MIMO-OFDM Symbol Detection via Echo State Networks

Zhou Zhou

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical Engineering

Lingjia Liu, Chair
Richard M. Buehrer
Steven W. Ellingson

Oct 18, 2019
Blacksburg, Virginia

Keywords: Echo State Network, MIMO-OFDM, Symbol Detection, Bit Error Rate

MIMO-OFDM Symbol Detection via Echo State Networks

Zhou Zhou

(ABSTRACT)

Echo state network (ESN) is a specific neural network structure composed of high dimensional nonlinear dynamics and learned readout weights. This thesis considers applying ESN for symbol detection in multiple-input, multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems. A new ESN structure, namely, windowed echo state networks (WESN) is introduced to further improve the symbol detection performance. Theoretical analysis justifies WESN has an enhanced short-term memory (STM) compared with the standard ESN such that WESN can offer better computing ability. Additionally, the bandwidth spent as the training set is the same as the demodulation reference signals defined in 3GPP LTE/LTE-Advanced systems for the ESN/WESN based symbol detection. Meanwhile, a unified training framework is developed for both comb and scattered pilot patterns. Complexity analysis demonstrates the advantages of ESN/WESN based symbol detector compared to conventional symbol detectors such as linear minimum mean square error (LMMSE) and sphere decoder when the system is employed with a large number of OFDM sub-carriers. Numerical evaluations show that ESN/WESN has an improvement of symbol detection performance as opposed to conventional methods in both low SNR regime and power amplifier (PA) nonlinear regime. Finally, it demonstrates that WESN can generate a better symbol detection result over ESN.

MIMO-OFDM Symbol Detection via Echo State Networks

Zhou Zhou

(GENERAL AUDIENCE ABSTRACT)

Artificial neural networks (ANN) are widely used in recognition tasks such as recommendation systems, robotics path planning, self-driving, video tracking, image classifications, etc. To further explore the applications of ANN, this thesis considers using a specific ANN, echo state network (ESN) for a wireless communications task: MIMO-OFDM symbol detection. Furthermore, it proposed an enhanced version of the standard ESN, namely, windowed echo state network (WESN). Theoretical analyses on the short term memory (STM) of ESN and WESN show that the later one has a longer STM. Besides, the training set size of this ESN/WESN based method is chosen the same as the pilot symbols used in conventional communications systems. The algorithm complexity analysis demonstrates the ESN/WESN based method performs with lower complexity compared with conventional methods, such as linear mean square error (LMMSE) and sphere decoding. Comprehensive simulations examine how the symbol detection performance can be improved by using ESN and its variant WESN when the transmission link is non-ideal.

Contents

List of Figures

List of Tables

List of Abbreviations

4G The Fourth Generation Cellular Network Technology

5G The Fifth Generation Cellular Network Technology

ANN Artificial Neural Network

BER Bit Error Rate

BPTT Backpropagation Through Time

BS Base Station

CP Cyclic-Prefix

CSI Channel State Information

DMRS Demodulation Reference Signal

DPD Digital Pre-Distortion

ESN Echo State Network

FLOPS Floating Point Operations per Second

LMMSE Linear Minimum Mean Square Error

LTE Long Term Evolution

MIMO Multiple-Input and Multiple-Output

MS Mobile Station

OFDM Orthogonal Frequency-Division Multiplexing

PAPR Peak Average Power Ratio

RB Resource Blocks

RNN Recurrent Neural Network

Rx Receiver

SD Sphere Decoding

SISO Single Input Single Output

STM Short Term Memory

Tx Transmitter

WESN Windowed Echo State Network

Chapter 1

Introduction

In the fourth and fifth-generation (4G and 5G) cellular networks, multiple-input, multiple-output, orthogonal frequency-division multiplexing (MIMO-OFDM) is the primary air interface technology. The powerful combination of MIMO and OFDM, MIMO-OFDM, enables the wireless system to achieve high spectral-efficiency over broadband channels with simplified transceiver architectures. This is one of the main reasons why MIMO-OFDM is not only adopted in most modern wireless systems but also envisioned for future networks. MIMO brings further spatial degrees of freedom into the transmission link which enables the system having flexible transmission modes, such as spatial-multiplexing and transmit-receive diversity [17].

In spatial-multiplexing based transmission, different transmitting antennas send different data streams to the receiver simultaneously. Thus, the received signal at a given receiving antenna contains interference from other transmitting antennas. In order to successfully communicate, the interference needs to be canceled out through a symbol detection procedure. In other words, the symbol detection is a critical step for realizing the MIMO spatial multiplexing transmission. The methods of MIMO symbol detection can generally be categorized into coherent detection and noncoherent detection [32].

For coherent MIMO detection, a two-stage approach is deployed. In the first stage, the explicit model-based channel state information (CSI) is estimated using embedded pilots in the transmission signals. In the second stage, transmitted symbols are estimated via solving

an inverse problem using the estimated CSI obtained in the first stage. Alternatively, the noncoherent symbol detection uses implicit CSI or completely without CSI. However, this method heavily relies on channel statistics or other prior assumptions. Therefore, cellular wireless systems utilize coherent MIMO detection rather than the noncoherent way.

OFDM technology essentially converts a wide-band frequency-selective fading channel into narrowband orthogonal channels. This fundamentally reduces the hardware complexity of the underlying transceiver architecture and enables a high spectrum efficiency. However, the OFDM waveform has a high peak-to-average power ratio (PAPR) due to the superposition from multi-carriers. The high PAPR feature brings a large dynamic range to the power amplifier (PA) input which increases the budget of PA energy consumption [24]. Meanwhile, transmitting the OFDM signal through the linear region of PA has a very low efficiency. Therefore, PAPR reduction is often employed together with input back-off to avoid the PA non-linear distortion. At the same time, digital pre-distortion (DPD) is often utilized to extend the scope of the PA linear region while keeping the spectral regrowth under control [18]. To this end, neural networks are proposed as the DPD blocks in [25, 30]. However, the perfect knowledge of the PA model and measurement bias are required to facilitate the DPD based non-linear compensation. These knowledge and measurements are very challenging to be perfectly obtained in reality [13]. Thus, a robust MIMO-OFDM detection method is meaningful to be explored in order to alleviate the non-linear distortion.

1.1 Contributions and Attribution

This thesis studies the symbol detection problem in MIMO-OFDM. Rather than conventional model-based approaches, this thesis provides a way of using echo state networks as the symbol detector. The main-body of this thesis is mainly reproduced from the author's paper

"Learning for Detection: MIMO-OFDM Symbol Detection through Downlink Pilots" and presented in Chapter 2, where the early access can be found in [35]. Shortly, the major contribution of this thesis is summarized as follows

- An enhanced echo state network, namely, windowed echo state network (WESN) is introduced to solve the MIMO-OFDM symbol detection problem.
- A training framework of the WESN based symbol detection for different pilot patterns is developed. Especially, it can be utilized for the pilot patterns defined in communications standards.
- A theoretical analysis for the short-term memory of WESN is derived. The result states that WESN has longer short-term memory than ESN.
- Complexity analysis for ESN/WESN based symbol detection and conventional methods are studied.

Chapter 2

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

2.1 Introduction

The emerging technology artificial neural networks (ANNs) offers new aspects to solve conventional communications problems. For instance, in [6, 21, 22], auto-encoder is considered as a way to interpret the symbol modulation in communications system, where the transmitter and receiver are learned through an end to end training. However, in practical systems, this end to end learning strategy requires heavy feedbacks and computations when the transmission channel cannot be analytically characterized. Therefore, employing ANNs as functional components in a communications link instead of replacing the whole link as ANNs is an alternative way in the applications. Previous work [7] uses RNN as the receiver in molecular communications system, where the underlying channel is unavailable to be modeled. Moreover, in optical fiber systems [14, 15, 16], ANNs are utilized as a channel equalizer as well as a network monitor. Especially, in [14], the proposed method is verified through a lab experiment.

In light of the challenge in MIMO-OFDM symbol detection, ANNs provide an ideal framework to conduct the symbol detection even under the non-linear distortion. In [33], it introduces a deep neural network for OFDM symbol detection without using explicit channel state information (CSI). The learning procedure is composed of an offline and an online training. The result shows the deep ANN performs a lower bit error rate (BER) than the minimum mean squared estimator (MMSE) under a non-linear distortion channel. However, the offline training stage is conducted according to an assumption on channel statistics. In [34], it introduced a neural network based method for the receiver design in a cyclic prefix (CP)-free OFDM system. In [12], a fully connected neural network based OFDM receiver is tested over the air. However, the extension to the MIMO-OFDM system has not been investigated in these aforementioned works. To study the MIMO case, in [28], a deep neural network based detector is introduced based on unfolding the standard belief propagation algorithm. The neural network parameters are tuned via an offline training which requires enormous training for various antenna configurations. In [20], the feature of the residual signals is applied to construct a layered neural network for symbol detection. Meanwhile, the loss function is conducted on multiple layers in order to avoid the gradient vanishing [27]. 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 neural networks coefficients or inputs which cannot be perfectly obtained due to channel distortion.

The aforementioned examples often utilize feedforward neural networks for the symbol detection by dividing the received signal into independent batches. However, the signal in communications system is sequentially constructed. Therefore, using recurrent neural networks (RNNs) allows the temporal dynamical behaviors[8] to be learned. In RNNs, the current outputs are fed to inputs for the future processing. Usually, the weights of RNN are

learned through the backpropagation through time algorithm (BPTT) [31]. However, the training of RNNs is very challenging, especially when the sequence is inherent with long-range temporal dependencies. This is because of the vanishing and exploding gradients of BPTT, i.e., a small change at the current iteration can result in a very large deviation for later iterations [23]. In order to solve this issue, specific RNN structures are introduced, such as the long short-term memory network [9] which uses "memory units" and "gating units" to control the gradient flow.

