Channel Equalization Through Reservoir Computing: A Theoretical Perspective

Shashank Jere, Ramin Safavinejad, Lizhong Zheng, *Fellow, IEEE*, and Lingjia Liu, *Senior Member, IEEE*

***Abstract*—Deep learning practice, including in wireless communications, often relies on trial and error to optimize neural network (NN) structures and their corresponding hyperparameters. We show that Reservoir Computing, especially the Echo State Network (ESN), is an ideal learning-based equalizer for a general fading channel and for an ESN equalizing a channel with known statistics, theoretically derive its optimum reservoir weights which are randomly initialized in state-of-the-art and lack interpretability. The theoretical results are validated with simulations. In contrast to existing literature, this letter analytically adapts the NN structure to the problem being addressed, guaranteeing optimum equalization under known channel statistics.**

***Index Terms*—Reservoir computing, echo state network, neural network, channel equalization, and receive processing.**

I. Introduction

DEEP learning has been adopted at a rapid pace in a variety of problems, including in wireless communications in recent times, e.g., in Dynamic Spectrum Sharing [1], among other applications. As end-to-end wireless links become increasingly complicated due to nonlinear device components, e.g., power amplifier and low-resolution analog-to-digital converters (ADCs), their analytical modeling in a tractable and accurate manner becomes difficult. Therefore, standard model-based approaches for receiver tasks such as channel equalization, which rely on accurate channel state information (CSI) can result in degraded performance, especially in the low signal-to-interference-plus-noise (SINR) ratio regime, thereby making machine learning based and specifically neural network (NN)-based approaches attractive. When applying deep learning models to wireless physical layer problems in particular, the typical offline learning approach is to train a NN model using a large amount of offline data with the implicit expectation of limited generalization ability of the model to channel configurations or statistics that are not seen in the training data. Examples of this approach applied to the receive symbol detection task specifically can be seen in strategies such as *DetNet* [2] and *MMNet* [3]. Often in such methods, the choice of the NN model is not adapted to the problem at hand, and hyperparameters are tuned via trial and error. Online learning attempts to address this problem of "uncertainty in generalization" by updating NN weights in an adaptive fashion. Reservoir Computing (RC) is a particularly attractive framework for online learning due to its inherent low-complexity training and is a prime candidate for applying learning-based techniques in receive processing, e.g., symbol detection, where over-the-air training data in the form of pilots is extremely limited, as shown in [4], [5], [6]. These methods have demonstrated superior performance compared to traditional model-based techniques and other offline learning strategies.

Amongst studies investigating the theoretical underpinnings of RC's effectiveness in time-series forecasting and regression problems, a notable mention is [7] which shows that an RC, specifically an ESN without non-linear activation is equivalent to vector autoregression (VAR). Reference [8] makes the case for ESNs being universal approximators for ergodic dynamical systems. The effectiveness of RC in predicting complex nonlinear dynamical systems such as the Lorenz and the Rössler systems was studied in [9], while [10] investigated tuning and optimizing the length of the fading memory of RC systems. A survey of hardware implementations of RC systems based on optoelectronic and photonic systems with time-delayed feedback was provided in [11], while [12] proposed a proof-of-principle experimental microwave reverberant based RC system. On the other hand, statistical learning theory-based works [13] attempt to bound the generalization error using measures such as Rademacher complexity. While these works shed light on the effectiveness of RC, they do not provide much insight into how an ESN should be designed or set up to exploit any available domain knowledge. In contrast to existing NN applications to wireless communications problems, we would like the NN structure to be designed according to the problem at hand. In this letter, we develop a traditional signal processing understanding of the ESN from first principles and show that for wireless signal propagation which can be viewed as a filtering operation, the basic ESN is the ideal learning-based structure to perform its inverse equalization operation. Given first-order statistical knowledge of the channel, we show that there is an "optimum" way to set up the reservoir of the ESN for equalization. The primary contribution of this letter is the analytical optimization of the ESN reservoir weights, providing much needed NN model interpretability and a systematic and flexible way to design the ESN for predictable equalization performance. *Notation:* $\mathfrak{Re}(\cdot)$ denotes the real part operator. $(*)$ denotes linear convolution. $\mathsf{VAR}[\cdot]$ denotes the variance of a random variable. "W.L.O.G." stands for "without loss of generality".

Manuscript received 7 December 2022; accepted 29 December 2022. Date of publication 6 January 2023; date of current version 10 May 2023. This work was supported in part by the U.S. National Science Foundation (NSF) under Grant CNS-2003059 and Grant CNS-2002908. The associate editor coordinating the review of this article and approving it for publication was C. K. Wen. *(Corresponding author: Lingjia Liu.)*

Shashank Jere, Ramin Safavinejad, and Lingjia Liu are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: ljliu@ieee.org).

Lizhong Zheng is with the EECS Department, Massachussets Institute of Technology, Cambridge, MA 02139 USA.

