and fundamental problem for designing a communication system using an autoencoder approach. This is an emerging research area with relatively a few research results. The revised work now presents a coherent story to address the low data rate issue associated with the conventional one-hot vector. Therefore, we believe that important

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

13

findings and the offered analysis have offered a significant advancement of the knowledge for this important and new research field.

14

REFERENCES

[R1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart. 2016.

[R2] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[R3] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Information Fusion*, vol. 37, pp. 132–156, 2017.