

Received March 17, 2021, accepted March 26, 2021, date of publication April 5, 2021, date of current version April 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070336

ELM-Based Frame Synchronization in Nonlinear Distortion Scenario Using Superimposed Training

CHAOJIN QING¹, (Member, IEEE), WANG YU¹, SHUHAI TANG¹, CHUANGUI RAO¹,
AND JIAFAN WANG², (Student Member, IEEE)

¹School of Electrical Engineering and Electronic Information, Xihua University, Chengdu 610039, China

²Synopsys Inc., Hillsboro, OR 97124, USA

Corresponding author: Chaojin Qing (qingchj@uestc.edu.cn)

This work was supported in part by the Key Projects of Education Department of Sichuan Province under Grant 15ZA0134, in part by the Sichuan Science and Technology Program under Grant 2021JDR0003, in part by the Demonstration Project of Chengdu Major Science and Technology Application under Grant 2020-YF09-00048-SN, in part by the Special Funds of Industry Development of Sichuan Province under Grant zyf-2018-056, in part by the Major Special Funds of Science and Technology of Sichuan Science and Technology Plan Project under Grant 19ZDZX0016/Grant 2019YFG0395, and in part by the Innovation Fund of Postgraduate, Xihua University under Grant ycyj2020116.

ABSTRACT The requirement of high spectrum efficiency puts forward higher requirements on frame synchronization (FS) in wireless communication systems. Meanwhile, a large number of nonlinear devices or blocks will inevitably cause nonlinear distortion. To avoid the occupation of bandwidth resources and overcome the difficulty of nonlinear distortion, an extreme learning machine (ELM)-based network is introduced into the superimposed training-based FS with nonlinear distortion. Firstly, a preprocessing procedure is utilized to reap the features of synchronization metric (SM). Then, based on the rough features of SM, an ELM network is constructed to estimate the offset of frame boundary. The analysis and experiment results show that, compared with existing methods, the proposed method can improve the error probability of FS and bit error rate (BER) of symbol detection (SD). In addition, this improvement has its robustness against the impacts of parameter variations.

INDEX TERMS Frame synchronization, extreme learning machine, nonlinear distortion, superimposed training.

I. INTRODUCTION

Due to the limited bandwidth resources, wireless communication systems have pursued high spectrum efficiency in the past few decades [1]. As we can see, the spectrum efficiency of the fifth generation (5G) wireless communication system is many times higher than that of the fourth generation (4G) wireless communication system [2], [3]. In wireless communication systems, frame synchronization (FS) is a fundamental and essential task to guarantee the overall system performance [4], which usually consumes substantial bandwidth resources to overcome the synchronization challenge [5]. Thus, during the FS phase, the contradiction between the high bandwidth resources consumption and the high spectrum efficiency requirement needs to be resolved. Meanwhile, the wireless communication system has a large number of nonlinear devices or blocks, e.g., high power

amplifier (HPA), digital to analog converter (DAC), etc., inevitably causing nonlinear distortion [6], [7]. With limited considerations for nonlinear distortion, the classical methods (e.g., correlation-based FS [8]) and the recent solutions (e.g., compressed sensing-based FS in [9]) are usually difficult to apply in nonlinear distortion scenarios [10]. Therefore, the FS is facing challenges from not only the spectrum efficiency but also the nonlinear distortion.

To cope with nonlinear distortion, machine learning (ML), in particular, deep learning (DL) has shown its prominent ability [11], [12]. In recent years, DL has been applied in wireless communication, e.g., signal detection [13], precoding [14], channel state information (CSI) feedback [15], channel estimation [16], [17], mobile Internet of Things (IoT) [18], etc. However, these DL-based approaches exist weaknesses such as long-time training, complex parameter tuning [10], [19] etc. Different from DL-based scheme, as a single hidden layer feed-forward neural network, extreme learning machine (ELM) can learn quickly, randomly

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenzhou Tang¹.

generate for input weight and hidden bias, require no gradient back-propagation, and has good generalization performance [10], [20], [21]. As one of the effective options, using ELM to deal with nonlinear distortion is a promising solution.

For saving bandwidth resources and thus improving spectrum efficiency, the FS using superimposed training sequence is an attractive scheme. Without any occupation of bandwidth resources, this FS superimposes the training sequence on data symbols, yielding more transmission symbols than that of non-superposition mode [22] in the same transmission interval. The superimposed training-based FS has been investigated in past years, e.g., [22]–[25]. These promising FSs promote us to develop further explorations, especially for the scenarios with nonlinear distortion.

Inspired by those advantages of ELM networks and superimposed training, we investigate an ELM-based FS by using superimposed training, which overcomes the challenges from spectrum efficiency and nonlinear distortion during the FS phase. In our work, the merits to cope with nonlinear distortion can be reaped by ELM networks, and then high spectrum efficiency can be achieved by using superimposed training. The combination of ELM network and superimposed training further improves the FS performance in the scenarios of nonlinear distortion, e.g., the error probability of FS. To our best knowledge, for ELM-based FS, there are limited works to focus on nonlinear distortion, much less to focus on superimposed training.

A. RELATED WORKS

We respectively present the related works of DL-based FS and ELM-based FS as follows.

The DL-based FS has been investigated in [26]–[29]. In [26], an artificial neural network (ANN)-based synchronization method was proposed. For the end-to-end communication systems, [27] and [28] investigated the FS based on neural network (NN). In [27], to achieve the task of FS, a deep neural network (DNN) was employed to auto-encoder, and a convolutional neural network (CNN) was developed in [28] to compensate impairments introduced by timing offset and timing error of sampling. As for [29], a CNN-based FS method was proposed to convert the one dimension (1D) correlator to two dimension (2D) matrix to find the frame offset. From [26]–[29], the DL technology provides effective approaches for FS. Nevertheless, these DL-based FSs are still challenged by many issues, such as long training time, complex parameter tuning, large memory requirements, etc.

Relative to DL networks, the ELM network features many advantages [10], [20], [21], e.g., the gradient back-propagation can be avoided, the output weight can be obtained by solving the least square (LS) question, and a single-hidden layer is employed for the feed-forward neural network, etc. In [10], an ELM-based time-division FS was proposed with the consideration of nonlinear distortion. In [30], the ELM network was employed to compensate the residual time offset with time-division mode. Although the error probability of FS is lower than conventional method,

the training sequence for FS still occupies the bandwidth resources, which reduces systems' spectrum efficiency. To avoid the occupation of bandwidth resources, an ELM-based FS method using superimposed training is proposed in this paper to reduce the error probability of FS.

B. CONTRIBUTIONS

To overcome the challenges of spectrum efficiency and nonlinear distortion during the FS phase, the ELM-based FS using superimposed training is investigated in this paper. The main contributions of this paper are summarized as follows.