Furthermore, rooted from the backpropagation-decorrelation learning rule, echo state network (ESN) is one type of RNNs which can naturally avoid the issues of conventional RNN training. Meanwhile it has high computational efficiency [10]. The learning is neither conducted on inputs nor hidden layers, but only on the output. The untrained layers are sampled from a distribution with well-designed conditions. In [19], it introduced an ESN based MIMO-OFDM symbol detector. It shows that the ESN-based symbol detector can effectively combat the non-linear distortion of PA with limited amount of training. However, this ESN-based symbol detector has poor 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 is interesting to see if an ESN-based symbol detector with additional short-term memory (STM) can improve the interference cancellation performance. Therefore, the windowed echo state network (WESN) is introduced in this chapter. In this chapter, the main contents of the WESN based MIMO-OFDM receiver are summarized as follows,

- Windowed echo state network (WESN) is constructed by adding buffers in the input layer of the standard ESN. Theoretical analysis proves that the added buffer can improve the short term memory of the underlying ESN. Numerical evaluations also demonstrate a positive correlation between the detection performance and the im-

proved 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.

- A unified training method is introduced for the WESN receiver using the pilot pattern which is compatible with the demodulation reference signal (DMRS) proposed in LTE/LTE-Advanced standards. In this way, it is able to demonstrate the effectiveness of the introduced symbol detector under a very limited training set. Meanwhile, simulations demonstrated the WESN can detect symbols using non-orthogonal pilots through numerical evaluations.

- It 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 [1, 26]. 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.

2.1.1 Chapter Outlines

This chapter is organized into six sections:

- Section 2 provides a background on the transmission model of MIMO-OFDM system, PA non-linearity and conventional symbol detection methods.

- Section 3 discusses the ESN/WESN based symbol detection as well as the pilot design.

- Section 4 derives the complexity of both conventional methods and the ESN based method.

- Section 5 gives the simulation results.
- Section 6 summarizes the this chapter

.

2.2 MIMO-OFDM Transceiver Link

2.2.1 Transmitter Architecture

Consider the point-to-point MIMO-OFDM system, where the number of Tx and Rx antennas are respectively denoted as N_t and N_r. At the pth transmitted antenna, the ith OFDM symbol is expressed as

$$u_i^{(p)}(t) = \sum_{n=0}^{N_c-1} x_i^{(p)}[n] \exp(2\pi jnt/\Delta t), t \in [i\Delta t, (i+1)\Delta t), \tag{2.1}$$

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[1]. At the qth antenna, the corresponding received OFDM symbol is given by

$$y_i^{(q)}(t) = \sum_{p=0}^{N_t-1} h_i^{(q,p)}(t) \circledast g(u_i^{(p)}(t)) + n(t), \tag{2.2}$$

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 representing

[1]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.

the waveform distortion, 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. (2.2) can be rewritten in the digital frequency domain as

$$\tilde{y}_i^{(q)}[n] = \sum_{p=0}^{N_t-1} \tilde{h}_i^{(p,q)}[n]\tilde{g}^{(p)}[n] + \tilde{n}[n], \tag{2.3}$$

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

$$\tilde{g}^{(p)}[n] = \int_{\Delta t} g(u_i^{(p)}(t))e^{-2\pi jtn/\Delta t}dt \tag{2.4}$$

$$\tilde{h}_i^{(p,q)}[n] = \int_{\Delta t} h_i^{(p,q)}(\tau)e^{-2\pi nj\tau/\Delta t}d\tau. \tag{2.5}$$

When $g(z_i^{(p)}(t)) = z_i^{(p)}(t)$, (2.3) is equal to

$$\tilde{y}_i^{(q)}[n] = \sum_{p=0}^{N_t-1} \tilde{h}_i^{(p,q)}[n]x_i^{(p)}[n] + \tilde{n}[n]. \tag{2.6}$$

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 by using the RAPP model [13]:

$$g(u(t)) = \frac{G_0 u(t)}{\left[1 + \left(\frac{|u(t)|}{u_{sat}}\right)^{2p}\right]^{1/2p}}, \tag{2.7}$$

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 is

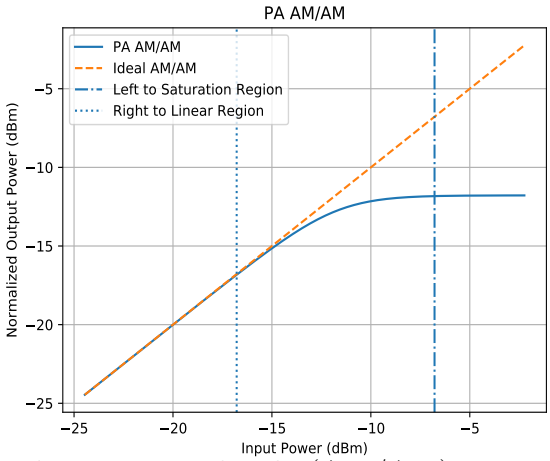


Figure 2.1: The input and output amplitude (AM/AM) curve of PA : $p = 3$ and $|u_{sat}|^2 = -11.78\mathrm{d}B$.

shown in Fig. 2.1 where the region is generally divided into three parts: the linear region $|u(t)| \ll u_{sat}$, the non-linear region $|u(t)| \sim u_{sat}$, and the 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 avoid the distortion while maintaining relatively high efficiency, the PA operational point is set close to the nonlinear region. Meanwhile, PAPR reduction is employed before signal is fed into PA[24]. However, these operations can result in a poor transmission reliability due to the waveform distortion.

2.2.2 Conventional Methods

Coherent symbol detection methods are conducted by two steps: channel estimation and symbol detection. In the channel estimation stage, 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 pilot patterns is designed based on resource blocks (RB) as shown in Fig. 2.2. For the single input single output (SISO) OFDM system, the pilot structures are

depicted in Fig. 2.2 (a). In the first sub-figure, Ω_t equals to the first OFDM symbol and Ω_f occupies all the sub-carriers. Accordingly, this comb pattern can be applied to the block fading channel. The size of Ω_f can be further reduced as shown in the second sub-figure of Fig. 2.2 (a). In the third subfigure of Fig. 2.2 (a), the scattered pilot pattern is introduced which can facilitate the channel tracking when Doppler shift exists. For the MIMO case, the pilot pattern is shown in Fig. 2.2 (b) and 2.2 (c). In these two figures, the pilot symbols at different antenna ports are non-overlapping, and the cross marker represents null pilot symbols. Thus, the pilot interference is designed to be free in these two figures.

In general, the channel coefficients on the corresponding resource elements (REs) are obtained through (2.3) solving

$$\min_{\tilde{h}_i^{(p,q)}[n]} l(\tilde{y}_i^{(q)}[n], \bar{x}_i^{(p)}[n] | (i,p,n) \in \Omega_t \times \Omega_s \times \Omega_f), \tag{2.8}$$

where $l(\cdot)$ is a pre-defined loss-function, such as likelihood function and mean square error, etc.. The CSI on the remained REs is inferred through an interpolation method. By substituting the estimated $\hat{h}_i^{(p,q)}[n]$ into (2.3), 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 by

$$\min_{x_i^p[n]} l(\tilde{y}_i^{(q)}[n], \hat{h}_i^{(p,q)}[n] | (i,p,n) \in (\Omega_t \times \Omega_s \times \Omega_f)^c). \tag{2.9}$$

However, the optimal solutions for (2.8) and (2.9) are not usually guaranteed due to the nonlinear distortion $g(\cdot)$. An improper assumption on $g(\cdot)$ can cause the model mismatch which can negatively affect the accuracy of the channel estimation (2.8) and symbol detection (2.9). To circumvent the dependence on the model assumption, neural network based methods can be employed as alternative solutions.

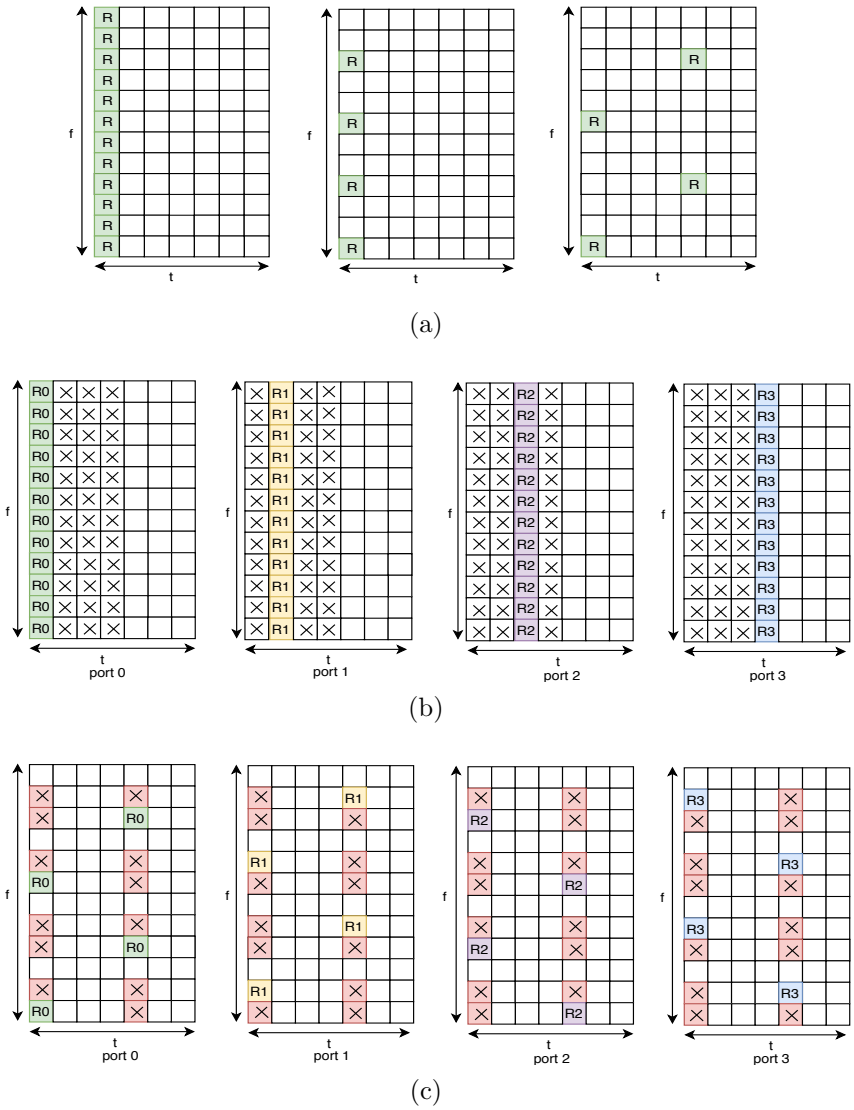


Figure 2.2: OFDM pilots structures in one RB (a) SISO-OFDM pilots (b) comb structured MIMO-OFDM pilots (c) scattered structured MIMO-OFDM pilots.

