Learning for Detection: MIMO-OFDM Symbol Detection through Downlink Pilots

Zhou Zhou, Lingjia Liu, *Senior Member, IEEE*, and Hao-Hsuan Chang

***Abstract*— In this paper, we introduce a reservoir computing (RC) structure, namely, windowed echo state network (WESN), for multiple-input-multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) symbol detection. We show that adding buffers in input layers is able to bring an enhanced short-term memory (STM) to the standard echo state network. A unified training framework is developed for the introduced WESN MIMO-OFDM symbol detector using both comb and scattered patterns, where the training set size is compatible with those adopted in 3GPP LTE/LTE-Advanced standards. Complexity analysis demonstrates the advantages of WESN based symbol detector over state-of-the-art symbol detectors when the number of OFDM sub-carriers is large, where the benchmark methods are chosen as linear minimum mean square error (LMMSE) detection and sphere decoder. Numerical evaluations suggest that WESN can significantly improve the symbol detection performance as well as effectively mitigate model mismatch effects using very limited training symbols.**

***Index Terms*— Machine learning, OFDM, MIMO, symbol detection, recurrent neural network, reservoir computing, echo state network, limited training sets.**

I. Introduction

MULTIPLE-INPUT-MULTIPLE-OUTPUT, orthogonal frequency-division multiplexing (MIMO-OFDM) is the dominant wireless access technology for 4G and 5G cellular networks. MIMO technology introduces additional spatial degrees of freedom and enables various multi-antenna transmission strategies such as transmit diversity, spatial multiplexing, and multi-user MIMO operations [1] to improve overall network performance. To achieve the spatial-multiplexing gain in MIMO systems, different data streams are transmitted from different antennas causing inter-streams interference at the receiver. Accordingly, the symbol detection of multiple transmitted symbols from receiving antennas becomes critical for MIMO to realize its promise. In general, MIMO detection is classified into coherent detection and non-coherent detection [2]. In the coherent MIMO detection, the instantaneous channel matrix is obtained at the receiver through explicit channel estimation. In this way, a two-step approach is adopted where the instantaneous channel matrix is estimated in the first step while MIMO symbol detection is conducted in the second step based on the estimated channel matrix as well as the received signals. On the other hand, in the non-coherent detection, the channel estimation is either performed implicitly or is completely avoided where differential encoding is usually applied on input symbols leading to higher computational complexity. Therefore, most modern systems use coherent MIMO detection.

OFDM combats the effect of frequency-selective fading by breaking a wide-band channel into multiple orthogonal flat-fading narrow-band channels to significantly simplify the transceiver architecture. However, the underlying time-domain waveform usually has a high peak-to-average power ratio (PAPR) leading to reduced power amplifier's (PA) efficiency. Meanwhile, it produces input signal excursions into the PA's non-linear operation region resulting in signal distortions and spectral regrowth [3]. The non-linear distortion has a significant negative impact on the MIMO-OFDM channel estimation and symbol detection. To address the non-linear distortion as well as the clipping noise, additional system resources are required to recover the distortion [4]. Alternatively, the digital pre-distortion (DPD) can be introduced ahead of the PA to compensate for PA's non-linearity effects [5]. Note that a perfect knowledge of PA modeling and measurement bias is required for the DPD-based compensation. However, obtaining this knowledge is very challenging in reality [6]. Therefore, it is desirable to have a robust MIMO-OFDM symbol detector against the non-linear distortion.

Artificial neural networks (NN) as an emerging technology provides new aspects for communication systems [7]. For instance, an auto-encoder is introduced in [8], [9] to conduct the symbol modulation. However, this end-to-end learning strategy often relies on a good channel model to facilitate the application. [10] employs recurrent neural network (RNN) as the receiver in molecular communication systems, where the underlying channel model is not available. Furthermore, in optical fiber systems [11]–[13], NNs are utilized as channel equalizers as well as network monitors. In [11], the introduced method is verified through a lab experiment.

In light of the challenges in MIMO-OFDM symbol detection, NNs provide an ideal framework to conduct symbol detection even under the non-linear distortion. In [14], a deep neural network (DNN) is introduced for

Manuscript received June 25, 2019; revised October 12, 2019, December 13, 2019, and February 9, 2020; accepted February 9, 2020. Date of publication March 2, 2020; date of current version June 10, 2020. This work was supported in part by the U.S. National Science Foundation (NSF) under Grant ECCS-1811497 and Grant CCF-1937487. The associate editor coordinating the review of this article and approving it for publication was C.-K. Wen. *(Corresponding author: Lingjia Liu.)*

The authors are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: ljliu@ieee.org).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2020.2976004

OFDM symbol detection without using explicit channel state information (CSI) where the offline training is conducted using available channel statistics. A NN-based method is introduced in [15] for the receiver design of a cyclic prefix (CP)-free OFDM system and a fully connected NN-based OFDM receiver is tested over the air [16]. However, none of these works investigate MIMO-OFDM systems. A DNN-based detector is introduced through unfolding the standard belief propagation algorithm [17]. The parameters of the underlying DNN are required to be trained for different antenna configurations in an offline manner. In [18], the feature of residual signals after layered processing is used to construct a NN for symbol detection. Meanwhile, the loss function is conducted on multiple layers in order to avoid the gradient vanishing [19]. It demonstrates the introduced network can perform as well as the spherical decoding while achieving lower computational complexity. However, these methods require pre-known CSI as the coefficients or input of the underlying NN which cannot be perfectly obtained when non-linear distortion presents.

From the aforementioned examples, feedforward neural networks are employed for symbol detection by dividing the received signal into independent batches. On the other hand, communication signals are usually temporally correlated. RNNs allow us to learn the temporal dynamic behaviors [20] making it a better tool for symbol detection. For standard RNNs, the coefficients are often calculated via backpropagation through time (BPTT) [21]. However, when the sequence is inherent with long-range temporal dependencies, the training cannot converge due to the vanishing and exploding gradient, i.e., a small change at the current iteration can result in a very large deviation for later iterations [22]. To resolve the issue, RNNs are introduced with specific structures, such as the long short-term memory network [23] which uses "memory units" and "gating units" to control the gradient flow in order to avoid the gradient vanishing. Although RNNs with resolved gradient issues have been considered for communication system design in [10] and [12], these methods rely on a large training set for a well-fitted RNN model. However, the available training set for cellular networks especially in the physical layer is usually very limited due to the fact that the size of the training set is associated with the underlying system control overhead. For example, in 3GPP LTE/LTE-Advanced systems, the pilot overhead is specified and is fixed for different MIMO configurations [24]: The training set (demodulation reference signals) for SISO-OFDM is around 5% of all the resource elements. On the other hand, for a 2×2 MIMO-OFDM system, the overhead for reference signals is around 10%. Therefore, how to effectively conduct RNN-based symbol detection for cellular networks under very limited training sets becomes important for realizing the promise of RNN in practical wireless networks.

Rooted in the backpropagation-decorrelation learning rule, reservoir computing (RC) is a type of RNNs, where the gradient issues of RNN training can be avoided. More importantly, it can offer high computational efficiency with very limited training set [25]. This is achieved by conducting learning only on the output layer, whereas the untrained layers are sampled from a well-designed distribution. This makes RC an ideal tool for conducting symbol detection for cellular networks where the training set is extremely limited. In fact, an RC-based MIMO-OFDM symbol detector is first introduced in our previous work [26], [27]. With limited training set, [26] shows that the RC-based symbol detector can effectively combat the non-linear distortion caused by PA. However, our previous introduced RC-based symbol detector has limited performance using practical pilot patterns, such as the reference signal defined in LTE/LTE-Advanced standards. Since the wireless channel memory can introduce multi-path interference to the received signal, it motivates us to consider if an RC-based symbol detector with additional short term memory (STM) can improve the interference cancellation performance. Thus, the windowed echo state network (WESN) is introduced. The contributions of our paper are summarized as follows

- We incorporated buffers[1] in the input layer of RC, i.e, WESN. Through theoretical analysis, we showed that the added buffers can improve the short-term memory of the underlying RC. Numerical evaluations also demonstrate a positive correlation between the detection performance and the improved short-term memory: WESN with improved short-term memory can perform better interference cancellation. A trade-off between the buffer length and the size of neurons is identified.
- We introduced a unified training for WESN-based on the pilot patterns that are compatible with the demodulation reference signal (DMRS) adopted in LTE/LTE-Advanced standards. In this way, we are able to demonstrate the fact that the introduced symbol detector can be effective under a very limited training set. To the best of our knowledge, this is the first work in the literature of conducting machine learning-based symbol detection using LTE/LTE-Advanced compatible pilot patterns. Meanwhile, we demonstrated WESN can detect symbols using non-orthogonal pilots through numerical evaluations.
- We analyzed the complexity of the WESN-based symbol detector compared to conventional MIMO-OFDM receivers, such as linear minimum mean square error (LMMSE) and sphere decoding which is an approximation to the maximum likelihood estimator [28], [29]. The results suggest that the WESN-based detector has less computational complexity than conventional methods, especially when a large number of sub-carriers are utilized.

The structure of this paper is organized as follows: In Section II, the system model of MIMO-OFDM and conventional symbol detection methods are introduced. Meanwhile, the basic knowledge of RC is reviewed. Section III contains the WESN-based MIMO-OFDM symbol detector as well as pilot structures. In addition, the analysis of the short term memory of WESN is presented in this section. The complexity comparison between conventional methods and the WESN-based method is discussed in Section IV. Section V conducts the performance evaluation. Finally, Section VI concludes remarks and future work.

[1]the buffer represents a linear shift register without any feedback tap

II. System Model and Preliminaries

A. Channel Model and Transmitter Architecture

We consider a point-to-point MIMO-OFDM system, where the number of transmitter (Tx) and receiver (Rx) antennas are denoted as N_t and N_r. The ith OFDM symbol of the pth Tx antenna can be expressed as

$$u_i^{(p)}(t) = \sum_{n=0}^{N_c-1} x_i^{(p)}[n] \exp(2\pi jnt/\Delta t), \quad t \in [i\Delta t, (i+1)\Delta t), \tag{1}$$

where $x_i^{(p)}[n]$ is the transmitted symbol at the nth sub-carrier, N_c stands for the number of sub-carriers, Δt is the time length of one OFDM symbol.[2] At the qth antenna, the corresponding received OFDM symbol is given by

$$y_i^{(q)}(t) = \sum_{p=0}^{N_t-1} h_i^{(q,p)}(t) \circledast g(u_i^{(p)}(t)) + n(t), \tag{2}$$

where $n(t)$ represents the additive noise, $\circledast$ stands for the circular convolution which is translated by the circular prefix of an OFDM symbol, $g(\cdot)$ is a general function of the waveform distortion which is discussed later in this section, and $h_i^{(q,p)}(t)$ is the channel response from the pth Tx antenna to the qth Rx antenna for the ith OFDM symbol.