- Firstly, an ELM-based FS method by using superimposed training is proposed. In contrast to the ELM-based time-division FS scheme in [9], not only the occupation of bandwidth resources is avoided in the proposed FS method, but also the smaller error probability of FS is achieved with the same energy cost.
- Secondly, the superimposed training-based FS is investigated in the scenarios of nonlinear distortion. Our investigation remedies the deficiencies of the existing superimposed training-based FS, which cannot work well in the scenarios of nonlinear distortion and is suitable for practical application.
- Thirdly, extensive experiments are given to verify the effectiveness of the proposed method in this paper. Compared with the classical correlation method in [8] and the time division method in [10], both the FS's error probability and the symbol detection (SD)'s bit error rate BER) are reduced with the same energy consumption. In addition, the proposed FS presents its robustness against the impacts of parameters.

The remainder of this paper is structured as follows: In Section II, we describe the system model. The ELM-based FS using superimposed training method is presented in Section III, followed by the experimental results and analysis are illustrated in Section IV. Finally, Section V concludes our work.

Notations: Bold face upper case and lower case letters denote matrix and vector respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, denote the transpose, conjugate transpose, and matrix pseudo-inverse, respectively. $\|\cdot\|_2$ is the Frobenius norm. $|x|$ denotes the absolute value of x and $|\mathbf{x}|$ denotes the absolute value operation to the entry-wise of vector \mathbf{x} .

II. SYSTEM MODEL

At the receiver, the received $M \times 1$ complex-valued signal vector, denoted as \mathbf{y} , can be expressed as

$$\mathbf{y} = \mathbf{H}\tilde{\mathbf{x}}_{\text{ext}} + \mathbf{n}, \quad (1)$$

where $\mathbf{n} \in \mathbb{C}^{M \times 1}$ represents the complex additive white Gaussian noise (AWGN) vector with zero-mean and σ^2 -variance entries. $\tilde{\mathbf{x}}_{\text{ext}} \in \mathbb{C}^{(2N-L+1) \times 1}$ denotes the extended vector of transmitted signal with nonlinear

distortion, which can be given by

$$\tilde{\mathbf{x}}_{\text{ext}} = \begin{bmatrix} 0, \dots, 0, \tilde{\mathbf{x}}^T, 0, \dots, 0 \\ \tau \qquad \qquad \qquad N-L-\tau+1 \end{bmatrix}^T, \quad (2)$$

where N and L are the size of search window and the number of multi-path, respectively. The unknown frame boundary offset is denoted as τ , whose range is $0 \leq \tau \leq N - L + 1$. In (2), $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N]^T$ denotes the distorted transmitted signal and can be expressed as

$$\tilde{\mathbf{x}} = f_{\text{dis}}(\mathbf{x}), \quad (3)$$

where $f_{\text{dis}}(\cdot)$ represents the influence of nonlinear distortion. $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ is the superimposed transmitting signal without nonlinear distortion, which can be given by

$$\mathbf{x} = \sqrt{\rho E} \mathbf{s} + \sqrt{(1 - \rho) E} \mathbf{c}, \quad (4)$$

where $\rho \in [0, 1]$ represents the power proportional coefficient (PPC), E denotes the transmitted power of superimposed transmitting signal. $\mathbf{s} \in \mathbb{C}^{N \times 1}$ is the training sequence and $\mathbf{c} \in \mathbb{C}^{N \times 1}$ is the modulated data symbol.

The complex matrix $\mathbf{H} \in \mathbb{C}^{M \times (2N-L+1)}$ given in (1) is an $M \times (2N - L + 1)$ cyclic matrix, which can be defined as

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 & \dots \\ \vdots & h_1 & \ddots \\ h_L & \vdots & \ddots \\ 0 & h_L & \ddots \\ \vdots & 0 & \ddots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (5)$$

where $\mathbf{h} = [h_1, h_2, \dots, h_L]^T$ denotes the finite CIR vector of L samples memory, and h_l represents the complex-valued channel impulse response (CIR) of the l th path, $l = 1, 2, \dots, L$.

With the received signal \mathbf{y} given in (1), we employ ELM network to implement FS using superimposed training, as well as reducing the influence of nonlinear distortion.

III. ELM-BASED FS

The performance of FS is seriously degraded by the influence of nonlinear distortion. To conquer this difficulty, the ELM network is introduced into FS using superimposed training due to its prominent ability to cope with nonlinear distortion [10]. In the following subsections, we first present the pre-processing of FS in Section III-A. Then, the ELM-based FS method is given in Section III-B.

A. PRE-PROCESSING FOR FS

In wireless systems, the FS is usually difficult to obtain, especially for nonlinear distortion scenarios. In particular, the error probability of DL-based timing synchronization is far higher than that of matched filtering in [26]. Similar behaviors are also observed in ELM-based FS experiments.

TABLE 1. Training procedure.

<p>A training set $\{(\bar{\mathbf{g}}_i, \mathbf{T}_i) i = 1, 2, \dots, N_t\}$ generated by SMs and one-hot label and these two are corresponding to the input and desired output of the ELM, respectively.</p> <p>step 1: Randomly choose the input weights $\mathbf{W} \in \mathbb{R}^{\tilde{N} \times N}$ and hidden layer biases $\mathbf{b} \in \mathbb{R}^{\tilde{N} \times 1}$, respectively, set the hidden neuron number \tilde{N}.</p> <p>step 2: With $\bar{\mathbf{g}}, \mathbf{W}$ and \mathbf{b}, the output matrix of hidden layer $\mathbf{H} \in \mathbb{R}^{\tilde{N} \times N_t}$ is calculated through activation function $\sigma(\cdot)$ according to (12).</p> <p>step 3: Construct the target output matrix \mathbf{T} using offset label vector \mathbf{T}_i according to (10), compute the output weights $\Upsilon \in \mathbb{R}^{N \times \tilde{N}}$ according to (14).</p>
--

Thus, a pre-processing of FS is employed to coarsely capture the features of SM. According to [8], by using the cross-correlation based method, the SM vector $\mathbf{g} \in \mathbb{R}^{N \times 1}$ can be expressed by

$$\mathbf{g} = \left| \mathbf{S}^H \mathbf{y} \right|^2. \quad (6)$$

Here, the $M \times N$ complex matrix \mathbf{S} can be written as

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \dots \\ \vdots & s_1 & \ddots \\ s_N & \vdots & \ddots \\ 0 & s_N & \ddots \\ \vdots & 0 & \ddots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (7)$$

where $s_i, i = 1, 2, \dots, N$, represents the i th entry of training sequence \mathbf{s} . Need to be mentioned that, besides the cross-correlation based SM in (6), other SMs can also be applied in our method with the similar processing.

In order to standardize the training of ELM network, the \mathbf{g} given in (6) is normalized as

$$\bar{\mathbf{g}} = \mathbf{g} / \|\mathbf{g}\|_2. \quad (8)$$

With the normalized SM (i.e., $\bar{\mathbf{g}}$), an ELM network is utilized to conquer nonlinear distortion and improve SMs for superimposed training-based FS, which is described in the following subsection.

