

Reliable and Secure Deep Learning-Based OFDM-DCSK Transceiver Design Without Delivery of Reference Chaotic Sequences

Haotian Zhang¹, Lin Zhang², Senior Member, IEEE, Yuan Jiang³, Member, IEEE, and Zhiqiang Wu⁴, Senior Member, IEEE

Abstract—Chaos-based communications can be applied to high-speed vehicular information transmissions thanks to the anti-jamming and anti-interference capabilities of chaotic transmissions. In traditional practical chaotic systems, reference chaotic signals are required to be delivered to remove complex chaotic synchronization circuits. However, the direct transmission of reference signals will degrade the security performances, while interferences and noises imposed on the reference signals due to imperfect channel conditions will deteriorate the reliability performances. In order to enhance the reliability and security performances over vehicular channels such as the railway channel and the channels undergoing fast fadings, in this paper, we propose a deep learning (DL) aided intelligent OFDM-DCSK transceiver. In this design, no reference chaotic signals are delivered, and we propose to utilize the time-delay neural network (TDNN) to learn the chaotic maps, followed by the long short-term memory (LSTM) units to extract and exploit the correlations between chaotic modulated signals, and multiple fully connected layers (FCLs) to estimate the user bit data. With the aid of the constructed deep neural network (DNN), after the offline neural network training, the receiver can recover the transmitted information with lower bit error rate (BER) and enhance security performances. Theoretical performance is then analyzed for the proposed intelligent transceiver. Simulation results validate the proposed design, and demonstrate that the intelligent DL-based OFDM-DCSK system can achieve better BER and security performances over fast fading and railway channels compared with the benchmark systems.

Index Terms—Deep learning (DL), long short-term memory (LSTM) network, orthogonal frequency division multiplexing-aided differential chaos shift keying (OFDM-DCSK), reliability and security, time-delay neural network (TDNN).

I. INTRODUCTION

OWING to the non-periodic, noise-like and initial value sensitive characteristics of chaotic sequences, chaos-based communications have been widely applied to provide secure and anti-jamming transmissions for wireless systems. For instance, the aperiodic chaotic sequences have been applied to modulate user data in practical systems such as ultra-wide-band (UWB) communication systems and power line communication systems [1], [2] to enhance security and reliability performances.

In recent years, empowered by the sixth-generation mobile communication system (6G), high-speed vehicular communications have been proposed to be applied in higher data rate applications, such as on-board and wayside high definition (HD) video surveillance, on-board real-time high-data-rate connectivity, and train operation information transmission, etc., to promote the development of intelligent transportation [3]. However, the vehicular information transmissions have to combat complicated time changing fadings induced by the mobility of users, which can be well modeled by fast fading channels such as Rayleigh fading channel [4] or Rician fading channel [5], [6]. Besides, due to the broadcasting property of vehicular channels, the mobile users might suffer from eavesdropping or malicious attacks.

Chaos-based transmissions can effectively mitigate the interferences and improve the security performances, which provide the promising solutions for reliable and secure high-speed vehicular information transmissions. Chaotic modulation methods can be classified into coherent and non-coherent schemes. Since non-coherent systems do not require the complicated synchronization circuit that is difficult to be implemented in practical systems, they have attracted more research interests than coherent systems. Among non-coherent chaotic modulation schemes, differential chaos shift keying (DCSK) [7] has been widely studied, which can provide reliable transmissions with low complexity. However, since half of the symbol duration is used to transmit reference chaotic sequences, DCSK systems suffer from the low spectrum efficiency and require delay lines.

Manuscript received 22 August 2021; revised 14 March 2022; accepted 15 May 2022. Date of publication 25 May 2022; date of current version 15 August 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1802300, in part by the National Science Foundation of China under Grant U20A20162, and in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2020A1515010703. The review of this article was coordinated by Dr. Haijun Zhang. (Corresponding author: Lin Zhang.)

Haotian Zhang is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China (e-mail: sam_3132@qq.com).

Lin Zhang is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China, and also with the Southern Marine Science and Engineering Guangdong Laboratory, Guangdong 519000, China (e-mail: isszl@mail.sysu.edu.cn).

Yuan Jiang is with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518000, China (e-mail: jiangyuan3@mail.sysu.edu.cn).

Zhiqiang Wu is with the Department of Electrical Engineering, Tibet University, Lhasa 850000, China, and also with the Wright State University, Dayton 45435 USA (e-mail: Zhiqiang.Wu@wright.edu).

Digital Object Identifier 10.1109/TVT.2022.3175968

To improve the efficiency and remove the delay line, the orthogonal frequency division multiplexing (OFDM) has been applied in DCSK systems to compose the OFDM-DCSK scheme [8]. In OFDM-DCSK systems, the reference chaotic signals are delivered via specific one or more subcarriers. Due to the naturally broadcasting property of wireless channels, malicious users or eavesdroppers might retrieve the data, and thus the direct delivery of chaotic signals will degrade the security of chaotic communication systems. Moreover, multicarrier interferences will increase due to the delivery of real-valued chaotic signals, which lead to reliability degradations [9], especially over fast fading channels.

In order to enhance the reliability and the security performances of chaotic systems, research works have been performed to improve the transceiver structures. For example, in [10] and [11], the chaotic sequences were scrambled in the time domain to enhance the security performances of the DCSK systems. However, the delay line circuits are still required for the demodulation, thus the practicality is weak. Besides, [12] proposed to scramble the chaotic sequences in the frequency domain instead of the time domain, but it suffers from low spectrum efficiency since only half of spectrum bands are exploited to transmit the information-bearing symbols. In addition, the chaotic sequences are cyclic shifted in [13] to enhance the system security. Besides, in [14], the information bits are transmitted by specific indices of selected Walsh codes implicitly, thus no reference chaotic signal is required and the energy efficiency can be improved. In our recent research works, we exploited the frequency hopping to scramble the chaotic modulated sequences to increase the security and the reliability performances [15]. Furthermore, in [16], we utilized chaotic sequences to implement the position modulation without transmitting reference chaotic signals, thus both the efficiency and the security performances can be improved.

However, most of these improved chaotic modulation schemes still require to deliver reference chaotic signals. For those few schemes [14], [16] dispensing with the reference signals, they need to add signal processing modules to guarantee the reliability performances, which increase the complexity and hardware cost. Our objective is to remove the requirement of the delivery of chaotic signals with high adaptability to existing schemes, and then to improve both the security and the reliability performances. Different from the traditional chaotic transceivers, we propose to remove the delivery of reference chaotic signals at the transmitter, then we propose a deep neural network (DNN) architecture to intelligently extract the characteristics of reference chaotic signals embedded in received chaotic modulated signals, which are delivered via multiple subcarriers. Thus the received signals can be recovered intelligently with enhanced reliability and security performances.

In this design, we utilize the powerful optimization and classification capabilities of DNN to determine the estimates for received data. Different from the DNN applied for the channel estimation [17], modulation recognition [18], demodulation [19], and end-to-end communication [20], and different from our recent work on the DNN aided chaotic receiver design [21], in this paper, we propose a DNN architecture

constituted by one time-delay neural network (TDNN) [22], two recursive long short-term memory (LSTM) units [23] and multiple fully-connected layers (FCLs). The TDNN consists of three one-dimensional convolutional layers (1D-CLs), which aim to find the mapping pattern of chaotic sequences and to extract the features. Then LSTM units are utilized to find and exploit the correlation relationship between the received chaotic modulated signals. Subsequently, the estimates can be recovered and output from FCLs.

For the proposed intelligent chaotic transceiver without delivery of reference chaotic signals, at the training stage, the DNN will gradually learn to extract the chaotic sequence features and to formulate an optimized demodulation mapping. Then, at the online deployment stage, with the configured trainable parameters, the user data can be reliably recovered in real time from the received signals with no use of reference chaotic signals. Since no reference chaotic sequences are transmitted, both the efficiency and the security can be improved. Moreover, thanks to the powerful learning and generalization capabilities of DNN, the proposed design can adapt to complicated time changing conditions, thus better reliability performances can be obtained over fast fading channels.

Briefly, the main contributions include the following:

- 1) We propose an intelligent deep learning (DL)-based OFDM-DCSK transceiver with no need of delivery of reference chaotic sequences to improve the efficiency, reliability and security performances over vehicular channels such as the railway channel.
- 2) We construct the DNN architecture for demodulations with considerations of the chaotic modulation characteristics. In the proposed architecture, the TDNN is applied to learn and extract the features of the mapping pattern of generating the chaotic sequences, while the LSTM units are utilized to find and exploit the correlation relationship between the received chaotic modulated signals to recover the received user data.
- 3) We propose the hyperparameter selection criterion. Then theoretical performances including the spectrum efficiency, the energy efficiency, the security performance and the complexity are analyzed. Moreover, we provide the data set collection method considering the fairness of performance comparisons.

This paper is organized as follows. Section II presents the transceiver structure of the DNN-based OFDM-DCSK transceiver. Then, in Section III, we describe the architecture of the proposed DNN and the operating principle of each layer, while the hyper-parameter selection criterion is also provided. Next, Section IV analyzes the spectrum efficiency, energy efficiency and security performances of the proposed system. Subsequently, the data set generation is discussed in Section V, followed by the simulation results for analysis of BER, security and robustness performances. Finally, Section VI concludes our findings.

II. DL-BASED OFDM-DCSK TRANSCEIVER

In this section, we will present the structure of the proposed DL-based OFDM-DCSK transceiver.

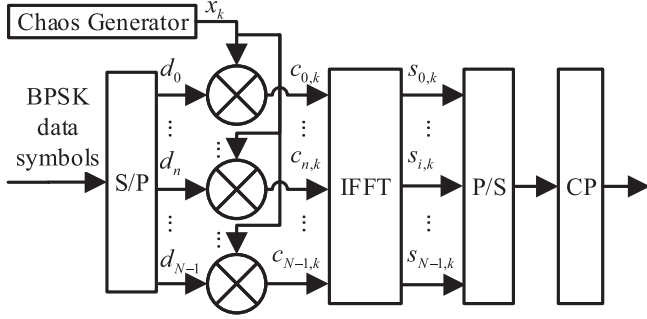


Fig. 1. The proposed transmitter structure.

A. The Transmitter Structure

Fig. 1 illustrates the transmitter structure of the proposed system. The user data are modulated by the binary phase shift keying (BPSK) at first. In the meantime, considering that the second-order Chebyshev polynomial function (CPF) possesses complex dynamic properties and low computational complexity [24], we apply the CPF to generate the chaotic sequence $\mathbf{x}$ of length K . The generated chaotic sequence is expressed as

$$x_k = 1 - 2x_{k-1}^2, \quad 0 \leq k \leq K - 1, \quad (1)$$

where x_k denotes the k th chip of $\mathbf{x}$ and is in a range of -1 to 1 , while the initial value x_{-1} is uniformly distributed between -1 and 1 .

Then, after the serial to parallel (S/P) conversion, at the k th chip time slot, the n th BPSK symbol is modulated by the k th chaotic chip as

$$c_{n,k} = d_n x_k, \quad (2)$$

where d_n represents the n th BPSK user data symbol with $0 \leq n \leq N - 1$.

Subsequently, the inverse fast Fourier transform (IFFT) operations are performed on the chaotic modulated symbols $c_{0,k}$ to $c_{N-1,k}$ to modulate the information-bearing symbols onto N available subcarriers, then we have

$$s_{i,k} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} c_{n,k} e^{j \frac{2\pi n}{N} i}, \quad (3)$$

where $s_{i,k}$ is the resultant i th OFDM symbol at the k th chip time slot with $0 \leq i \leq N - 1$, j is the imaginary unit with $j^2 = -1$. After the IFFT operations and the parallel to serial (P/S) conversion, a copy of the tail of OFDM-DCSK modulated signal $\mathbf{s}$, named as the cyclic prefix (CP), is added in front of $\mathbf{s}$ to mitigate the possible inter-symbol interference (ISI). Subsequently, the signals are transmitted over the channel.