Equivalently, the signal in Eq. (2) can be rewritten in the digital frequency domain as

$$\tilde{y}_i^{(q)}[n] = \sum_{p=0}^{N_t-1} \tilde{h}_i^{(p,q)}[n]\tilde{g}^{(p)}[n] + \tilde{n}[n], \tag{3}$$

where $\tilde{n}[n]$ is the additive noise in the frequency domain, and

$$\tilde{g}^{(p)}[n] = \int_{\Delta t} g(u_i^{(p)}(t)) e^{-2\pi jtn/\Delta t} dt \tag{4}$$

$$\tilde{h}_i^{(p,q)}[n] = \int_{\Delta t} h_i^{(p,q)}(\tau) e^{-2\pi nj\tau/\Delta t} d\tau. \tag{5}$$

When we set $g(z_i^{(p)}(t)) = z_i^{(p)}(t)$, we have

$$\tilde{y}_i^{(q)}[n] = \sum_{p=0}^{N_t-1} \tilde{h}_i^{(p,q)}[n] x_i^{(p)}[n] + \tilde{n}[n]. \tag{6}$$

In the OFDM system, after the waveform is converted into the analog domain, it passes through RF circuits, such as power amplifiers, filters, and delay lines. These analog components are usually nonlinear systems due to practical constraints (e.g., circuit spaces and power consumption). For instance, the input-output relation of the power amplifier (PA) can be represented using the RAPP model [6]:

$$g(u(t)) = \frac{G_0 u(t)}{\left[1 + \left(\frac{|u(t)|}{u_{sat}}\right)^{2p}\right]^{1/2p}} \tag{7}$$

where $u(t)$ is the input signal of PA, G_0 stands for the power gain of PA, u_{sat} is the saturation level, and $p > 0$ is the smooth factor. The corresponding operational region of the PA model is shown in Fig. 1, which is divided into three parts: a linear region $|u(t)| \ll u_{sat}$, a non-linear region $|u(t)| \sim u_{sat}$, and a saturation region $|u(t)| \gg u_{sat}$. Even though the signal waveform is perfectly retained in the linear region, the power efficiency is low. Therefore, to reduce the distortion while maintaining relatively high efficiency, the PA operational point is set closely to the nonlinear region. Meanwhile, due to the high PAPR of OFDM signal, PAPR reduction is also employed to guarantee a certain level of PA efficiency [3]. However, improving PA efficiency will lead to the deficiency in transmission reliability due to the underlying waveform distortion. In this paper, we denote the resulting distortion as a function $g(\cdot)$.

[2]For simplicity, the index t used in this paper can represent both analog and digital time index based on the context. When the t is related to digital processing components, an ADC is assumed as a prior to the processing. Otherwise, it represents the analog domain time index.

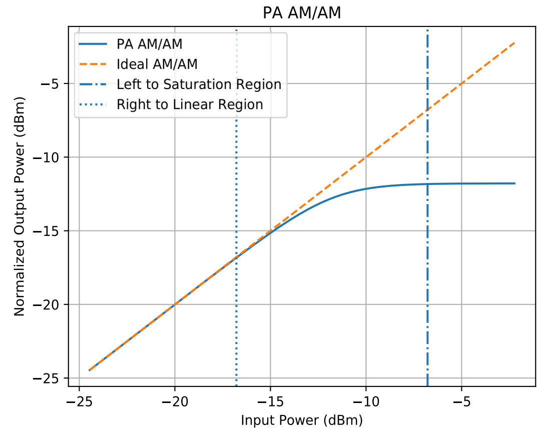


Fig. 1. The input and output amplitude (AM/AM) curve of PA: $p = 3$ and $|u_{sat}|^2 = -11.78dB$.

B. Conventional Methods

Coherent symbol detection methods are conducted by two steps: channel estimation and symbol detection. In the channel estimation, a series of pre-known pilots $\bar{x}_i^p[n]$ is sent to Rx, where $i \in \Omega_t$, $p \in \Omega_s$, $n \in \Omega_f$ in which Ω_t, Ω_s and Ω_f respectively represent the pilot index sets of OFDM symbols, antennas, and sub-carriers. Specifically, in LTE/LTE-Advanced systems, the design of pilot patterns is based on resource blocks (RBs) as shown in Fig. 2. For single input single output (SISO) OFDM systems, the pilot structures are depicted in Fig. 2 (a). The first sub-figure illustrates that Ω_t equals to the first OFDM symbol and Ω_f occupies all the sub-carriers. This comb pattern can be applied to the block fading channel assumption which is used in [26]. The size of Ω_f can be further reduced as shown in the second sub-figure of Fig. 2 (a) where the channel interpolation can be incorporated using frequency coherence. In the third subfigure of Fig. 2 (a), the scattered pilot pattern is applied on a Doppler channel which facilitates the channel tracking with very limited pilot overhead. For the MIMO channel, the pilot pattern is shown in Fig. 2 (b) and 2 (c). In Fig. 2 (b), the pilot symbols at different antenna ports are non-overlapping since they are allocated to different OFDM symbols. In Fig. 2 (c), the cross marker represents the null pilot symbols. Therefore, the pilot

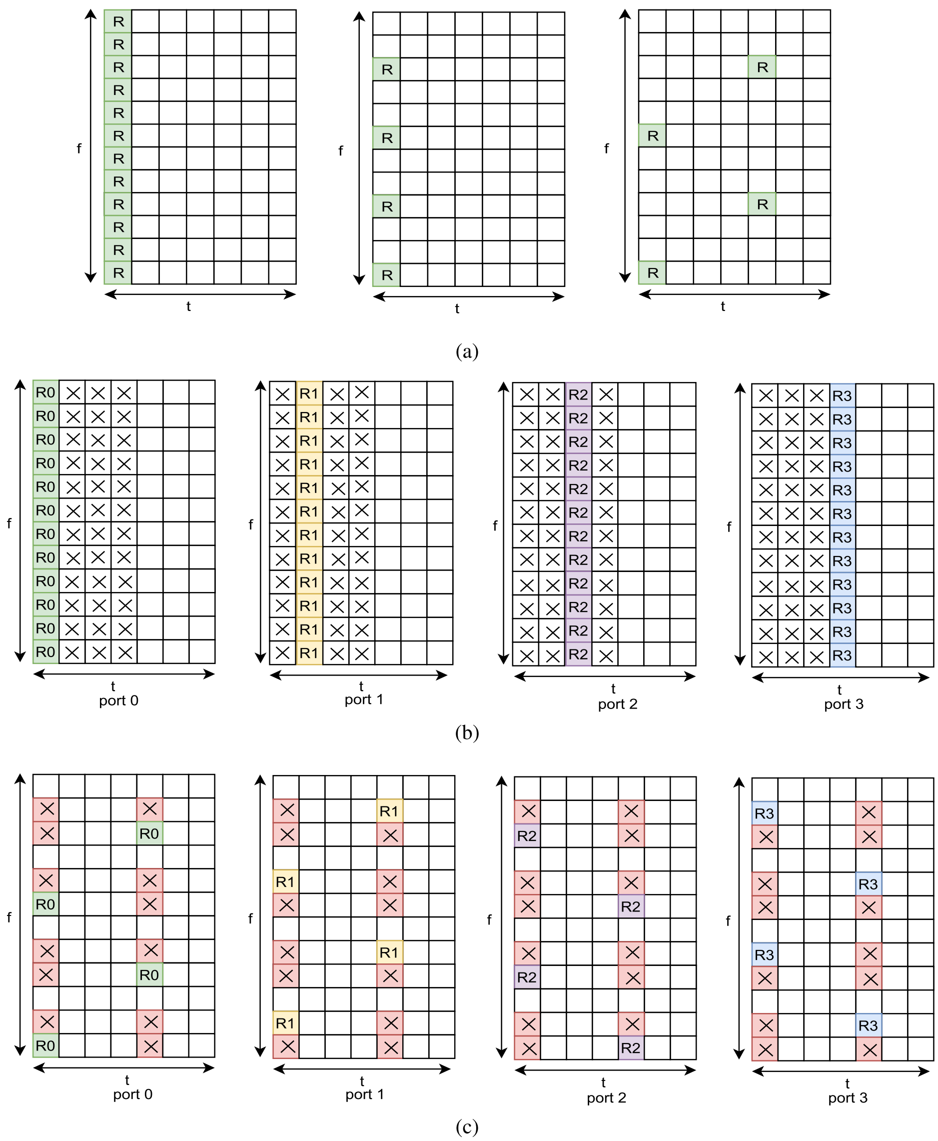


Fig. 2. Pilot structure per RB (a) SISO-OFDM (b) comb structured MIMO-OFDM (c) scattered structured MIMO-OFDM.

interference is eliminated during the channel estimation stage for MIMO.

Using pilots, the channel coefficients on the corresponding resource elements (REs) are obtained through (3) by solving

$$\min_{\tilde{h}_i^{(p,q)}[n]} l(\tilde{y}_i^{(q)}[n], \bar{x}_i^{(p)}[n] | (i,p,n) \in \Omega_t \times \Omega_s \times \Omega_f) \quad (8)$$

where $l(\cdot)$ is a pre-defined loss-function, such as likelihood function, mean square error, etc.. The CSI on the rest of the REs is inferred through an interpolation method. By substituting the estimated $\hat{h}_i^{(p,q)}[n]$ into (3), the rest symbols $\{x_i^{(p)}[n] | (i,p,n) \in (\Omega_t \times \Omega_s \times \Omega_f)^c\}$ (where Ω^c stands for the complementary set of Ω) are estimated using

$$\min_{x_i^p[n]} l(\tilde{y}_i^{(q)}[n], \hat{h}_i^{(p,q)}[n] | (i,p,n) \in (\Omega_t \times \Omega_s \times \Omega_f)^c). \quad (9)$$

However, the optimal solutions for (8) and (9) are not usually guaranteed due to the nonlinear distortion $g(\cdot)$. An improper assumption on $g(\cdot)$ can cause a model mismatch which deteriorates the accuracy on solving the channel estimation (8) and the symbol detection (9). To circumvent this dilemma, i.e., the dependence on the model assumption, RC-based method can be employed as an alternative approach.

C. Reservoir Computing

RC is one category of RNNs which consists of an input mapping, a fixed dynamic system, and a trained readout network. In general, there are two types of RC network architectures: echo state network (ESN) and liquid state machine (LSM). The network architecture of ESN [25] is illustrated in Fig. 3. The underlying network dynamics can be described by the following equation

$$\boldsymbol{s}(t+1) = f_{states}(\boldsymbol{W}'[\boldsymbol{y}^T(t+1), \boldsymbol{s}^T(t), \boldsymbol{x}^T(t)]^T), \quad (10)$$

where $\boldsymbol{s}(t) \in \mathbb{C}^{N_n}$ represents the inner states, N_n is the number of neurons inside the reservoir, $\boldsymbol{y}(t)$ is the

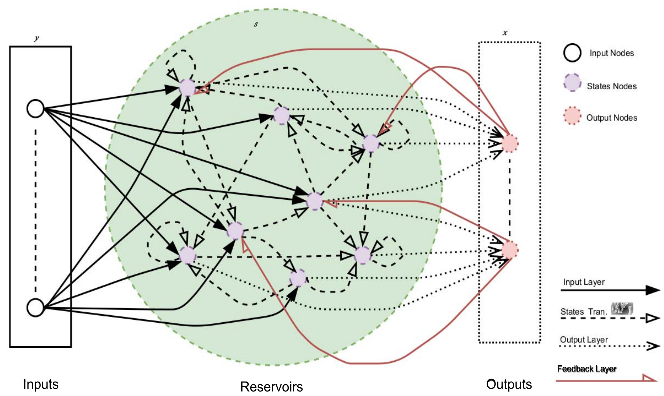


Fig. 3. An example of reservoir computing, the echo state network.

input signal, f_{states} represents the states activation function, $\boldsymbol{W}' = [\boldsymbol{W}_{in}, \boldsymbol{W}_1, \boldsymbol{W}_{f_1}]$, where $\boldsymbol{W}_{in}$ is weights of the input layer, $\boldsymbol{W}_1$ is the inner state transition weights, and $\boldsymbol{W}_{f1}$ is weights of the feedback layer. Moreover, $\boldsymbol{W}_{f1}$ can be omitted when feedback is not required. The output equation is given by

$$\boldsymbol{x}(t+1) = f_{out}(\boldsymbol{W}_{out}\boldsymbol{s}^T(t+1)) \tag{11}$$

where f_{out} is the activation function, and $\boldsymbol{W}_{out}$ represents the output layer. $\boldsymbol{W}'$ is designed following the *echo state property*.