B. ELM-BASED FS

The ELM-based network is employed to improve SMs and decrease the influence of nonlinear distortion. The ELM-based FS includes offline and online procedures, which are elaborated in TABLE 1 and TABLE 2, respectively. The offline training procedure is described as follows.

1) OFFLINE TRAINING SPECIFICATION

The offline procedure is elaborated in TABLE 1. In the following, we first describe the data collection for ELM-net training.

TABLE 2. Online procedure.

step 1 :	From (6) to (8), perform the preprocessing to obtain metric vector $\bar{\mathbf{q}}$ as the input of the trained ELM-net.
step 2 :	With input metric vector $\bar{\mathbf{q}}$, the learned output weights Υ , the randomly chosen input weights \mathbf{W} and hidden layer biases \mathbf{b} , the ELM network output \mathbf{O} can be obtain according to (15).
step 3 :	Find the location of the maximum square of the absolute value of the element in \mathbf{O} , (i.e., estimation of frame boundary offset $\hat{\tau}$) according to (16).

• Data Collection For ELM-Net training

For ELM-net training, N_t samples of input signals and offset labels, denoted by $\{(\bar{\mathbf{g}}_i, \mathbf{T}_i)\}, i = 1, 2, \dots, N_t$ are collected to form a training set. According to the N_t collected $\bar{\mathbf{g}}_i$ forms the input matrix $\bar{\mathbf{g}} \in \mathbb{R}^{N \times N_t}$, which can be written as

$$\bar{\mathbf{g}} = [\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2, \dots, \bar{\mathbf{g}}_{N_t}]. \tag{9}$$

Similarly, N_t offset label vectors \mathbf{T}_i are converted to construct the target output matrix $\mathbf{T} \in \mathbb{R}^{N \times N_t}$, which can be given by

$$\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_{N_t}], \tag{10}$$

where the label \mathbf{T}_i can be encoded according to one-hot mode, i.e.,

$$T_i = \begin{bmatrix} 0, \dots, 0, 1, 0, \dots, 0 \\ \tau_i \qquad \qquad \qquad N - \tau_i - 1 \end{bmatrix}^T, \tag{11}$$

where τ_i is the i th sample's frame boundary offset.

In this paper, the input weights $\mathbf{W} \in \mathbb{R}^{\tilde{N} \times N}$ and hidden layer biases $\mathbf{b} \in \mathbb{R}^{\tilde{N} \times 1}$ of ELM network (with \tilde{N} hidden neuron number) are respectively randomly chosen, which is similar to the standard process of ELM network [21]. It should be noted that the initial values of \mathbf{W} and \mathbf{b} have a certain impact on the FS performance, we mainly investigate the FS method in this paper. Admittedly, initial values can further improve the FS's error probability and SD's BER. Then, the input weights \mathbf{W} and hidden layer biases \mathbf{b} are saved in storage space for later use in the offline and online procedure.

• Networking Training

As shown in TABLE 1, during the training procedure, the training data-set $\{(\bar{\mathbf{g}}, \mathbf{T})\}$, input wights \mathbf{W} and hidden layer biases \mathbf{b} are gradually loaded from storage space. Then, with $\{(\bar{\mathbf{g}}, \mathbf{T})\}$, \mathbf{W} and \mathbf{b} , the output matrix of the hidden layer $\mathbf{H} \in \mathbb{R}^{N \times N_t}$ can be given by

$$\mathbf{H} = \sigma(\mathbf{W}\bar{\mathbf{g}} + \mathbf{b}), \tag{12}$$

where $\sigma(\cdot)$ represents the activation function, e.g., sigmoid [31], hyperbolic tangent [32], rectified linear units

(ReLU) [33], etc. The sigmoid is used in this ELM network [21]. The objective of ELM network is actually to find the suitable output weights $\Upsilon \in \mathbb{R}^{N \times \tilde{N}}$ to approximate the target output matrix \mathbf{T} , which can be expressed as [20]

$$\mathbf{T} = \Upsilon\mathbf{H}, \tag{13}$$

where $\Upsilon = [\Upsilon_1, \dots, \Upsilon_k, \dots, \Upsilon_{\tilde{N}}]$, Υ_k denotes the output weighting vector connecting the k th hidden neuron and the output neurons. The LS solution $\hat{\Upsilon}$ of $\mathbf{T} = \Upsilon\mathbf{H}$ with minimum norm of output weights Υ is given by [20]

$$\hat{\Upsilon} = \min_{\Upsilon} \|\Upsilon\mathbf{H} - \mathbf{T}\| = \mathbf{T}\mathbf{H}^\dagger, \tag{14}$$

the output weighting matrix Υ is learned from offline training of ELM network, which is saved in storage space for online running.

2) ONLINE DEPLOYMENT

With the learned ELM-based network parameters, the online running procedure can be implemented in this subsection, which is shown in TABLE 2.

With the input metric vector $\bar{\mathbf{q}}$, which can be obtained by employing the preprocessing procedure according to (6)-(8), the learned output weights Υ , the random input weights \mathbf{W} and hidden layer biases \mathbf{b} , the ELM network output $\mathbf{O} \in \mathbb{R}^{N \times 1}$ can be written as

$$\mathbf{O} = \Upsilon \cdot \sigma(\mathbf{W}\bar{\mathbf{q}} + \mathbf{b}), \tag{15}$$

where $\mathbf{O} = [o_1, o_2, \dots, o_N]^T$, and the estimation of frame boundary offset can be expressed as

$$\hat{\tau} = \arg \max_{1 \leq j \leq N} |o_j|^2. \tag{16}$$

With the estimation from (16), the FS is completed by acquiring the frame's starting point $\hat{\tau}$. After FS, the SD is performed, and the detected symbol $\hat{\mathbf{c}}$ can be represented as

$$\hat{\mathbf{c}} = \frac{\tilde{\mathbf{x}}_{\text{est}} - \sqrt{\rho E_s}}{\sqrt{1 - \rho}}, \tag{17}$$

where $\tilde{\mathbf{x}}_{\text{est}}$ denotes the estimation of superimposed transmitting signal in the scenario of nonlinear distortion, and can be obtained by

$$\tilde{\mathbf{x}}_{\text{est}} = \mathbf{H}^\dagger \mathbf{y}. \tag{18}$$

The FS and SD can be achieved according to (6)-(18), which improve SMs and address the issues about multi-path interfere and nonlinear distortion.

IV. EXPERIMENTAL ANALYSIS

In this section, numerical results of the proposed ELM-based FS using superimposed training are given. Firstly, basic parameters and definitions involved in simulations are given in Section IV-A. Then, in Section IV-B, the FS's error probability and SD's BER of the proposed scheme with nonlinear distortion are shown to verify the effectiveness of the proposed ELM-based FS, followed by the robustness of improvement with different parameters is discussed in Section IV-C.

