

Universality of Real Minimal Complexity Reservoir

Robert Simon Fong¹, Boyu Li², Peter Tiño¹

¹School of Computer Science, University of Birmingham, Birmingham, B15 2TT, UK

²Department of Mathematical Sciences, New Mexico State University, Las Cruces, New Mexico, 88003, USA
r.s.fong@bham.ac.uk, boyuli@nmsu.edu, p.tino@bham.ac.uk.com

Abstract

Reservoir Computing (RC) models, a subclass of recurrent neural networks, are distinguished by their fixed, non-trainable input layer and dynamically coupled reservoir, with only the static readout layer being trained. This design circumvents the issues associated with backpropagating error signals through time, thereby enhancing both stability and training efficiency. RC models have been successfully applied across a broad range of application domains. Crucially, they have been demonstrated to be universal approximators of time-invariant dynamic filters with fading memory, under various settings of approximation norms and input driving sources.

Simple Cycle Reservoirs (SCR) represent a specialized class of RC models with a highly constrained reservoir architecture, characterized by uniform ring connectivity and binary input-to-reservoir weights with an aperiodic sign pattern. For linear reservoirs, given the reservoir size, the reservoir construction has only one degree of freedom – the reservoir cycle weight. Such architectures are particularly amenable to hardware implementations without significant performance degradation in many practical tasks. In this study we endow these observations with solid theoretical foundations by proving that SCRs operating in real domain are universal approximators of time-invariant dynamic filters with fading memory. Our results supplement recent research showing that SCRs in the complex domain can approximate, to arbitrary precision, any unrestricted linear reservoir with a non-linear readout. We furthermore introduce a novel method to drastically reduce the number of SCR units, making such highly constrained architectures natural candidates for low-complexity hardware implementations. Our findings are supported by empirical studies on real-world time series datasets.

Introduction

Reservoir Computing (RC) is a subclass of Recurrent Neural Network defined by a fixed parametric state space representation (the reservoir) and a static trained readout map. This distinctive approach not only simplifies the training process by focusing adjustments solely on the static readout layer but also enhances computational efficiency. The simplest recurrent neural network realization of RC (Jaeger 2001; Maass, Natschläger, and Markram 2002; Tiño and Dorffner 2001; Lukosevicius and Jaeger 2009) are known as Echo

State Networks (ESN) (Jaeger 2001, 2002a,b; Jaeger and Haas 2004). The representation capacity of ESNs have been demonstrated in a series of papers, showing *existentially* that ESNs can approximate any time-invariant dynamic filters with fading memory in a variety of settings (Grigoryeva and Ortega 2018a,b,b; Gonon and Ortega 2019).

Simple Cycle Reservoirs (SCRs) (Rodan and Tiño 2010) are RC models characterized by a highly restricted architecture: a uniform weight ring connectivity among reservoir neurons and binary uniformly weighted input-to-reservoir connections. This simplicity is particularly advantageous for hardware implementations, reducing implementation and computational costs, as well as enhancing real-time processing capabilities without degrading performance. However, while cyclic reservoir topology with a single connection weight has been adopted in a variety of hardware implementations (Appeltant et al. 2011; Nakajima, Tanaka, and Hashimoto 2021; Coarer et al. 2018; Abe et al. 2024), its theoretical foundations have been missing. To rectify this situation we rigorously prove that SCRs operating in real domain are universal approximators of time-invariant dynamic filters with fading memory. Our results supplement recent research (Li, Fong, and Tino 2024) showing that SCRs in the complex domain can approximate, to arbitrary precision, any unrestricted linear reservoir with a non-linear readout, opening the door to a broad range of practical scenarios involving real-valued computations. Furthermore, based on our constructive proof arguments, we formulate a novel method to drastically reduce the number of SCR units, making such highly constrained architectures natural candidates for low-complexity hardware implementations.