Digital Object Identifier 10.1109/LWC.2023.3234239

II. System Model

A. Wireless Propagation

The transmission of a signal through a wireless channel can be viewed as linear filtering of that signal through a

time-varying finite impulse response (FIR) filter. The discrete-time baseband input-output relationship for a channel with L resolvable taps is a convolution operation according to

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell[m]x[m-\ell] + w[m], \tag{1}$$

where $x[m] \in \mathbb{C}$ is the baseband transmitted symbol, $y[m] \in \mathbb{C}$ is the baseband received symbol and $h_\ell[m] \in \mathbb{C}$ is the ℓ^{th} channel filter tap value respectively, all at time index m. The noise sample $w[m]$ is circularly symmetric Gaussian distributed, i.e., $w[m] \sim \mathcal{CN}(0, \sigma_n^2)$. In general, the received signal suffers from inter-symbol interference (ISI), i.e., the received symbol $y[m]$ is affected by symbols transmitted before time index m and not only by the symbol $x[m]$ transmitted at index m.

B. The Conventional Echo State Network (ESN)

Consider the Echo State Network (ESN), which is a prominent framework within the Reservoir Computing (RC) paradigm. Specifically, consider an ESN with a reservoir containing K randomly inter-connected neurons and a single output (readout) weights layer. The relevant parameters for this ESN structure are defined below:

- $\mathbf{x}_{\text{in}}[m] \in \mathbb{C}^{d_{\text{in}}}$: ESN input at time index m.
- $\mathbf{x}_{\text{res}}[m] \in \mathbb{C}^{K}$: Reservoir state vector at time index m.
- $\mathbf{W}_{\text{in}} \in \mathbb{C}^{K \times d_{\text{in}}}$: Input weights matrix.
- $\mathbf{W}_{\text{res}} \in \mathbb{C}^{K \times K}$: Reservoir (recurrent) weights matrix.
- $\mathbf{W}_{\text{out}} \in \mathbb{C}^{d_{\text{out}} \times K}$: Output weights matrix.
- $\mathbf{x}_{\text{out}}[m] \in \mathbb{C}^{d_{\text{out}}}$: ESN output at time index m.

d_{in} and d_{out} are the dimensions of the ESN's input and output respectively. For a point-wise non-linear activation function $\sigma(\cdot)$, the state update and output equations are respectively:

$$\mathbf{x}_{\text{res}}[m] = \sigma\big(\mathbf{W}_{\text{res}}\mathbf{x}_{\text{res}}[m-1] + \mathbf{W}_{\text{in}}\mathbf{x}_{\text{in}}[m]\big), \tag{2}$$

$$\mathbf{x}_{\text{out}}[m] = \mathbf{W}_{\text{out}}\mathbf{x}_{\text{res}}[m]. \tag{3}$$

To make our analysis tractable, we consider a "linear" ESN, i.e., where the activation function $\sigma(\cdot)$ is an identity mapping, so that the state update equation becomes

$$\mathbf{x}_{\text{res}}[m] = \mathbf{W}_{\text{res}}\mathbf{x}_{\text{res}}[m-1] + \mathbf{W}_{\text{in}}\mathbf{x}_{\text{in}}[m]. \tag{4}$$

A linearized formulation like the above lends itself well to tractable analysis, such as that in [7]. In the state-of-the-art, only $\mathbf{W}_{\text{out}}$ is trained using a Least Squares approach, while $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{res}}$ are initialized according to a certain pre-determined distribution and then fixed throughout the training as well as the inference stages. However, since our objective in this letter is to understand why ESNs are particularly effective in channel equalization, we view this as an optimization problem where the goal is to also optimize $\mathbf{W}_{\text{res}}$. We hypothesize that the advantage of having nonlinearity in the reservoir and the optimization of the reservoir weights are either orthogonal or at least separable issues, and therefore we limit our scope in this study to optimizing the reservoir weights in the linear setting. The validity of this hypothesis is confirmed via the simulation results of Section IV. The question we aim to answer is: *Given knowledge of the channel statistics and for conventionally trained* $\mathbf{W}_{out}$, *what is the optimum choice of* $\mathbf{W}_{res}$ *that gives the minimum expected error across channel realizations?* In this letter, we answer this question fully for the simple case of a two-tap fading channel and a single-neuron ESN equalizer. The general answer for a more complicated fading channel and reservoir construction with interconnected neurons is part of future letter.