2.3 ESN/WESN Based Symbol Detection

The ESN based symbol detection approach will be introduced in this section. The concept of neural network based symbol detection will be first explained. Then, the basic structure of ESN will be introduced. In the next, the WESN based MIMO-OFDM symbol detector will be elaborated. Finally, the short-term memory of the underlying ESN/WESN will be

discussed.

2.3.1 Neural Network Based Symbol Detection Framework

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

$$\min_{\mathcal{D}} f(\mathcal{D}(\bar{y}_i^{(p)}(t)), \bar{x}_i^{(p)}[n]|(i,p,n) \in \Omega_t \times \Omega_s \times \Omega_f), \tag{2.10}$$

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 the MS in the training stage. For instance, $f(\cdot)$ can be the mean squared error or the cross-entropy; $\mathcal{D}$ can be the fully connected, convolution 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., $\hat{\mathcal{D}}(y_i^{(p)}(t))$. Accordingly, the symbol detection performance and implementation complexity are determined by the selected neural network structure 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 not carefully designed.

2.3.2 Echo State Networks

ESN is one category of RNNs which consists of an input layer, a fixed dynamic system, and a trained readout network. The network architecture of the ESN [10] is illustrated in Fig. 2.3 where the underlying network dynamics can be described by the following equation,

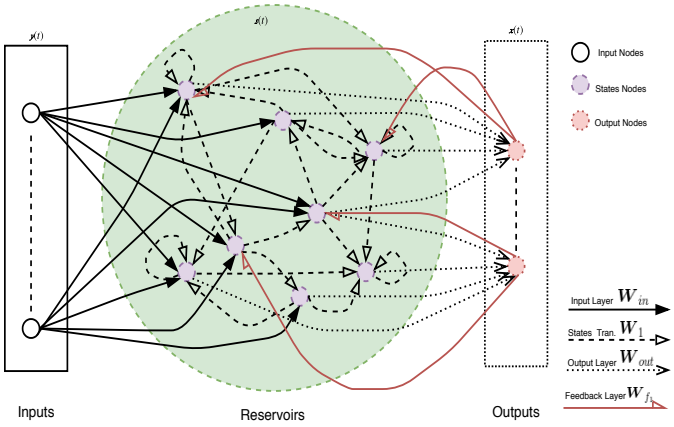


Figure 2.3: Echo state network.

$$\boldsymbol{s}(t+1) = f_{states}(\boldsymbol{W}'[\boldsymbol{y}^T(t+1), \boldsymbol{s}^T(t), \boldsymbol{x}^T(t)]^T), \tag{2.11}$$

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 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 the input layer weight, $\boldsymbol{W}_1$ is the inner state transition weight, and $\boldsymbol{W}_{f1}$ is the feedback layer weight. Moreover, $\boldsymbol{W}_{f1}$ can be omitted when feedback is not required. The output equation is given by

$$\boldsymbol{x}(t+1) = f_{out}(\boldsymbol{W}_{out}\boldsymbol{s}^T(t+1)), \tag{2.12}$$

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

Definition 2.1. We consider an ESN following the state transition equation (2.11). 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 2.2 (Proposition 3 in [10]). *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.*

Moreover, the output weight $\boldsymbol{W}_{out}$ is learned through the following two steps:

- Generating 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 (2.11), 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 (2.12), the weight $\boldsymbol{W}_{out}$ is calculated by

$$\min_{\boldsymbol{W}_{out}} L(\{\bar{\boldsymbol{y}}(t)\}_{t=0}^{T}, \{f_{out}(\boldsymbol{W}_{out}\bar{\boldsymbol{s}}^T(t))\}_{t=0}^{T}). \tag{2.13}$$

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

$$\min_{\boldsymbol{W}_{out}} \sum_{t=0}^{T} \|\bar{\boldsymbol{y}}(t) - \boldsymbol{W}_{out}\bar{\boldsymbol{s}}(t)\|_F^2, \tag{2.14}$$

which has a closed-form solution as follows,

$$\boldsymbol{W}_{out} = \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{+}, \tag{2.15}$$

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 $\bar{\boldsymbol{S}}$.

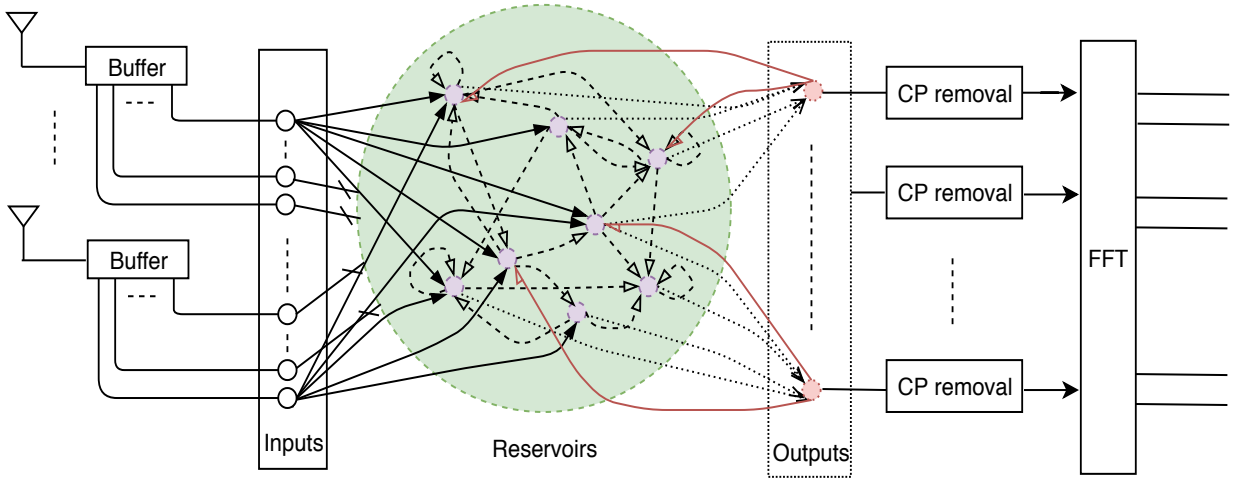


Figure 2.4: The architecture of WESN based MIMO-OFDM symbol detector.

2.3.3 Windowed Echo State Networks

The introduced WESN based MIMO-OFDM symbol detector is shown in Fig. 2.4. We see the receiving link is concatenated by a WESN, a cyclic prefix (CP) removal and an FFT block, where the dimension of the WESN output 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 $[y_i^{(j)}(t-N_{bf}), y_i^{(j)}(t-N_{bf}+1), \cdots, y_i^{(j)}(t)]^T$. Thereafter, 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 quantize the resulting frequency signal into modulation symbols according to the constellation $\mathcal{C}$, i.e., $\mathcal{Q}_{\mathcal{C}}(\boldsymbol{Z}_i\boldsymbol{F})$, where $\boldsymbol{F}$ represents the Fourier transform matrix.

2.3.4 Training of WESN

We begin by considering the training the WESN receiver under the SISO channel with zero Doppler shift, i.e., $f_D = 0$. As discussed in Sec. 2.2.2, we assume the first OFDM symbol is the training set. According to (2.10), 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. 2.3.2, we have the output of WESN as $\boldsymbol{W}\boldsymbol{S}$, 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 buffers. Therefore, similarly as (2.14), the readout weights of the WESN are updated by solving

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}\boldsymbol{S}\boldsymbol{F} - \bar{\boldsymbol{x}}_0^T\|_2, \tag{2.16}$$

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,

$$\boldsymbol{W} \overset{(a)}{=} \bar{\boldsymbol{x}}_0^T(\boldsymbol{S}\boldsymbol{F})^+ \overset{(b)}{=} (\bar{\boldsymbol{x}}_0^T\boldsymbol{F}^H)\boldsymbol{S}^+, \tag{2.17}$$

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 with a zero Doppler shift. 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. 2.5a shows the

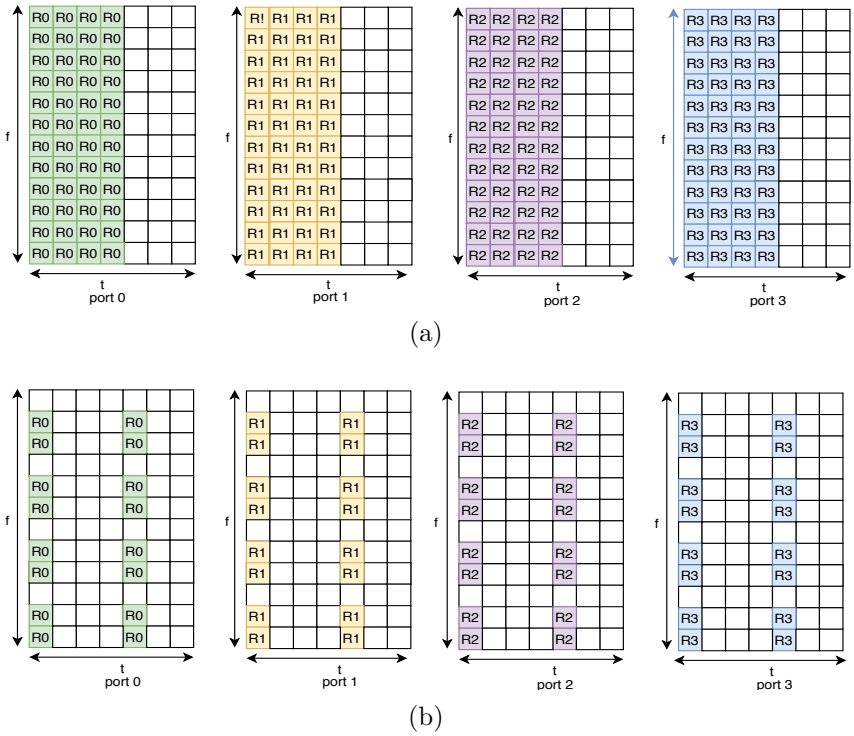


Figure 2.5: The OFDM pilots structures for WESN in one RB: (a) block (b) scattered.

case of $N_t = N_r = 4$. This pattern occupies the same number of REs as the comb structured MIMO-OFDM pilots in Fig. 2.2b. There is a slight difference between these two patterns: in Fig. 2.2b, the pilots from different antennas are orthogonal to each other, while those in Fig. 2.5a are overlapping. This is due to the fundamental difference between learning-based methods and conventional channel estimation-based methods: In the learning-based methods, the neural networks need to learn the interference situation of the transmission; in conventional methods, received pilots should be interference-free to improve the channel estimation performance.