Definition 1: We consider an ESN following the state transition of (10). Given an input sequence $\boldsymbol{y}(t)$ and two finite initial states $\boldsymbol{s}_1(0)$ and $\boldsymbol{s}_2(0)$, for any $\epsilon > 0$ and $\boldsymbol{y}(t)$, if we have $\|\boldsymbol{s}_1(t) - \boldsymbol{s}_2(t)\| < \epsilon$ when $t > \kappa(\epsilon)$, where $\xi(\epsilon)$ is a positive number, then the ESN satisfies the echo state property. Nevertheless, the echo state property of a given ESN cannot be easily justified from the above definition. For ease of application, the following sufficient condition is usually applied.

Theorem 1 (Proposition 3 in [25]): Assume an ESN with $\tanh(\cdot)$ as the activation function. If the maximum singular value of the inner states transition weight matrix $\boldsymbol{W}$ is smaller than 1, i.e., $\sigma(\boldsymbol{W})_{max} < 1$, then for all input $\boldsymbol{y}(t)$ and initial states $\boldsymbol{s} \in [-1, 1]^N$, the ESN satisfies the echo state property.

Learning output weights, $\boldsymbol{W}_{out}$, contains the following:

- Generation of the states trajectory: By feeding the training input $\{\bar{\boldsymbol{y}}(t)\}_{t=0}^T$ into ESN with target $\{\bar{\boldsymbol{x}}(t)\}_{t=0}^T$, the states set $\{\bar{\boldsymbol{s}}(t)\}_{t=0}^T$ is obtained by (10), where T represents the sequence length of the training input.
- Regression on the output weights: Substituting the generated states $\{\bar{\boldsymbol{s}}(t)\}_{t=0}^T$ into (11), we can calculate the weights $\boldsymbol{W}_{out}$ through

$$\min_{\boldsymbol{W}_{out}} L(\{\bar{\boldsymbol{y}}(t)\}_{t=0}^T, \{f_{out}(\boldsymbol{W}_{out}\bar{\boldsymbol{s}}^T(t))\}_{t=0}^T). \tag{12}$$

Specifically, when we choose f_{out} as an identity function, L as Frobenius norm, the output weights can be solved by

$$\min_{\boldsymbol{W}_{out}} \sum_{t=0}^{T} \|\bar{\boldsymbol{y}}(t) - \boldsymbol{W}_{out}\bar{\boldsymbol{s}}(t)\|_F^2 \tag{13}$$

which has a closed-form solution as follows

$$\boldsymbol{W}_{out} = \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^+, \tag{14}$$

where $\bar{\boldsymbol{Y}} = [\bar{\boldsymbol{y}}(0), \cdots, \bar{\boldsymbol{y}}(T)]$, $\bar{\boldsymbol{S}} = [\bar{\boldsymbol{s}}^T(0), \cdots, \bar{\boldsymbol{s}}^T(T)]$, and $\bar{\boldsymbol{S}}^+$ is the Moore-Penrose inverse of $\bar{\boldsymbol{S}}$.

III. Symbol Detection

A. Neural Network-Based Approach

The neural network-based symbol detection consists of two steps: training and testing. In the training stage, base station (BS) sends pre-defined symbols $\{\bar{x}_i^{(p)}[n] | (i,p,n) \in \Omega_t \times \Omega_s \times \Omega_f\}$ to mobile stations (MSs). Then, MSs train a neural network receiver $\mathcal{D}$ by solving

$$\min_{\mathcal{D}} f(\mathcal{D}(\bar{y}_i^{(p)}(t)), \bar{x}_i^{(p)}[n] | (i,p,n) \in \Omega_t \times \Omega_s \times \Omega_f), \tag{15}$$

where $f(\cdot)$ is the training objective function and $\mathcal{D}$ is the neural network; $\bar{y}_i^{(p)}(t)$ represents the received signal at a MS in the training stage. For instance, $f(\cdot)$ can be mean squared error or cross-entropy; $\mathcal{D}$ can be fully connected, convolutional or recurrent neural networks. In the testing stage, the symbols are estimated by feeding the observation $y_i^{(p)}(t)$ to the learned neural network $\hat{\mathcal{D}}$, i.e., $\hat{\mathcal{D}}(y_i^{(p)}(t))$. Consequently, the symbol detection performance and implementation complexity are determined by the utilized neural network and learning method. However, in wireless communications, the resources allocated to pilots are much less than the transmitted data symbols. Therefore, overfitting can occur if the adopted NN structure is unconformable.

B. Windowed Echo State Network

1) ESN Short Term Memory: For RNN, the output features are expected as a function of the memory encoded from inputs. A longer memory allows wider time-spanned features to be learned. Intuitively, the memory size can be characterized as the ability of recovering historical inputs. Thus, the memory capacity of ESN is defined as follows:

Definition 2 (Short Term Memory [30]): Given an ESN with fixed coefficients of the inner state transient matrix, input layer, and activation function, we first define the following *self-delay reconstruction correlation*

$$d(m, \boldsymbol{w}_{out}) = \frac{cov(y(n-m), x(n))}{\sigma(y(n-m))\sigma(x(n))}, \tag{16}$$

where $\boldsymbol{w}_{out}$ is the output weight for the ESN with a single input and a single output; With a slight abuse of notations, in this subsection, n represents the time sequence index, m is the input delay degree, and $x(n)$ is the ESN output when input is $y(n)$. Then, relying on the self-delay reconstruction correlation, we have the following definitions,

- The m-th delay STM capacity:

$$MC_m = \max_{\boldsymbol{w}_{out}} d(m, \boldsymbol{w}_{out}). \tag{17}$$

- The STM capacity:

$$MC = \sum_{m=1,2,\ldots} MC_m. \tag{18}$$

Remark that the above definition is only for ESN with a single input and single output. The general definition of STM for ESN with multiple inputs and multiple outputs is obtained by

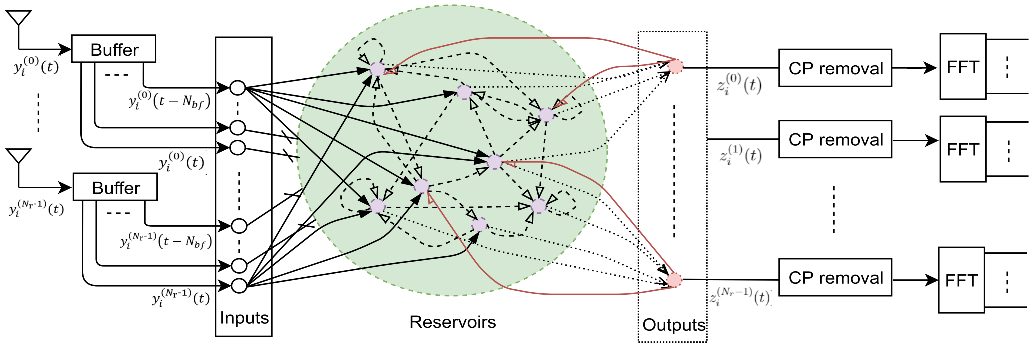


Fig. 4. The architecture of WESN-based MIMO-OFDM symbol detector.

extending the concept to each input-output pair. Furthermore, the metric in (17) can be approximately calculated through a self-delay training procedure defined as follows: 1) Input the zero mean sequence $\{y(n)\}_{n=0}^{N-1}$ to ESN; 2) Train the output $\{x(n)\}_{n=m}^{N-1}$ using the target $\{y(n)\}_{n=0}^{N-m-1}$, where $x(n) = \boldsymbol{w}_{out}\boldsymbol{s}(n)$ and $\boldsymbol{s}(n)$ is the state of the ESN. Therefore, the *self-delay reconstruction correlation* can be rewritten as

$$d(m, \boldsymbol{w}_{out}) = \frac{\sum_{n=0}^{N-m-1} y(n)x(n+m)}{\sqrt{\sum_{n=0}^{N-m-1}|y(n)|^2}\sqrt{\sum_{n=m}^{N-1}|x(n)|^2} \tag{19}$$

$$\propto -\|\tilde{\boldsymbol{x}}(m:N-1) - \tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2, \tag{20}$$

where $\propto$ stands for in a relation of proportionality; $\tilde{\boldsymbol{x}}(m : N-1)$ is a normalized vector stacked by the samples from $x(m)$ to $x(N-1)$; and $\tilde{\boldsymbol{y}}(0 : N-m-1)$ is stacked by samples from $y(0)$ to $y(N-m-1)$. According to the output equation of ESN in (11), $\tilde{\boldsymbol{x}}(m : N-1)$ can be expressed as

$$\tilde{\boldsymbol{x}}(m:N-1) = \boldsymbol{w}_{out}\tilde{\boldsymbol{S}}, \tag{21}$$

where $\tilde{\boldsymbol{S}} = [\tilde{\boldsymbol{s}}^T(m), \tilde{\boldsymbol{s}}^T(m+1), \cdots, \tilde{\boldsymbol{s}}^T(N-1)]$ and $\tilde{\boldsymbol{s}}(n)$ denotes the scaled states such that $\tilde{\boldsymbol{s}}(n) = \boldsymbol{s}(n)/\|\boldsymbol{w}_{out}\boldsymbol{s}(n)\|_2$. From the above definition, we can obtain the STM capacity of the buffer as follows

Theorem 2: The memory capacity of a buffer is greater than M, where M is the buffer's size.

Proof: With buffers we have $MC_m = 1$ if $0 \leq m \leq M$. When $m > M$, we have $MC_m \geq 0$ as the signal can be self-correlated. Therefore, we have $MC_W \geq M$. ■

Furthermore, we have the following upper bound for the STM capacity of ESN

Theorem 3 (Proposition 2 in [30]): The memory capacity of ESN is bounded by neuron number, i.e., $MC_{ESN} < N_n$.

Note that the above conclusion can only be made when the network is with an identity output activation and an i.i.d input. However, this theorem can give us a general guide on setting the number of neurons. Comparing Theorem 2 to Theorem 3, we see a buffer has a higher STM capacity than ESN when the buffer size is the same as the number of neurons of the ESN. However, a higher STM capacity does not necessary stand for a better nonlinear feature mapping ability. This is because reservoirs process the input history through a highly nonlinear recursive procedure rather than simply preserve the input. In our extension, adding a buffer at the input of ESN as depicted in Fig. 4, we can obtain the WESN. Its STM can be characterized using the following theorem.

Theorem 4: Assume the STM capacity of the buffer and the ESN are MC_W and MC, the STM capacity of the WESN, MC_{WESN}, is given by

$$\frac{1}{2}MC_{WESN} \geq \lambda MC_W + (1-\lambda)MC_{ESN}, \quad \lambda \in (0,1).$$

Proof: See Appendix for details. ■

The result suggests that WESN can achieve a higher STM capacity than the convex combination of the buffer and ESN.