At last, the computational time complexity is analyzed in Section IV-D.

A. PARAMETER SETTING

In the simulations, the basic parameters are set as $N = 512$, $M = 2N = 1024$, $\tilde{N} = 10N = 5120$ [20], [34], $L = 8$, and $N_t = 10^5$. The Zadoff-Chu sequence [35] is employed as the training sequence \mathbf{s} . For the time-division method in [10], $N_s = 16$ is considered as the length of training sequence. By referencing [22] and considering the total performance of FS’s error probability and SD’s BER, $\rho = 0.3$ is adopted in this paper. The modulated data symbol \mathbf{c} is formed according to the symbol of quadrature-phase-shift-keying (QPSK) modulation. For the channel model, the multipath Rayleigh fading channel with an exponentially-decayed power coefficient $\eta = 0.2$ is considered, where each of the following $L - 1$ paths is set as zero-valued with a probability of 0.5 beside the first path to keep the same situation as [9] and [10]. For the sake of fair comparison, we assume the superimposed FS and the time-division FS consume the same energy for transmitting symbols.

Definitions involved are listed as follows. The signal-to-noise ratio (SNR) in decibel (dB) is defined as [36]

$$SNR = 10\log_{10} \left(\frac{E}{\sigma^2} \right). \tag{19}$$

For nonlinear distortion, the HPA effect is taken into account in these simulations. The nonlinear amplitude $A(x)$ and phase $\Phi(x)$ are obtained by [37]

$$A(x) = \frac{\alpha_a x}{1 + \beta_a x^2}, \Phi(x) = \frac{\alpha_\phi x^2}{1 + \beta_\phi x^2}, \tag{20}$$

where $\alpha_a = 1.96$, $\beta_a = 0.99$, $\alpha_\phi = 2.53$, and $\beta_\phi = 2.82$ are considered in the experiments according to [37]. To measure the distortion intensity, the error vector magnitude (EVM) is used in this paper, which is expressed as [38]

$$EVM(\%) = \sqrt{\frac{\sum_{n=1}^N |\tilde{x}_n - R_n|^2}{\sum_{n=1}^N |R_n|^2}}, \tag{21}$$

where \tilde{x}_n is the n -th distorted symbol through HPA, which denotes the HPA workings in saturated region. R_n denotes the desired linear outputs of HPA given the same input without amplification distortion. In this paper, the EVM is set as $EVM = 35\%$ except for the robustness analysis against EVMs.

For simplicity, “Prop”, “TD_Corr”, “TD_ELM”, and “Sup_Corr” are used to denote the proposed ELM-based superimposed FS, the correlation-based time-division FS in [8], the ELM-based time-division FS in [10], and the correlation-based superimposed FS method in [22], respectively.

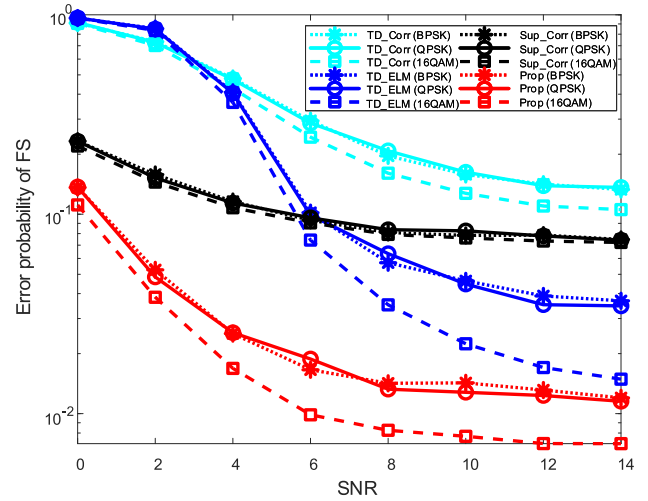


FIGURE 1. Error probability of FS vs. SNR, where $N = 512$, $L = 8$ and $EVM = 35\%$.

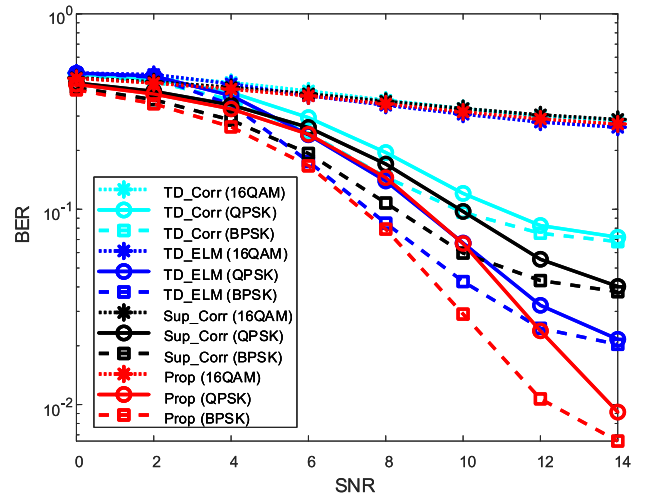


FIGURE 2. BER vs. SNR, where $N = 512$, $L = 8$ and $EVM = 35\%$.

B. FS AND SD PERFORMANCE

To validate the effectiveness of proposed ELM-based FS using superimposed training, the error probability of FS and BER of SD under different SNRs are illustrated in Fig. 1 and Fig. 2, respectively.

The effectiveness of the error probability of FS is presented in Fig. 1. It could be observed that the “Prop” reaches the smallest error probability among different methods. This reflects the “Prop” obtains the best performance of error probability for FS, and thus can work well in the scenarios of nonlinear distortion. In addition, some insights of FS with nonlinear distortion can be achieved in Fig. 1. Firstly, the error probability of “Sup_Corr” is smaller than that of “TD_Corr” with the same energy consumption, which embodies the superiority of superimposed FS compared with time division FS. Secondly, for relatively high SNR (e.g., $SNR > 6$ dB), the “TD_ELM” has a smaller error probability of FS than