By using this pilot pattern, the output 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 learned by solving

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}\boldsymbol{S}\boldsymbol{F}' - \bar{\boldsymbol{X}}\|_2, \tag{2.18}$$

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 $\boldsymbol{F}$; $\bar{\boldsymbol{X}} = [\bar{\boldsymbol{X}}_0, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \bar{\boldsymbol{X}}_3] \in \mathcal{C}^{N_r \times 4N_c}$ represents the concatenation of pilot symbols. Accordingly, we have

$$\begin{aligned} \boldsymbol{W} &= \bar{\boldsymbol{X}}[\boldsymbol{S}_0\boldsymbol{F}, \boldsymbol{S}_1\boldsymbol{F}, \boldsymbol{S}_2\boldsymbol{F}, \boldsymbol{S}_3\boldsymbol{F}]^+ \\ &\overset{(a)}{=} [\bar{\boldsymbol{X}}_0\boldsymbol{F}^H, \bar{\boldsymbol{X}}_1\boldsymbol{F}^H, \bar{\boldsymbol{X}}_2\boldsymbol{F}^H, \bar{\boldsymbol{X}}_3\boldsymbol{F}^H][\boldsymbol{S}_0, \boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3]^+. \end{aligned} \tag{2.19}$$

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

Now, we consider the MIMO channel with a non-zero Doppler shift. To be compatible with the conventional pilots design in SISO depicted in the third sub-figure of Fig. 2.2a, we directly utilize this scattered pilots pattern as the training set of WESN. Therefore, the weights of the outputs are updated by

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}[\boldsymbol{S}_0\boldsymbol{F}(:, \Omega_{f_0}), \boldsymbol{S}_4\boldsymbol{F}(:, \Omega_{f_4})] - [\bar{\boldsymbol{x}}_0^T(\Omega_{f_0}), \bar{\boldsymbol{x}}_4^T(\Omega_{f_4})]\|_2, \tag{2.20}$$

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

$$\min_{\boldsymbol{W}} \|\boldsymbol{W}[\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}] - [\bar{\boldsymbol{x}}_{0,\Omega_{f_0}}^T, \bar{\boldsymbol{x}}_{4,\Omega_{f_4}}^T]\|_2, \tag{2.21}$$

where

$$\bar{\boldsymbol{x}}_{t,\Omega_f}(n) \triangleq \begin{cases} \bar{\boldsymbol{x}}_t(n), & n \in \Omega_f \\ 0, & n \notin \Omega_f \end{cases} \tag{2.22}$$

$$\boldsymbol{F}_{\Omega_f}(n) \triangleq \begin{cases} \boldsymbol{F}(n), & n \in \Omega_f \\ \boldsymbol{0}, & n \notin \Omega_f \end{cases}. \tag{2.23}$$

Therefore, the output weight is given by

$$\boldsymbol{W} = [\bar{\boldsymbol{x}}^T_{0,\Omega_{f_0}}, \bar{\boldsymbol{x}}^T_{4,\Omega_{f_4}}][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}]^+, \tag{2.24}$$

which can be rewritten as

$$\boldsymbol{W} = [\bar{\boldsymbol{x}}^T_{0,\Omega_{f_0}}\boldsymbol{F}^H, \bar{\boldsymbol{x}}^T_{4,\Omega_{f_4}}\boldsymbol{F}^H][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}\boldsymbol{F}^H, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}\boldsymbol{F}^H]^+, \tag{2.25}$$

where $\bar{\boldsymbol{x}}^T_{\Omega_f}\boldsymbol{F}^H$ represents the time domain OFDM waveform transformed merely from the symbols defined on the sub-carriers $\bar{\Omega}_f$. It means that the output weight also can be obtained by fitting the waveform of scattered pilots. Similarly, using the scattered pilots plotted in Fig. 2.5b, we have the following learning rule,

$$\boldsymbol{W} = [\bar{\boldsymbol{X}}^T_{0,\Omega_{f_0}}\boldsymbol{F}^H, \bar{\boldsymbol{X}}^T_{4,\Omega_{f_4}}\boldsymbol{F}^H][\boldsymbol{S}_0\boldsymbol{F}_{\Omega_{f_0}}\boldsymbol{F}^H, \boldsymbol{S}_4\boldsymbol{F}_{\Omega_{f_4}}\boldsymbol{F}^H]^+, \tag{2.26}$$

where the definition of $\bar{\boldsymbol{X}}_{t,\Omega_f}$ is similar to (2.22).

2.3.5 WESN Short Term Memory

For RNN, the output features are expected as a certain 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 2.3 (Short Term Memory [11])**.** Given an ESN with fixed coefficients of the inner state transient matrix, input layer, and activation function, the following *self-delay reconstruction correlation* is defined

$$d(m, \boldsymbol{w}_{out}) = \frac{cov(y(n-m), x(n))}{\sigma(y(n-m))\sigma(x(n))} \tag{2.27}$$

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 represent the time sequence index, m is the input delay degree, and $x(n)$ is the ESN output when the input is $y(n)$. According to the self-delay reconstruction correlation, the STM capacity can be defined as follows,

- The m-th delay STM capacity:

$$MC_m = \max_{\boldsymbol{w}_{out}} d(m, \boldsymbol{w}_{out}) \tag{2.28}$$

- The STM capacity:

$$MC = \sum_{m=1,2,\cdots} MC_m \tag{2.29}$$

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

by extending the concept to each input-output pair. Furthermore, the metric in (2.28) 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

$$d(m, \boldsymbol{w}_{out}) = \frac{\sum_{n=0}^{N-m-1} y(n)x(n+m)}{\sqrt{\sum_{n=0}^{N-m-1} |y(n)|^2}\sqrt{\sum_{n=m}^{N-1} |x(n)|^2}} \tag{2.30}$$

$$\propto -\|\tilde{\boldsymbol{x}}(m : N-1) - \tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2, \tag{2.31}$$

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 (2.12), $\tilde{\boldsymbol{x}}(m : N-1)$ can be equivalently expressed as

$$\tilde{\boldsymbol{x}}(m : N-1) = \boldsymbol{w}_{out}\tilde{\boldsymbol{S}}, \tag{2.32}$$

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 2.4. *The memory capacity of a buffer is greater than M, where M is the buffer's size.*

Proof. For a buffer, it is easily known that $MC_m = 1$ if $0 \leq m \leq M$. When $m > M$, it has $MC_m \geq 0$ as the signal can be self-correlated. Therefore, $MC_W \geq M$. □

Furthermore, the following upper bound for the STM capacity of ESN can be obtained.

Theorem 2.5 (Proposition 2 in [11]). *The memory capacity of ESN is bounded by the number of neurons, 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 the configuration of neuron numbers. Comparing Theorem 2.4 to Theorem 2.5, we see the 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. The STM of WESN is characterized by

Theorem 2.6. *Given a WESN, suppose the STM capacity of the buffer and the ESN component are MC_W and MC respectively. Then, the STM capacity of the WESN, MC_{WESN}, is followed by*

$$\frac{1}{2}MC_{WESN} \geq \lambda MC_W + (1-\lambda)MC_{ESN}, \lambda \in (0,1) \tag{2.33}$$

Proof. The output weights for the m-th delay capacity can be calculated by

$$\min_{\boldsymbol{w}_{out}} \|\tilde{\boldsymbol{x}}(m:N-1) - \tilde{\boldsymbol{y}}(0:N-m-1)\|_2^2 \tag{2.34}$$

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 [10]. By splitting $\boldsymbol{w}_{out}$ into $[\boldsymbol{w}_1, \boldsymbol{w}_2]$, it has

$$\|[\boldsymbol{w}_1, \boldsymbol{w}_2][\boldsymbol{Y}^T, \boldsymbol{S}_{ESN}^T]^T - \tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.35}$$

$$= \|\boldsymbol{w}_1\boldsymbol{Y} - \lambda\tilde{\boldsymbol{y}}(0 : N-m-1) + \boldsymbol{w}_2\boldsymbol{S}_{ESN} - (1-\lambda)\tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.36}$$

$$\leq 2\|\boldsymbol{w}_1\boldsymbol{Y} - \lambda\tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 + 2\|\boldsymbol{w}_2\boldsymbol{S}_{ESN} - (1-\lambda)\tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.37}$$

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

$$\min_{\boldsymbol{w}_1, \boldsymbol{w}_2} \frac{1}{2}\|[\boldsymbol{w}_1, \boldsymbol{w}_2][\boldsymbol{Y}^T, \boldsymbol{S}_{ESN}^T]^T - \tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.38}$$

$$\leq \min_{\boldsymbol{w}_1} \|\boldsymbol{w}_1\boldsymbol{Y} - \lambda\tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 + \min_{\boldsymbol{w}_2} \|\boldsymbol{w}_2\boldsymbol{S}_{ESN} - (1-\lambda)\tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.39}$$

$$= \lambda^2 r_W + (1-\lambda)^2 r_{ESN} \tag{2.40}$$

where

$$r_W = \min_{\boldsymbol{w}_1} \|(1/\lambda)\boldsymbol{w}_1\boldsymbol{Y} - \tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.41}$$

$$r_{ESN} = \min_{\boldsymbol{w}_2} \|(1/(1-\lambda))\boldsymbol{w}_2\boldsymbol{S}_{ESN} - \tilde{\boldsymbol{y}}(0 : N-m-1)\|_2^2 \tag{2.42}$$

According to the definition of STM, it results in

$$\frac{1}{2}MC_{WESN} \geq \lambda^2 MC_W + (1-\lambda)^2 MC_{ESN} \tag{2.43}$$

$$\geq \lambda^2 MC_W + (1-\lambda^2) MC_{ESN} \tag{2.44}$$

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

The above result shows that WESN can achieve a higher STM capacity than the sum of the STMs of the buffer and the ESN.