We emphasize that proving that SCR architecture retains universality when moving from the complex to the real domain is far from straightforward. Indeed, as shown in (Li, Fong, and Tino 2024), attempts to even partially restrict SCR in the complex domain to the real one result in more complex multi-reservoir structures.¹

We conclude the paper with numerical experiments that

¹As we will show, to retain advantages of the single simple SCR structure in the real domain *and* retain universality one needs to consider orthogonal similarity throughout the approximation pipeline, as well as completion of the set of roots of unity in the canonical form of orthogonal matrices for cyclic dilation of orthogonally dilated coupling matrices.

validate the structural approximation properties demonstrated in our theoretical analysis. In particular, we begin with an arbitrary linear reservoir system with linear readout and demonstrate, on real-life datasets, that the approximation SCR counterparts will gradually approach the original system as the number of reservoir neurons increase, reinforcing the theoretical insights provided.

The Setup

We first introduce the basic building blocks needed for the developments in this study.

Definition 1. A **linear reservoir system over** \mathbb{R} is formally defined as the triplet $R := (W, V, h)$ where the **dynamic coupling** W is an $n \times n$ real-valued matrix, the **input-to-state coupling** V is an $n \times m$ real-valued matrix, and the state-to-output mapping (**readout**) $h : \mathbb{R}^n \rightarrow \mathbb{R}^d$ is a (trainable) continuous function.

The corresponding linear dynamical system is given by:

$$\begin{cases} \mathbf{x}_t = W\mathbf{x}_{t-1} + V\mathbf{u}_t \\ \mathbf{y}_t = h(\mathbf{x}_t) \end{cases} \quad (1)$$

where $\{\mathbf{u}_t\}_{t \in \mathbb{Z}_-} \subset \mathbb{R}^m$, $\{\mathbf{x}_t\}_{t \in \mathbb{Z}_-} \subset \mathbb{R}^n$, and $\{\mathbf{y}_t\}_{t \in \mathbb{Z}_-} \subset \mathbb{R}^d$ are the external inputs, states and outputs, respectively. We abbreviate the dimensions of R by (n, m, d) .

We make the same assumption as in (Li, Fong, and Tino 2024, Definition 1), that

1. The input stream $\{\mathbf{u}_t\}$ is uniformly bounded by a constant M .
2. The dynamic coupling matrix W has its operator norm $\|W\| < 1$.

The only difference in this present work is the requirement that all the matrices and vectors are over \mathbb{R} . Under the assumptions, for each left infinite time series $u = \{\mathbf{u}_t\}_{t \in \mathbb{Z}_-}$, the system (1) has a unique solution given by

$$\begin{aligned} \mathbf{x}_t(u) &= \sum_{n \geq 0} W^n V \mathbf{u}_{t-n}, \\ \mathbf{y}_t(u) &= h(\mathbf{x}_t(u)). \end{aligned}$$

We refer to the solution simply as $\{(\mathbf{x}_t, \mathbf{y}_t)\}_t$.

Definition 2. For two reservoir systems $R = (W, V, h)$ (with dimensions (n, m, d)) and $R' = (W', V', h')$ (with dimensions (n', m, d)):

1. We say the two systems are **equivalent** if for any input stream $u = \{\mathbf{u}_t\}_{t \in \mathbb{Z}_-}$, the solutions $\{(\mathbf{x}_t, \mathbf{y}_t)\}_t$ and $\{(\mathbf{x}'_t, \mathbf{y}'_t)\}_t$ for systems R and R' satisfy $\mathbf{y}_t = \mathbf{y}'_t$ for all t .
2. For $\epsilon > 0$, we say the **two systems are ϵ -close** if for any input stream $u = \{\mathbf{u}_t\}_{t \in \mathbb{Z}_-}$, the solutions of the two systems (under the notation above) satisfy $\|\mathbf{y}_t - \mathbf{y}'_t\|_2 < \epsilon$ for all t .

We now define the main object of interest in this paper. We begin by the following definition.

Definition 3. Let $P = [p_{ij}]$ be an $n \times n$ matrix. We say P is a **permutation matrix** if there exists a permutation σ in the symmetric group S_n such that

$$p_{ij} = \begin{cases} 1, & \text{if } \sigma(i) = j, \\ 0, & \text{if otherwise.} \end{cases}$$

We say a permutation matrix P is a **full-cycle permutation**² if its corresponding permutation $\sigma \in S_n$ is a cycle permutation of length n . Finally, a matrix $W = cP$ is called a **contractive full-cycle permutation** if $c \in (0, 1)$ and P is a full-cycle permutation.