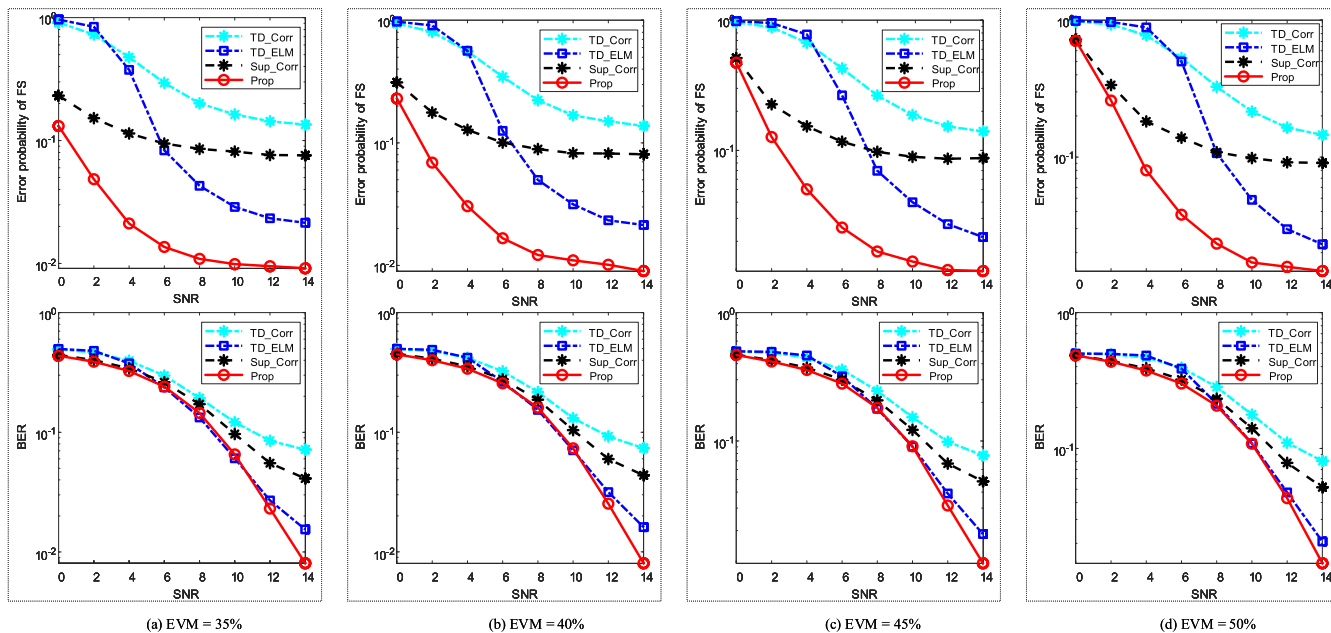


FIGURE 3. Error probability of FS and BER against the impact of EVM, where EVM = 35%, EVM = 40%, EVM = 45%, and EVM = 50% are considered, respectively.

those of “Sup_Corr” and “TD_Corr”. That is, the ELM network effectively suppresses the nonlinear distortion for time division FS in [8], and the ELM-based time-division FS can further obtain better performance of error probability than that of superimposed FS in [22]. Although the ELM-based time-division FS has shown its effectiveness to suppress the nonlinear distortion, it can only be effective in a relatively high SNR region, and the merits of superimposed FS have not been developed. The “Prop” develops the merits of superimposed training and ELM network, and thus obtains the smallest error probability of FS in all given SNR region. Therefore, the combination of superimposed training and ELM network in “Prop” possesses its effectiveness to improve the error probability of FS in the scenarios of nonlinear distortion.

Since the training sequence s is superimposed on the modulated data symbol c , it needs to be verified whether the superimposed interference (from the superimposed training) degrades the detection performance of data symbols. In this paper, the BER of SD is used to measure the detection performance and is plotted in Fig. 2. From Fig. 2, the “Prop” achieves the smallest BER for almost all given SNRs. Thus, for the same energy consumption, the superimposed interference does not degrade the BER performance. On the contrary, the BER of “Prop” effectively benefits from the proposed FS, especially for the relatively high SNR region (e.g., $SNR > 10\text{dB}$). That is, with simple processing of interference cancellation given in (17), the “Prop” achieves the best BER performance among all given FS methods. In particular, the advantages of superimposed training and ELM network can be separately demonstrated from Fig. 2. Without using superimposition approaches, “TD_ELM” obtains smaller BER than that of “TD_Corr”, which reflects the effectiveness of ELM network to deal with nonlinear

distortion. We can also observe that the BER of “Sup_Corr” is smaller than that of “TD_Corr”. That is, the superimposed training used in “Sup_Corr” is useful to improve the BER performance of “TD_Corr”. Thus, by combining the superimposed training and ELM network, the BER performance is improved.

As a whole, compared with the “TD_Corr”, “TD_ELM”, and “Sup_Corr”, both the FS’s error probability and the SD’s BER in the scenarios of nonlinear distortion are improved by “Prop”. Especially, compared with the “TD_Corr” and “TD_ELM”, the “Prop” can obtain the chance to transmit more data symbols, and thus can further improve the spectrum efficiency. By the way, with different modulation conditions (e.g., BPSK and 16QAM), the proposed method still improves the FS’s error probability and SD’s BER from Fig. 1 and Fig. 2.

C. ANALYSIS OF PARAMETER IMPACT

In this subsection, the robustness of the proposed scheme against parameter variation is analysed. The impact of EVM is first discussed, followed by the number of multi-path (i.e., L), the transmitted frame-length N , and the PPC ρ . It is worth noting that, besides the change of the impact parameter (i.e., EVM, L , N , and ρ), other basic parameters remain the same as those given in Section IV-A during the simulations.

1) IMPACT OF EVM

EVM is usually used to measure the distortion intensity. To analyze the robustness of the proposed method against different distortion intensities, Fig. 3 respectively plots the curves of FS’s error probability and SD’s BER with different EVMs (i.e., EVM = 35%, EVM = 40%, EVM = 45%, and EVM = 50%).

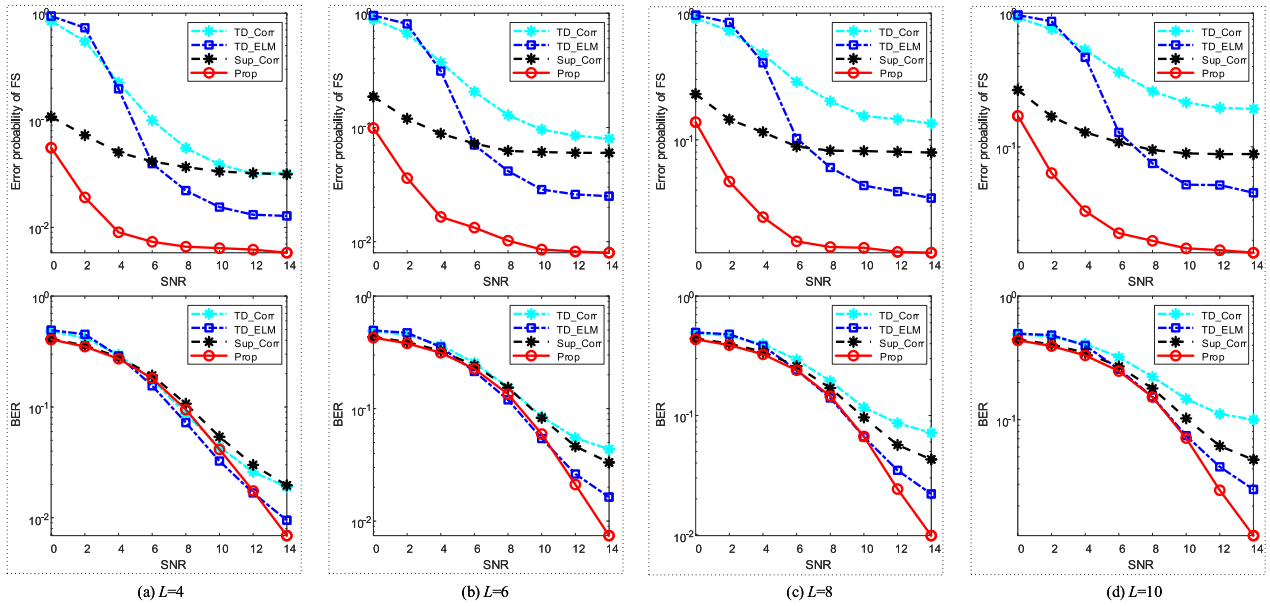


FIGURE 4. Error probability of FS and BER against the impact of L , where $L = 4, L = 6, L = 8$, and $L = 10$ are considered, respectively.