C. WESN-based MIMO-OFDM receiver

The introduced WESN-based MIMO-OFDM symbol detector is shown in Fig. 4. The receiving link is concatenated by a WESN, a cyclic prefix (CP) removal and an FFT block, where the dimension of the WESN outputs is the same as the number of the transmission streams. The received ith OFDM symbol $\boldsymbol{y}_i(t) = [y_i^{(0)}(t), y_i^{(1)}(t), \cdots, y_i^{(N_r-1)}(t)]^T$ is first fed into the buffers. At the jth antenna's buffer, it collects N_{bf} samples from $y_i^{(j)}(t)$ to create a vector $[y_i^{(j)}(t-N_{bf}), y_i^{(j)}(t-N_{bf}+1), \cdots, y_i^{(j)}(t)]^T$. The vector is mapped into reservoirs through the input layer. Reservoirs update their inner states and generate an output vector $\boldsymbol{z}_i(t) = [z_i^{(0)}(t), z_i^{(1)}(t), \cdots, z_i^{(N_r-1)}(t)]^T$, where $\boldsymbol{z}_i(t) \in \mathcal{C}^{N_r}$ and $\mathcal{C}$ represents the modulation constellation. Finally, it converts $\boldsymbol{Z}_i$ into the frequency domain, where $\boldsymbol{Z}_i = [\boldsymbol{z}_i(0), \boldsymbol{z}_i(1), \cdots, \boldsymbol{z}_i(N_c-1)] \in \mathbb{C}^{N_r \times N_c}$, and maps the frequency signal into modulation symbols according to the constellation $\mathcal{C}$, i.e., $\mathcal{Q}_{\mathcal{C}}(\boldsymbol{Z}_i\boldsymbol{F})$, where $\boldsymbol{F}$ is the Fourier matrix.

D. Training of WESN

We begin by considering the training of the WESN receiver under the SISO channel without Doppler shifts, i.e., $f_D = 0$. As discussed in Sec. II-B, the first OFDM symbol contains the training set. According to (15), we select the objective function f as the Frobenius norm induced distance and $\mathcal{D}$ as the WESN. Using the ESN's dynamics and output equations in Sec. II-C, we have the output of WESN as $\boldsymbol{WS}$, where $\boldsymbol{S} \in \mathbb{C}^{(N_n) \times N_c}$ stands for the reservoir states, and $\boldsymbol{W} \in \mathbb{C}^{1 \times (1+N_n)}$ is the readout weights. With a slight generalization in our notations, here N_n stands for the number of neurons plus the length of

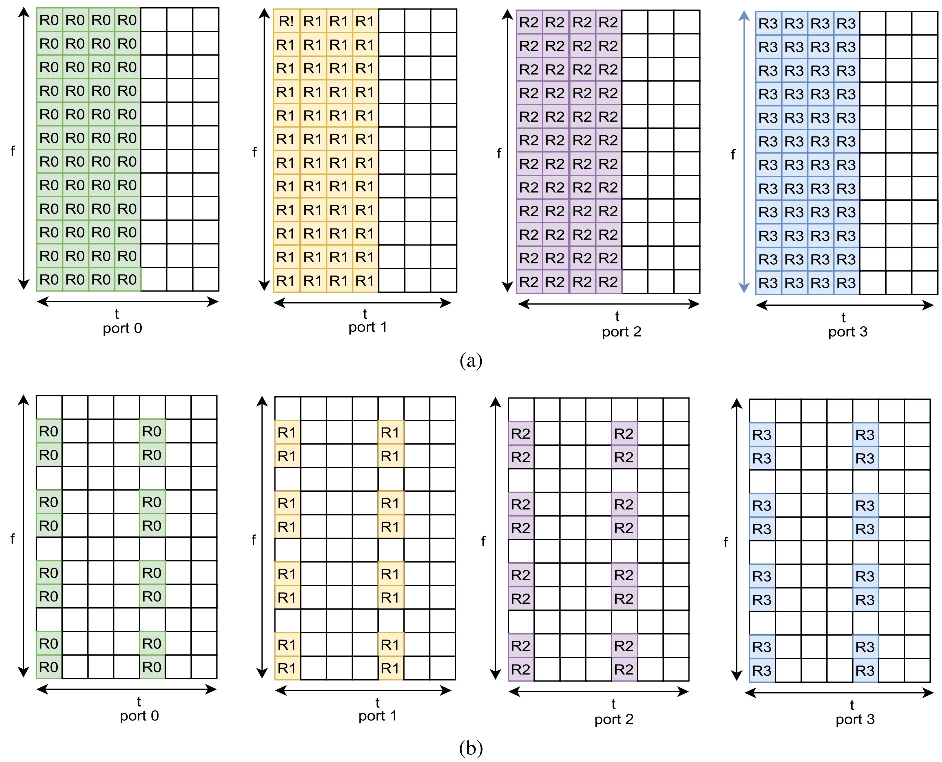


Fig. 5. The OFDM pilots structures for WESN in one RB: (a) block (b) scattered.

buffers. Therefore, similarly as (13), the readout weights of WESN are updated by solving

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}\boldsymbol{S}\boldsymbol{F} - \bar{\boldsymbol{x}}_0^T\|_2, \tag{22}$$

where $\boldsymbol{F} \in \mathbb{C}^{N_c \times N_c}$ represents the Fourier transform matrix, and $\bar{\boldsymbol{x}}_0 \in \mathcal{C}^{N_c}$ is the pilot symbols in which the subscript stands for the first OFDM symbol. The solution can be further written as the following closed-form,

$$\boldsymbol{W} \overset{(a)}{=} \bar{\boldsymbol{x}}_0^T(\boldsymbol{S}\boldsymbol{F})^+ \overset{(b)}{=} (\bar{\boldsymbol{x}}_0^T\boldsymbol{F}^H)\boldsymbol{S}^+, \tag{23}$$

where (a) holds when we assume the number of training symbols is greater than the number of neurons plus inputs. Alternatively, through (b), the weights learning can be interpreted as fitting the output of WESN to the waveform of the target OFDM symbols $\bar{\boldsymbol{x}}_0^T\boldsymbol{F}^H$.

We then extend the symbol detection method to the MIMO channel without Doppler shifts. Rather than SISO, the MIMO receiver needs to mitigate the inter-streams interference. To realize this, a tailored training pilot pattern is introduced, where Fig. 5a shows the case of $N_t = N_r = 4$. This pattern occupies the same number of REs as the comb structured MIMO-OFDM pilots in Fig. 2b. There is a slight difference between the two patterns: in Fig. 2b, the pilots from different antennas are orthogonal while those in Fig. 5a are overlapping. This is due to the fundamental difference between learning-based methods and conventional channel estimation-based methods: In learning-based methods, NNs need to learn the interference situation of the transmission; in conventional methods, received pilots should be interference-free to improve channel estimation performance.

By using this pilot pattern, the outputs of the WESN can be expressed as the matrix $\boldsymbol{Z} = [\boldsymbol{Z}_0, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3]$, where the subscripts represent the indices of the OFDM symbols allocated as pilots. Similarly, we have $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{S}$, where $\boldsymbol{S} = [\boldsymbol{S}_0, \boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3] \in \mathbb{C}^{N_n \times 4N_c}$ represents the state matrix of WESN. Thus, the output layer is solved by

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}\boldsymbol{S}\boldsymbol{F}' - \bar{\boldsymbol{X}}\|_2, \tag{24}$$

where $\boldsymbol{F}' = diag(\boldsymbol{F}, \boldsymbol{F}, \boldsymbol{F}, \boldsymbol{F}) \in \mathbb{C}^{4N_c \times 4N_c}$ is a block diagonal matrix in which the diagonal element is $\boldsymbol{F}$; $\bar{\boldsymbol{X}} = [\bar{\boldsymbol{X}}_0, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \bar{\boldsymbol{X}}_3] \in \mathcal{C}^{N_r \times 4N_c}$ is the pilot symbols. Accordingly, we have

$$\begin{aligned} \boldsymbol{W} &= \bar{\boldsymbol{X}}[\boldsymbol{S}_0\boldsymbol{F}, \boldsymbol{S}_1\boldsymbol{F}, \boldsymbol{S}_2\boldsymbol{F}, \boldsymbol{S}_3\boldsymbol{F}]^+ \\ &\overset{(a)}{=} [\bar{\boldsymbol{X}}_0\boldsymbol{F}^H, \bar{\boldsymbol{X}}_1\boldsymbol{F}^H, \bar{\boldsymbol{X}}_2\boldsymbol{F}^H, \bar{\boldsymbol{X}}_3\boldsymbol{F}^H][\boldsymbol{S}_0, \boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3]^+. \end{aligned} \tag{25}$$

From (a), we know that the weight learning can be conducted in the time domain as well.

Now, we consider MIMO channels with non-zero Doppler shifts. To be compatible with the conventional pilots design in SISO depicted in the third sub-figure of Fig. 2a, we directly utilize this scattered pilots pattern as the training set of WESN. Therefore, the weights of the outputs are updated by

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}[\boldsymbol{S}_0\boldsymbol{F}(:, \Omega_{f_0}), \boldsymbol{S}_4\boldsymbol{F}(:, \Omega_{f_4})] - [\bar{\boldsymbol{x}}_0^T(\Omega_{f_0}), \bar{\boldsymbol{x}}_4^T(\Omega_{f_4})]\|_2, \tag{26}$$

where Ω_{f_0} and Ω_{f_4} respectively represents the sub-carriers allocated to the pilot symbols at $t = 0$ and $t = 4$ in the figure. Alternatively, the above minimization problem can be expressed as

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}[\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}] - [\bar{\boldsymbol{x}}_{0,\Omega_{f_0}}^T, \bar{\boldsymbol{x}}_{4,\Omega_{f_4}}^T]\|_2, \quad (27)$$

where

$$\bar{\boldsymbol{x}}_{t,\Omega_f}(n) \triangleq \begin{cases} \bar{\boldsymbol{x}}_t(n), & n \in \Omega_f \\ 0, & n \notin \Omega_f \end{cases} \quad (28)$$

$$\boldsymbol{F}_{\Omega_f}(n) \triangleq \begin{cases} \boldsymbol{F}(n), & n \in \Omega_f \\ \boldsymbol{0}, & n \notin \Omega_f \end{cases}. \quad (29)$$

Therefore, the output weight is given by

$$\boldsymbol{W} = [\bar{\boldsymbol{x}}_{0,\Omega_{f_0}}^T, \bar{\boldsymbol{x}}_{4,\Omega_{f_4}}^T][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}]^+, \quad (30)$$

which can be rewritten as

$$\boldsymbol{W} = [\bar{\boldsymbol{x}}_{0,\Omega_{f_0}}^T\boldsymbol{F}^H, \bar{\boldsymbol{x}}_{4,\Omega_{f_4}}^T\boldsymbol{F}^H][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}\boldsymbol{F}^H, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}\boldsymbol{F}^H]^+, \quad (31)$$

where $\bar{\boldsymbol{x}}_{\Omega_f}^T\boldsymbol{F}^H$ represents the time domain OFDM waveform transformed merely from the symbols defined on the sub-carriers $\bar{\Omega}_f$. It demonstrates the output weight is also obtained by fitting the waveform of the scattered pilots. Similarly, using the MIMO scattered pilots in Fig. 5b, we have,

$$\boldsymbol{W} = [\bar{\boldsymbol{X}}_{0,\Omega_{f_0}}^T\boldsymbol{F}^H, \bar{\boldsymbol{X}}_{4,\Omega_{f_4}}^T\boldsymbol{F}^H][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}\boldsymbol{F}^H, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}\boldsymbol{F}^H]^+, \quad (32)$$

where $\bar{\boldsymbol{X}}_{t,\Omega_f}$ represents MIMO pilots similar to (28).