Simple cycle reservoir systems originate from the minimum complexity reservoir systems introduced in (Rodan and Tino 2010):

Definition 4. A linear reservoir system $R = (W, V, h)$ with dimensions (n, m, d) is called a **Simple Cycle Reservoir (SCR) over** \mathbb{R} ³ if

1. W is a contractive full-cycle permutation, and
2. $V \in \mathbb{M}_{n \times m}(\{-1, 1\})$.

Our goal is to show that every linear reservoir system is ϵ -close to a SCR over \mathbb{R} . This is done in a way that does not increase the complexity of the readout function h . To be precise:

Definition 5. We say that a function g is h with **linearly transformed domain** if $g(\mathbf{x}) = h(A\mathbf{x})$ for some matrix A .

Universality of Orthogonal Dynamical Coupling

In (Halmos 1950), Halmos raised the question of what kind of operators can be embedded as a corner of a normal operator. He observed that one can embed a contractive operator W inside a unitary operator:

$$\tilde{U} = \begin{bmatrix} W & D_{W^\top} \\ D_W & -W^\top \end{bmatrix}$$

where $D_W = (I - W^\top W)^{1/2}$ and $D_{W^\top} = (I - WW^\top)^{1/2}$. This motivated the rich study of the dilation theory of linear operators. One may refer to (Paulsen 2002) for more comprehensive background on dilation theory. In the recent study (Li, Fong, and Tino 2024), the authors used a dilation theory technique to obtain an ϵ -close reservoir system with a unitary matrix as the dynamic coupling. A key idea is the theorem of Egerváry, which states that for any $n \times n$ matrix W with $\|W\| \leq 1$ and $N > 1$, there exists a $(N+1)n \times (N+1)n$ unitary matrix U such that the upper left $n \times n$ corner of U^k is W^k for all $1 \leq k \leq N$. In fact, this matrix U can be constructed explicitly as:

$$U = \begin{bmatrix} W & 0 & 0 & \dots & 0 & D_{W^\top} \\ D_W & 0 & 0 & \dots & 0 & -W^\top \\ 0 & I & 0 & \dots & 0 & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ 0 & \dots & & & I & 0 \end{bmatrix}.$$

We first notice that when W is a matrix over \mathbb{R} , this dilation matrix U is over \mathbb{R} as well. Therefore, U is an orthogonal matrix. This technique allows us to obtain an ϵ -close reservoir system with an orthogonal dynamic coupling matrix.

²Also called left circular shift or cyclic permutation in the literature

³We note that the assumption on the aperiodicity of the sign pattern in V is not required for this study

Theorem 6. Let $R = (W, V, h)$ be a reservoir system defined by contraction W with $\|W\| =: \lambda \in (0, 1)$. Given $\epsilon > 0$, there exists a reservoir system $R' = (W', V', h')$ that is ϵ -close to R , with dynamic coupling $W' = \lambda U$, where U is orthogonal. Moreover, h' is h with linearly transformed domain.

Proof. The proof follows that of an analogous statement in the complex domain in (Li, Fong, and Tino 2024, Theorem 11). The arguments follow through by replacing unitary matrices by orthogonal matrices and conjugate transpose by regular transpose. \square

Universality of Cyclic Permutation Dynamical Coupling

Having established universality of orthogonal dynamic coupling in the reservoir domain, we now continue to show that for any reservoir system with orthogonal state coupling, we can construct an equivalent reservoir system with cyclic coupling of state units weighted by a single common connection weight value. In broad terms we will employ the strategy of (Li, Fong, and Tino 2024), but here we need to pay special attention to maintain all the matrices in the real domain.