2.4 Complexity Analysis

This section compares the computational complexity of the ESN/WESN-based symbol detector to the conventional methods discussed in Sec. 2.2.2. The discussions are divided into SISO and MIMO. Moreover, the time computational complexity is evaluated by the floating-point operations per second (FLOPS).

2.4.1 SISO

Channel Estimation

In the channel estimation problem (2.8), $g(\cdot)$ is assumed as a linear function. When $l(\cdot)$ is chosen as the mean squared error (MSE), the channel estimation methods can be branched according to the pilot patterns plotted in Fig. 2.2 (a). For the comb pilots, the objective function in (2.8) is rewritten as

$$\min_{\tilde{\boldsymbol{h}}} \mathbb{E}\|\tilde{\boldsymbol{y}}_i - \bar{\boldsymbol{x}}_i \odot \tilde{\boldsymbol{h}}\|_F^2 \tag{2.45}$$

where $\odot$ denotes the Hadamard product. From [3], the solution is given by

$$\tilde{\boldsymbol{h}} = \boldsymbol{R}_{hy}\boldsymbol{R}_{yy}^{-1}\boldsymbol{y}_i \tag{2.46}$$

where $\boldsymbol{R}_{hy} = \boldsymbol{F}\boldsymbol{R}_{hh}\boldsymbol{F}^H\bar{\boldsymbol{X}}_i^H$, $\boldsymbol{X}_i = diag(\boldsymbol{x}_i)$, $\boldsymbol{R}_{yy} = \boldsymbol{X}_i\boldsymbol{F}\boldsymbol{R}_{hh}\boldsymbol{F}^H\boldsymbol{X}_i^H + \sigma^2\boldsymbol{I}$, and $\boldsymbol{R}_{hh}$ is the channel covariance matrix. For the scattered pilots, the channel coefficients on the pilot

time-frequency grids are calculated by

$$\min_{\boldsymbol{h}(\Omega_f)} \mathbb{E}\|\tilde{\boldsymbol{y}}_i[\Omega_f] - \bar{\boldsymbol{X}}_i(\Omega_f)\boldsymbol{h}[\Omega_f]\|_F^2 \tag{2.47}$$

which has the closed-form solution as follows,

$$\boldsymbol{h}[\Omega_f] = \boldsymbol{R}_{hY}(\Omega_f)\boldsymbol{R}_{yy}(\Omega_f)^{-1}\tilde{\boldsymbol{y}}_i[\Omega_f] \tag{2.48}$$

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 $\boldsymbol{R}_{yy}(\Omega_f) = \boldsymbol{X}_i(\Omega_f)\boldsymbol{F}(\Omega_f,:)\boldsymbol{R}_{hh}\boldsymbol{F}(\Omega_f,:)^H\boldsymbol{X}(\Omega_f,:)_i^H + \sigma^2\boldsymbol{I}$. Thereafter, the channels on the rest grids are inferred by the interpolation as discussed in [4]. Specifically, when the channel tap is assumed to be uncorrelated i.e., $\boldsymbol{R}_{hh} = \boldsymbol{I}$, it has

$$\tilde{h}[n] = \bar{x}_i^*[n] * \tilde{y}_i[n]/(|\bar{x}_i[n]|^2 + \sigma^2) \tag{2.49}$$

where n stands for the index of sub-carriers.

Symbol Detection

For the symbol detection problem (2.9), when $l(\cdot)$ is selected as MSE, it has

$$\min_{\boldsymbol{x}_i} \mathbb{E}\|\tilde{\boldsymbol{y}}_i - \boldsymbol{x}_i \odot \hat{\boldsymbol{h}}_i\|_F^2 \tag{2.50}$$

When the transmission symbols are uncorrelated between sub-carriers, (2.50) 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 the solution as follows

$$\hat{x}_i[n] = \hat{h}_i^*[n] * \tilde{y}_i[n]/(|\hat{h}_i[n]|^2 + \sigma^2) \tag{2.51}$$

Complexity

In order to conduct the complexity analysis, the FLOPS of standard matrix operations is first reviewed. 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 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. 2.2b, the FLOPS of the LMMSE channel estimation (2.46) 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 (2.51), 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, for the scattered pilot pattern in Fig. 2.2c, 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 (2.17), the FLOPS for the output weights learning is $2N_c(N_n+1) + 3N_cN_n^2 + 2N_n^3$. Meanwhile, the calculation at the symbol detection stage is merely conducting 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

$2(\kappa N_c)(N_n+1)+3\kappa N_c N_n^2+2N_n^3$. Therefore, the total number of FLOPS is proportional to $\delta(2(\kappa N_c)(N_n+1)+3\kappa N_c N_n^2+2N_n^3)+N_n N_c$. The resulting complexity of the ESN/WESN 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 this analysis does not consider the coplexity inside ESN. This is because ESN are usually implemented through analog circuits which perform faster than the digital circuit[5, 29] with less energy consumption.

2.4.2 MIMO

By using the comb and scattered pilots respectively plotted in Fig. 2.2b and Fig. 2.2c, the FLOPS of 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 to 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, consider the LMMSE MIMO symbol detection using (2.9). When the transmitted symbols on different sub-carriers are independent to each others, 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

$$\min_{\tilde{\boldsymbol{x}}_i(n)} \mathbb{E}\|\tilde{\boldsymbol{y}}_i(n)-\hat{\boldsymbol{H}}_i(n)\tilde{\boldsymbol{x}}_i(n)\|_F^2 \tag{2.52}$$

which has the following closed-form solution

$$\tilde{\boldsymbol{x}}_i(n)=(\hat{\boldsymbol{H}}_i^H(n)\hat{\boldsymbol{H}}_i(n)+\sigma^2\boldsymbol{I})^{-1}\hat{\boldsymbol{H}}_i^H(n)\tilde{\boldsymbol{y}}_i(n) \tag{2.53}$$

Thus, the number of FLOPS is $2N_c(N^3+N^2+N)$, 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,

$$\min_{\boldsymbol{x}_i(n)\in\mathcal{C}^{N_r}} \|\tilde{\boldsymbol{y}}_i(n) - \hat{\boldsymbol{H}}(n)\tilde{\boldsymbol{x}}(n)\|_2 \tag{2.54}$$

where $\mathcal{C}$ represents the modulation constellation of the transmitted symbols. Since the standard sphere decoding usually has high redundancy in the implementation. A complexity reduced sphere decoding algorithm proposed in [2] is selected 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 the output weight learning is $2N^2N_c(N_n+1)+3N_cNN_n^2+2N_n^3$ according to (2.19), where the number of training OFDM symbols is the same as the transmission antennas. At the symbol detection stage, FLOPS is N_cNN_n. Similarly, it can arrive at the FLOPS using the scattered pilot. The results of complexity comparison is summarized in Table 2.1, where N is the number of antennas at Tx and Rx; δ stands for the ratio of OFDM pilots; N_c: the number of sub-carriers; κ represents the ratio of pilots over all sub-carriers; N_n is the number of neurons and the length of buffer together. 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. 2.5.

Table 2.1: Computational Complexity of Symbol Detection Methods

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

2.5 Numerical Results

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 four times up-sampling upon the baseband signal. We assume that the channel is given by the following tap-delay model:

$$h_i^{(q,p)}(\tau) = \sum_{l=0}^{L-1} a_i^{(p,q)}(l) p_w(\tau - \tau_l), \tag{2.55}$$

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

$$\mathbb{E}(a_i(l)a_{(i+1)}(l)) = \sigma_l^2 J_0(2\pi f_D \Delta t), \tag{2.56}$$

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 or channel models also can be utilized without changing the training framework. The number of paths, L, in the channel model (2.55) is set as 6. The base-band modulation order is selected as 16-QAM. For the conventional methods using scattered pilots, the CSI is obtained by linear interpolation.

Furthermore, it is important to 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 [20, 28, 33]. 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 systems.

2.5.1 Overfitting Issue

Before proceeding on the comparison between the ESN based receiver and the conventional methods, the overfitting issue on selecting the number of neurons of the underlying ESN based receiver is first discussed. As shown in Fig. 2.6, the BER of the training set decreases as the model becomes more complicated. At the same time, the BER gap between the testing set and the training set is enlarged as the number of neurons increases. Therefore, in order to achieve low generalization error (i.e., low BER on testing set), it requires a proper

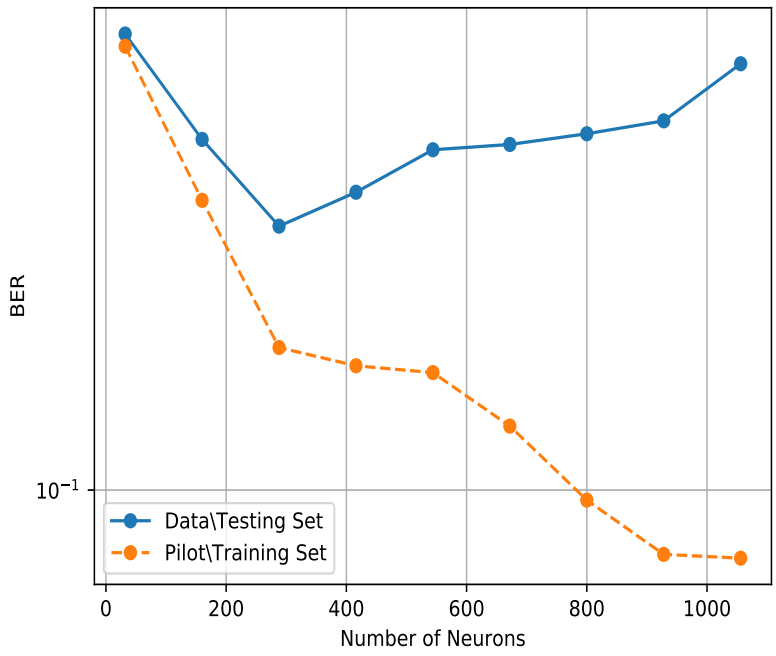


Figure 2.6: The over-fitting issue of changing the number of neurons in ESN under the MIMO block fading channel.

selection on the number of neurons.