C. Signal Processing View of the ESN

The channel frequency response $H_{\text{ch}}(j\omega)$ can be found by taking the Discrete-Time Fourier Transform (DTFT) of the channel impulse response (CIR) with a tapped delay line (TDL) model $\mathbf{h} := \begin{bmatrix} h_0 & h_1 & \dots & h_{L-1} \end{bmatrix}$ and is written as $H_{\text{ch}}(j\omega) = \sum_{\ell=0}^{L-1} h_\ell e^{-j\ell\omega\Delta\tau_s}$, where $\omega \in [0, 2\pi]$ rad/sec and $\Delta\tau_s$ is the tap-spacing or the sampling time interval in sec/sample, in the TDL model of the CIR. For an input sequence $x[m]$ with DTFT $X(j\omega)$, the channel output $y[m]$ with DTFT $Y(j\omega)$ is given by $Y(j\omega) = H_{\text{ch}}(j\omega)X(j\omega)$. If an ESN with a frequency response $H_{\text{RC}}(j\omega)$ is employed to recover the transmitted symbol $x[m]$ from $y[m]$, the DTFT of the ESN output is $\widehat{X}(j\omega) = H_{\text{RC}}(j\omega)\,Y(j\omega)$.

Consider a reservoir without any interconnections between neurons. In this case, each neuron with a delayed feedback connection to itself can be viewed as a simple infinite impulse response (IIR) filter. The Z-transform of an IIR filter with a tap weight a on the feedback path is well-known to be $H(z) = \frac{1}{1-az^{-1}}$. Based on this perspective, the frequency response of the IIR filter corresponding to the i^{th} neuron with a feedback connection having a delay $\Delta\tau_{\text{RC}}$ and without interconnections to other neurons can be written as $H_i(j\omega) = \frac{W_{\text{in},i}}{1-W_{r,i}e^{-j\omega\Delta\tau_{\text{RC}}}}$, where $W_{r,i}$ is the recurrent weight in the feedback path of the IIR filter modeled by the i^{th} neuron, and $W_{\text{in},i}$ is the corresponding input weight. We make the simplifying assumption that $\Delta\tau_s = \Delta\tau_{\text{RC}} = \Delta\tau$. The overall frequency response of the ESN is a weighted sum of the individual frequency responses, i.e., $H_{\text{RC}}(j\omega) = \sum_{i=1}^{K} W_i H_i(j\omega)$, where W_i is the output weight for the $i^{\overline{th}}$ neuron within the reservoir. Thus, the linearized ESN is an IIR filter which effectively inverts the FIR filtering effect of the wireless channel, or in other words, performs a 1-D linear deconvolution. Intuitively, we want $\widehat{X}(j\omega)$ to approach $X(j\omega)$, which is achieved when $H_{\text{RC}}(j\omega) \approx H_{\text{ch}}^{-1}(j\omega)$. From a geometric perspective, the K non-interconnected neurons inside the reservoir constitute a subspace $\boldsymbol{S}$ which is a linear combination of the basis functions $H_i(j\omega)$, i.e., $\boldsymbol{S} = \text{span}\{H_1(j\omega), H_2(j\omega), \dots, H_K(j\omega)\}$. The inverse channel response $H_{\text{ch}}^{-1}(j\omega)$ can be projected onto this subspace $\boldsymbol{S}$, leaving a residual (regression) error of ε as shown in Fig. 1. For a given $H_{\text{ch}}(j\omega)$ and hence $H_{\text{ch}}^{-1}(j\omega)$, an optimally designed reservoir would span the subspace $\boldsymbol{S}$ via the basis functions $\{H_i(j\omega)\}$ such that the error ε is minimized, or alternatively, the orthogonal projection $\text{proj}_{\boldsymbol{S}}\{H_{\text{ch}}^{-1}(j\omega)\}$ of $H_{\text{ch}}^{-1}(j\omega)$ onto $\boldsymbol{S}$ is maximized.

In conventional RC (ESN)-based symbol detector design, the reservoir is randomly initialized and then kept fixed throughout, with the recurrent weights $W_{r,i}$ chosen from a pre-determined distribution without consideration of the channel statistics. The ESN's conventional training involves selecting its output weights W_i to solve the following optimization problem such that the squared loss given below is minimized:

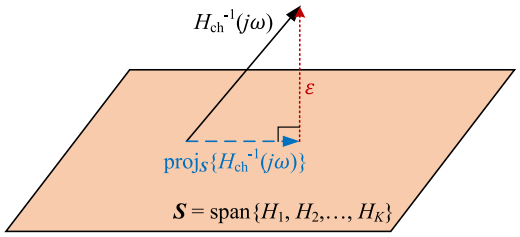

Fig. 1. Channel inverse frequency response projected onto the subspace spanned by the reservoir basis functions.

$$[W_i] = \underset{[W_i]_{i=1}^K}{\arg\min} \int_0^{2\pi} \left| 1 - H_{\text{ch}}(j\omega) \sum_{i=1}^{K} W_i H_i(j\omega) \right|^2 d\omega. \quad (5)$$

However, it is more natural to conceptualize the equalization performed by the ESN as an inversion of the channel filtering operation. The formulation that captures this is given below:

$$[W_i] = \underset{[W_i]_{i=1}^K}{\arg\min} \int_0^{2\pi} \left| H_{\text{ch}}^{-1}(j\omega) - \sum_{i=1}^{K} W_i H_i(j\omega) \right|^2 d\omega. \quad (6)$$