From Fig. 3, compared with those of “TD_Corr”, “TD_ELM”, and “Sup_Corr”, the “Prop” method achieves the smallest error probability for each given EVM. That is, relative to the existing methods, the proposed FS scheme still improves the FS’s error probability against varying EVMs. With the increase of EVM, the FS’s error probabilities for all curves in Fig. 3 (i.e., “TD_Corr”, “TD_ELM”, “Sup_Corr”, and “Prop”) increase due to the rise of distortion intensity. However, the FS’s error probability of “Prop” is smaller than those of “TD_Corr”, “TD_ELM”, and “Sup_Corr”, especially for the high SNR region. This reflects the proposed scheme can improve the FS’s error probability against varying EVMs.

For the SD, Fig. 3 shows the BER of “Prop” is smaller than those of “TD_Corr” and “Sup_Corr” in almost all given SNR regions. Especially, in a relatively high SNR region, e.g., $SNR \geq 12\text{dB}$, we can observe the BER of “Prop” is smaller than those of “TD_Corr”, “Sup_Corr”, and “TD_ELM”. Thus, for different EVMs, the “Prop” achieves similar or better BER performance.

As a result, against the impact of EVM, the “Prop” possesses its robustness for improving FS’s error probability and SD’s BER.

2) IMPACT OF L

The FS’s error probability and SD’s BER are usually impacted by the number of multi-path (i.e., L). To demonstrate the robustness of the proposed FS scheme against the impact of L , the error probability of FS and BER of SD curves are given in Fig. 4, where $L = 4, L = 6, L = 8$, and $L = 10$ are considered, respectively.

From Fig. 4, relative to the “TD_Corr”, “TD_ELM”, and “Sup_Corr”, the “Prop” achieves the minimal error probability of FS for each given L . This reflects the “Prop”

improves the FS’s error probability of the existing methods with the variations of L . In addition, with the increase of L , the FS’s error probabilities of “Prop”, “TD_Corr”, “TD_ELM”, and “Sup_Corr” rise with the enlargement of multi-path interference. Even so, the “Prop” still presents the ability to cope with the nonlinear distortion and multi-path interference under different values of L , and thus obtains the smallest FS’s error probability. As a whole, against the impact of L , the “Prop” can robustly reduce FS’s error probability.

From the curves of SD’s BER in Fig. 4, the “Prop” achieves the smallest BER for relatively large L (e.g., $L \geq 8$). This reflects the “Prop” possesses better BER performance compared with the existing methods when $L \geq 8$. For the case where $4\text{dB} \leq SNR \leq 12\text{dB}$ and $L = 4$, the BER of “Prop” is slightly higher than that of “TD_ELM”. This is because the “TD_ELM” consumes additional bandwidth resources to avoid the superimposed interference of “Prop”. Moreover, the ELM network is also employed by “TD_ELM”, and thus effectively suppresses the nonlinear distortion as well. Nevertheless, compared with “TD_ELM” for the case where $4\text{dB} \leq SNR \leq 8\text{dB}$, the “Prop” has only slightly higher BER, while transmitting more data symbols to obtain higher spectrum efficiency. Besides, the BER of “Prop” is still smaller than that of “TD_ELM” when $SNR \geq 12\text{dB}$.

To sum up, against the impact of L , the “Prop” can effectively reduce the FS’s error probability and SD’s BER, especially for the cases of relatively large L and relatively high SNR.

3) IMPACT OF N

Usually, the FS’s error probability and SD’s BER performance are influenced by the frame-length, i.e., N . To validate the robustness against the impact of N , the error probability of FS and BER of SD are illustrated in Fig. 5 with different

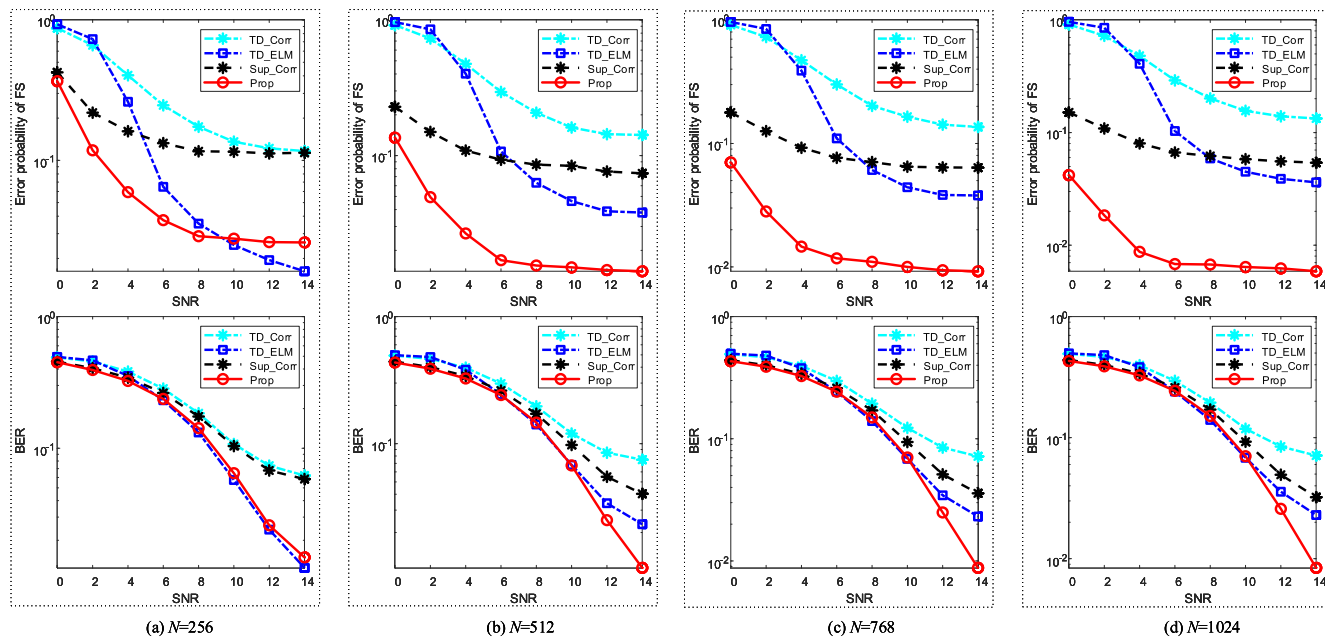


FIGURE 5. Error probability of FS and BER against the impact of N , where $N = 256$, $N = 512$, $N = 768$, and $N = 1024$ are considered, respectively.

values of N (i.e., $N = 256$, $N = 512$, $N = 768$, and $N = 1024$).