2.5.2 SISO

The WESN in the SISO channel under different operation regions of PA is represented. Fig. 2.7 shows the BER results when the Doppler shift is 0Hz, where the threshold for PA linear region is set as 3dB up to the boundary of the linear region as depicted in Fig. 2.1. 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 the LMMSE using the CSI obtained from the LMMSE channel estimation. It shows that these three methods have comparable performance among the linear region. Moreover, for WESN, the BER performance is the best in PA nonlinear

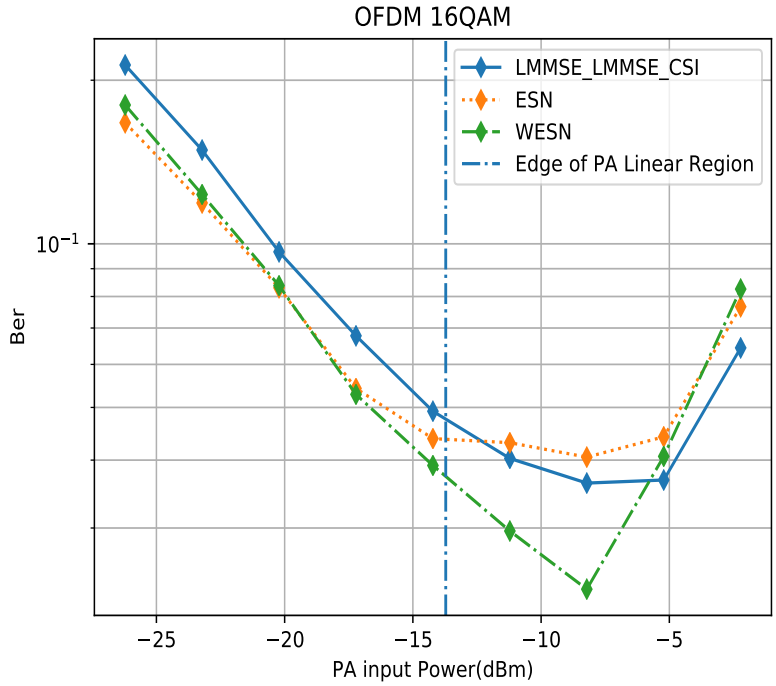


Figure 2.7: The BER comparison of the ESN symbol detector, WESN symbol detector and LMMSE method under the SISO block fading channel, where the number of neurons is set as 64 and the length of buffers is 30.

region when the optimal PA input power is selected. It demonstrates that the WESN can considerably compensate for the non-linear waveform distortion. It also can be concluded that the symbol detection using the estimated CSI does not necessarily lead to the optimality in BER performance.

Nevertheless, the performance of the WESN is highly related to the settings of neural network parameters, especially the number of internal neurons and the buffer length. Therefore, how the length of buffer and the number of neurons can jointly impact the BER performance is inverstigated in the next. In Fig. 2.8, it shows 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 length of buffer, the resulting BER declines. However, due to the overfitting, BER increases again when the number of neurons becomes

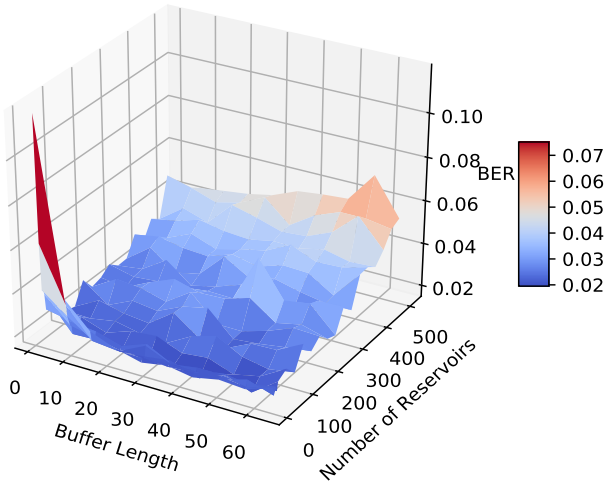
greater. Furthermore, it shows that compared to the WESN configured with more neurons, the WESN with few neurons and longer buffers can achieve the same performance. This is because the memory capacity of WESN is jointly determined by the configuration of the neurons and the buffers. Moreover, it reveals the overfitting issue in Fig. 2.8b is severer than Fig. 2.8a when the number of neurons is increased. The reason is the neural network structure for anti-distortion is expected to be simpler when the PA input power is closer to the linear region. Thus, the overfitting is anticipated to happen at smaller neurons numbers compared to the case with stronger PA nonlinearity.

Moreover, the BER performance under different Doppler shifts in the SISO channel is shown in Fig. 2.9. It shows that these three methods are comparable in the BER as well.

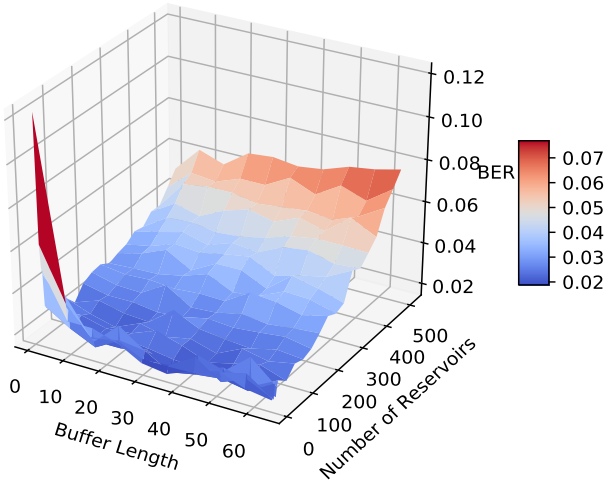
In Fig. 2.10, the comparison between ESN and WESN under different Doppler shifts is investigated. The results show that the WESN always perform better than the ESN under different Doppler shifts. Fig. 2.11 shows the BER distribution by varying the buffer length and neurons number. It reveals that increasing buffer size can significantly decrease BER which indicates that WESN can gain more advantages over the Doppler shift channel compared to the standard ESN. Meanwhile, adding more neurons can always lead to model overfitting.

2.5.3 MIMO

In Fig. 2.12 compares the BER performance of WESN to the conventional methods, i.e., LMMSE and sphere decoding (SD) under block fading channel. For the conventional methods, the CSI is obtained by LMMSE using the pilot patterns depicted in Fig. 2.2c. It shows that the performance gap between WESN and the conventional methods is enlarged compared to the SISO case. Especially, for SD, the BER performance deteriorates quickly when



(a) 3D surface when the PA input power is -8 dBm



(b) 3D surface when the PA input power is -11 dBm

Figure 2.8: The average BER performance of the WESN symbol detector under the SISO block fading channel by varying the length of buffers and the number of neurons, where the number of neurons varies from 8 to 512 and the length of buffers ranges from 1 to 64.

the PA input power is in the non-linear region. This is because SD requires more accurate CSI for symbol detection.

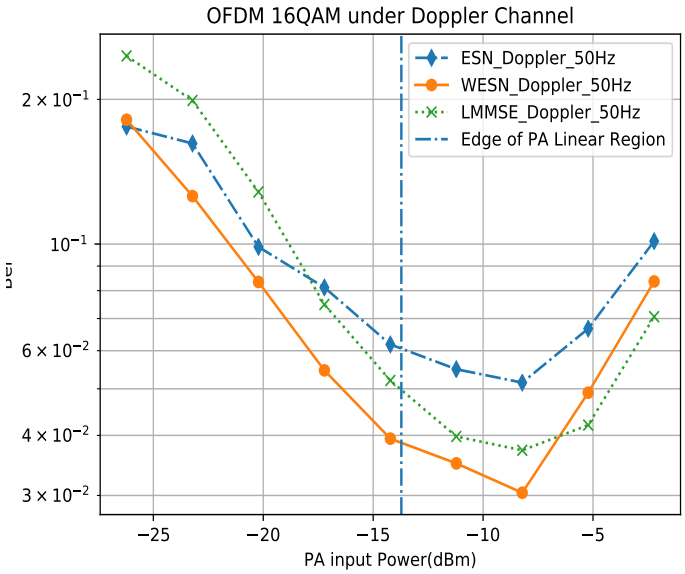


Figure 2.9: The BER comparison of the ESN symbol detector, the WESN detector and the LMMSE method under the SISO Doppler channel with different Doppler shifts, where the number of neurons is set as 64 and the length of buffers is 30.