Note that the two formulations in (5) and (6) are *not* entirely equivalent since their respective objective functions are different. In order to use the objective function in (6) but achieve the same results for the optimum W_i's as in (5), we define a new norm measure called the "channel" norm. The definitions of the channel norm $\|\cdot\|_c$ and the channel dot product $\langle \cdot, \cdot \rangle_c$ are provided in Appendix A. Based on these definitions, the training problem in (5) can be rewritten as

$$[W_i] = \underset{[W_i]_{i=1}^K}{\arg\min} \left\| H_{\text{ch}}^{-1}(j\omega) - \sum_{i=1}^{K} W_i H_i(j\omega) \right\|_c^2. \quad (7)$$

Equivalently, (7) can be reformulated into (5) in a matrix form as $\mathbf{1} = \bar{\mathbf{H}}_{\text{RC}} \mathbf{w}_{\text{out}}$, where $\mathbf{1} \in \mathbb{C}^{N\times 1} := [1, \ldots, 1]^T$, $\mathbf{w}_{\text{out}} \in \mathbb{C}^{K\times 1} := [W_1, W_2, \ldots, W_K]^T$ and the k^{th} column $(k = 1, \ldots, K)$ of $\bar{\mathbf{H}}_{\text{RC}} \in \mathbb{C}^{N\times K}$ is defined as $\bar{\mathbf{H}}_{\text{RC}}(k) \in \mathbb{C}^{N\times 1} := [\ldots, H_{\text{ch}}(j\omega_n) H_k(j\omega_n), \ldots]^T$ for $n = -\infty, \ldots, \infty$, i.e., $N \to \infty$. The Least Squares solution for $\mathbf{w}_{\text{out}}$ is $\widehat{\mathbf{w}}_{\text{out}} = \bar{\mathbf{H}}_{\text{RC}}^{\dagger} \mathbf{1}$, where $\bar{\mathbf{H}}_{\text{RC}}^{\dagger} = (\bar{\mathbf{H}}_{\text{RC}}^H \bar{\mathbf{H}}_{\text{RC}})^{-1} \bar{\mathbf{H}}_{\text{RC}}^H$ is the left Moore-Penrose pseudo-inverse. The $(p, q)^{\text{th}}$ element of the $K \times K$ matrix $\bar{\mathbf{H}}_{\text{RC}}^H \bar{\mathbf{H}}_{\text{RC}}$ can be written as $(\bar{\mathbf{H}}_{\text{RC}}^H \bar{\mathbf{H}}_{\text{RC}})^{(p,q)} = \langle H_q(j\omega), H_p(j\omega) \rangle_c$ which can be evaluated through complex integration and employing Cauchy's Residue Theorem. The regression error ε in general is $\varepsilon = \|\mathbf{1} - \bar{\mathbf{H}}_{\text{RC}} \widehat{\mathbf{w}}_{\text{out}}\|_2^2$.

III. A Simple Example: Reservoir Design Insight

We consider a single-input single-output (SISO) wireless communications system, i.e., $d_{\text{in}} = d_{\text{out}} = 1$ from Section II-B. A simple wireless channel modeled as a frequency-selective tapped delay line (TDL) with two taps is considered, with each tap representing an aggregation of multiple physical paths. The corresponding channel impulse response (CIR) vector is then $\mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix}$. W.L.O.G., this can alternatively be written as $\mathbf{h} = \begin{bmatrix} 1 & \alpha \end{bmatrix}$, where $\alpha = {}^{h_1}/_{h_0} \in \mathbb{C}$ in general, is a random variable with a known distribution. While the derived optimum output weight W_1 and the derived closed-form error expression hold for $\alpha \in \mathbb{C}$ in general, we assume $\alpha \in \mathbb{R}$ in order to make subsequent approximations of the error expression more tractable. Additionally, for the error approximation only, we will assume $\alpha \sim \mathcal{N}(\alpha_0, \sigma_c^2) = \alpha_0(1 + u)$, i.e., α is centered around a known statistical value α_0 and the variation around α_0 is captured by $u \sim \mathcal{N}(0, \sigma_c^2/\alpha_0^2)$.

For the CIR $\mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix}$, the channel frequency response can be obtained as $H_{\text{ch}}(j\omega) = h_0 + h_1 e^{-j\omega\Delta\tau}$. Recall the assumption that the tap delay is the same as the IIR feedback delay, both being $\Delta\tau$ and W.L.O.G., set $\Delta\tau = 1$. For a single-neuron $(K = 1)$ reservoir trying to invert a two-tap wireless channel with ISI, the output weight expression simplifies to

$$\widehat{W}_1 = \frac{\left\langle H_{\text{ch}}^{-1}(j\omega), H_1(j\omega) \right\rangle_c}{\|H_1(j\omega)\|_c^2} = \frac{\int_0^{2\pi} \frac{h_0^* + h_1^* e^{j\omega}}{1 - W_{r,1}^* e^{j\omega}} d\omega}{\int_0^{2\pi} \left| \frac{h_0 + h_1 e^{-j\omega}}{1 - W_{r,1} e^{-j\omega}} \right|^2 d\omega}, \quad (8)$$