IV. Complexity Analysis

In this section, we compare the computational complexity of the RC-based symbol detector to the conventional methods discussed in Sec. II-B, where the complexity is evaluated by floating-point operations per second (FLOPS).

A. Single-Input-Single-Output Systems

1) Channel Estimation: For the conventional methods of solving the channel estimation problem (8), $g(\cdot)$ is assumed as a linear function. When $l(\cdot)$ is chosen as mean squared error (MSE), we branch the discussion according to the pilot patterns plotted in Fig. 2 (a). For the comb pilots, the objective function in (8) is rewritten as

$$\min_{\tilde{\boldsymbol{h}}} \mathbb{E}\|\tilde{\boldsymbol{y}}_i - \bar{\boldsymbol{x}}_i \odot \tilde{\boldsymbol{h}}\|_F^2, \quad (33)$$

where $\odot$ denotes the Hadamard product. From [31], we know that the solution is given by

$$\tilde{\boldsymbol{h}} = \boldsymbol{R}_{hy}\boldsymbol{R}_{yy}^{-1}\boldsymbol{y}_i, \quad (34)$$

where $\boldsymbol{R}_{hy} = \boldsymbol{F}\boldsymbol{R}_{hh}\boldsymbol{F}^H\bar{\boldsymbol{X}}_i^H$, $\boldsymbol{X}_i = \mathrm{diag}(\boldsymbol{x}_i)$, $\boldsymbol{R}_{yy} = \boldsymbol{X}_i\boldsymbol{F}\boldsymbol{R}_{hh}\boldsymbol{F}^H\boldsymbol{X}_i^H + \sigma^2\boldsymbol{I}$, and $\boldsymbol{R}_{hh}$ is the channel covariance matrix. For the scattered pilots, the channel coefficients on the time-frequency grids allocated as pilots are calculated by

$$\min_{\boldsymbol{h}(\Omega_f)} \mathbb{E}\|\tilde{\boldsymbol{y}}_i[\Omega_f] - \bar{\boldsymbol{X}}_i(\Omega_f)\boldsymbol{h}[\Omega_f]\|_F^2, \quad (35)$$

which has a closed-form solution as follows

$$\boldsymbol{h}[\Omega_f] = \boldsymbol{R}_{hY}(\Omega_f)\boldsymbol{R}_{yy}(\Omega_f)^{-1}\tilde{\boldsymbol{y}}_i[\Omega_f], \quad (36)$$

where $\boldsymbol{R}_{hY}(\Omega_f) = \boldsymbol{F}(\Omega_f,:)\boldsymbol{R}_{hh}\boldsymbol{F}(\Omega_f,:)^H\bar{\boldsymbol{X}}_i(\Omega_f)^H$ and $\boldsymbol{R}_{yy}(\Omega_f) = \boldsymbol{X}_i(\Omega_f)\boldsymbol{F}(\Omega_f,:)\boldsymbol{R}_{hh}\boldsymbol{F}(\Omega_f,:)^H\boldsymbol{X}(\Omega_f,:)_i^H + \sigma^2\boldsymbol{I}$. The channel gains on the rest grids are inferred by the interpolation as discussed in [32]. When the channel tap is assumed to be uncorrelated i.e., $\boldsymbol{R}_{hh} = \boldsymbol{I}$, we have

$$\tilde{h}[n] = \bar{x}_i^*[n] \cdot \tilde{y}_i[n]/(|\bar{x}_i[n]|^2 + \sigma^2), \quad (37)$$

where n stands for the index of sub-carriers.

2) Symbol Detection: For the symbol detection problem (9), when $l(\cdot)$ is selected as MSE, we have

$$\min_{\boldsymbol{x}_i} \mathbb{E}\|\tilde{\boldsymbol{y}}_i - \boldsymbol{x}_i \odot \hat{\boldsymbol{h}}_i\|_F^2. \quad (38)$$

When the transmission symbols are uncorrelated between sub-carriers, (38) becomes

$$\min_{x_i[n]} \sum_{n=0}^{N_c-1} \mathbb{E}|\tilde{y}_i[n] - x_i[n]\hat{h}_i[n]|^2,$$

which has a following solution

$$\hat{x}_i[n] = \hat{h}_i^*[n] * \tilde{y}_i[n]/(|\hat{h}_i[n]|^2 + \sigma^2). \quad (39)$$

3) Complexity: For complexity analysis, we first review the FLOPS of standard matrix operations. Given two matrices $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{C}^{n\times p}$, the matrix product $\boldsymbol{AB}$ requires $N_{FLOPS}(\boldsymbol{AB}) = 2mnp$ for the summations and additions. For any invertible matrix $\boldsymbol{C} \in \mathbb{C}^{n\times n}$, FLOPS of the inverse is $N_{FLOPS}(\boldsymbol{C}^{-1}) = n^3 + n^2 + n$. When $\boldsymbol{C} \in \mathbb{C}^{m\times n}$ is with full column rank, FLOPS of the MP-inverse $\boldsymbol{C}^+$ is given by $3mn^2 + 2n^3$. Therefore, for the comb pilot pattern in Fig. 2b, the FLOPS of the LMMSE channel estimation (34) is $2N_c^2$, in which the calculation of the covariance matrices $\boldsymbol{R}_{hy}$ and $\boldsymbol{R}_{yy}$ are ommited. In the symbol detection stage (39), the FLOPS is proportional to N_c. Thus, the total FLOPS for the LMMSE channel estimation plus the symbol detection is on the scale of $\delta N_c^2 + (1-\delta)N_c$, where δ represents the ratio of the pilot symbols to all the transmission symbols in the OFDM system. Moreover, when we consider the scattered pilot pattern in Fig. 2c, the complexity of interpolation needs to be included. For the standard linear interpolation method, the FLOPS is on the scale of $7N_c(1-\kappa)$, where κ is the ratio of pilot sub-carriers over all sub-carriers. Thus, the total FLOPS for the LMMSE channel estimation with LMMSE symbol detection using scattered pilot is $\delta(\kappa N_c)^2 + \delta 7N_c(1-\kappa) + (1-\delta)N_c + \delta(1-\kappa)N_c$.

For the ESN/WESN using comb pilots, according to (23), the FLOPS for the output weights learning is $2N_c(N_n+1) + 3N_cN_n^2 + 2N_n^3$. Meanwhile, the computation at the symbol detection stage is merely on the output layer mapping, where the FLOPS is N_nN_c. Thus, the overall FLOPS for the ESN/WESN-based symbol detection is $\delta(2N_c(N_n+1) + 3N_cN_n^2 + 2N_n^3) + (1-\delta)N_nN_c$. For scattered pilots, FLOPS at the learning stage is $2(\kappa N_c)(N_n+1) + 3\kappa N_cN_n^2 + 2N_n^3$. Therefore, the total number of FLOPS is proportional to $\delta(2(\kappa N_c)(N_n+1) + 3\kappa N_cN_n^2 + 2N_n^3) + N_nN_c$. It indicates the resulting complexity of the ESN/WESN receiver is linearly proportional to the number of subcarriers. It suggests that the ESN/WESN has less computational burden than the LMMSE method when the number of subcarriers is large. Remark that we do not consider the computations inside the reservoirs in this analysis. This is because the reservoirs are usually implemented through analog circuits which perform faster than the digital circuit [33], [34] with less energy consumption.

B. Multiple-Input-Multiple-Output Systems

By using the comb and scattered pilots respectively plotted in Fig. 2b and Fig. 2c, the FLOPS of the LMMSE channel estimation on each antenna pair is the same as the SISO case due to free interference. Therefore, the complexity of the MIMO channel estimation is $N_t N_r$ times more than the SISO case. However, for the symbol detection, the interference caused by multiple transmitted antennas are required to be annihilated. Thus, the MIMO symbol detection demands more computations than the SISO case.

Now, we consider the LMMSE MIMO symbol detection using (9). When the transmitted symbols on different sub-carriers are independent, the symbol detection can be conducted in sub-carrier-wise. Therefore, at the nth sub-carrier of the tth OFDM symbol, the symbol detection is solved by

$$\min_{\tilde{\boldsymbol{x}}_i(n)} \mathbb{E}\|\tilde{\boldsymbol{y}}_i(n) - \hat{\boldsymbol{H}}_i(n)\tilde{\boldsymbol{x}}_i(n)\|_F^2, \tag{40}$$

which has the following closed-form solution

$$\tilde{\boldsymbol{x}}_i(n) = (\hat{\boldsymbol{H}}_i^H(n)\hat{\boldsymbol{H}}_i(n) + \sigma^2\boldsymbol{I})^{-1}\hat{\boldsymbol{H}}_i^H(n)\tilde{\boldsymbol{y}}_i(n). \tag{41}$$

It leads the FLOPS to $2N_c(N^3+N^2+N)$, where N denotes the number of antennas at Tx and Rx when $N_t = N_r$.

For the MIMO sphere decoding, it is an approximation of solving the following maximum likelihood estimation,

$$\min_{\boldsymbol{x}_i(n)\in\mathcal{C}^{N_r}} \|\tilde{\boldsymbol{y}}_i(n) - \hat{\boldsymbol{H}}(n)\tilde{\boldsymbol{x}}(n)\|_2, \tag{42}$$

where $\mathcal{C}$ represents the modulation constellation of the transmitted symbols. Since the standard sphere decoding usually has high redundancy in the implementation. We choose a complexity reduced sphere decoding algorithm proposed in [35] for the evaluation. It shows that the FLOPS is proportional to $N_c|\mathcal{C}|^N(2N^2+2N-1)$ which implies the sphere decoding is extremely complicated when a high order modulation is adopted. Using the comb pilot for ESN/WESN, the FLOPS for output weight learning is $2N^2N_c(N_n+1)+3N_cNN_n^2+2N_n^3$ according to (25), where the number of training OFDM symbols is the same as the transmission antennas. At the symbol detection stage, the FLOPS is N_cNN_n. Similarly, we can calculate the FLOPS using the scattered pilots. The results of complexity comparison are summarized in Table I. We see that the computational complexity of ESN/WESN is dominated by the number of neurons which is smaller than N_c through the numerical experiments in Sec. V.