We begin by invoking (Li, Fong, and Tino 2024, Proposition 12), which stated that matrix similarity of dynamical coupling implies reservoir equivalence. It therefore remains to be shown that for any given *orthogonal* state coupling we can always find a full-cycle permutation that is close to it to arbitrary precision. Specifically, when given an orthogonal matrix, the goal is to perturb it to another orthogonal matrix that is *orthogonally* equivalent to a permutation matrix. Here we **cannot** adopt the strategy in (Li, Fong, and Tino 2024, Section 5) because it would inevitably involve a unitary matrix over \mathbb{C} during the diagonalization process. Instead, we convert an orthogonal matrix to its canonical form via a (real) orthogonal matrix.

The core idea of our approach is schematically illustrated in Figure 1. Given a reservoir system with dynamic coupling W , we first find its equivalent with orthogonal coupling U . Rotational angles in the canonical form of U are shown as red dots (a). Roots of unity corresponding to a sufficiently large cyclic dilation coupling can approximate the rotational angles to arbitrary precision ϵ (b).

We begin by recalling some elementary results of orthogonal and permutation matrices. For each $\theta \in [0, 2\pi)$, define the following rotation matrix:

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The eigenvalues of R_θ are precisely $e^{\pm i\theta}$. Moreover, notice that:

$$\begin{aligned} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R_{-\theta} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = R_\theta. \end{aligned}$$

Therefore, R_θ and $R_{-\theta}$ are orthogonally similar. Employing the real version of Schur decomposition, for any orthogonal

matrix $C \in O(n)$, there exists an orthogonal matrix S such that the product $S^\top C S$ has the following form:

$$\begin{aligned} S^\top C S &= \begin{bmatrix} R_{\theta_1} & & & & & \\ & \ddots & & & & \\ & & R_{\theta_k} & & & \\ & & & \pm 1 & & \\ & & & & \ddots & \\ & & & & & \pm 1 \end{bmatrix} \\ &= \begin{bmatrix} R_{\theta_1} & & & & & \\ & \ddots & & & & \\ & & R_{\theta_k} & & & \\ & & & \Upsilon & & \end{bmatrix}, \end{aligned}$$

where $\theta_i \in (0, \pi)$, and $\Upsilon := \text{diag}\{a_1, a_2, \dots, a_q\}$, $a_i \in \{-1, +1\}$, $i = 1, 2, \dots, q$, is a diagonal matrix with q entries of ± 1 's. For simplicity, **we will assume for the rest of the paper an even dimension n** , which inherently implies that q is also even. The case when n is odd is analogous and follows from similar arguments.

Note that without loss of generality, Υ can always take the form where $+1$'s (if any) preceded -1 's (if any). This can be achieved by permuting rows of Υ which is an orthogonality preserving operation and invoking (Li, Fong, and Tino 2024, Proposition 12). This can be further simplified by observing:

$$R_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_\pi = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Hence, pairs of $+1$'s (and -1 's) can therefore be grouped as blocks of R_0 (and R_π). Therefore, without loss of generality, $S^\top C S$ is a block diagonal matrix consisting of $\{R_{\theta_1}, \dots, R_{\theta_m}\}$, $\theta_i \in [0, \pi]$, $i = 1, 2, \dots, m$, and at most one block of the form $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. In the literature this is known as the **canonical form of the orthogonal matrix C** .

Given the inherent orthogonality of permutation matrices, their corresponding canonical forms can be computed. Given an integer $\ell \geq 1$, the complete set of ℓ -th roots of unity is a collection of uniformly positioned points along the complex circle, denoted by $\{e^{i\frac{2j\pi}{\ell}} : 0 \leq j \leq \ell - 1\}$. It is well-known from elementary matrix analysis that the eigenvalues of a full-cycle permutation matrix form a complete set of roots of unity.