where W.L.O.G., we set $W_{\text{in},1} = 1$. The integrals are evaluated in Appendix B with the final expression for $\widehat{W}_1$ found as

$$\widehat{W}_1 = \frac{h_0^*(1 - |W_{r,1}|^2)}{|h_0|^2 + |h_1|^2 + 2\mathfrak{Re}\{h_1^* h_0 W_{r,1}\}} := \frac{h_0^* A}{B}, \quad (9)$$

with A and B defined as shown. Adapting the general error expression to $L = 2$ and $K = 1$, it follows that

$$\varepsilon = \int_0^{2\pi} \left| 1 - \widehat{W}_1 \frac{(h_0 + h_1 e^{-j\omega})}{(1 - W_{r,1} e^{-j\omega})} \right|^2 d\omega. \quad (10)$$

We evaluate (10) in Appendix C and provide a condensed version of the derivation of the final expression in (11) therein:

$$\varepsilon = \frac{2\pi |W_{r,1} + \alpha|^2}{1 + |\alpha|^2 + 2\mathfrak{Re}\{\alpha^* W_{r,1}\}}. \quad (11)$$

Note that the error $\varepsilon \in \mathbb{R}^+ \cup \{0\}$ is random and varies with the channel realization and therefore that of $\alpha = h_1/h_0$. Therefore, to design a reservoir (choose $W_{r,1}$) that guarantees the lowest error ε across channel realizations and by extension, the lowest detection symbol error rate (SER) or bit error rate (BER) subsequent to hard-decision decoding, we must solve the following minimization problem:

$$\underset{W_{r,1}}{\arg\min}\, \mathbb{E}_{\mathbf{h}}[\varepsilon(\mathbf{h}, W_{r,1})] = \underset{W_{r,1}}{\arg\min}\, \mathbb{E}_{\alpha}[\varepsilon(\alpha, W_{r,1})]. \quad (12)$$

Computing this expectation using the complete expression for ε is non-trivial. To simplify ε while noting that (9) and (11) hold in general for $\alpha \in \mathbb{C}$, assume a real-valued $\alpha = \alpha_0(1+u)$. We can inspect from (11) that if $W_{r,1} = -\alpha$ in general, then $\varepsilon = 0$. However, since α changes randomly across channel realizations due to the randomness in u, setting $W_{r,1} = -\alpha$ is not possible. In the trivial case where $u = 0$, the optimum reservoir weight is $W_{r,1} = -\alpha_0$. Therefore, we can generally formulate the problem as setting $W_{r,1} = -\alpha_0 + \zeta$ and finding the optimum $\zeta = \zeta_{\text{opt}}$ that minimizes the expected error across channel realizations. With these approximations in (11) and retaining terms up to u^2, the first-order approximation $\hat{\varepsilon}$ is

$$\hat{\varepsilon} = 2\pi \left(\frac{\zeta^2 + 2\alpha_0 \zeta u + \alpha_0^2 u^2}{1 - \alpha_0^2} \right), \quad (13)$$

where $1 + u \approx 1$ is assumed for small u to simplify the denominator. To simplify further, we take the natural logarithm on

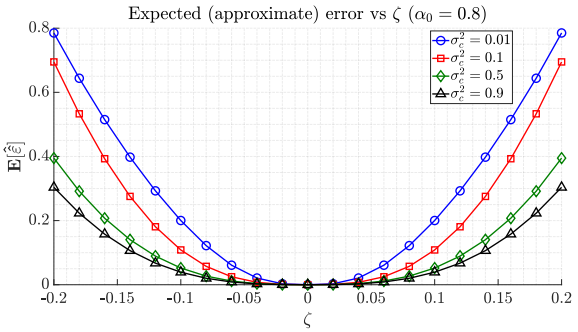

Fig. 2. Expected approximate error versus ζ for a given $\alpha_0 = \mathbb{E}[\alpha]$.

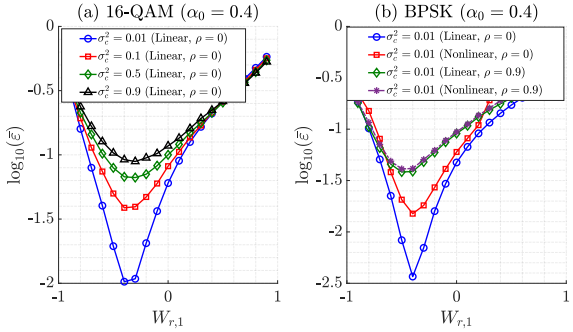

Fig. 3. Empirical MSE vs $W_{r,1}$ for Gaussian distributed real-valued α.