V. Performance Evaluation

In this section, we evaluate the performance of the WESN-based symbol detection. Through our numerical experiments, we incorporate the model of RF circuits, such as up/downsamplers, PA, and anti-interference/alias filters into the link simulation. To simulate the analog domain, we apply a four times up-sampling upon the baseband signal. We assume that the channel is given by the following tap-delay model:

$$h_i^{(q,p)}(\tau) = \sum_{l=0}^{L-1} a_i^{(p,q)}(l)p_w(\tau-\tau_l), \tag{43}$$

where L is the maximum number of resolvable paths and $p_w(\tau)$ is the pulse shaping function which is chosen as the ideal rectangular shaped filter in the frequency domain. At the lth delay tap, we assume $a_i(l)$ is generated by the circular Gaussian distribution,

$$a_i(l) \sim \mathcal{NC}(0, \sigma_l^2),$$

where σ_l^2 is assumed to be an exponential power delay profile, i.e., $\sigma_l^2 = exp(-\alpha\tau_l/\tau_{\max})$. Moreover, between two adjacent OFDM symbols, the correlation is assumed to be

$$\mathbb{E}(a_i(l)a_{(i+1)}(l)) = \sigma_l^2 J_0(2\pi f_D \Delta t), \tag{44}$$

where J_0 stands for the Bessel function of the first kind with parameter 0. Note that, for simplicity, we set the path-coefficients for any two different Tx-Rx antenna pairs to be independent. In general, other spatial correlation models

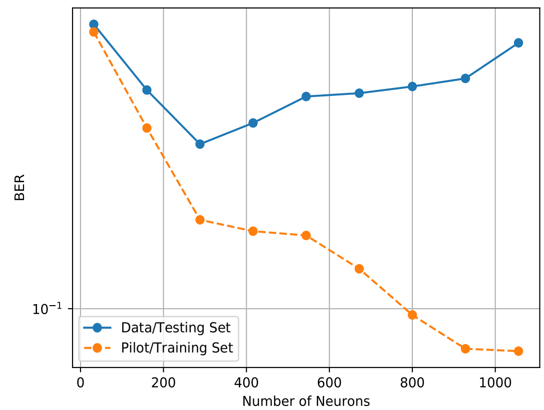


Fig. 6. Overfitting issue of ESN-based Symbol detector.

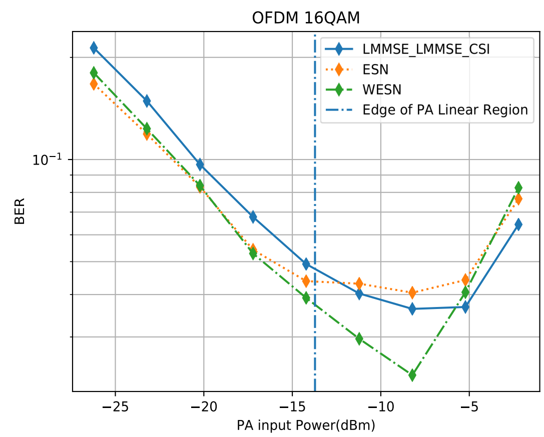
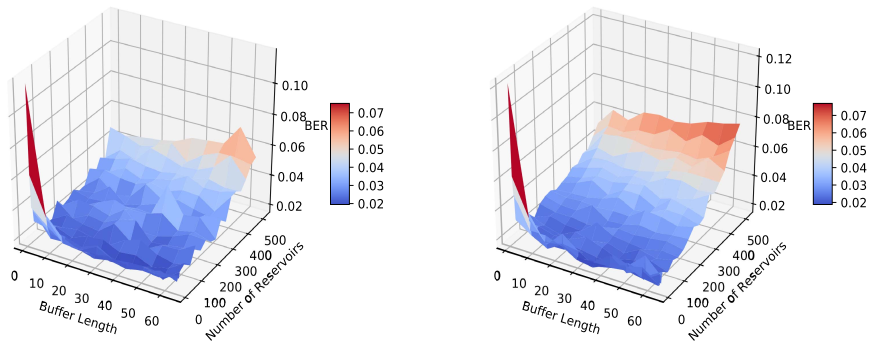


Fig. 7. BER comparison of ESN symbol detector, WESN symbol detector and LMMSE method under the SISO block fading channel: the number of neurons is 64 and the length of buffers is 30.

TABLE I
COMPUTATIONAL COMPLEXITY OF SYMBOL DETECTION METHODS

Symbol Detection Method	Number of FLOPS
SISO LMMSE CSI with LMMSE	$\delta(\kappa N_c)^2 + \delta 7N_c(1-\kappa) + (1-\delta)N_c + \delta(1-\kappa)N_c$
SISO ESN/WESN	$\delta(2(\kappa N_c)(N_n+1) + 3\kappa N_c N_n^2 + 2N_n^3) + N_n N_c$
MIMO LMMSE CSI with LMMSE	$N^2(\delta(\kappa N_c)^2 + \delta 7N_c(1-\kappa)) + (1-\kappa\delta)N_c(N^3+N^2+N)$
MIMO LMMSE CSI with SD	$N^2(\delta(\kappa N_c)^2 + \delta 7N_c(1-\kappa)) + (1-\kappa\delta)N_c\lvert\mathcal{C}\rvert^N(2N^2+2N-1)$
MIMO ESN/WESN	$\delta(2N^2\kappa N_c(N_n+1) + 3\kappa N_c N N_n^2 + 2N_n^3) + N_c N N_n$



(a) (b)

Fig. 8. Average BER of WESN symbol detector under SISO block fading channel by varying the length of buffers (from 1 to 64) and the number of neurons (from 8 to 512):(a) 3D surface when the PA input power is -8 dBm, (b) 3D surface when the PA input power is -11 dBm.

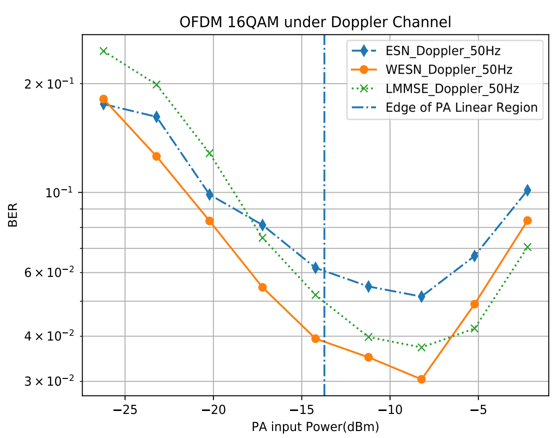


Fig. 9. BER comparison of ESN symbol detector, WESN detector and LMMSE method under SISO Doppler channels with different Doppler shifts, the number of neurons is 64 and the length of buffers is 30.

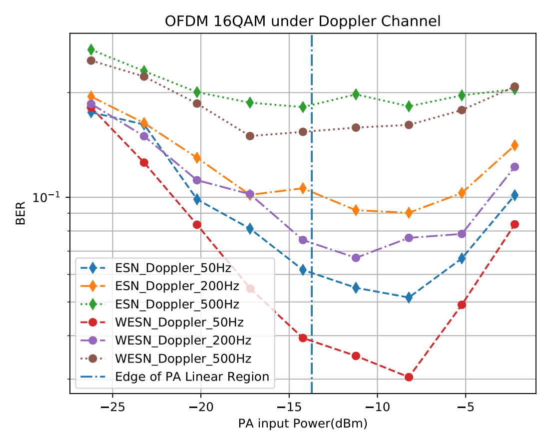


Fig. 10. BER comparison between ESN symbol detector and WESN symbol detector under SISO Doppler channels with different Doppler shifts: the number of neurons is 64 and the length of buffers is 30.

or channel models also can be utilized without changing the training framework. The number of paths, L, in the channel model (43) is set as 6. The baseband modulation order is selected as 16-QAM. For the conventional methods using scattered pilots, the CSI is obtained by linear interpolation.

Overfitting is an important issue for NN-based approaches. Fig. 6 shows the BER of both the pilot/training set and the data/testing set of the ESN-based symbol detector as the number of neurons changes. It shows that the BER of the training set decreases as the model becomes more complicated. However, the BER gap between testing and training set is enlarged as the number of neurons increases. Therefore, a proper selection on the number of neurons is needed to achieve low generalization error (i.e., low BER on testing set).

A. Single-Input-Single-Output Systems

We first evaluate the WESN receiver in the SISO channel under different operation regions of PA. Fig. 7 shows the BER results when the Doppler shift is 0Hz, where the threshold for the PA linear region is set as 3dB up to the boundary of the linear region as depicted in Fig. 1. Here the ESN is referred to as the buffer length of WESN is set as 1. The number of neurons for ESN and WESN is chosen to be the same, 64. The buffer length of WESN is set as 30. For the labeled “LMMSE-LMMSE-CSI” method, the symbol detection is conducted by LMMSE using the CSI obtained from the LMMSE channel estimation. We can observe that these three methods have comparable performance among the linear region. Moreover, for WESN, the BER performance is the best in PA nonlinear region when the optimal PA input power

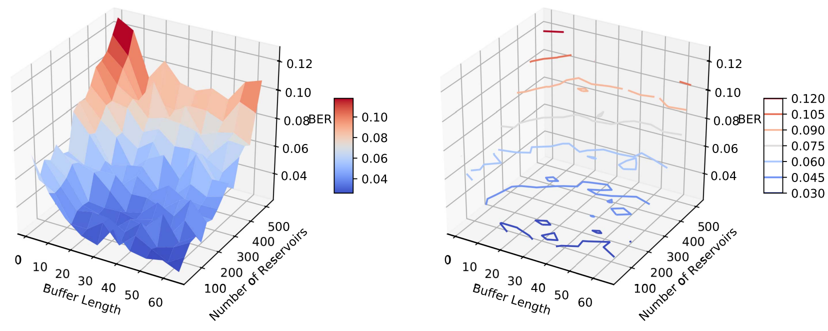


Fig. 11. Average BER performance of the WESN symbol detector under the SISO Doppler channel by varying the length of buffers and the number of neurons when the PA input power is -8 dBm: (a) 3D surface (b) 3D contour version, where the number of neurons varies from 8 to 512, the length of buffers ranges from 1 to 64 and the Doppler shift is 50 Hz.

is selected. It demonstrates that the WESN can considerably compensate for the non-linear waveform distortion. We can also conclude that the symbol detection using the estimated CSI does not necessarily lead to the optimality in BER performance.

Nevertheless, the performance of the WESN receiver is highly related to the settings of neural network parameters, especially the number of internal reservoirs and the buffer length. We further investigate how the length of buffer and the number of neurons can jointly impact the BER performance. In Fig. 8. we observe that the length of buffer brings another degree of freedom to improve the symbol detection performance. From this figure, it shows that by either increasing the number of neurons or the length of buffer, the resulting BER declines. However, due to overfitting, BER increases again when the number of neurons becomes greater. Furthermore, it shows that compared to the WESN configured with more neurons, the WESN with a few numbers of neurons but longer buffers can achieve the same performance. This is because the memory capacity of WESN is jointly determined by the configuration of neurons and buffers. Furthermore, we see the overfitting issue in Fig. 8b is slightly different from that in Fig. 8a. When the number of neurons is large (such as close to 500), the BER in Fig. 8b is higher than that in Fig. 8a. This is because when the input power is closer to the linear region, the transmitted signal is less distorted. Therefore, the size of the employed neural network is expected to be smaller. On the other hand, using more neuron states (more complicated models) can result in worse BER performance (overfitting) when the input power is close to the linear region.