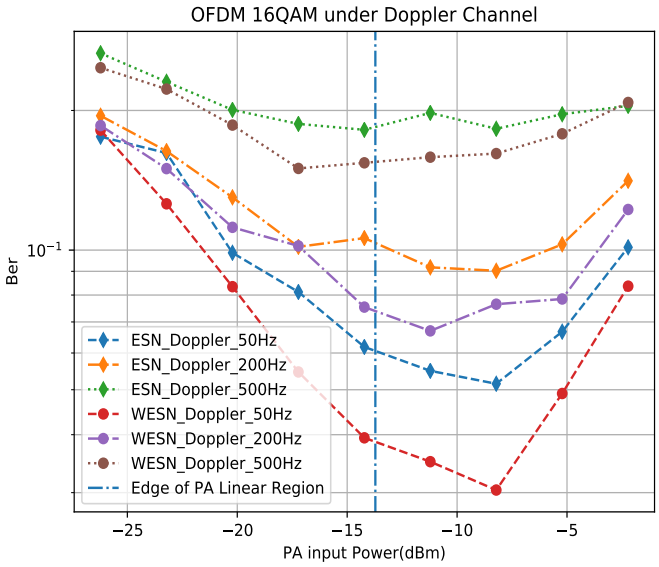
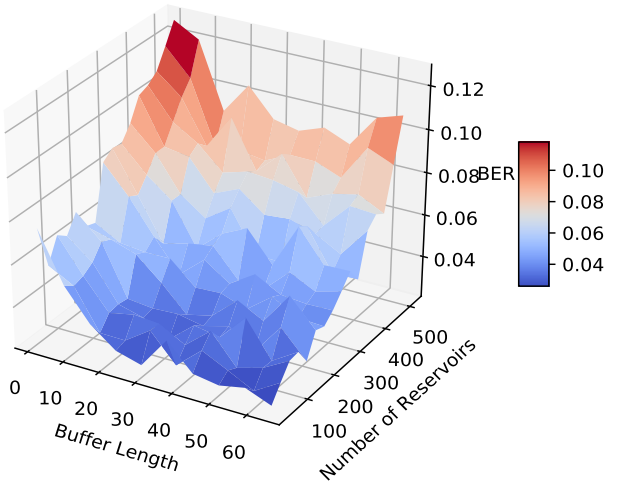
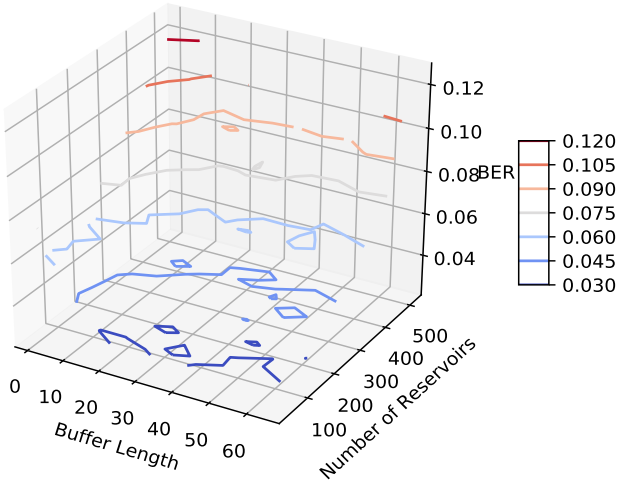


Figure 2.10: The BER comparison between the ESN symbol detector and the WESN symbol detector under SISO Doppler channel with different Doppler shifts, where the number of neurons is set as 64 and the length of buffers is 30.



(a) 3D surface



(b) 3D contour version

Figure 2.11: The 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, 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.

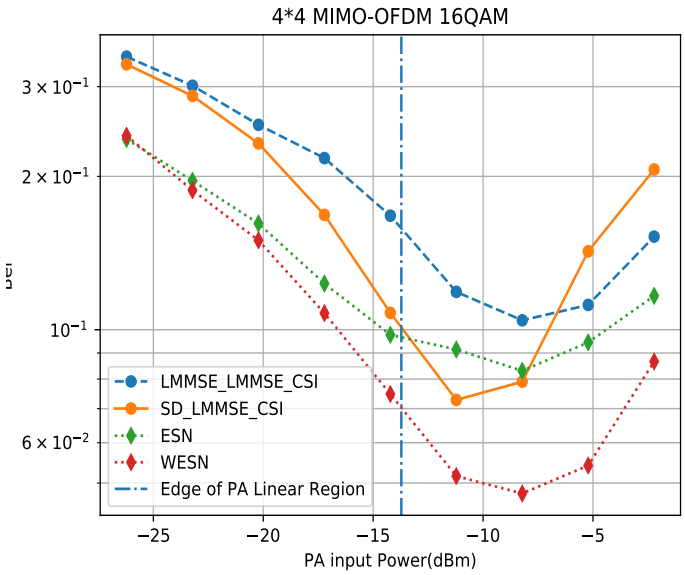
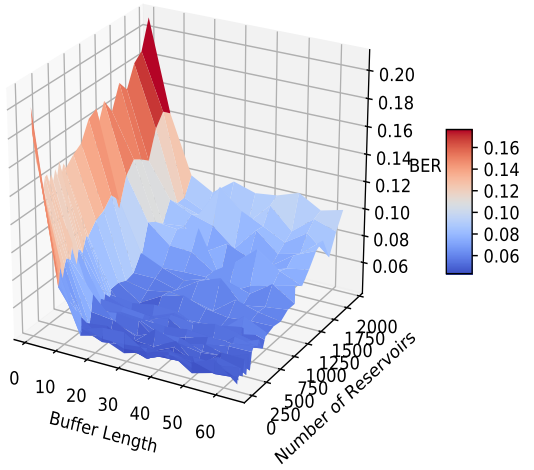


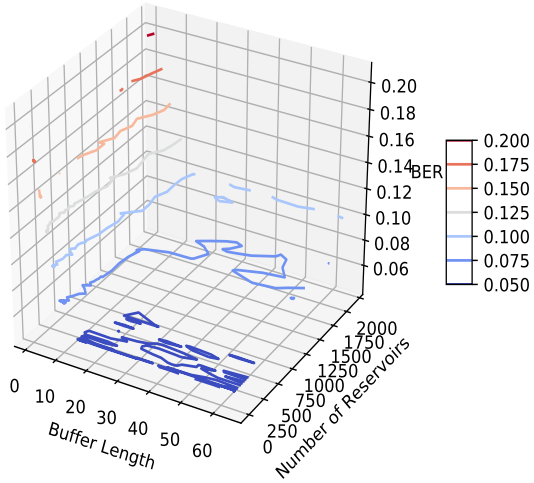
Figure 2.12: The BER comparison of the ESN symbol detector, the WESN symbol detector, the LMMSE method and sphere decoding under the MIMO block fading channel, where the number of neurons is set as 64 and the length of buffers is 30.

Again, it plots the BER distribution by varying the buffer length and the number of neurons as shown in Fig. 2.13. The advantages of the introduced buffer are more obvious compared to the SISO case by looking at Fig. 2.8.

Moreover, by using the pilot pattern in Fig.2.5a, the number of pilot symbols in training can be flexibly adjusted. Fig. 2.14 shows the BER performance by varying the number of pilots, i.e., the number of OFDM symbols allocated as pilots, where the PA input power is chosen as -9dBm, 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.. For instance, Fig. 2.5a 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. In conventional methods, using non-orthogonal pilot is not enough to avoid the pilot interference during the channel estimation stage. This means the conventional



(a) 3D surface



(b) 3D contour version

Figure 2.13: The 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, where the number of neurons varies from 8 to 512 and the length of buffers ranges from 1 to 64.

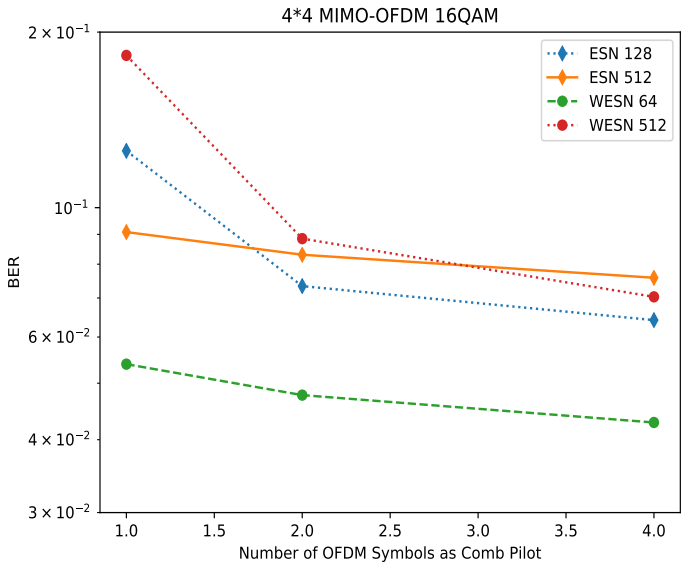


Figure 2.14: The BER performance of the ESN symbol detector and the WESN symbol detector under the MIMO block fading channel by varying the number of pilots OFDM symbols.

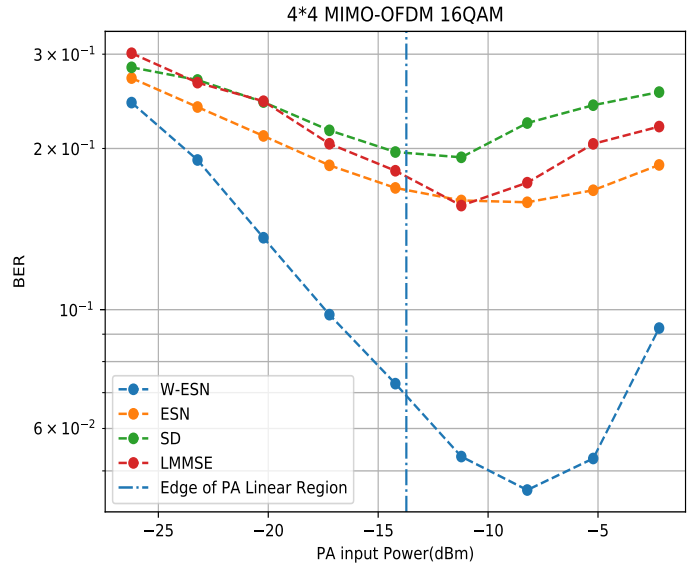
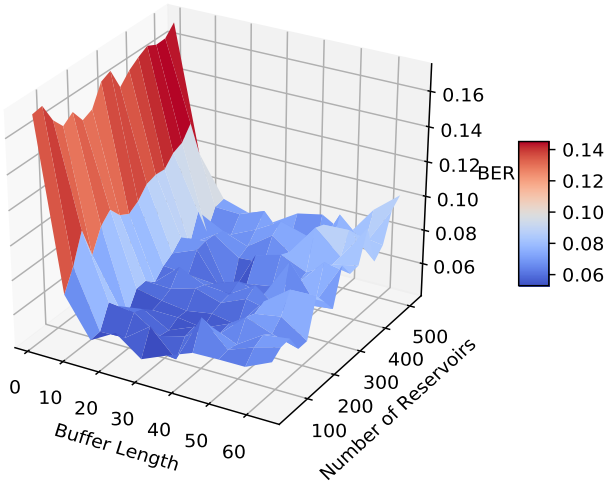
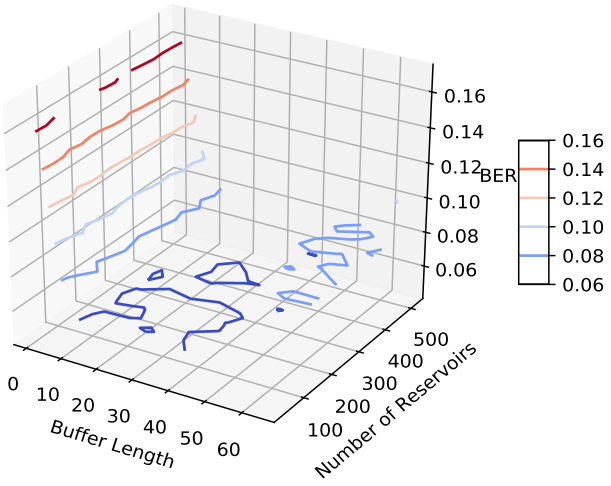


Figure 2.15: The BER comparison of the ESN symbol detector, the WESN symbol detector, the LMMSE method and sphere decoding under the MIMO Doppler channel, where the length of buffers is 30, the number of neurons is 64 and the Doppler shift is 50Hz.