From Fig. 5, for the cases where $N = 512$, $N = 768$, and $N = 1024$, the “Prop” obtains a smaller FS’s error probability than those of “TD_Corr”, “TD_ELM”, and “Sup_Corr”. This reflects the “Prop” could reduce the error probability of FS against the varying N . When $N = 256$ and $SNR \leq 8$ dB, the “Prop” can still obtain the minimum error probability of FS to embodies the robustness against the impact of N . However, for $SNR \geq 10$ dB, such situation could not be held. As can be seen in Fig. 5(a), the FS’s error probability of “Prop” is higher than that of “TD_ELM”. This is because the “TD_ELM” can also suppress the nonlinear distortion as that of the “Prop”, while the relatively short N brings “Prop” the difficulty to combat the superimposed interference from the data symbol c . Even so, compared with the “TD_ELM”, the “Prop” saves the bandwidth resources and significantly reduces FS’s error probability in a relatively low SNR region (e.g., $SNR \leq 8$ dB). In particular, relative to the “TD_Corr” and “Sup_Corr”, the “Prop” clearly reduces the error probability of FS for all given SNRs. In addition, with the increase of N , the FS’s error probability of “Prop” decreases. This reflects the elongated N can effectively suppress the superimposed interference of the “Prop”, due to the elongation of the superimposed training sequence s .

Compared with “TD_ELM”, “TD_Corr”, and “Sup_Corr”, the “Prop” has similar or smaller BER, which demonstrates the “Prop” obtains the similar or better SD’s BER performance with different values of N . Relative to “TD_Corr” and “Sup_Corr”, the “Prop” reduces the BER for each given N , especially in relatively high SNR region (e.g., $SNR \geq 10$ dB). Meanwhile, for the case where

TABLE 3. Time complexity comparison.

	TD_ELM	Proposed
Training time	about 19.6 minutes	about 20.2 minutes
Online running time	about 11.6 minutes	about 11.7 minutes

$N = 256$ and $0\text{dB} \leq SNR \leq 10\text{dB}$, the “Prop” obtains similar BER performance as that of “TD_ELM”. When $SNR \geq 10$ dB, the BER of “Prop” is slightly higher than that of “TD_ELM”. The reasons are given as follows. On the one hand, the nonlinear distortion is also suppressed by employing the ELM network in “TD_ELM”. On the other hand, the “TD_ELM” consumes additional bandwidth to avoid superimposed interference. Moreover, the corresponding FS’s error probability of “TD_ELM” is smaller than that of “Prop”, which also deteriorates the SD’s BER performance. Even so, the “Prop” can save the bandwidth resource and obtains similar BER performance as that of “TD_ELM” when $N = 256$. Especially, for relatively large N (e.g., $N \geq 512$), the “Prop” can achieve lower BER than that of “TD_ELM” when $SNR \geq 12$ dB.

On a whole, with the varying of N , the “Prop” can improve the performance of FS’s error probability and SD’s BER, and this improvement possesses good robustness.

D. COMPLEXITY ANALYSIS

The training time and online running time between the correlation-based superimposed FS method in [10] (i.e., “TD_ELM”) and the proposed ELM-based superimposed FS (i.e., “Prop”) are illustrated in TABLE 3 to compare the computational time complexity.

For a fair comparison, 10^5 experiments are conducted for “Prop” and “TD_ELM” on the same server with Intel

Xeon(R) E5-2620 CPU 2.1GHz_16 by using Matlab software, respectively. During the experiments, only running time is considered, (i.e., the time to generate data sets and through channel is not included). From TABLE 3, in the training phase, the “TD_ELM” consumes about 19.6 minutes, while the “Prop” costs about 20.2 minutes. In the online running stage, the “TD_ELM” consumes about 11.6 minutes, and the “Prop” costs about 11.7 minutes. We can see that the average training and online running time of “Prop” is slightly more than that of “TD_ELM” in each experiment, while the “Prop” improves the spectrum efficiency with similar or better performance of FS’s error probability and SD’s BER relative to the existing methods.

V. CONCLUSION

In this work, we integrated superimposed training-based FS and ELM network to investigate an ELM-based FS scheme using superimposed training in nonlinear distortion scenarios. Firstly, a preprocessing procedure is employed to coarsely reap the features of SM. Then, an ELM network is introduced to conquer the impact of nonlinear distortion and to obtain the estimation of frame boundary offset. Compared with some existing methods, the proposed method can improve the error probability of FS and BER of SD, and those improvements are robust against parameter variation. In future works, we will investigate the generalization method of ELM-based FS to alleviate the difference between the data set from simulation and the data set in real scenarios.