The BER performance under different Doppler shifts in the SISO channel is shown in Fig. 9. We see that these three methods are comparable in the BER as well. In Fig. 10, the comparison between ESN and WESN under different Doppler shifts is investigated. We can see that the WESN always performs better than the ESN under different Doppler shifts. From Fig. 11, we again investigate the BER distribution by varying buffer length and neurons number. We see that increasing buffer size can significantly decrease BER which indicates that WESN can get more advantages over

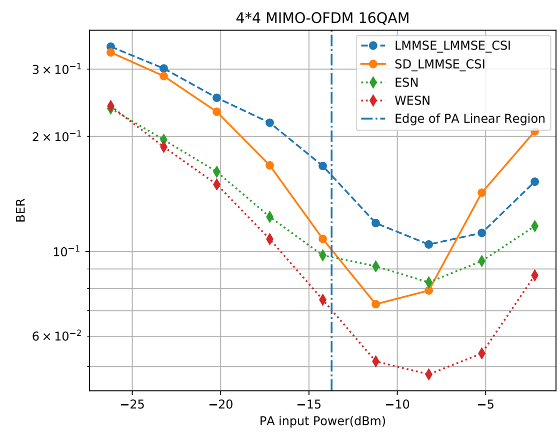


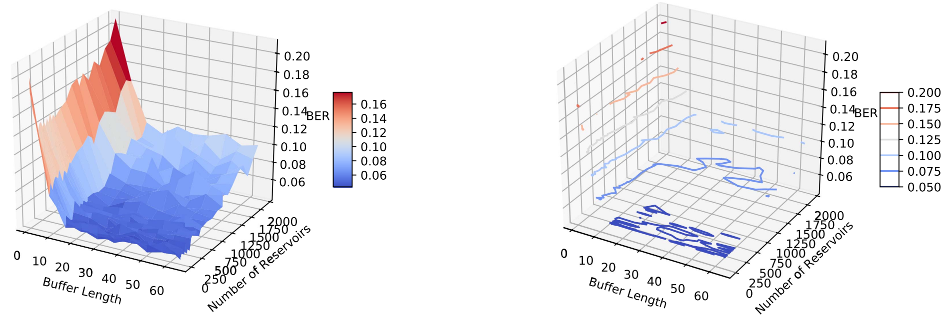
Fig. 12. BER comparison of ESN symbol detector, WESN symbol detector, LMMSE method and sphere decoding under MIMO block fading channels (the number of neurons is 64 and the length of buffers is 30).

the Doppler shift channel compared to the standard ESN. Meanwhile, adding more neurons can always lead to model overfitting.

B. Multiple-Input-Multiple-Output Systems

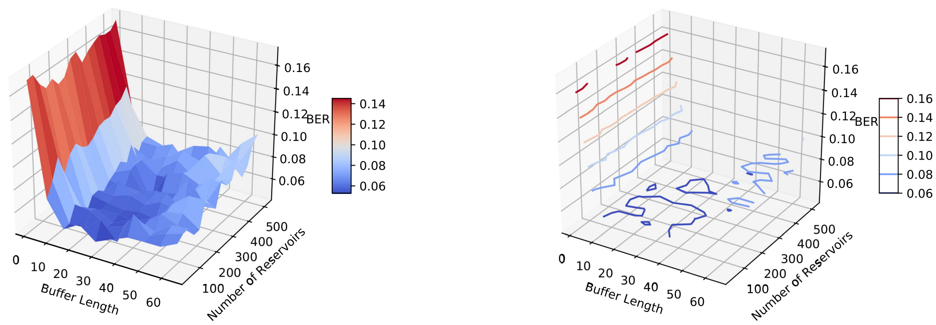
In Fig. 12, we compare the BER performance of WESN to conventional methods, i.e., LMMSE and sphere decoding (SD) under block fading channels where conventional methods obtain CSI through LMMSE using pilot patterns depicted in Fig. 2c. We see that the performance gap between WESN and the conventional methods is enlarged compared to the SISO case. For SD the BER performance deteriorates quickly when the PA input power is in the non-linear region. This is because SD requires more accurate CSI for symbol detection.

Again, we plot the BER distribution by varying the buffer length and the number of neurons as shown in Fig. 13. The advantages of the introduced buffer are more obvious compared to the SISO case by looking at Fig. 8. Moreover, by using the pilot pattern in Fig.5a, the number of pilot symbols in training can be flexibly adjusted. In Fig. 15, we show the BER performance by varying the number of pilots, i.e., the number of OFDM symbols allocated as pilots. To be clarified,



(a) (b)

Fig. 13. Average BER performance of the WESN symbol detector under the MIMO block fading channel by varying the length of buffers and the number of neurons when the PA input power is -8 dBm: (a) 3D surface (b) 3D contour version, where the number of neurons varies from 8 to 512 and the length of buffers ranges from 1 to 64.



(a) (b)

Fig. 14. Average BER performance of WESN symbol detector under the MIMO Doppler channel by varying the length of buffers (from 1 to 64) and the number of neurons (from 8 to 512), the PA input power is -8 dBm, and the Doppler shift is 50 Hz: (a) 3D surface, (b) 3D contour version.

Fig. 5a shows the number of OFDM pilot symbols is equal to 4. Specifically, when $T < 4$, it is non-orthogonal pilots as the number of pilot OFDM symbols is smaller than the number of Tx antennas, 4. When we employ the conventional methods, using non-orthogonal pilot is not enough to avoid the pilot interference during the channel estimation stage. This means the conventional channel estimation method cannot be directly applied using non-orthogonal pilots. However, by using the RC-based method, we can observe that the BER performance is almost invariant compared to orthogonal pilots. It is because that the learning-based symbol detection can extract important features underlying the channel which are the inherent sparsity in the time-delay domain. Meanwhile, by increasing the number of neurons, we can also observe the deterioration of the BER performance due to overfitting.

In Fig. 16, we plotted the performance using the scattered pilot of MIMO under the Doppler shift channel. The 2D BER distribution under the Doppler channel is shown in Fig. 14 which has a similar distribution as Fig. 13. All these results suggest that the RC-based approach can outperform conventional methods in low SNR regime and under nonlinear distortion. Furthermore, note that the WESN/ESN symbol detector is trained using compatible pilot patterns of LTE/LTE-Advanced systems, making it completely different from most of the existing literature, such as [14], [17], [18]. In fact, almost all other work in the field assumes a large training set to train the underlying neural networks for symbol detection while we are focusing on using the extremely limited training overhead provided by LTE/LTE-Advanced networks. As shown in the 3rd sub-figure of Fig. 2(a), the training set (demodulation reference signals) for SISO-OFDM is around 5% of all the REs: one resource block has $12 \times 7 = 84$ REs with 4 of them being reference signals. On the other hand, for 4×4 MIMO-OFDM systems, one resource block has $12 \times 7 \times 4 = 336$ REs with 16 of them being reference signals (the overhead is around 20%). To evaluate activation functions other than the identity function and to compare the introduced symbol detector with other neural network-based approaches under limited training symbols, we conduct the following:

- We use the long short-term memory (LSTM) [36] to replace WESN in the receiver. The LSTM configuration follows the standard interface provided by Keras [37]: The activation function of the recurrent step is chosen as the hard sigmoid; other activation functions are chosen as the hyperbolic tangent; bias is added in each layer; bias for the forget gate and other parts are all initialized as zero; the number of LSTM units is set as 64; bias regularizer, kernel and recurrent regularizer are not configured. Meanwhile, we utilize one dense layer as the output to read out cell states. The training object is selected as l_2 norm as well, where the target is the time-domain OFDM symbols. Furthermore, to make a fair comparison,

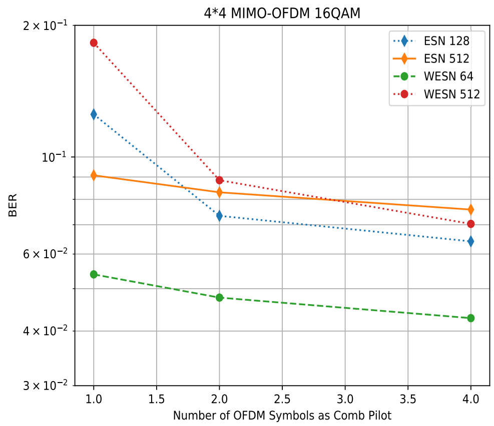


Fig. 15. BER performance of ESN symbol detector and WESN symbol detector under MIMO block fading channels by varying the number of pilots OFDM symbols, the PA input power is chosen as −9 dBm, the number of neurons for ESN is equal to 128 and 512, the number of neurons for WESN is equal to 64 and 512, and the length of buffers is 30.

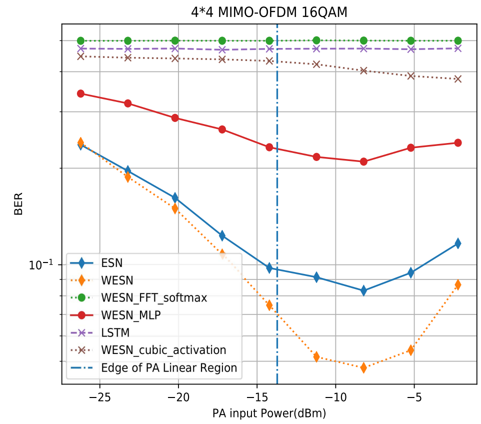


Fig. 17. BER comparison of the ESN symbol detector, the WESN symbol detector, the WESN-MLP detector, the WESN-FFT-softmax detector, the LSTM-based detector, and WESN with cubic output activation under the MIMO block fading channel, the number of neurons is set as 64 and the length of buffers is 30.

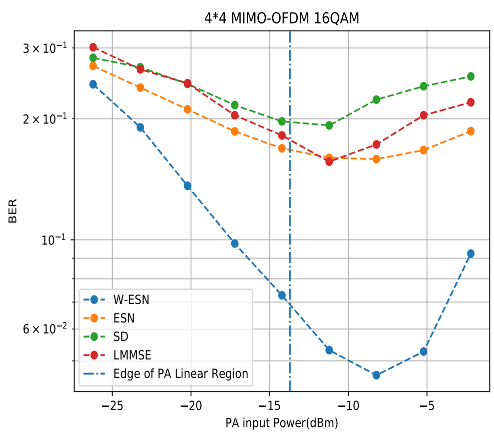


Fig. 16. BER comparison of ESN symbol detector, WESN symbol detector, LMMSE method and sphere decoding under MIMO Doppler channels: length of buffers is 30, number of neurons is 64, and Doppler shift is 50Hz.

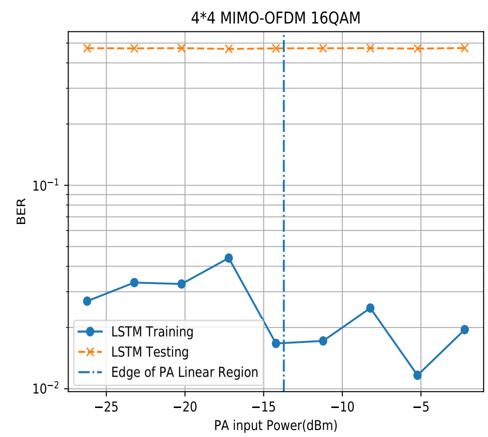


Fig. 18. Training and testing BER using LSTM in 4×4 MIMO-OFDM.

we add a sliding window in the same way as the WESN to increase the training batches of LSTM.

- We choose SoftMax associated with a one-hot coding as the learnable layer after the FFT of WESN's inner states. Each stream of a subcarrier is operated with one SoftMax instead of setting one SoftMax for all subcarriers.
- We add a multilayer perceptron (MLP) at the output of the WESN to allow an arbitrary activation function. The added MLP has three layers and "arctan" is selected as the activation of intermediate layers.
- We use a fixed nonlinear function, cubic function, as the output activation function of WESN to check the impact of changing the activation functions.