Notably, different from the traditional OFDM-DCSK transmitter [8], all subcarriers are used to deliver chaotic modulated signals, and no reference chaotic sequences are transmitted. Thus both the efficiency and the security performances can be improved.

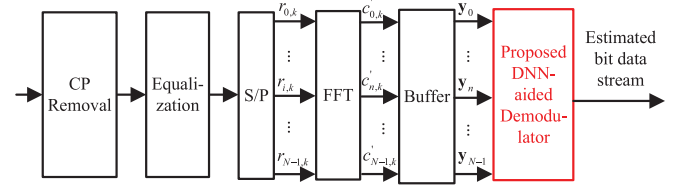


Fig. 2. The intelligent DNN-aided receiver structure.

B. The Receiver Structure

At the receiver side, as shown in Fig. 2, after the CP removal, the equalization is performed to suppress the inter symbol interference, and zero-forcing and time domain equalizers are applied in this paper. Subsequently, the serially received signal at the k th chip time slot are converted to parallel information subsets $r_{0,k}$ to $r_{N-1,k}$. Then, fast Fourier transform (FFT) operations are performed to extract the information-bearing chaotic modulated symbols in the frequency domain, which is expressed as

$$c'_{n,k} = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} r_{i,k} e^{-j \frac{2\pi i}{N} n}, \quad (4)$$

where $c'_{n,k}$ denotes the recovered n th chaotic modulated symbol at the k th chip slot.

Next, $c'_{0,k}$ to $c'_{N-1,k}$ are buffered, and then they are concatenated with the chaotic modulated symbols in the whole K chip slots to constitute the information-bearing sequences $\mathbf{y}_n = [c'_{n,0}, \dots, c'_{n,k}, \dots, c'_{n,K-1}]$ ($0 \leq n \leq N - 1$). After that, $\mathbf{y}_n$ acts as the inputs to the proposed intelligent DNN-based demodulator to attain the data estimates. Note that as mentioned above, with the proposed DNN, the received data can be recovered intelligently and reliably. More details about the proposed DNN-aided intelligent demodulator are presented as follows.

III. LSTM-BASED DEEP LEARNING DEMODULATOR

In this section, we present the DNN architecture of the proposed DNN-aided demodulator, and describe the operation principle of each layer. Subsequently, the training procedure and the hyper-parameter configurations for the DNN are introduced.

A. Architecture of DNN-Aided Demodulator

Fig. 3 illustrates the proposed architecture of the DNN employed in the DNN-aided demodulator. The DNN is composed of three 1D-CLs, two recursive LSTM units, two FCLs, and one batch normalization (BN) layer. As mentioned above, at the receiver, after buffering and concatenating the recovered chaotic modulated symbols $c'_{n,k}$ in K chip time slots, the resultant recovered information-bearing chaotic sequences $\mathbf{y}_0$ to $\mathbf{y}_{N-1}$, which contain the information of N data bits, are input to the DNN for the data recovery.

To be more explicit, the input vector $\mathbf{y}_n$ is first processed by the TDNN consisting of three 1D-CLs to find the mapping pattern and extract the features of chaotic sequences. The reason why we select the TDNN is that the TDNN is capable of processing one-dimension data, while most convolutional neural networks are proposed to process two-dimensional data. With

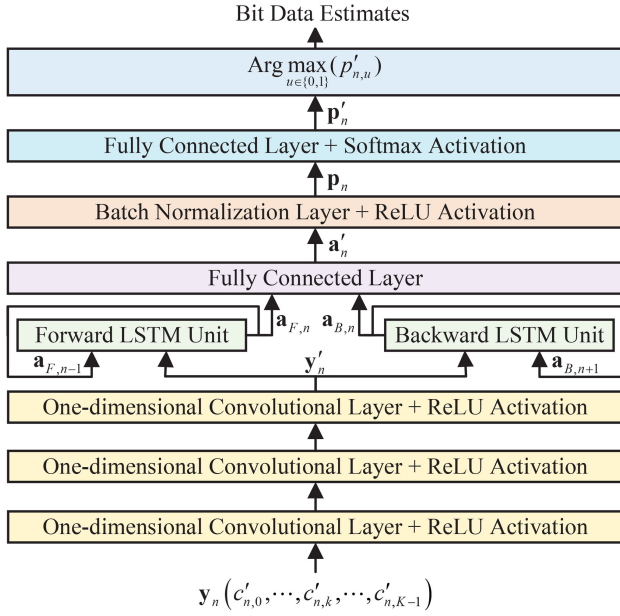


Fig. 3. The neural network architecture of the DNN-aided demodulator.

the aid of the TDNN, the proposed DNN aided demodulator is able to capture the temporal features of chaotic modulated sequences having the property of one dimension. Owing to the powerful temporal dynamic behavior capture capability, the TDNN can intelligently formulate the chaotic dynamics characteristics. The resultant feature vector $\mathbf{y}'_n$ is then input to the following bi-directional LSTM layer.

As shown in Fig. 3, the bi-directional LSTM layer includes one forward LSTM unit and one backward LSTM unit, which respectively perform the forward and backward recursive non-linear operations on the feature vectors in turn to capture the correlations between chaotic modulated sequences. By exploiting the memory cells in the LSTM unit, the features of the strong correlations between the input feature vectors which are correlated to the same reference chaotic sequence can be extracted while the noises can be suppressed.

Next the output feature vectors $[\mathbf{a}_{F,n}, \mathbf{a}_{B,n}]$ of the forward and backward LSTM units are concatenated and sent to the subsequent two FCLs to determine the estimates of information-bearing vectors, while one BN layer is applied between them to mitigate the vanishing and exploding gradient effects and accelerate the convergence. After multiple linear and nonlinear operations are conducted, the probability of data bit will be calculated and the probability vector $\mathbf{p}'_n$ will be generated. Note that $\mathbf{p}'_n$ contains two elements $p'_{n,0}$ and $p'_{n,1}$, which separately denote the probability that the n th transmitted bit data is 0 or 1. Then according to the index of the largest element of $\mathbf{p}'_n$, the proposed DNN can provide the estimate of the user data bits. More details about the operations conducted in DNN layers are presented as follows.

B. Operations Conducted in DNN Layers

1) *1D-CL*: In this paper, we apply three 1D-CLs to compose a TDNN similar to [22], where C_{out} means the output

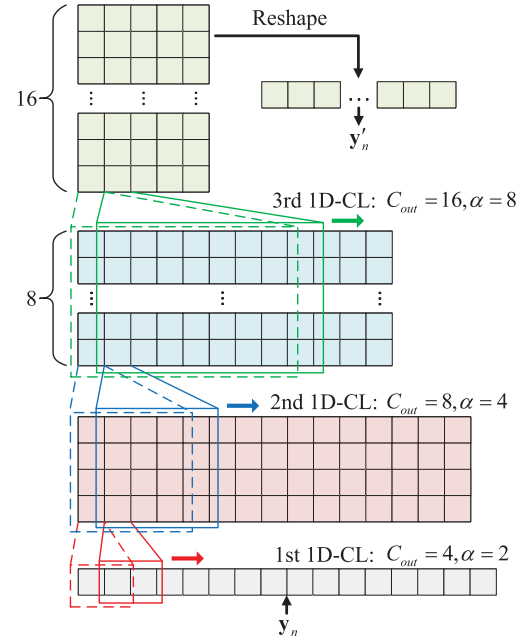


Fig. 4. The structure of the TDNN.

feature dimension. Note that in our proposed DNN, we use $C_{out} = 4, 8, 16$. As shown in Fig. 4, each convolution kernel will process α adjacent elements of the input along the time axis in turn to capture the temporal dynamic behavior within α time slot, which matches the characteristics that most chaotic sequences are generated from difference or differential equations and have specific temporal features. Besides, the 1D-CL in the higher layer of the employed TDNN will apply convolution kernels with larger quantity and width. Therefore, both the short-term and long-term temporal behaviors of $\mathbf{y}_n$ will be learnt and exploited to compress the time dimension and enlarge the feature dimension. Finally, the output matrix is reshaped to a one-dimensional feature vector $\mathbf{y}'_n$ and sent to the following LSTM units.

To be specific, as shown in Fig. 4, in the TDNN, the inputs of a 1D-CL are usually two-dimensional matrices which consist of a time axis and a feature axis, and the convolution kernels in the 1D-CL will sequentially perform the convolutional operations on each adjacent elements of the input along the time axis. For brevity, in this paper, the row denotes the feature axis and the column represents the time axis. Notably, the width in the time axis of a convolution kernel is decided by the user while the height in the feature axis is always the same as the feature dimension of the inputs. Let $\mathbf{m}$ denote the input of 1D-CL and $\mathbf{m}'$ represent the output, then each element $m'_{i,j}$ of $\mathbf{m}'$ can be calculated as

$$m'_{i,j} = \sigma_{\mathbf{m}'} \left(\sum_{v=0}^{C_{in}-1} \sum_{u=0}^{\alpha-1} w_{CK,v,u}^{(i)} m_{v,j+u} + b_{\mathbf{m}',i,j} \right), \quad (5)$$

where $m'_{i,j}$ denotes the element of i th row and j th column of $\mathbf{m}'$ and $m_{v,j+u}$ represents the element of v th row and $(j+u)$ th column of $\mathbf{m}$, $w_{CK,v,u}^{(i)}$ is the element of v th row and u th column of the i th convolution kernel, $b_{\mathbf{m}',i,j}$ means the bias of $m'_{i,j}$, α

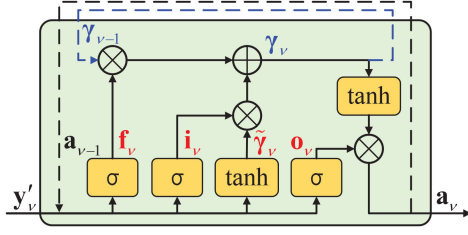


Fig. 5. The structure of an LSTM unit.

denotes the width in time axis of the convolution kernel, C_{in} represents the feature dimension of the input, and $\sigma_{m'}(\cdot)$ is the activation function applied in the 1D-CL. Notably, $w_{CK,v,u}^{(i)}$ and $b_{m',i,j}$ are the trainable parameters that can be learnt and adjusted through the neural network training. Besides, the rectified linear unit (ReLU) activation function is applied in all three 1D-CLs to enable complicated nonlinear mappings and mitigate vanishing gradient effects [25], which is expressed as

$$\text{ReLU}(\chi) = \max(0, \chi), \quad (6)$$

Because of the convolution operations, the width of $\mathbf{m}'$ will be smaller than that of $\mathbf{m}$ as long as $\alpha > 1$. Thus as shown in (5), one convolution kernel will exploit the whole feature dimension and compress the time dimension of the input, and then output a vector with smaller time dimension value and feature dimension value equaling to 1. It is worth mentioning that since one convolution kernel generates one vector along the time axis, we can compress the output feature dimension, i.e. the maximum value of i , by using fewer convolution kernels, or enlarge it by applying more convolution kernels.