both sides and perform a Taylor series expansion subsequently to get a second-order approximation $\hat{\hat{\varepsilon}}$. The i^{th} term in the Taylor series expansion for $\log \hat{\varepsilon}$ evaluated at $u = 0$ is given by $c_i = \frac{1}{(i-1)!} \frac{d^{(i)}(\log \hat{\varepsilon})}{du^{(i)}}|_{u=0}$. Accordingly, the first and second terms are found as $c_1 = {}^{2\alpha_0}/_{\zeta}$ and $c_2 = -\alpha_0^2/\zeta^2$ respectively. Expanding the Taylor series only up to u^2, we get

$$\hat{\hat{\varepsilon}} = \left(\frac{2\pi\zeta^2}{1-\alpha_0^2}\right) \exp\left(c_1 u + c_2 u^2\right). \tag{14}$$

Inspecting (14), it is intuitively clear that if u is distributed around 0 with a variance $\sigma_u^2 (= \sigma_c^2/\alpha_0^2)$ that is sufficiently small, then "on average" $\alpha \approx \alpha_0$ and the optimum value of $W_{r,1}$ should be set as $W_{r,1}^{(\text{opt})} = -\alpha_0$, i.e., $\zeta_{\text{opt}} = 0$. However, we prove that this is true even when σ_u^2 is not small so that $\zeta_{\text{opt}} = 0$ still holds. The proof of this claim along with the complete expression for $\mathbb{E}_{\mathbf{h}}[\hat{\varepsilon}] (= \mathbb{E}_u[\hat{\varepsilon}])$ is provided in Appendix D. Therefore, the optimum reservoir weight $W_{r,1}^{(\text{opt})}$ minimizing the expectation of the approximate error is $W_{r,1}^{(\text{opt})} = -\alpha_0 + \zeta_{\text{opt}} = -\alpha_0$. This expectation versus ζ is shown in Fig. 2 for $\alpha_0 = 0.8$.

In summary, we have analytically derived the optimum choice for the reservoir weight of a single-neuron ESN equalizer equalizing a two-tap fading ISI channel. Although the reservoirs used in general RC-based symbol detectors, e.g., [5], [6] contain multiple neurons with random interconnections, our analysis can be extrapolated to the extreme case of a reservoir with multiple neurons but without interconnections, as the example in Section IV demonstrates. A tractable analysis of general ESN reservoirs with multiple neurons and random interconnections is part of future work. Although our analysis is for a single-carrier system, the ESN as an equalizer is equally applicable to multicarrier (e.g., OFDM) based systems [4], [5], [6].

IV. Numerical Results and Discussion

In this section, we empirically validate the optimum reservoir weight derived in Section III for a two-tap frequency-selective channel and a single-neuron reservoir ESN following linear dynamics when employed as a channel equalizer. However, to mimic a more realistic channel, we consider a TDL channel model with $L = 5$ taps with CIR $\mathbf{h} = [h_0, \alpha_{c_1}, \alpha_{c_2}, \alpha_{c_3}, \alpha_{c_4}]$, where $h_0 \sim \mathcal{N}(1, 10^{-4})$ and equalize it using a reservoir with $K = 4$ non-interconnected neurons, such that the delay of each neuron is configured to account for a different NLOS tap of the CIR. Therefore, the channel effectively has only one ISI tap and the reservoir effectively has a single neuron whose recurrent weight must be optimized to mitigate the effective tap's ISI. The means and variances of the NLOS taps are set as $\boldsymbol{\mu}_{\text{sim}} = [0.4, -0.2, 0.1, -0.1]$ and $\boldsymbol{\sigma}^2_{\text{sim}} = [0.05, 0.03, 0.02, 0.01]$ respectively. The reservoir weight for the k^{th} neuron $(2 \leqslant k \leqslant 4)$ is fixed to its theoretical optimum as $W_{r,k} = -\mathbb{E}[\alpha_{c_k}]$ while only $W_{r,1}$ is swept. The output weight for each neuron is set according to (9) before summing the weighted outputs of all the neurons. The input sequence $\boldsymbol{x} \in \mathbb{C}^{T\times 1}$ consists of T 16-QAM modulated symbols, with $T = 1000$ to align with standardized wireless systems such as 5G NR. The received sequence $\boldsymbol{y} = \mathbf{h} * \boldsymbol{x}$ (without additive white Gaussian noise) is input to the ESN for equalization, which outputs the estimated sequence $\widehat{\boldsymbol{x}}$. The empirical MSE $\bar{\varepsilon}$ plotted as a function of $W_{r,1}$ in Fig. 3a is computed as $\bar{\varepsilon} = ({}^1/_{TN_{\text{sim}}}) \sum_{i=1}^{N_{\text{sim}}} \|\widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2^2$, where N_{sim} is the number of independent Monte-Carlo runs. It can be seen from Fig. 3 that $\bar{\varepsilon}$ over $N_{\text{sim}} = 10^5$ channel realizations is minimized when the recurrent weight $W_{r,1}$ is chosen to be $W_{r,1}^{(\text{opt})} = -\alpha_0 = -\mathbb{E}[\alpha_{c_1}]$ as previously derived. Due to the approximations made to ε in the derivation of the theoretical $W_{r,1}^{(\text{opt})}$, the true optimum which minimizes $\bar{\varepsilon}$ may not exactly match the theoretical optimum of $-\alpha_0$ especially at high σ_c^2. However, the sensitivity of $\bar{\varepsilon}$ to these approximations is low, so that setting $W_{r,1} = -\alpha_0$ causes a deviation of only 1.37% and 3.35% from the corresponding true minimum error for $\sigma_c^2 = 0.5$ and $\sigma_c^2 = 0.9$ respectively, where $\sigma_c^2 = \mathsf{VAR}[\alpha_{c_1}]$. Fig. 3b also includes results for the reservoir with a nonlinear function (Tanh) applied (with and without channel tap correlation, controlled by the correlation coefficient ρ), when $\boldsymbol{x}$ consists of BPSK modulated symbols. When the channel taps are independent (implying $\rho = 0$), the theoretically derived $W_{r,1}^{(\text{opt})}$ in the linear setting still holds with the nonlinearity applied. With correlation, e.g., $\rho = 0.9$, the gap between the linear and nonlinear reservoir cases is minimal with the theoretical $W_{r,1}^{(\text{opt})}$ causing a deviation of only 1% from the true optimum, strengthening our hypothesis of reservoir optimization and nonlinear activation being orthogonal or separable issues.