The same training set as the WESN-based receiver is used to evaluate all the above methods. Fig. 17 shows the performance of all these methods compared to the WESN/ESN approach. It can be seen that none of these methods are providing performance even close to the introduced WESN symbol detector. In fact, the identity output activation function of the WESN/ESN-based approach can be replaced by any other function which has a closed-form expression. However, the output design cannot be easily optimized under arbitrary fixed activation functions unlike the identity function case where the optimal output weights can be analytically characterized. Therefore, the performance of using cubic activation cannot be guaranteed. Even though the use of SoftMax and MLP will allow us to learn arbitrary output function, it significantly increases the number of trainable parameters which is not desirable for symbol detection under limited training sets. On the other hand, the available training symbols based on the reference signals/pilots provided by LTE/LTE-Advanced networks are too small to train a proper-fitted general neural network. That is why we see the LSTM performs poorly in this situation. To gain further insights on this, we also evaluated the BER for both training and testing of LSTM as shown in Fig. 18. From this figure, we see during the training stage, LSTM can achieve a good training BER. However, due to the limited training set, it cannot be generalized to a good BER performance at the testing stage.

VI. Conclusion and Future Work

In this paper, we consider the application of reservoir computing to MIMO-OFDM symbol detection. Compared with our previous work [26], a new RC-based detector, WESN, is introduced which can significantly improve the performance of interference cancellation and nonlinear compensation. As an advanced ESN, WESN is proved to be able to fundamentally enhance the short term memory. In addition, numerical evaluation demonstrates that WESN offers a great performance improvement over conventional approaches even using the same amount of pilots defined in 3GPP LTE standards under both static and dynamic MIMO channels. Moreover, through complexity analysis, we prove that WESN requires fewer FLOPS than conventional methods. For future work, this symbol detection method can be extended to a joint symbol demodulation and channel decoding framework. It is also interesting to explore other activation functions and neural network architectures such as extending the shallow RC architecture to deep RNNs.

Appendix

Proof of Theorem 4: The output weights for the m-th delay capacity can be calculated by

$$\min_{\boldsymbol{w}_{out}} \|\tilde{\boldsymbol{x}}(m:N-1)-\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2.$$

Suppose $\tilde{\boldsymbol{x}} = \boldsymbol{w}_{out}\boldsymbol{S}_{WESN}$, where $\boldsymbol{S}_{WESN} = [[\tilde{\boldsymbol{y}}^T(m : m-M), \tilde{\boldsymbol{s}}^T(m)]^T, [\tilde{\boldsymbol{y}}^T(m+1 : m+1-M), \tilde{\boldsymbol{s}}^T(m+1)], \cdots, [\tilde{\boldsymbol{y}}^T(N-1-M:N), \tilde{\boldsymbol{s}}^T(N-1)]]^T$ represents the extended states as introduced in [25]. By splitting $\boldsymbol{w}_{out}$ into $[\boldsymbol{w}_1, \boldsymbol{w}_2]$, we have

$$\begin{aligned}
&\|[\boldsymbol{w}_1,\boldsymbol{w}_2][\boldsymbol{Y}^T,\boldsymbol{S}_{ESN}^T]^T-\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\quad=\|\boldsymbol{w}_1\boldsymbol{Y}-\lambda\tilde{\boldsymbol{y}}(0:N-m-1)+\boldsymbol{w}_2\boldsymbol{S}_{ESN}\\
&\qquad-(1-\lambda)\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\quad\le 2\|\boldsymbol{w}_1\boldsymbol{Y}-\lambda\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\qquad+2\|\boldsymbol{w}_2\boldsymbol{S}_{ESN}-(1-\lambda)\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2,
\end{aligned}$$

where $\lambda \in (0,1)$. Thus,

$$\begin{aligned}
&\min_{\boldsymbol{w}_1,\boldsymbol{w}_2}\frac{1}{2}\|[\boldsymbol{w}_1,\boldsymbol{w}_2][\boldsymbol{Y}^T,\boldsymbol{S}_{ESN}^T]^T-\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\quad\le\min_{\boldsymbol{w}_1}\|\boldsymbol{w}_1\boldsymbol{Y}-\lambda\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\qquad+\min_{\boldsymbol{w}_2}\|\boldsymbol{w}_2\boldsymbol{S}_{ESN}-(1-\lambda)\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2\\
&\quad=\lambda^2\ r_W+(1-\lambda)^2 r_{ESN},
\end{aligned}$$

where

$$\begin{aligned}
r_W &= \min_{\boldsymbol{w}_1}\|(1/\lambda)\boldsymbol{w}_1\boldsymbol{Y}-\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2,\\
r_{ESN} &= \min_{\boldsymbol{w}_2}\|(1/(1-\lambda))\boldsymbol{w}_2\boldsymbol{S}_{ESN}-\tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2.
\end{aligned}$$

According to the definition of STM, we have

$$\begin{aligned}
\frac{1}{2}MC_{WESN} &\ge \lambda^2\ MC_W+(1-\lambda)^2 MC_{ESN}\\
&\ge \lambda^2\ MC_W+(1-\lambda^2)MC_{ESN}.
\end{aligned}$$

Finally, the theorem is proved by substituting λ^2 as λ.

References

[1] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou, "Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 140–147, Feb. 2012.

[2] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart., 2015.

[3] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1567–1592, 4th Quart., 2013.

[4] H. Chen and A. M. Haimovich, "Iterative estimation and cancellation of clipping noise for OFDM signals," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 305–307, Jul. 2003.

[5] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[6] J. Joung, C. K. Ho, K. Adachi, and S. Sun, "A survey on power-amplifier-centric techniques for spectrum- and energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 315–333, 1st Quart., 2015.

[7] R. Shafin, L. Liu, V. Chandrasekhar, H. Chen, J. H. Reed, and J. Zhang, "Artificial intelligence-enabled cellular networks: A critical path to beyond-5G and 6G," *IEEE Wireless Commun. Mag.*, to be published.

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

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

[10] N. Farsad and A. Goldsmith, "Neural network detectors for molecular communication systems," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[11] B. Karanov *et al.*, "End-to-End deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, Oct. 15, 2018.

[12] B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," *Opt. Express*, vol. 27, no. 14, pp. 19650–19663, Jul. 2019.

[13] F. N. Khan, C. Lu, and A. P. T. Lau, "Machine learning methods for optical communication systems," in *Proc. Adv. Photon. (IPR, NOMA, Sensors, Netw., SPPCom, PS)*, 2017, pp. 1–3, Paper SpW2F-3.

[14] 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.

[15] J. Zhang, C.-K. Wen, S. Jin, and G. Ye Li, "Artificial intelligence-aided receiver for a CP-free OFDM system: Design, simulation, and experimental test," 2019, *arXiv:1903.04766*. [Online]. Available: http://arxiv.org/abs/1903.04766

[16] P. Jiang *et al.*, "Artificial intelligence-aided OFDM receiver: Design and experimental results," 2018, *arXiv:1812.06638*. [Online]. Available: http://arxiv.org/abs/1812.06638

[17] X. Tan, W. Xu, Y. Be'ery, Z. Zhang, X. You, and C. Zhang, "Improving massive MIMO belief propagation detector with deep neural network," 2018, *arXiv:1804.01002*. [Online]. Available: http://arxiv.org/abs/1804.01002

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

[19] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

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

[21] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[22] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, pp. 1310–1318.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] *Physical Channels and Modulation*, document 3GPP TS 36.211, Rev. 13.2.0, 2016.

[25] H. Jaeger, "The Echo state approach to analysing and training recurrent neural networks-with an erratum note," *Bonn, Germany, German Nat. Res. Center Inf. Technol. GMD Tech. Rep.*, vol. 148, no. 34, p. 13, 2001.

[26] S. S. Mosleh, L. Liu, C. Sahin, Y. R. Zheng, and Y. Yi, "Brain-inspired wireless communications: Where reservoir computing meets MIMO-OFDM," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4694–4708, Oct. 2018.

[27] S. Mosleht, C. Sahint, L. Liut, R. Y. Zheng, and Y. Yit, "An energy efficient decoding scheme for nonlinear MIMO-OFDM network using reservoir computing," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1166–1173.

[28] R. Shafin *et al.*, "Realizing green symbol detection via reservoir computing: An energy-efficiency perspective," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[29] L. Barbero and J. Thompson, "Fixing the complexity of the sphere decoder for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2131–2142, Jun. 2008.

[30] H. Jaeger, "Short term memory in echo state networks," German Nat. Res. Inst. Comput. Sci. (GMD), Germany, Tech. Rep. 152, 2001, vol. 5.

[31] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[32] X. Dong, W.-S. Lu, and A. Soong, "Linear interpolation in pilot symbol assisted channel estimation for OFDM," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1910–1920, May 2007.

[33] F. Duport, A. Smerieri, A. Akrout, M. Haelterman, and S. Massar, "Fully analogue photonic reservoir computer," *Sci. Rep.*, vol. 6, no. 1, p. 22381, Mar. 2016.

[34] K. Vandoorne *et al.*, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Commun.*, vol. 5, no. 1, p. 3541, Mar. 2014.

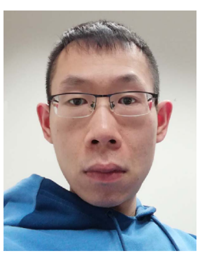
[35] R. Y. Chang, W.-H. Chung, and S.-J. Lin, "A algorithm inspired memory-efficient detection for MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 508–511, Oct. 2012.

[36] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. Int. Conf. Artif. Neural Netw.*, vol. 2, Sep. 1999, pp. 850–855.

[37] A. Gulli and S. Pal, *Deep Learning With Keras*. Birmingham, U.K.: Packt Publishing, 2017.

Lingjia Liu (Senior Member, IEEE) received the B.S. degree in electronic engineering from Shanghai Jiao Tong University and the Ph.D. degree in electrical and computer engineering from Texas A&M University. He was an Associate Professor with the EECS Department, University of Kansas (KU). He spent more than four years working at the Mitsubishi Electric Research Laboratory (MERL) and the Standards and Mobility Innovation Lab, Samsung Research America (SRA), where he received the Global Samsung Best Paper Award in 2008 and 2010. He was leading Samsung's efforts on multiuser MIMO, CoMP, and HetNets in LTE/LTE-Advanced standards. He is currently an Associate Professor with the Bradley Department of Electrical and Computer Engineering, Virginia Tech (VT), and the Associate Director of the Wireless@Virginia Tech. His general research interests mainly lie in emerging technologies for 5G cellular networks, including machine learning for wireless networks, massive MIMO, massive MTC communications, and mm-Wave communications.

He received Air Force Summer Faculty Fellowship from 2013 to 2017, Miller Scholarship at KU in 2014, the Miller Professional Development Award for Distinguished Research at KU in 2015, the 2016 IEEE GLOBECOM Best Paper Award, the 2018 IEEE ISQED Best Paper Award, and the 2018 IEEE TAOS Best Paper Award.



Zhou Zhou received the B.S. degree in communications engineering from the University of Electronic Science and Technology of China in 2011. Since 2018, he has been with the ECE Department, Virginia Tech, as a Research Assistant. His current research interests are in the broad area of neural networks, machine intelligence, and wireless communications.

Hao-Hsuan Chang received the B.Sc. degree in electrical engineering and the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2014 and 2015, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. His research interests include wireless communications, deep learning, and machine learning.