2) *LSTM Unit*: As illustrated in Fig. 5, after obtaining the ν th external input vector $\mathbf{y}'_\nu$ and the $(\nu - 1)$ th output vector $\mathbf{a}_{\nu-1}$, the ν th output vector $\mathbf{a}_\nu$ of an LSTM unit is calculated as [23]

$$\mathbf{f}_\nu = \text{sigmoid}(\mathbf{W}_f \cdot [\mathbf{a}_{\nu-1}, \mathbf{y}'_\nu] + \mathbf{b}_f), \quad (7a)$$

$$\tilde{\gamma}_\nu = \tanh(\mathbf{W}_\gamma \cdot [\mathbf{a}_{\nu-1}, \mathbf{y}'_\nu] + \mathbf{b}_\gamma), \quad (7b)$$

$$\mathbf{i}_\nu = \text{sigmoid}(\mathbf{W}_i \cdot [\mathbf{a}_{\nu-1}, \mathbf{y}'_\nu] + \mathbf{b}_i), \quad (7c)$$

$$\gamma_\nu = \mathbf{f}_\nu \odot \gamma_{\nu-1} + \mathbf{i}_\nu \odot \tilde{\gamma}_\nu, \quad (7d)$$

$$\mathbf{o}_\nu = \text{sigmoid}(\mathbf{W}_o \cdot [\mathbf{a}_{\nu-1}, \mathbf{y}'_\nu] + \mathbf{b}_o), \quad (7e)$$

$$\mathbf{a}_\nu = \mathbf{o}_\nu \odot \tanh(\gamma_\nu), \quad (7f)$$

where $[\mathbf{a}_{\nu-1}, \mathbf{y}'_\nu]$ is the concatenation of $\mathbf{a}_{\nu-1}$ and $\mathbf{y}'_\nu$ and acts as the ν th input vector of the LSTM unit, and $\gamma_{\nu-1}$ denotes the memory cell that stores the useful information in previous operations. Meanwhile, $\mathbf{f}_\nu$, $\mathbf{i}_\nu$, $\mathbf{o}_\nu$ are the Forget Gate, Input Gate, Output Gate respectively and $\tilde{\gamma}_\nu$ represents the optional update value, all of which are generated from the ν th input vector and used to update the memory cell and generate the output vector in the current iteration. To be more explicit, $\mathbf{f}_\nu$ eliminates the unnecessary parts of $\gamma_{\nu-1}$ and retains the valuable parts, $\tilde{\gamma}_\nu$ provides the useful information extracted from the ν th input vector, and $\mathbf{i}_\nu$ controls the exploited parts in $\tilde{\gamma}_\nu$ to update $\gamma_{\nu-1}$. Then, as shown in (7d), the updated memory cell γ_ν in this iteration is calculated. After that, the ν th output vector $\mathbf{a}_\nu$ is generated from $\tanh(\gamma_\nu)$ with the aid of $\mathbf{o}_\nu$ that determines the

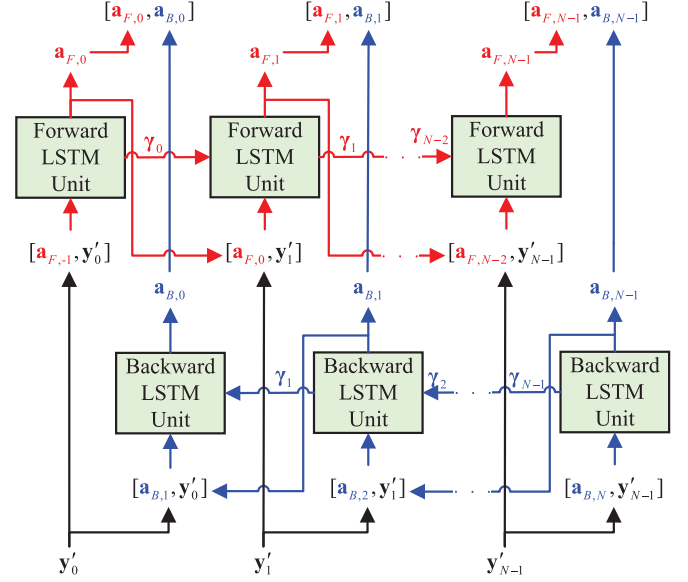


Fig. 6. The unfolded operations of the employed forward and backward LSTM units.

output part of $\tanh(\gamma_\nu)$. At last, $\mathbf{a}_\nu$ is also utilized to compose the next input vector and calculate the output vector in $(\nu + 1)$ th iteration.

Except for the forementioned symbols, $\mathbf{W}_f$, $\mathbf{b}_f$, $\mathbf{W}_\gamma$, $\mathbf{b}_\gamma$, $\mathbf{W}_i$, $\mathbf{b}_i$, $\mathbf{W}_o$, $\mathbf{b}_o$ are all trainable parameters that are learnt and adjusted through the neural network training, $\text{sigmoid}(\cdot)$ represents the sigmoid activation function where $\text{sigmoid}(\chi) = \frac{1}{1 + \exp(-\chi)}$, $\tanh(\cdot)$ means the tanh activation function where $\tanh(\chi) = \frac{\exp(\chi) - \exp(-\chi)}{\exp(\chi) + \exp(-\chi)}$, and $\odot$ denotes the element-wise product operation. To facilitate the illustration of the employed bi-directional LSTM layer, we visualize and unfold the recursive operations of the forward and backward LSTM units as Fig. 6. At each iteration, one external input vector is concatenated with the last output vector and acts as the input for the current iteration. Then, the output vector is generated with the aid of the useful information recorded in previous iterations and stored in the inherited memory cell. Meanwhile, the memory cell is also updated for the next iteration. Owing to the recursive working procedure and the memory cell, the LSTM unit can exploit the knowledge extracted in previous operations to better learn and capture the dynamic behavior and the correlations between the input vectors.

However, if only one LSTM unit is employed, less knowledge can be utilized by the input. To address this issue, we apply two LSTM units to process the external input vectors in different directions. The operations conducted by the forward and the backward LSTM units are the same as mentioned above, while the only difference between them is the processing order for the external input. As shown in Fig. 6, the first external input vector of the backward LSTM unit is $\mathbf{y}'_{N-1}$ and the last one is $\mathbf{y}'_0$, which is opposite to the forward LSTM unit. By this means, all the input vectors can utilize complete knowledge generated from the other external input vectors. Notably, both the $\mathbf{a}_{F,-1}$ and $\mathbf{a}_{B,N}$ are set to be 0 to compose the initial

input vectors for the forward and the backward LSTM units separately.

Besides, from Fig. 6, we can notice that the output feature vector of the bi-directional LSTM layer is not only dependent on the current external input vector, but also decided by the information recorded in different iterations and stored in the memory cells, which is in consistent with the characteristics of the proposed OFDM-DCSK system that the information-bearing chaotic sequences are correlated to each other owing to the chaotic modulation with the same reference chaotic sequence $\mathbf{x}$. Therefore, with the aid of the bi-directional LSTM layer, the correlation features between the recovered information-bearing chaotic sequences can be extracted to enhance the reliability performances.

3) *Fcl*: In the following two FCLs, the output vectors of the LSTM units will be further processed to implement the data bit estimation. Let β denote the index of the FCL, i.e., $\beta = 1, 2$, then for the input vector $\mathbf{z}_\beta$ of the β th FCL, the output vector $\mathbf{z}'_\beta$ is calculated as

$$\mathbf{z}'_\beta = \sigma_\beta(\mathbf{W}_\beta \cdot \mathbf{z}_\beta + \mathbf{b}_\beta), \quad (8)$$

where $\mathbf{W}_\beta$ and $\mathbf{b}_\beta$ are the trainable parameters of the β th FCL that will be learnt and adjusted through the neural network training, and $\sigma_\beta(\cdot)$ represents the activation function employed in the β th FCL.

To be more explicit, when $\beta = 1$, the input vector $\mathbf{z}_1 = [\mathbf{a}_{F,n}, \mathbf{a}_{B,n}]$, which is the concatenation of the output feature vectors of the forward and backward LSTM units, while the output vector $\mathbf{z}'_1 = \mathbf{a}'_n$. Besides, no activation function is applied in this FCL, i.e., $\sigma_1(\chi) = \chi$. Additionally, when $\beta = 2$, the input vector $\mathbf{z}_2 = \mathbf{p}_n$, where $\mathbf{p}_n$ is the standardized vector from the BN layer which will be introduced later, while the output vector $\mathbf{z}'_2 = \mathbf{p}'_n$. It is worth mentioning that the softmax activation function is applied in the 2nd FCL, which is expressed as

$$\text{softmax}(\chi_u) = \frac{\exp(\chi_u)}{\sum_{\chi_{u'} \in \chi} \exp(\chi_{u'})}, \quad (9)$$

where χ_u is the u th element of the vector χ . After processing by the softmax activation function, the value of each element of $\mathbf{p}'_n$ will be constrained between 0 and 1 while maintaining the sum as 1. In other words, the softmax activation function applied in the last FCL enables it to learn to calculate the probability of the data bit carried by the information-bearing sequences, then the generated probability vector $\mathbf{p}'_n$ is exploited to determine the estimates of the user data bits.

4) *BN Layer*: A BN layer is employed between the two FCLs to standardize the input vector for efficient neural network training. As shown in Fig. 3, let $\mathbf{a}'_n$ denote the input vector of the BN layer and $\mathbf{p}_n$ represent the output vector, then the standardization process is expressed as [26]

$$p_{n,\zeta} = \sigma_{BN} \left(\delta \times \frac{a'_{n,\zeta} - \text{mean}(a'_{n,\zeta})}{\sqrt{\text{var}(a'_{n,\zeta}) + \epsilon}} + \eta \right), \quad (10)$$

where $a'_{n,\zeta}$ and $p_{n,\zeta}$ are the ζ th elements of $\mathbf{a}'_n$ and $\mathbf{p}_n$, while $\text{mean}(a'_{n,\zeta})$ and $\text{var}(a'_{n,\zeta})$ represent the mean and variance of

the ζ th element of the input vector $\mathbf{a}'_n$ respectively, which are learnt and estimated during the training stage and then directly used at the deployment stage. In addition, $\epsilon = 10^{-5}$ is a small constant applied to avoid the denominator from becoming zero, $\sigma_{BN}(\cdot)$ is the activation function employed in the BN layer, which is set as the ReLU activation function (6) in this paper, and δ and η are the trainable scale and shift factors used for possible distribution recovery if needed.

After performing standardizations, each element of the input vector is normalized to the range with zero mean and unit variance, wherein the activation functions applied in the DNN, such as sigmoid, tanh and softmax, will have high gradients when updating the trainable parameters by the back propagation algorithm during the training. Therefore, the vanishing and exploding gradient problem can be alleviated and the convergence process can be accelerated. However, the standardizations will change the distributions of input vectors, leading to the loss of information. To address this issue, we apply the scale and shift factors to recover the original distributions, which can be adjusted during the training, thus the strength of standardizations can be controlled to achieve the trade off between convergence acceleration and information preservation.

C. Neural Network Training of DNN

At the offline training stage, the training data set provides the known input and output data for the DNN to learn and formulate the parameter configurations for the performance optimization. During the training, by using the back propagation algorithm, the DNN calculates the derivatives of the differences between the actual output and known knowledge with respect to trainable parameters which are then utilized to update the parameters by the selected optimization algorithm. Along with the iterative updating of parameters, the differences will decrease gradually and converge to a previously specified threshold.

Explicitly, in our design, the received chaotic sequences $\mathbf{y}_n = [c'_{n,0}, c'_{n,1}, \dots, c'_{n,K-1}]$, ($0 \leq n \leq N-1$) are applied as the training samples and the data bits $\mathbf{d} = [d_1, d_2, \dots, d_{N-1}]$ act as the known output. After the neural network training, we can formulate and establish the mapping from $\mathbf{y}_n$ to the estimates $\hat{d}_n$ with the aim of minimizing the differences between $\mathbf{d}$ and $\hat{\mathbf{d}} = [\hat{d}_0, \dots, \hat{d}_n, \dots, \hat{d}_{N-1}]$, which can be evaluated by the loss function. Here we generate the training data set and the test data set via simulations. Thanks to the statistical randomness of the noises and fading amplitudes, the data splitting and the data augmentation are not required to enrich the data set. More details about the data set generation used for simulations are discussed in Section V-B.

D. Hyper-Parameter Selection

The hyper-parameter selections have great impacts on the learning capability, the complexity, the convergence rate, etc., of the DNN. In this subsection, we select four typical hyper-parameters for discussions of the influence of the hyper-parameter selection on the performances.