V. Conclusion

In this letter, we introduce a signal processing view of the Echo State Network (ESN), a Reservoir Computing (RC) implementation and show its resulting utility as an ideal and

configurable learning-based channel equalizer. Going beyond convention, we analytically optimize the ESN's randomly initialized reservoir weight, when used as an equalizer for a simple fading channel with known statistics. Corroborating theory with simulation, this letter builds the ground for extension to reservoirs with greater complexity used to equalize higher-order fading channels.

Appendix A
Definition of the Channel Dot Product and Norm

The "channel dot product" can be defined as

$$\langle g_1(j\omega), g_2(j\omega)\rangle_c := \int_0^{2\pi} H_{\text{ch}}(j\omega) g_1(j\omega) H_{\text{ch}}^*(j\omega) g_2^*(j\omega) d\omega, = \int_0^{2\pi} |H_{\text{ch}}(j\omega)|^2 \times g_1(j\omega) g_2^*(j\omega) d\omega. \quad (15)$$

Similarly, the "channel norm" can be defined as $\|g(j\omega)\|_c^2 := \int_0^{2\pi} |H_{\text{ch}}(j\omega)|^2 \times |g(j\omega)|^2 d\omega$.

Appendix B
Derivation of Optimum Output Weight

Define I_1 and I_2 as the definite integrals in the numerator and denominator of (8) respectively. Substituting $z = e^{j\omega}$ in (8), the integral I_1 becomes

$$I_1 = \oint_{|z|=1} \frac{h_0^* + h_1^* z}{1 - W_{r,1}^* z} \frac{dz}{jz} = -j \oint_{|z|=1} \left(\frac{A_1}{z} + \frac{A_2}{1 - W_{r,1}^* z} \right) dz, \overset{(a)}{=} -j[2\pi j A_1] = 2\pi A_1, \quad (16)$$

where (a) follows from Cauchy's Residue Theorem. Using the cover-up method, $A_1 = h_0^*$ and therefore $I_1 = 2\pi h_0^*$. Similarly, I_2 can be rewritten as

$$I_2 = \oint_{|z|=1} \frac{(h_0 + h_1 z^{-1})(h_0^* + h_1^* z)}{(1 - W_{r,1} z^{-1})(1 - W_{r,1}^* z)} \frac{dz}{jz}, \overset{(b)}{=} -j[2\pi j (C_1 + C_2)] = 2\pi (C_1 + C_2), \quad (17)$$

where (b) follows after partial fraction expansion and applying the Residue Theorem. From the cover-up method, $C_1 = {}^{(-h_0^* h_1}/_{W_{r,1})}$ and $C_2 = \frac{(h_1 + h_0 W_{r,1})(h_0^* + h_1^* W_{r,1})}{W_{r,1}(1 - |W_{r,1}|^2)}$, giving,

$$I_2 = 2\pi \frac{|h_0|^2 + |h_1|^2 + 2\Re\{h_0 h_1^* W_{r,1}\}}{1 - |W_{r,1}|^2}. \quad (18)$$

Dividing I_1 by I_2 yields the expression for $\widehat{W}_1$ given in (9).