(a) 3D surface



(b) 3D contour version

Figure 2.16: The average BER performance of the WESN symbol detector under the MIMO Doppler channel by varying the length of buffers and the number of neurons when the PA input power is -8 dBm, where the number of neurons varies from 8 to 512 and the length of buffers ranges from 1 to 64 and the Doppler shift is 50 Hz.

channel estimation method cannot be directly applied using non-orthogonal pilot. However, by using ESN based method, it can be observed 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, it has the deterioration of BER performance due to the overfitting.

Fig. 2.15 plots the performance of using the scattered pilot of MIMO under the Doppler shift channel. The 2D BER distribution under the Doppler channel is shown in Fig. 2.16 which has a similar distribution as Fig. 2.13.

Thus, we conclude that the ESN based approach performs better than conventional methods in low SNR regime and nonlinear distortion channel. Furthermore, note that 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%). These training sets are too small to train a proper-fitted deep neural network from scratch. However, experiments show that the ESN-based approach can achieve a good generalization result using such a few numbers of training.

2.6 Conclusion

This chapter elaborated the application of ESN, a special RNN, to MIMO-OFDM symbol detection. A new ESN based detector, WESN, is introduced as the receiver to significantly improve the performance of interference cancellation. As an improvement upon previous ESN, the WESN is proved to be able to fundamentally enhance the short term memory of

the underlying ESN. Additionally, compared to conventional coherent MIMO-OFDM symbol detection strategies as well as ESN, numerical evaluation demonstrates that WESN offers great performance improvement even under the constraint of using compatible pilot patterns defined in 3GPP LTE standards in both static and dynamic MIMO channel. Moreover, in complexity analysis, it proves that WESN performs relatively few FLOPS compared with conventional methods.

Chapter 3

Summary and Future Work

The problem addressed in this thesis is how to use neural networks to mitigate the interference and distortion of the received signal. However, the major challenge in this application is the available training set is very limited. This is because the pilots in the physical layer of cellular systems cannot exceed a certain level of the overall transmission bandwidth. To facilitate the training, this thesis introduced and evaluated the ESN/WESN based framework. The resulting training method is compatible with both comb and scattered pilot patterns used in conventional communications systems. Moreover, theoretical analyses of the short-term memory for the underlying ESN/WESN are studied in this thesis. The result shows WESN can achieve longer short-term memory than the standard ESN. Simulation results demonstrate the advantage of the ESN/WESN based symbol detection as opposed to conventional methods.

For future work, the extension of the shallow ESN structure to deep structures can be considered. Meanwhile, other nonlinear distortions, such as quantization, can be incorporated into this symbol detection framework for evaluation. Finally, this framework can be considered to add on top of channel coding/decoding strategies for joint processing.

Bibliography

[1] Luis G Barbero and John S Thompson. Fixing the complexity of the sphere decoder for MIMO detection. *IEEE Trans. Wireless Commun.*, 7(6), 2008.

[2] Ronald Y Chang, Wei-Ho Chung, and Sian-Jheng Lin. A* algorithm inspired memory-efficient detection for MIMO systems. *IEEE Wireless Commun. Lett.*, 1(5):508–511, 2012.

[3] Sinem Coleri, Mustafa Ergen, Anuj Puri, and Ahmad Bahai. Channel estimation techniques based on pilot arrangement in OFDM systems. *IEEE Trans. Broadcast.*, 48(3): 223–229, 2002.

[4] Xiaodai Dong, Wu-Sheng Lu, and Anthony CK Soong. Linear interpolation in pilot symbol assisted channel estimation for ofdm. *IEEE Trans. Wireless Commun.*, 6(5), 2007.

[5] François Duport, Anteo Smerieri, Akram Akrout, Marc Haelterman, and Serge Massar. Fully analogue photonic reservoir computer. *Scientific Reports*, 6:22381, 2016.

[6] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink. Deep learning based communication over the air. *IEEE J. Sel. Topics Signal Process.*, 12(1):132–143, Feb 2018.

[7] Nariman Farsad and Andrea Goldsmith. Neural network detectors for molecular communication systems. In *IEEE 19th Intl. Workshop on Signal Process. Adv. in Wireless Commun. (SPAWC)*, pages 1–5, 2018.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.

[10] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Inf. Technol. GMD Technical Report*, 148(34):13, 2001.

[11] Herbert Jaeger. *Short term memory in echo state networks*, volume 5. German National Research Institute for Computer Science (GMD) Report, 2001.

[12] Peiwen Jiang, Tianqi Wang, Bin Han, Xuanxuan Gao, Jing Zhang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Artificial intelligence-aided OFDM receiver: Design and experimental results. *arXiv preprint arXiv:1812.06638*, 2018.

[13] 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.*, 17(1):315–333, First Quarter 2015.

[14] Boris Karanov, Mathieu Chagnon, Félix Thouin, Tobias A Eriksson, Henning Bülow, Domaniç Lavery, Polina Bayvel, and Laurent Schmalen. End-to-end deep learning of optical fiber communications. *J. Lightw. Technol.*, 36(20):4843–4855, 2018.

[15] Boris Karanov, Domaniç Lavery, Polina Bayvel, and Laurent Schmalen. End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks. *Opt. Express*, 27(14):19650–19663, Jul 2019.

[16] Faisal Nadeem Khan, Chao Lu, and Alan Pak Tao Lau. Machine learning methods for optical communication systems. In *Signal Process. in Photon. Commun.*, pages SpW2F–3. Opt. Society of America, 2017.

[17] 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.*, 50(2):140–147, February 2012. ISSN 0163-6804.

[18] Dennis R Morgan, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan. A generalized memory polynomial model for digital predistortion of rf power amplifiers. *IEEE Trans. Signal Process.*, 54(10):3852–3860, 2006.

[19] Somayeh Mosleh, Lingjia Liu, Cenk Sahin, Yahong Rosa Zheng, and Yang Yi. Brain-inspired wireless communications: Where reservoir computing meets MIMO-OFDM. *IEEE Trans. Neural Netw. Learn. Syst.*, 29(10):4694–4708, Oct 2018.

[20] Ami Wiesel Neev Samuel, Tzvi Diskin. Learning to detect. *arXiv preprint arXiv:1805.07631*, 2018.

[21] Timothy J O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Trans. on Cogn. Commun. Netw.*, 3(4):563–575, 2017.

[22] Timothy J O'Shea, Tugba Erpek, and T Charles Clancy. Deep learning based MIMO communications. *arXiv preprint arXiv:1707.07980*, 2017.

[23] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *Int. Conf. on Machine Learning (ICML)*, pages 1310–1318, 2013.

[24] Yasir Rahmatallah and Seshadri Mohan. Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy. *IEEE Commun. Surveys Tuts.*, 15(4):1567–1592, 2013.

[25] Meenakshi Rawat, Karun Rawat, and Fadhel M Ghannouchi. Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused

time-delay line neural networks. *IEEE Trans. Microw. Theory Techn.*, 58(1):95–104, 2009.

[26] R. Shafin, L. Liu, J. Ashdown, J. Matyjas, M. Medley, B. Wysocki, and Y. Yi. Realizing green symbol detection via reservoir computing: An energy-efficiency perspective. In *IEEE Int. Conf. on Commun. (ICC)*, May 2018.

[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proc. IEEE conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[28] Xiaosi Tan, Weihong Xu, Yair Be'ery, Zaichen Zhang, Xiaohu You, and Chuan Zhang. Improving massive MIMO belief propagation detector with deep neural network. *arXiv preprint arXiv:1804.01002*, 2018.

[29] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nature Commun.*, 5:3541, 2014.

[30] Bruce E Watkins and Richard North. Predistortion of nonlinear amplifiers using neural networks. In *Proc. IEEE Mil. Commun. Conf. (MILCOM'96)*, volume 1, pages 316–320, 1996.

[31] Paul J Werbos et al. Backpropagation through time: what it does and how to do it. *Proc. IEEE*, 78(10):1550–1560, 1990.

[32] Shaoshi Yang and Lajos Hanzo. Fifty years of MIMO detection: The road to large-scale MIMOs. *IEEE Commun. Surveys Tuts.*, 17(4):1941–1988, 2015.

[33] H. Ye, G. Y. Li, and B. Juang. Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Commun. Lett.*, 7(1):114–117, Feb 2018.

[34] Jing Zhang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Artificial intelligence-aided receiver for a CP-free OFDM system: Design, simulation, and experimental test. *arXiv preprint arXiv:1903.04766*, 2019.

[35] Zhou Zhou, Lingjia Liu, and Hao-Hsuan Chang. Learning for detection: MIMO-OFDM symbol detection through downlink pilots. *arXiv preprint arXiv:1907.01516*, 2019.