1) *Output Dimension of Each Layer*: The output dimension of each layer for the employed DNN is presented in Table I,

TABLE I
OUTPUT DIMENSION OF EACH LAYER FOR THE PROPOSED DNN-AIDED DEMODULATOR

Layer	Output Dimension			
	$K = 8$	$K = 16$	$K = 32$	$K = 256$
Input	16	32	64	512
1st 1D-CL	4×15	4×31	4×63	4×511
2nd 1D-CL	8×12	8×28	8×60	8×508
3rd 1D-CL	16×5	16×21	16×53	16×501
F. LSTM	16	32	64	64
B. LSTM	16	32	64	64
1st FC	24	48	96	96
BN	24	48	96	96
2nd FC	2	2	2	2

TABLE II
NUMBER OF TRAINABLE PARAMETERS OF EACH LAYER FOR THE PROPOSED DNN-AIDED DEMODULATOR

Layer	Number of Trainable Parameters			
	$K = 8$	$K = 16$	$K = 32$	$K = 256$
Input	0	0	0	0
1st 1D-CL	12	12	12	12
2nd 1D-CL	136	136	136	136
3rd 1D-CL	1040	1040	1040	1040
F. LSTM	6272	47360	233984	2068992
B. LSTM	6272	47360	233984	2068992
1st FC	792	3120	12384	12384
BN	48	96	192	192
2nd FC	50	98	194	194

which are determined by the dimension of input vectors, the learning capacity, the complexity and the convergence rate of the DNN. When the dimension becomes larger, the learning capability can be enhanced and the DNN can learn more complicated mappings. However, the computational complexity of the DNN also increases, and the convergence time will be longer. With the aim to get a better tradeoff between the learning capability and the complexity, we train and test the proposed DNN with different sets of output dimensions, including different output feature dimensions C_{out} , convolution kernel widths α , etc., to select the set to achieve the performances approaching to the saturation with the lowest output dimensions.

2) *Number of Trainable Parameters*: The number of trainable parameters of each layer is highly related with the output dimension. Let I and O respectively represent the input and output dimension, then, when using the PyTorch DL framework, the number of trainable parameters can be calculated as $IO + O$ for the FCLs and $4(I + O)O + 8O$ for the LSTM units. On the other side, the parameter numbers of 1D-CLs are decided by the input and output feature dimensions and the width in the time axis of the convolution kernel. In this paper, as shown in Tables I and II, the output feature dimensions of the three 1D-CLs are $C_{out} = 4, 8, 16$, while the convolution kernel widths of them are $\alpha = 2, 4, 8$. Then, the number of trainable parameters of 1D-CL can be calculated as $C_{out}(C_{in}\alpha + 1)$. Besides, the BN layer needs $2O$ trainable parameters to learn the scale and shift factors for each element of the output vectors, and the parameter number of the input layer is zero since it is applied only to receive and group the input data.

3) E_b/N_0 : When applying DL methods to communication applications, the energy per bit to noise power spectral density ratio (E_b/N_0) of the simulated channel environment at the offline training stage is a key hyper-parameter, which can significantly affect the detection performances of the DNN. Excessively high

value of the training E_b/N_0 will induce the overfitting problem, while the too small value of E_b/N_0 will induce the underfitting problem [27]. To address this issue, we can apply the optimization algorithm, such as the simulated annealing optimization algorithm, to select the value of E_b/N_0 for the DNN training with the objective of maximizing the learning efficiency. For fairness of performance comparisons to be presented in the following section, the value of E_b/N_0 is uniformly set as 20 dB for the fast Rayleigh fading channel and 17 dB for the high-speed railway channel at the training stage.

4) *Loss Function and Optimization Algorithm Selection*: For the DNN, the loss function can quantitatively evaluate the difference between the actual DNN output and the known output, while the optimization algorithm can formulate the trainable parameter configurations with the aid of the gradients calculated by the back propagation algorithm. Obviously, appropriate loss function and optimization algorithm helps to improve the convergence rate and the accuracy of the DNN. In our design, considering that the chaotic demodulation process is equivalent to solve a classification problem, we propose to apply the categorical cross-entropy loss function, which can provide a reliable measure of the difference between two probability distributions [28], to act as the loss function and to solve this classification problem [29]. Besides, in order to efficiently adjust the trainable parameters in DNN according to the measured loss value and the corresponding derivatives, the Adam algorithm [30] is selected as the optimization algorithm due to the invariance to diagonal rescaling of the gradients, high computational efficiency and low memory requirements. At the training stage, we set the disturbance term and exponential decay rates for the Adam algorithm the same as those given in [30], and the batch size is set as 50 to mitigate the oscillation effect while maintaining satisfactory generalization capacity [31]. Additionally, with the objective to guarantee the learning efficiency without losing the convergence stability, we first set the learning rate of the Adam algorithm as 0.01, and then gradually lower the learning rate till 10^{-5} if the training loss does not decrease in an epoch [32].

IV. THEORETICAL ANALYSIS

In this section, we will analyze and compare the spectrum efficiency, energy efficiency, security performances and complexity of the proposed DL-based OFDM-DCSK system with the benchmark traditional OFDM-DCSK system [8], [9], the OFDM code shifted DCSK (OFDM-CS-DCSK) system [33] and the frequency-and-time hybrid-interleaving OFDM-DCSK (FH-OFDM-DCSK) system [15].

A. Spectrum Efficiency Analysis

The spectrum efficiency is defined as the ratio of the bit rate to the total bandwidth [34]. Let T_{OFDM} denote the duration of one OFDM symbol, and B_{OFDM} represent the occupied bandwidth of one OFDM symbol. Then, as presented in Section II, since one OFDM-DCSK symbol is composed of K modulated OFDM symbols, the duration of one OFDM-DCSK symbol is expressed as $T = KT_{OFDM}$. For the traditional OFDM-DCSK

TABLE III
SPECTRUM EFFICIENCY COMPARISONS

Scheme	Spectrum Efficiency
OFDM-DCSK [8]	$\frac{N-1}{TB_{OFDM}}$
OFDM-CS-DCSK [33]	$\frac{M}{TB_{OFDM}} (1 \leq M \leq N-1)$
FH-OFDM-DCSK [15]	$\frac{N-1}{TB_{OFDM}}$
DL-based OFDM-DCSK	$\frac{N}{TB_{OFDM}}$

TABLE IV
ENERGY EFFICIENCY COMPARISONS

Scheme	Energy Efficiency
DCSK [7]	$\frac{1}{N}$
OFDM-DCSK [8]	$\frac{N-1}{N}$
OFDM-CS-DCSK [33]	$\frac{M}{M+1} (1 \leq M \leq N-1)$
FH-OFDM-DCSK [15]	$\frac{N-1}{N}$
DL-based OFDM-DCSK	1

system and the FH-OFDM-DCSK system, $N-1$ bits will be transmitted by a modulated symbol, while our proposed system can transmit N bits in one symbol. Meanwhile, for the OFDM-CS-DCSK system, the interval of each subcarrier are set as the same as the forementioned three systems for comparison fairness, thus the occupied bandwidth will be KB_{OFDM} and the duration will be T_{OFDM} , and the number of transmitted bits will be M with $1 \leq M \leq N-1$.

Subsequently, we can evaluate the spectrum efficiency of the proposed system and benchmark systems. As shown in Table III, since no reference chaotic sequences are transmitted by any subcarrier, we can notice that the proposed DL-based OFDM-DCSK system achieves higher spectrum efficiency than the benchmark systems, especially when the number of exploited subcarriers N is small.

B. Energy Efficiency Analysis

Let E_b denote the transmitted energy for each bit, E_{data} and E_{ref} respectively represent the energy of the data and the reference signals. Then, the energy efficiency is defined as the data-energy-to-bit-energy-ratio (DBR) $\frac{E_{data}}{E_b}$ [15]. Table IV presents and compares the energy efficiency of the proposed design and benchmark schemes.

For the DCSK system, since one modulated symbol is the concatenation of one reference chaotic sequence and one information-bearing chaotic sequence, we have $E_{data} = E_{ref}$ and $E_b = E_{data} + E_{ref} = 2E_{data}$, thus the energy efficiency will be $\frac{1}{2}$. While in the traditional OFDM-DCSK system and the FH-OFDM-DCSK system, one reference sequence is shared by $N-1$ information-bearing sequences, thus $E_{ref} = \frac{1}{N-1}E_{data}$. Without loss of generality, we neglect the energy cost of CP, the energy efficiency is calculated as $\frac{E_{data}}{E_b} = \frac{E_{data}}{E_{data} + \frac{1}{N-1}E_{data}} = \frac{N-1}{N}$, which is higher than that of the DCSK system as presented in Section I. In the OFDM-CS-DCSK system, M information-bearing sequences share one reference chaotic signal, thus the energy efficiency will be $\frac{M}{M+1}$, where $1 \leq M \leq N-1$. Finally, for the proposed DL-based OFDM-DCSK system, since no reference sequences are needed to be transmitted, we have $E_b = E_{data}$ and the energy efficiency will be 1, which demonstrates

the superiority of the efficiency of our design, especially when N has small value in some vehicular Internet of Things (IoT).

C. Security Evaluation

Furthermore, we evaluate the security performances of the proposed intelligent transceiver in terms of the information leakage rate and the secrecy capacity. Without loss of generality, we assume that the probabilities of transmitting 0 and 1 are equal, then the information leakage Λ is calculated as [15]

$$\begin{aligned} \dot{I}_{E,n}(R_E, S) &= H_n(R_E) - H_n(R_E|S) \\ &= 1 + \rho_{E,n} \log_2(\rho_{E,n}) + (1 - \rho_{E,n}) \log_2(1 - \rho_{E,n}), \quad (11) \end{aligned}$$

$$\Lambda = \frac{1}{N} \sum_{n=0}^{N-1} \dot{I}_{E,n}(R_E, S), \quad (12)$$

where $\dot{I}_{E,n}$ denotes the mutual information between the transmitted data S and the data retrieved by the eavesdroppers R_E at the n th subcarrier, $H_n(\cdot)$ represents the entropy operation, and $\rho_{E,n}$ denotes the BER at the n th subcarrier for the eavesdropping receivers.

Based on the information leakage rate given by (12), the secrecy capacity of legitimate users can be derived as

$$\begin{aligned} C_{secrecy} &= \frac{1}{N} \sum_{n=0}^{N-1} \dot{I}_{L,n}(R_L, S) - \Lambda \\ &= H_n(R_L) - H_n(R_L|S) - \Lambda \\ &= 1 + \rho_{L,n} \log_2(\rho_{L,n}) \\ &\quad + (1 - \rho_{L,n}) \log_2(1 - \rho_{L,n}) - \Lambda, \quad (13) \end{aligned}$$

where $\dot{I}_{L,n}$ denotes the mutual information between the transmitted data S and the data retrieved by the legitimate users R_L at the n th subcarrier, and $\rho_{L,n}$ indicates the BER at the n th subcarrier for the legitimate receivers.

Notably, if the potential eavesdroppers have known the key features of signals such as the number of subcarriers, the symbol duration and the symbol starting points, they can easily retrieve the user data from the traditional OFDM-DCSK modulated symbols with the aid of transmitted reference chaotic sequences. On the other side, since no reference chaotic sequences are directly transmitted in our design, the eavesdroppers can not utilize them to retrieve the user data. Hence it becomes difficult for the eavesdroppers to crack the chaotic modulated signals through brute force methods thanks to the non-periodic, noise-like and initial value sensitive characteristics of chaotic sequences. As a result, the BER of the eavesdropping receivers will be high in the proposed system, and lower information leakage and higher secrecy capacity can be guaranteed by applying our design, which will be demonstrated in the following Section V.