REFERENCES

- [1] F. Lyu, N. Cheng, H. Zhu, H. Zhou, W. Xu, M. Li, and X. S. Shen, “Intelligent context-aware communication paradigm design for IoVs based on data analytics,” *IEEE Netw.*, vol. 32, no. 6, pp. 74–82, Nov./Dec. 2018.
- [2] J. Ni, X. Lin, and X. S. Shen, “Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 644–657, Mar. 2018.
- [3] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. W. Mark, and X. Shen, “Cloud assisted HetNets toward 5G wireless networks,” *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 59–65, Jun. 2015.
- [4] J. Blumenstein and M. Bobula, “Coarse time synchronization utilizing symmetric properties of Zadoff–Chu sequences,” *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1006–1009, May 2018.
- [5] Y. Li and L. Zhang, “An improved algorithm for OFDM symbol timing synchronization,” *J. Harbin Univ. Sci. Technol.*, vol. 17, no. 3, pp. 62–64, 2012.
- [6] J. Guerreiro, R. Dinis, and P. Montezuma, “Analytical performance evaluation of precoding techniques for nonlinear massive MIMO systems with channel estimation errors,” *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1440–1451, Apr. 2018.
- [7] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, “Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.
- [8] F. Ling and J. Proakis, *Synchronization in Digital Communication Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
- [9] B. Lopes, S. Catarino, N. M. B. Souto, R. Dinis, and F. Cercas, “Robust joint synchronization and channel estimation approach for frequency-selective environments,” *IEEE Access*, vol. 6, pp. 53180–53190, Sep. 2018.
- [10] C. Qing, W. Yu, B. Cai, J. Wang, and C. Huang, “ELM-based frame synchronization in burst-mode communication systems with nonlinear distortion,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 915–919, Jun. 2020.
- [11] J. Liu, K. Mei, X. Zhang, D. Ma, and J. Wei, “Online extreme learning machine-based channel estimation and equalization for OFDM systems,” *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1276–1279, Jul. 2019.
- [12] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, “Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions,” *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, Feb. 2020.
- [13] H. Ye, G. Y. Li, and B.-H. 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.
- [14] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, “Deep-learning-based millimeter-wave massive MIMO for hybrid precoding,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.
- [15] C. Qing, B. Cai, Q. Yang, J. Wang, and C. Huang, “Deep learning for CSI feedback based on superimposed coding,” *IEEE Access*, vol. 7, pp. 93723–93733, Jul. 2019.
- [16] H. He, C.-K. 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.
- [17] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, “Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
- [18] L. Xu, X. Yu, and T. A. Gulliver, “Intelligent outage probability prediction for mobile IoT networks based on an IGWO-Elman neural network,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1365–1375, Feb. 2021.
- [19] C. Qing, B. Cai, Q. Yang, J. Wang, and C. Huang, “ELM-based superimposed CSI feedback for FDD massive MIMO system,” *IEEE Access*, vol. 8, pp. 53408–53418, Mar. 2020.
- [20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
- [21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: A new learning scheme of feedforward neural networks,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Budapest, Hungary, vol. 2, Jul. 2004, pp. 985–990.
- [22] A. T. P. Nguyen, R. Le Bidan, and F. Guilloud, “Superimposed frame synchronization optimization for finite blocklength regime,” in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop (WCNCW)*, Apr. 2019, pp. 1–6.
- [23] F. Guilloud and R. Bidan, “On the optimization of resources for short frame synchronization,” *Ann. Telecommun.*, vol. 75, pp. 1–6, Aug. 2020.
- [24] S. Pfletschinger and P. Closas, “Non-orthogonal frame synchronization for low latency communication,” in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–6.
- [25] J.-K. Hwang, C.-F. Li, and C. Ma, “Efficient detection and synchronization of superimposed NB-IoT NPRACH preambles,” *IEEE Internet Things J.*, vol. 6, no. 1, pp. 1173–1182, Feb. 2019.
- [26] T. O’Shea, K. Karra, and T. Clancy, “Learning approximate neural estimators for wireless channel state information,” in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process.*, Tokyo, Japan, Sep. 2017, pp. 1–7.
- [27] 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.
- [28] H. Wu, Z. Sun, and X. Zhou, “Deep learning-based frame and timing synchronization for end-to-end communications,” *J. Phys., Conf. Ser.*, vol. 1169, no. 1, pp. 491–496, 2018.
- [29] E.-R. Jeong, E.-S. Lee, J. Joung, and H. Oh, “Convolutional neural network (CNN)-based frame synchronization method,” *Appl. Sci.*, vol. 10, no. 20, p. 7267, Oct. 2020.
- [30] J. Liu, K. Mei, X. Zhang, D. McLernon, D. Ma, J. Wei, and S. A. R. Zaidi, “Fine timing and frequency synchronization for MIMO-OFDM: An extreme learning approach,” *IEEE J. Sel. Areas Commun.* [Online]. Available: <https://arxiv.org/abs/2007.09248>
- [31] T. Fu, C. Wang, and N. Cheng, “Deep-learning-based joint optimization of renewable energy storage and routing in vehicular energy network,” *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6229–6241, Jul. 2020.
- [32] H. Peng and X. Shen, “Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 131–141, Jan. 2021.
- [33] G. Gui, Z. Zhou, J. Wang, F. Liu, and J. Sun, “Machine learning aided air traffic flow analysis based on aviation big data,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4817–4826, May 2020.

[34] Z. Chen, Y. Cao, D. Zou, and Q. Gu, "How much over-parameterization is sufficient to learn deep ReLU networks?" 2019, *arXiv:1911.12360*. [Online]. Available: <http://arxiv.org/abs/1911.12360>

[35] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.

[36] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," *IEEE Trans. Evol. Comput.*, vol. 3, no. 2, pp. 82–102, Jul. 1999.

[37] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.

[38] A. M. Angelotti, G. P. Gibiino, C. Florian, and A. Santarelli, "Broadband error vector magnitude characterization of a GaN power amplifier using a vector network analyzer," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Aug. 2020, pp. 747–750.



SHUHAI TANG received the B.S. degree from the School of Electrical Engineering and Electronic Information, Xihua University, Chengdu, China, in 2020, where he is currently pursuing the M.S. degree, under the supervision of Prof. Qing. His research interests include synchronization and channel estimation in OFDM systems, and machine learning applications in physical layer of wireless communications.



CHAOJIN QING (Member, IEEE) received the B.S. degree in communication engineering from the Chengdu University of Information Technology, Chengdu, China, in 2001, and the M.S. and Ph.D. degrees in communications and information systems from the University of Electronic Science and Technology of China, Chengdu, in 2006 and 2011, respectively.

From 2001 to 2004, he was a Teacher with the Communications Engineering Teaching and Research Office, Chengdu University of Information Technology. From November 2015 to December 2016, he was a Visiting Scholar with the Broadband Communication Research Group (BBCR), University of Waterloo, Waterloo, ON, Canada. Since 2011, he has been an Associate Professor with the School of Electrical Engineering and Electronic Information, Xihua University, Chengdu. He has authored more than 50 articles and more than 20 Chinese inventions. His research interests include detection and estimation, massive MIMO systems, and deep learning in physical layer of wireless communications.



CHUANGUI RAO received the B.S. degree in electrical engineering and information from Southwest Petroleum University, China, in 2017. He is currently pursuing the M.S. degree with the School of Electrical Engineering and Electronic Information, Xihua University, Chengdu, China, under the supervision of Prof. Qing. His research interests include frame synchronization and channel estimation, and deep learning applications in physical layer of wireless communications.



WANG YU received the B.S. degree from the School of Electrical Engineering and Electronic Information, Xihua University, Chengdu, China, in 2017, where he is currently pursuing the M.S. degree, under the supervision of Prof. Qing. His research interests include frame synchronization and frequency synchronization in wireless communications, and deep learning in physical layer of wireless communications.



JIAFAN WANG (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the University of Electronic Science and Technology of China, in 2006 and 2009, respectively, and the Ph.D. degree in computer engineering from Texas A&M University, College Station, TX, USA, in 2017.

He is currently working with Synopsys Inc., to develop the world's leading silicon chip design software in electronic design automation (EDA) industry. His major field of study is smart integrated circuit design, which includes multi-dimensional non-deterministic gate implementation with systematic optimization framework, self-training analog/digital mixed-system for circuit feature calibration, and configurable locking mechanism against analog IP piracy.

...