Appendix C
Closed-Form Error Expression

Starting with (10), set $z = e^{j\omega}$ and apply the Residue Theorem. Then, using partial fraction expansion, the error is

$$\varepsilon = 2\pi \left(\frac{B - A|h_0|^2}{B} \right)^2 [D_1 + D_2], \quad (19)$$

where D_1 and D_2 are found to be $D_1 = \frac{B W_{r,1} + A h_0^* h_1}{W_{r,1}(B - A|h_0|^2)}$ and $D_2 = \frac{(W_{r,1} - \frac{B W_{r,1} + A h_0^* h_1}{B - A|h_0|^2})(1 - \frac{B W_{r,1}^* + A h_0 h_1^*}{B - A|h_0|^2} W_{r,1})}{W_{r,1}(1 - |W_{r,1}|^2)}$.

Substituting for A, B, D_1 and D_2 in (19) after some simplification gives the final expression for ε in (11).

Appendix D
Expectation of the Approximate Error

The source of the randomness is the fractional deviation of α from α_0, $u = {}^{(\alpha - \alpha_0)}/_{\alpha_0}$. Assuming $u \sim \mathcal{N}(0, \sigma_u^2)$ where $\sigma_u = {}^{\sigma_c}/_{\alpha_0}$, the expectation $\mathbb{E}_{\mathbf{h}}[\hat{\varepsilon}] (= \mathbb{E}_u[\hat{\varepsilon}])$ is

$$\mathbb{E}_u[\hat{\varepsilon}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{u^2}{2\sigma_u^2}} \left(\frac{2\pi\zeta^2}{1 - \alpha_0^2} \right) e^{(c_1 u + c_2 u^2)} du, = \left(\frac{2\pi\zeta^2}{1 - \alpha_0^2} \right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{[c_1 u - c_2' u^2]} du, \quad (20)$$

where $c_2' = {}^1/_{2\sigma_u^2} - c_2$. Completing squares in the exponent,

$$\Rightarrow \mathbb{E}_u[\hat{\varepsilon}] = \sqrt{\frac{2}{c_2' \sigma_u^2}} \left(\frac{\pi\zeta^2}{1 - \alpha_0^2} \right) \exp\left(\frac{c_1^2}{4c_2'} \right), \quad (21)$$

where $\widetilde{\sigma}^2 = {}^1/_{2c_2'}$ and $\mu = {}^{c_1}/_{2c_2'}$. This simplifies to

$$\mathbb{E}_u[\hat{\varepsilon}] = \sqrt{\frac{4\zeta^2}{\zeta^2 + 2\alpha_0^2\sigma_u^2}} \left(\frac{\pi\zeta^2}{1 - \alpha_0^2} \right) \exp\left(\frac{2\alpha_0^2 \sigma_c^2}{\zeta^2 + 2\alpha_0^2 \sigma_u^2} \right), \quad (22)$$

which is plotted in Fig. 2 for $\alpha_0 = 0.8$.

References

[1] H.-H. Chang, L. Liu, and Y. Yi, "Deep echo state Q-network (DEQN) and its application in dynamic spectrum sharing for 5G and beyond," *IEEE Trans. Neur. Netw. Learn. Syst.*, vol. 33, no. 3, pp. 929–939, Mar. 2022.

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

[3] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5635–5648, Aug. 2020.

[4] Z. Zhou, L. Liu, and H.-H. Chang, "Learning for detection: MIMO-OFDM symbol detection through downlink pilots," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3712–3726, Jun. 2020.

[5] Z. Zhou, L. Liu, S. Jere, J. Zhang, and Y. Yi, "RCNet: Incorporating structural information into deep RNN for online MIMO-OFDM symbol detection with limited training," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3524–3537, Jun. 2021.

[6] J. Xu, Z. Zhou, L. Li, L. Zheng, and L. Liu, "RC-Struct: A structure-based neural network approach for MIMO-OFDM detection," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7181–7193, Sep. 2022.

[7] E. Bollt, "On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD," *Chaos*, vol. 31, no. 1, 2021, Art. no. 13108.

[8] A. G. Hart, J. L. Hook, and J. H. P. Dawes, "Echo state networks trained by Tikhonov least squares are L2(μ) approximators of ergodic dynamical systems," *Physica D Nonlinear Phenomena*, vol. 421, Jul. 2021, Art. no. 132882.

[9] A. Haluszczynski and C. Räth, "Good and bad predictions: Assessing and improving the replication of chaotic attractors by means of reservoir computing," *Chaos*, vol. 29, no. 10, 2019, Art. no. 103143.

[10] T. L. Carroll, "Optimizing memory in reservoir computers," *Chaos*, vol. 32, no. 2, 2022, Art. no. 23123.

[11] Y. K. Chembo, "Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems," *Chaos*, vol. 30, no. 1, 2020, Art. no. 13111.

[12] S. Ma, T. M. Antonsen, S. M. Anlage, and E. Ott, "Short-wavelength reverberant wave systems for physical realization of reservoir computing," *Phys. Rev. Res.*, vol. 4, May 2022, Art. no. 23167.

[13] L. Gonon, L. Grigoryeva, and J.-P. Ortega, "Risk bounds for reservoir computing," *J. Mach. Learn. Res.*, vol. 21, no. 240, pp. 1–61, 2020.