D. Complexity Analysis

Finally, we analyze the computational complexity of the DL-based OFDM-DCSK system. In this paper, for fairness of

TABLE V
 COMPUTATIONAL COMPLEXITY COMPARISONS

Scheme	Computational Complexity
OFDM-DCSK [8]	$O(NK \log_2 N)$
OFDM-CS-DCSK [33]	$O(N^2 K) + O(NK \log_2(NK))$
FH-OFDM-DCSK [15]	$O(NK \log_2 N)$
DL-based OFDM-DCSK	$O(\sum_{\theta=1}^3 NK C_{out,\theta-1} C_{out,\theta} \alpha_{\theta})$ $+ O(NK C_{out,3} L_{LSTM}) + O(N L_{LSTM}^2)$ $+ O(N L_{LSTM} L_{FC,1}) + O(NK \log_2 N)$

comparisons, we apply the same complexity evaluation method as that used in [15], [20]. Explicitly, the computational complexity is defined as the total number of operations per symbol conducted at both transmitters and receivers [15], [20], which is expressed in terms of the asymptotic upper bound $O(\cdot)$. Similar to the computational complexity analysis presented in [20] and [35], considering that the DNN training is carried out offline, we only evaluate the complexity at the online deployment stage at which the proposed design is applied for the real time detection of signals.

Let $C_{out,\theta}$ and α_{θ} respectively denote the output feature dimension and the convolution kernel width of the θ th 1D-CL, where $\theta = 1, 2, 3$ and $C_{out,0} = 1$, L_{LSTM} represent the output dimension of the LSTM unit, and $L_{FC,1}$ represent the output dimension of the first FC layer, then we evaluate the computational complexity of the proposed DL-based OFDM-DCSK system and the benchmarks as provided in Table V.

For the OFDM systems, the computational complexity of IFFT or FFT is $O(N \log_2 N)$, and K times IFFT operations are needed to generate an OFDM-DCSK symbol. For the proposed transmitter, to generate an OFDM-DCSK symbol, the computational complexity of IFFT or FFT is at the order of $O(N \log_2 N)$, and K times IFFT operations are needed. At the receiver, $N \times K$ multiplications are carried out to perform the channel equalization, and K times FFT operations are required to recover the chaotic modulated sequences from the received signals. Therefore, the complexity of modulation, equalization and information-bearing chaotic sequences recovery of the proposed DL-based OFDM-DCSK system is $O(2NK \log_2 N) + O(NK) = O(NK \log_2 N)$. Next, the chaotic demodulation is conducted to recover the user data from information-bearing chaotic modulated sequences without using reference sequences.

After recovering the chaotic modulated sequences, they are input to the proposed DNN for chaotic demodulation. The number of operations of the θ th 1D-CL has the complexity of $O(NK C_{out,\theta-1} C_{out,\theta} \alpha_{\theta})$, the complexity of the LSTM unit is $O(NK C_{out,3} L_{LSTM}) + O(N L_{LSTM}^2)$, the complexity of the first FCL is $O(N L_{LSTM} L_{FC,1})$, the complexity of the BN layer is $O(N L_{FC,1})$, and the complexity of the second FCL is $O(N L_{FC,1})$. Summing the computational complexity of modulation, equalization, information-bearing chaotic sequences recovery and DNN operations, we can obtain the complexity of the proposed system as $O(\sum_{\theta=1}^3 NK C_{out,\theta-1} C_{out,\theta} \alpha_{\theta}) + O(NK C_{out,3} L_{LSTM}) + O(N L_{LSTM}^2) + O(N L_{LSTM} L_{FC,1}) + O(NK \log_2 N)$.

Notably, at the online deployment stage, the dimension of each layer remains invariant. Thus we can see that the complexity of the proposed design is at the same level as that of the traditional OFDM-DCSK system and the FH-OFDM-DCSK system.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, simulation results are provided to demonstrate superior reliability and security performances of the proposed intelligent OFDM-DCSK transceiver over vehicular channels including fast fading channels and railway channels. Moreover, the robustness of our design is also analyzed.

A. Simulation Settings

Considering that vehicles have high mobility, in the simulations, the single-path fast Rayleigh fading channel, the multi-path fast Rayleigh fading channel and the high-speed railway channel are taken into account of simulated performance analysis.

To be explicit, for the single-path fast Rayleigh fading channel, the received signal $\mathbf{r}$ is expressed by $\mathbf{r} = \mathbf{g} \odot \mathbf{s} + \boldsymbol{\nu}$, where $\mathbf{s}$ denotes the transmitted signal, $\mathbf{g}$ represents the channel coefficients, $\boldsymbol{\nu}$ denotes the additive noise and $\odot$ means the element-wise product operation. In this channel model, $\mathbf{g} = \psi_1 + j\psi_2$, where ψ_1 and ψ_2 are independent and identically distributed Gaussian random vectors having zero mean and unit variance, thus the normalized amplitude $|\mathbf{g}| = \sqrt{\frac{\psi_1^2 + \psi_2^2}{2}}$ follows the Rayleigh distribution. Note that considering the mobility in high-speed vehicular communications, the channel coefficient is assumed to change at each time slot to reflect the fast fading property [4]. Besides, each channel coefficient in $\mathbf{g}$ is also assumed to be mutually independent because of large spatial and velocity variations in vehicular information transmissions. Thus the single coefficient exponential correlation matrix is an identity matrix and the correlation coefficient equals zero [36].

Next, for the multi-path fast Rayleigh fading channel, the received signal is modelled as $r_{\xi} = \sum_{\omega=1}^G g_{\xi,\omega} s_{\xi-\omega\tau} + \nu_{\xi}$, $\xi = 1, 2, \dots, N \times K$, where r_{ξ} and $s_{\xi-\omega\tau}$ respectively denote the ξ th and $(\xi - \omega\tau)$ th element of $\mathbf{r}$ and $\mathbf{s}$, $g_{\xi,\omega}$ represents the Rayleigh channel coefficient of the ω th path when causing r_{ξ} , ν_{ξ} is the additive noise, G indicates the path number and τ means the length of time delay. Besides, each channel path has the average power of $E\{(g_{\xi,\omega}^2)\} = \exp(1 - \omega)E\{(g_{\xi,1}^2)\}$, $\omega = 1, 2, \dots, G$ [15], where the summed average power of each path equals 1, i.e., $\sum_{\omega=1}^G E\{(g_{\xi,\omega}^2)\} = 1$.

Subsequently, for the high-speed railway channel, existing research works have demonstrated that the Rician distribution can better describe the channel conditions experienced by users. In this case, g is composed of a constant component caused by line-of-sight propagation and a Rayleigh component caused by non-line-of-sight propagation, which is defined as $\mathbf{g} = \sqrt{\frac{K_{Rician}}{K_{Rician} + 1}} + \sqrt{\frac{1}{K_{Rician} + 1}}(\psi_1 + j\psi_2)$, where the normalized amplitude $|\mathbf{g}|$ follows the Rician distribution. Besides, K_{Rician} represents the Rician K -factor, which can be modeled

as a double-slope linear function as [5]

$$K_{Rician}(dist) = \begin{cases} \varsigma_1 \times dist + \varsigma_2 \times dist, & dist \leq dist_{BP} \\ \varsigma_3 \times dist + \varsigma_4 \times dist, & dist > dist_{BP}, \end{cases} \quad (14)$$

where $dist$ and $dist_{BP}$ denote the transmitter to receiver distance and the break point distance respectively. In this paper, the suburban high-speed railway scenario is applied, thus $\varsigma_1 = -0.027\text{dB}$, $\varsigma_2 = 8.48\text{dB}$, $\varsigma_3 = -0.0023\text{dB}$, $\varsigma_4 = 4.024\text{dB}$, $dist_{BP} = 200\text{m}$, and the mean value of statistical Rician K -factor $K_{Rician} = 2.83\text{dB}$ [5] is adopted.

Moreover, in the following simulations, the length of CP is long enough to eliminate the ISI, while both perfect channel state information (CSI) and imperfect CSI are considered for the performance analysis of the reliability, the security and the robustness. Besides, in this paper, we use Python 3.6.4 as the programming software and apply PyTorch 1.0.0 as the DL framework, and the Intel Core i5-7300HQ is used for computations.

B. Data Set Discussion

As mentioned above, the data set is generated via simulations. For fairness of comparisons, different lengths of the chaotic sequences K , numbers of subcarriers N , numbers of paths G and lengths of time delay τ between each two adjacent paths over multi-path channels are applied to generate the data. Besides, typical vehicular channel conditions, such as the fast fading channels and railway channels, are considered to compose the data set.

To be explicit, with consideration of the mobility of vehicles, we respectively construct the single-path fast Rayleigh fading channel, the multi-path fast Rayleigh fading channel and the high-speed railway channel, wherein the fading coefficients are assumed to vary at each time slot. Besides, different values of G and τ and different channel conditions of the multi-path Rayleigh fading channel are considered. Moreover, both tropospheric and ionospheric signal propagation effects, as well as the urban environment, are also considered in the Rayleigh fading channels [37], [4], and the multi-path transmission, which typically exists in the coordinated multi-point transmission/reception scenario, is also simulated in the multi-path Rayleigh fading channel [38]. Furthermore, a fast Rician fading channel with the Rician factor $K_{Rician} = 2.83\text{ dB}$ is applied to model the fading behavior in the suburban high-speed railway scenario [5], [6]. In addition, in order to analyze the performances, different modulation parameters, i.e., K and N are applied to compose the different data set for performance comparisons.

Then we set the system parameters to generate information bits randomly. Explicitly, 50 thousand received signals are generated via simulations as the training set of one epoch, and 50 training epochs in total are used to implement the offline training of the DNN. At the online deployment stage, similar procedure is conducted to evaluate the performances for the proposed intelligent transceiver with the configured DNN based demodulator, and to compare with the benchmark schemes.

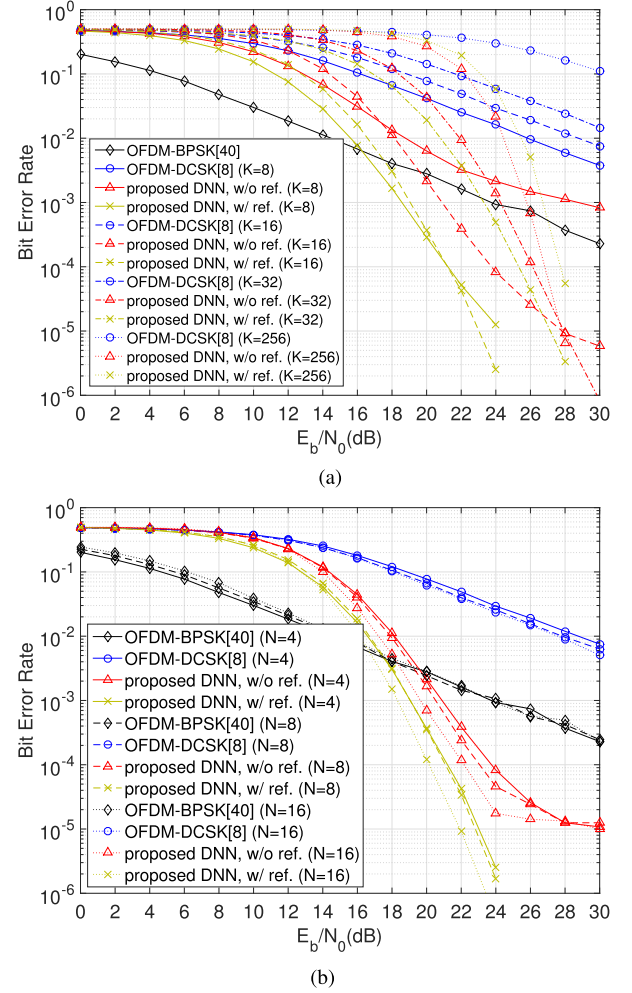


Fig. 7. BER performance comparisons over single-path fast Rayleigh fading channels with different values of K and N . (a) Single-path fast Rayleigh fading, $N = 4$. (b) Single-path fast Rayleigh fading, $K = 16$.

C. BER Performance Comparisons

1) *BER Performances Over Different Channels:* Figures 7 - 9 illustrate and compare the BER performances under different channel conditions and modulation parameters over single-path and multipath fast Rayleigh fading channels and high-speed railway channels. In the comparisons, we consider two cases that the transmitter delivers or does not deliver the reference chaotic signals, which are respectively labelled as “w/ ref.” and “w/o ref.”. Besides, we also compare the performances of the benchmark traditional OFDM-DCSK system [8] and the OFDM-BPSK system [39] with no chaotic interferences induced.

It can be observed from Fig. 7 and 9 that the proposed intelligent transceiver outperforms the OFDM-DCSK system [8] no matter where the reference chaotic signal is delivered or not over fast fading channels with different values of K and N . Besides, we can notice that at higher E_b/N_0 , the BER performances of the proposed system can even be better than those of the OFDM-BPSK system [39] thanks to the intelligently optimized DNN based demodulator. Moreover, it is also noticeable that

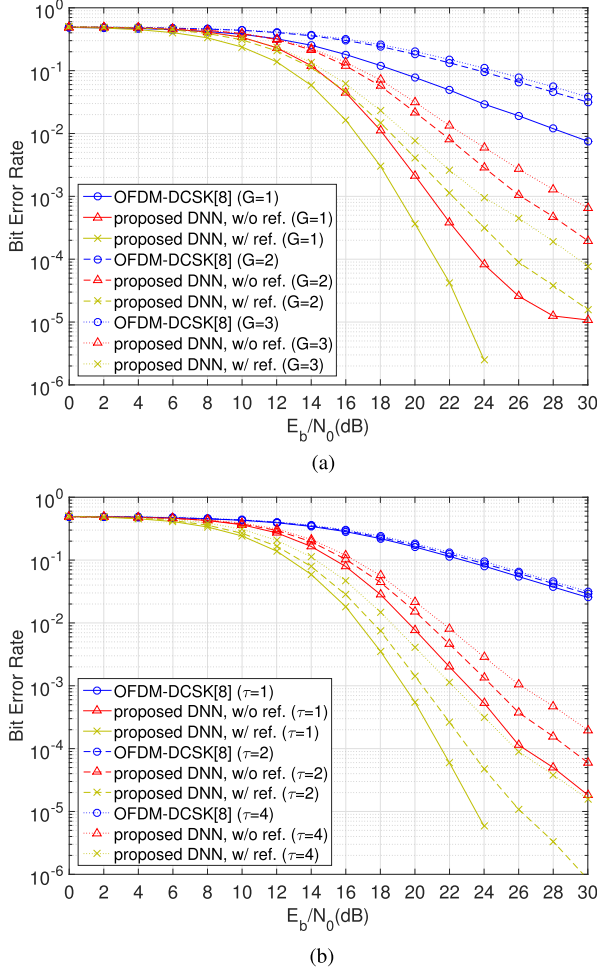


Fig. 8. BER performance comparisons over multi-path fast Rayleigh fading channels with different path numbers and time delays. (a) Multi-path fast Rayleigh fading, $N = 4$, $K = 16$, $\tau = 4$. (b) Multi-path fast Rayleigh fading, $N = 4$, $K = 16$, $G = 2$.

when the reference chaotic sequence is directly delivered to the receiver, thanks to the additional information input to the DNN, the proposed intelligent chaotic transceiver can achieve better BER and higher reliability performances for small to medium values of K , although the security performances are deteriorated. However, when K is large enough, we can notice that the BER performances of the system with no reference signals become better than those of the system delivering reference signals. The reasons are that on the one hand, the DNN can learn and extract sufficient features of mapping pattern for generating the chaotic sequence to mitigate the absence of the knowledge about the reference signal, on the other hand, the interferences from the reference signals are removed. Furthermore, it can be seen that for larger value of N , since more exact features can be extracted from signals received from larger number of subcarriers, lower BER can be attained. By contrast, for the OFDM-BPSK system, when N becomes larger, inter-carrier interferences increase accordingly, thereby leading to higher BER, especially at lower value of E_b/N_0 .

In addition, as illustrated in Fig. 8, we also investigate the BER performances of the proposed system with different G and

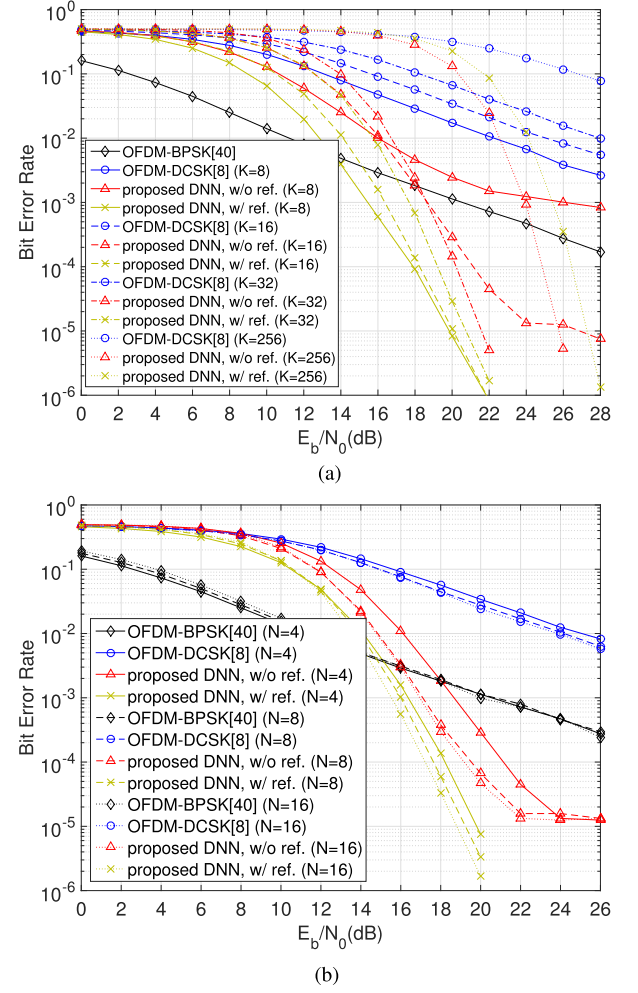


Fig. 9. BER performance comparisons over high-speed railway channels with different values of K and N . (a) High-speed railway channel, $N = 4$. (b) High-speed railway channel, $K = 16$.

τ over multi-path fast Rayleigh fading channels. It can be seen that the proposed DL-based systems outperform the traditional OFDM-DCSK system. Meanwhile, we can observe that when G or τ becomes larger, the BER performances degrade accordingly.

Notably, the noise floor might appear at high E_b/N_0 due to the removal of the reference chaotic signal. To address this issue, a promising method is enlarging the dimension or number of 1D-CLs to improve the feature extraction capability of DNN, while another choice is selecting larger K . As shown in Figs. 7(a) and 9(a), as K increases, the noise floor of the proposed system becomes lower and almost disappears.

2) *Performance Comparisons With the Benchmarks:* Subsequently, in Fig. 10, we compare the BER performances of the proposed design with those of the benchmark improved DCSK-based systems, including the OFDM-CS-DCSK system [33] and the FH-OFDM-DCSK system [15], over single-path and multipath fast Rayleigh fading channels and high-speed railway channels. From Fig. 10(a) to (c), we can notice that when transmitting signals over fast fading channels, the proposed DNN-aided demodulator can obtain better BER performances compared with the benchmark schemes, regardless of whether

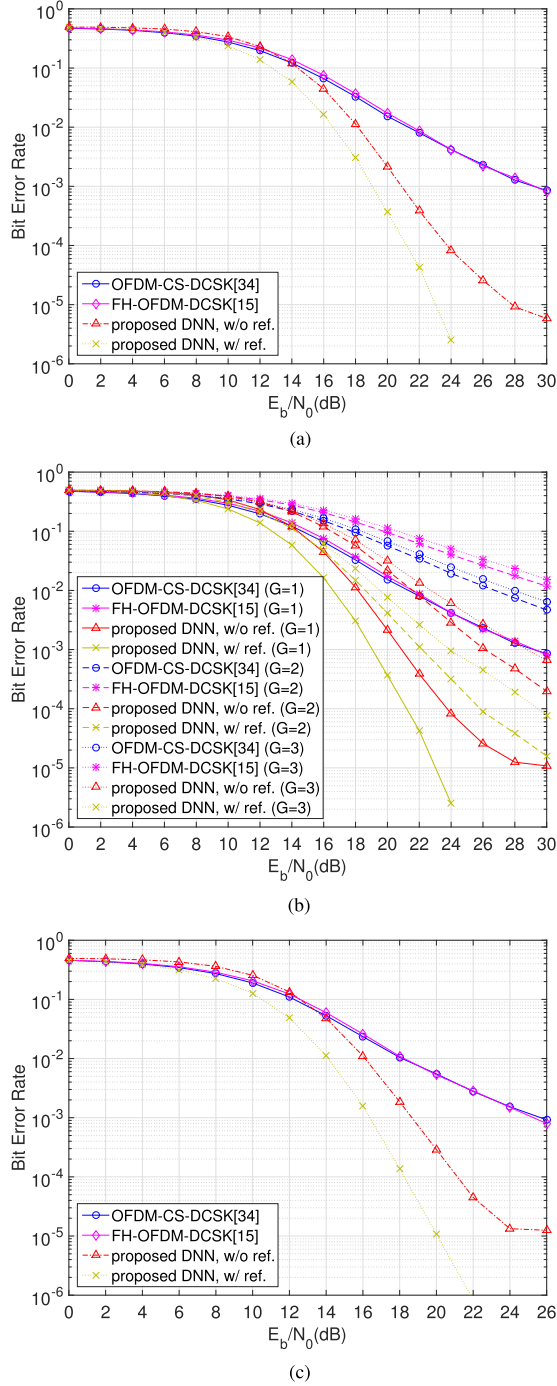


Fig. 10. BER performance comparisons with the benchmark modified systems over different channels. (a) Single-path fast Rayleigh fading, $N = 4$, $K = 16$. (b) Multi-path fast Rayleigh fading, $N = 4$, $K = 16$, $\tau = 4$. (c) High-speed railway channel, $N = 4$, $K = 16$.

the reference chaotic sequence is delivered or not. The reason is that the intelligent demodulator design has the powerful learning and feature extraction capabilities, thus exhibiting higher adaptability to high speed transportation and providing more reliable transmission performances.

It is worth mentioning that the diversity gain achieved by the FH-OFDM-DCSK system become negligible for users with high mobility, since the channel conditions vary so fast

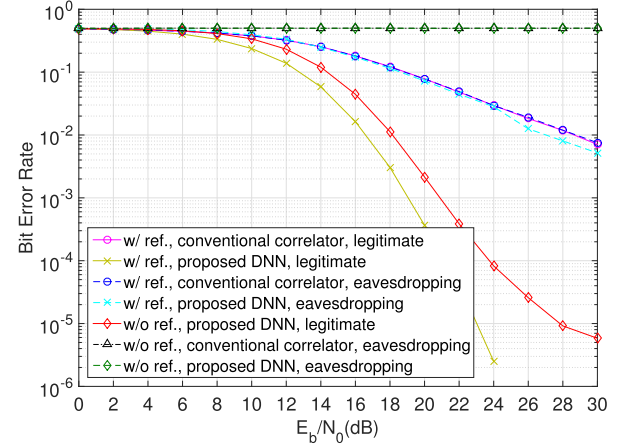


Fig. 11. BER performance comparisons of the legitimate users and the eavesdroppers in the OFDM-DCSK system with or without transmitting reference sequences, where $K = 16$ and $N = 4$.

that the transceiver can hardly attain the diversity. Hence we here propose the DNN-aided demodulator to improve the performances.

D. Security Performance Comparisons

1) *Security Performances in the OFDM-DCSK System:* Firstly, we investigate the BER performances of the legitimate users and the eavesdroppers in the OFDM-DCSK system with or without transmitting reference sequences. Here we assume that the key signal features such as the number of subcarriers, symbol duration and symbol starting points have been known by the potential eavesdroppers.

As shown in Fig. 11, for the traditional OFDM-DCSK system transmitting reference sequences, the eavesdroppers can obtain similar BER to the legitimate users when both of them use conventional correlators [8] for demodulation. The reason is that due to broadcasting property of vehicular channels, the eavesdroppers might retrieve the reference sequences and then obtain the user data. By contrast, when applying the proposed DNN for chaotic demodulation, the legitimate users can achieve reliable information recovery with low BER, which is much lower than the BER of the eavesdroppers using the same demodulation method. This is because the legitimate users can train the DNN with the correct training set, while the eavesdroppers have to first estimate the bits from the received signals and then utilize them as training set, which will cause accumulated error. Moreover, if no reference sequences are delivered, the eavesdroppers can hardly retrieve the user information even they have known the key signal features, because of the non-periodic, noise-like and initial value sensitive characteristics of chaotic sequences and the difficulty in cracking the chaotic modulated signals without the knowledge of the reference sequences. Note that the legitimate users can still achieve satisfactory BER performances.

Next we evaluate the secrecy capacity of legitimate users with (12) and (13). From Fig. 12(a), we can observe that the information leakage in the traditional OFDM-DCSK system gradually increases as the E_b/N_0 becomes higher and is close

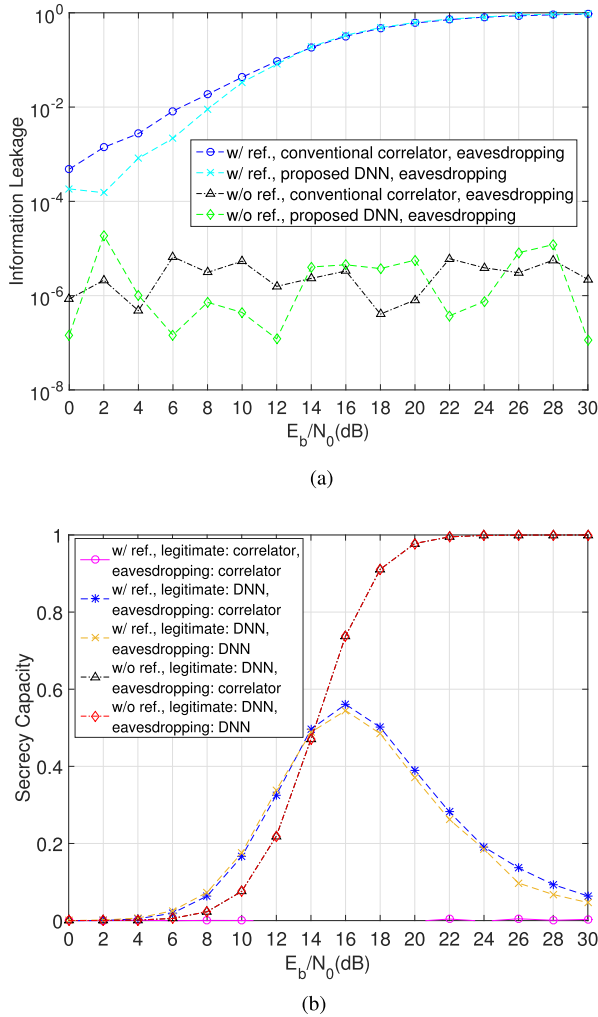


Fig. 12. Security performance comparisons in the OFDM-DCSK system with or without transmitting reference sequences, where $K = 16$ and $N = 4$. (a) Information leakage comparisons. (b) Secrecy capacity comparisons.

to 1 at last, which means the potential eavesdroppers can retrieve the user information at high E_b/N_0 with the aid of transmitted reference sequences. By contrast, for the proposed DL-based OFDM-DCSK system, the information leakage always maintains at a low level, which demonstrates that the eavesdroppers can hardly obtain the transmitted information since they have to directly crack the information-bearing chaotic sequences, thus the secure communication services can be provided.

To elaborate a bit further, as shown in Fig. 12(b), for the system transmitting reference sequences, when the legitimate users use the proposed DNN-aided demodulator, they can achieve relatively high secrecy capacity at low and medium E_b/N_0 thanks to better BER performances, while the secrecy capacity still decreases after 16 dB since the information leakage become aggravated at high E_b/N_0 . By contrast, if no reference sequences are delivered, the secrecy capacity of legitimate users can still maintain a high level at high E_b/N_0 . Finally, for the traditional OFDM-DCSK system, the secrecy capacity of legitimate users is always close to zero. Therefore,

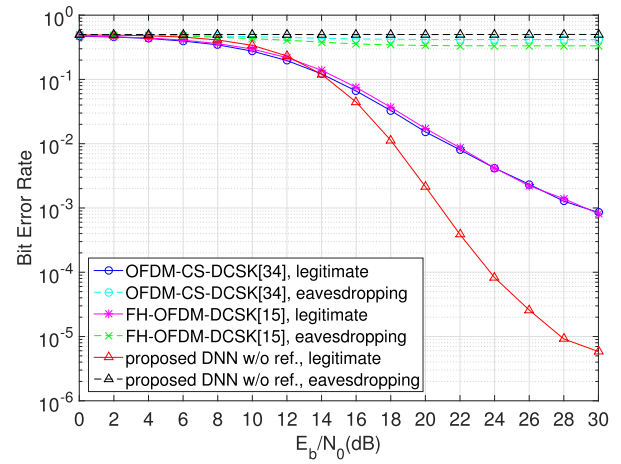


Fig. 13. BER performance comparisons of the legitimate users and the eavesdroppers in the proposed and the benchmark systems with $K = 16$ and $N = 4$.

the proposed DL-based OFDM-DCSK system can effectively increase the secured information rate and improve the security performances.

2) Security Performance Comparisons With the Benchmarks:

Next we compare the security performances with the benchmark OFDM-CS-DCSK [33] and FH-OFDM-DCSK [15] systems. For fairness of comparisons, we assume that the transmitter details can be learned by all users in considered systems. Explicitly, we assume that eavesdroppers can occasionally learn the Walsh code sequences utilized in OFDM-CS-DCSK system, and the frequency hopping pattern applied in the FH-OFDM-DCSK system. It can be observed from Fig. 13 that the proposed intelligent DNN aided system can provide better security performances than benchmark schemes.

Then, we evaluate and compare the secrecy capacity of the proposed system and the two benchmark systems. As shown in Fig. 14(a), the information leakage rate of the benchmark systems gradually increases to a certain value as the E_b/N_0 becomes higher since eavesdroppers can learn the key information. By contrast, the proposed design can keep the information leakage at lower rate. Besides, it can be observed from Fig. 14(b) that the proposed system can provide higher secrecy capacity at medium and high E_b/N_0 .

E. Robustness Analysis

In the simulations presented above, the CSI at both the offline training stage and the online deployment stage is assumed to be perfect. However, in practical systems, the CSI might be imperfect due to imperfect operating conditions of functional modules, which can be modeled as [40], [41]

$$\hat{g} = \varphi g + \sqrt{1 - \varphi^2} \varepsilon, \quad (15)$$

where $\hat{g}$ represents the estimated channel coefficient, g denotes the actual channel coefficient, φ indicates the correlation coefficient ranging from 0 to 1, which determines

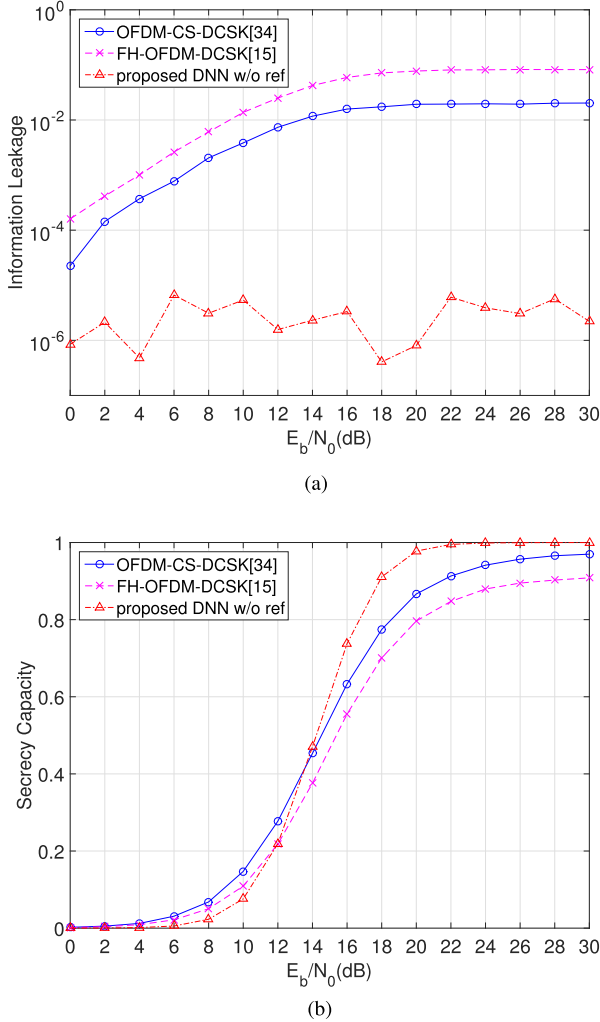


Fig. 14. Security performance comparisons of the proposed and the benchmark systems with $K = 16$ and $N = 4$. (a) Information leakage comparisons. (b) Secrecy capacity comparisons.

the accuracy of channel estimations, and ε is a complex Gaussian random variable with zero mean and unit variance.

Namely, when $\varphi = 1.0$, the perfect channel estimation and the exact CSI can be provided for the receivers. In this case, the channel conditions used at the training stage perfectly match the conditions applied for practical deployment. When $\varphi < 1.0$, the channel estimation is imperfect and will lead to the divergence between training and deployment conditions. Obviously, for smaller φ , the forementioned gap will become wider, leading to performance degradations.

It can be observed from Fig. 15 that the proposed DL-based OFDM-DCSK system can always maintain better BER performances than the traditional OFDM-DCSK system whatever the value of φ is. However, as φ decreases, the performance gain achieved by the DL-based systems also declines. The reason is that since the transmission characteristics learnt at the training stage do not match the transmission conditions at the deployment stage, the performance of the DNN-based

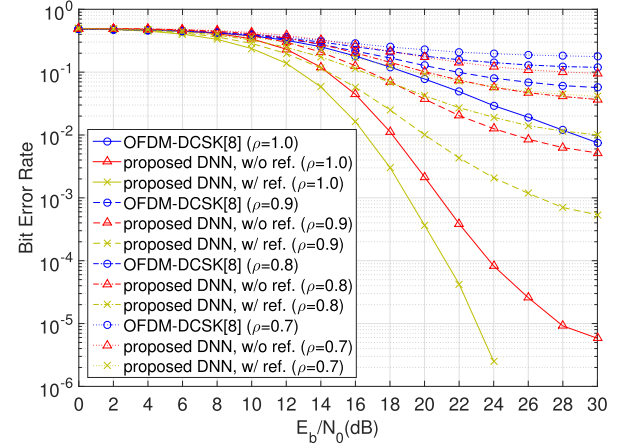


Fig. 15. BER performance comparisons over single-path fast Rayleigh fading channel with imperfect CSI, where $K = 16$ and $N = 4$.

receiver will degrade and the features can not be correctly extracted.

Last but not the least, it is worth mentioning that in practical systems, we usually have $\varphi > 0.95$ to guarantee the normalized mean-square error $E(\frac{\|\hat{g}-g\|^2}{\|g\|^2})$ be lower than 0.1 even when the signal to noise ratio is 0 dB [42], [43]. Hence, the robustness of the proposed design can meet practical user demands in vehicular communications.

VI. CONCLUSION

In this paper, considering that vehicular communications suffer from weak security due to the broadcasting property of vehicular channels, and interferences induced by mobility of users and fast fading of channels, we propose a reliable and secure intelligent chaotic receiver. In this design, we propose to remove the delivery of the reference chaotic signals at the transmitter to improve both the efficiency and the security performances. Then we propose a DNN to construct an intelligent demodulator to enhance the reliability performances. In the proposed DNN, we apply a TDNN composed of three 1D-CLs to learn and extract the features of mapping pattern for chaotic sequence generation, and then we employ a bi-directional LSTM layer to capture and formulate the correlation characteristics between chaotic modulated sequences, followed by two FCLs with one BN layer inserted between them to attain the data estimates. After the offline training, the proposed DL-based communication system is configured with the optimized trainable parameters. Then at the online deployment stage, the optimized transceiver can provide reliable and secure information transmissions for vehicular users. Simulation results under different conditions demonstrate that the proposed DL-based OFDM-DCSK system achieves better BER and security performances than the benchmark traditional OFDM-DCSK, OFDM-CS-DCSK and FH-OFDM-DCSK systems over vehicular channels such as high-speed railway channels. Therefore, the proposed intelligent design can provide more reliable, secure and efficient services for vehicular user equipments.

REFERENCES

- [1] H. Ma, G. Cai, Y. Fang, J. Wen, P. Chen, and S. Akhtar, "A new enhanced energy-detector-based FM-DCSK UWB system for tactile internet," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 3028–3039, May 2019.
- [2] M. Miao, L. Wang, G. Chen, and W. Xu, "Design and analysis of replica piecewise m-ary DCSK scheme for power line communications with asynchronous impulsive noise," *IEEE Trans. Circuits Syst. I*, vol. 67, no. 12, pp. 5443–5453, Dec. 2020.
- [3] K. Guan *et al.*, "Channel characterization and capacity analysis for THz communication enabled smart rail mobility," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4065–4080, May 2021.
- [4] L. Dong, G. Xu, and H. Ling, "Predictive downlink beamforming for wideband CDMA over rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 410–421, Mar. 2005.
- [5] T. Zhou, C. Tao, S. Salous, and L. Liu, "Measurements and analysis of short-term fading behavior in high-speed railway communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 101–112, Jan. 2019.
- [6] R. He, Z. Zhong, B. Ai, J. Ding, Y. Yang, and A. F. Molisch, "Short-term fading behavior in high-speed railway cutting scenario: Measurements, analysis, and statistical models," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 2209–2222, Apr. 2013.
- [7] G. Vizvari *et al.*, "Differential chaos shift keying: A robust coding for chaos communication," in *Proc. Nonlinear Dyn. Electron. Syst.*, 1996, pp. 87–92.
- [8] G. Kaddoum, "Design and performance analysis of a multiuser OFDM based differential chaos shift keying communication system," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 249–260, Jan. 2016.
- [9] S. Li, Y. Zhao, and Z. Wu, "Design and analysis of an OFDM-based differential chaos shift keying communication system," *J. Commun.*, vol. 10, no. 3, pp. 199–205, Mar. 2015.
- [10] M. Herceg, G. Kaddoum, D. Vranješ, and E. Soujeri, "Permutation index DCSK modulation technique for secure multiuser high-data-rate communication systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 2997–3011, Apr. 2018.
- [11] F. C. M. Lau, K. Y. Cheong, and C. K. Tse, "Permutation-based DCSK and multiple-access DCSK systems," *IEEE Trans. Circuits Syst. I*, vol. 50, no. 6, pp. 733–742, Jul. 2003.
- [12] G. Kaddoum, F. Gagnon, and F. Richardson, "Design of a secure multi-carrier DCSK system," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2012, pp. 964–968.
- [13] X. Cai, L. Hu, W. Xu, and L. Wang, "Design of an OFDM-based differential cyclic-shifted DCSK system for underwater acoustic communications," in *Proc. IEEE Asia-Pacific Conf. Commun. (APCC)*, 2021, pp. 304–309.
- [14] X. Cai, W. Xu, S. Hong, and L. Wang, "A trinal-code shifted differential chaos shift keying system," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 1000–1004, Mar. 2021.
- [15] Z. Liu, L. Zhang, Z. Wu, and J. Bian, "A secure and robust frequency and time diversity aided OFDM-DCSK modulation system not requiring channel state information," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1684–1697, Mar. 2011.
- [16] L. Zhang, Z. Chen, W. Rao, and Z. Wu, "Efficient and secure non-coherent OFDM-based overlapped chaotic chip position shift keying system: Design and performance analysis," *IEEE Trans. Circuits Syst. I*, vol. 67, no. 1, pp. 309–321, Jan. 2020.
- [17] Q. Bai, J. Wang, Y. Zhang, and J. Song, "Deep learning-based channel estimation algorithm over time selective fading channels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 125–134, Mar. 2020.
- [18] Y. Wang, J. Wang, W. Zhang, J. Yang, and G. Gui, "Deep learning-based cooperative automatic modulation classification method for MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4575–4579, Apr. 2020.
- [19] H. Liu, Y. Liu, M. Yang, and X. Li, "A joint demodulation and estimation algorithm for plasma sheath channel: Extract principal curves with deep learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 433–437, Apr. 2020.
- [20] H. Ye, L. Liang, G. Y. Li, and B. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3133–3143, May 2020.
- [21] L. Zhang, H. Zhang, Y. Jiang, and Z. Wu, "Intelligent and reliable deep learning LSTM neural networks-based OFDM-DCSK demodulation design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16 163–16 167, Dec. 2020.
- [22] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 328–339, Mar. 1989.
- [23] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017.
- [24] G. Kaddoum, P. Charge, D. Roviras, and D. Fournier-Prunaret, "A methodology for bit error rate prediction in chaos-based communication systems," *Circuits, Syst. Signal Process.*, vol. 28, no. 6, pp. 925–944, Aug. 2009.
- [25] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *J. Mach. Learn. Res.*, vol. 15, pp. 315–323, Jan. 2011.
- [26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [27] H. Zhang, L. Zhang, and Y. Jiang, "Overfitting and underfitting analysis for deep learning based end-to-end communication systems," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2019, pp. 1–6.
- [28] A. Sangari and W. Sethares, "Convergence analysis of two loss functions in soft-max regression," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1280–1288, Mar. 2016.
- [29] X. Li, L. Yu, D. Chang, Z. Ma, and J. Cao, "Dual cross-entropy loss for small-sample fine-grained vehicle classification," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4204–4212, May 2019.
- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [31] Y. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer, pp. 9–50, 1998. [Online]. Available: https://doi.org/10.1007/3-540-49430-8_2
- [32] B. Shi, W. J. Su, and M. I. Jordan, "On learning rates and Schrödinger operators," 2020, *arXiv:2004.06977*.
- [33] M. Chen, W. Xu, D. Wang, and L. Wang, "Design of a multi-carrier different chaos shift keying communication system in doubly selective fading channels," in *Proc. 23rd Asia-Pacific Conf. Commun. (APCC)*, 2017, pp. 1–6.
- [34] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Boston, MA, USA: Springer Science & Business Media, 1999.
- [35] G. de Veciana and A. Zakhor, "Neural net-based continuous phase modulation receivers," *IEEE Trans. Commun.*, vol. 40, no. 8, pp. 1396–1408, Aug. 1992.
- [36] L. Zhang, X. Yang, H. Liu, H. Zhang, and J. Cheng, "Efficient residual shrinkage CNN denoiser design for intelligent signal processing: Modulation recognition, detection, and decoding," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 97–111, Jan. 2022.
- [37] K. A. Norton, L. E. Vogler, W. V. Mansfield, and P. J. Short, "The probability distribution of the amplitude of a constant vector plus a rayleigh-distributed vector," in *Proc. IRE*, vol. 43, no. 10, pp. 1354–1361, Oct. 1955.
- [38] J. Seo, S. Kim, and V. C. M. Leung, "Outage probability characterization of CoMP-joint transmission with path-loss and Rayleigh fading," *IEEE Commun. Lett.*, vol. 19, no. 1, pp. 78–81, Jan. 2015.
- [39] Z. Du, J. Cheng, and N. C. Beaulieu, "Accurate error-rate performance analysis of OFDM on frequency-selective Nakagami-m fading channels," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 319–328, Feb. 2006.
- [40] Z. Liu, L. Zhang, and Z. Wu, "Non contiguous frequency hopping aided OFDM-DCSK design without requiring CSI," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–6.
- [41] H. Kim, H. Wang, S. Lim, and D. Hong, "On the impact of outdated channel information on the capacity of secondary user in spectrum sharing environments," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 284–295, Jan. 2012.
- [42] F. Bellili, F. Sotiraki, and W. Yu, "Generalized approximate message passing for massive MIMO mmWave channel estimation with laplacian prior," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3205–3219, May 2019.
- [43] H. Yuan and P. Kam, "Soft-decision-aided, smoothness-constrained channel estimation over time-varying fading channels with no channel model information," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 73–86, Jan. 2017.



Haotian Zhang received the B.Eng. degree in 2019 from the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China, where he is currently working toward the Ph.D. degree with the School of Electronics and Communication Engineering. His research interests include deep learning-aided intelligent communication and its application to wireless communication systems.



Lin Zhang (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Sun Yat-sen University, Guangzhou, China, in 2003. In 2003, she joined the School of Information Science and Technology, Sun Yat-sen University, where she has been an Associate Professor since 2007. From 2008 to 2009, she was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA. In 2016, she joined the School of Electronics and Information Technology, Sun Yatsen University. She is also with Southern Marine Science and Engineering Guangdong Laboratory. From 2021, she joined School of Information Science and Technology, Tibet University, Lhasa, China. Her research has been supported by the National Natural Science Foundation of China and the Science and Technology Program Project of Guangdong Province. Her current research interests include signal processing and their applications to wireless communication systems.



Zhiqiang Wu (Senior Member, IEEE) received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1993, the M.S. degree from Peking University, Beijing, in 1996, and the Ph.D. degree from Colorado State University, Fort Collins, CO, USA, in 2002, all in electrical engineering. From 2003 to 2005, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Institute of Technology, West Virginia University, Morgantown, WV, USA. In 2005, he joined Wright State University, Dayton, OH, USA, where he is currently a Full Professor with the Department of Electrical Engineering. He has also held visiting positions with Peking University, Harbin Engineering University, Harbin, China, Guizhou Normal University, Guiyang, China and Tibet University, Lhasa, China. His research has been supported by NSF, AFRL, ONR, AFOSR, and OFRN.



Yuan Jiang (Member, IEEE) received the B.Eng. degree in satellite communication and the M.Sc. degree in communication and electronic system from the PLA Information Engineering University, Zhengzhou, China, in 1991 and 1998, respectively, and the Ph.D. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2004. From 1991 to 2005, he was with the Department of Scientific Research, PLA Information Engineering University. From 2005 to 2008, he was with the Postdoctoral Workstation of

Computer Science and Engineering, South China University of Technology, Guangzhou, China. From 2008 to 2018, he was successively a Vice General Manager, Chief Engineer, and the Vice President with the Guangzhou Haige Communications Group Incorporated Company. He has presided more than or participated in more than 20 research projects, including the projects supported by the National Natural Science Foundation of China, National High Technology Research and Development Program ("863" Program) of China, and National Science and Technology Major Project of China. His research interests include cognitive communications, satellite communication networks, satellite navigation, and their applications to wireless communication systems.