Therefore, given an $\ell \times \ell$ full-cycle permutation P (ℓ even), we can find an orthogonal matrix Q such that $Q^\top P Q$ is a block diagonal matrix of $\{1, -1, R_{\frac{2\pi j}{\ell}} : 1 \leq j < \frac{\ell}{2}\}$. Here, note that for each $1 \leq j < \frac{\ell}{2}$, $R_{\frac{2\pi j}{\ell}}$ has two conjugate eigenvalues $e^{i\frac{2\pi j}{\ell}}$ and $e^{-i\frac{2\pi j}{\ell}}$. Hence, an $\ell \times \ell$ orthogonal matrix X is orthogonally equivalent to a full-cycle permutation if and only if its canonical form consists of:

1. A complete set of rotation matrices $\{R_{\frac{2\pi j}{\ell}} : 1 \leq j < \frac{\ell}{2}\}$, and
2. Two additional diagonal entries of 1 and -1 .

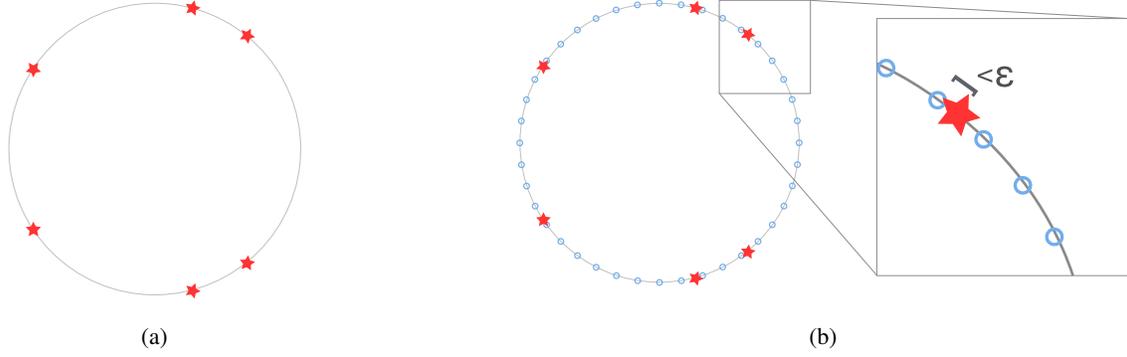


Figure 1: Schematic illustration of the core idea enabling us to prove universality of SCRs in the real domain. Given a reservoir system with dynamic coupling W and its equivalent with orthogonal coupling U , rotational angles in the canonical form of U are shown as red dots (a). Roots of unity corresponding to cyclic dilation approximate the rotational angles to a prescribed precision ϵ (b).

Theorem 7. *Let U be an $n \times n$ orthogonal matrix and $\delta > 0$ be an arbitrarily small positive number. There exists $n_1 \geq n$, an $n_1 \times n_1$ orthogonal matrix S , an $n_1 \times n_1$ full-cycle permutation P and an $(n_1 - n) \times (n_1 - n)$ orthogonal matrix D , such that*

$$\left\| S^\top P S - \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \right\| < \delta.$$

Proof. We only prove the case when the canonical form of U is in a block diagonal form consisting of $R_{\theta_1}, R_{\theta_2}, \dots, R_{\theta_k}$. The case when the canonical form contains additional entries of ± 1 is analogous. Let S_1 be the orthogonal matrix such that $S_1^\top U S_1$ is in the canonical form. For fixed $\delta > 0$, pick an integer $\ell_0 > 0$ such that: $|1 - e^{\frac{\pi i}{\ell_0}}| < \delta$.

For each $j = 1, \dots, k$, the interval $I_j := \left(\frac{\theta_j \ell_0 (k+1)}{\pi} - (k+1), \frac{\theta_j \ell_0 (k+1)}{\pi} + (k+1) \right)$ has length $2(k+1)$, and therefore contain $2(k+1)$ distinctive integers. Moreover, since $\theta_j \in [0, \pi]$, the interval I_j contains at least k distinct integers strictly within $(0, \ell_0 \cdot (k+1))$. From each interval I_1, \dots, I_k we can thus choose distinct integers a_1, \dots, a_k , with $a_j \in I_j$ and $0 < a_j < \ell_0 \cdot (k+1)$ such that for each $j = 1, \dots, k$:

$$0 < \left| \frac{\theta_j \ell_0 (k+1)}{\pi} - a_j \right| < (k+1),$$

or equivalently,

$$0 < \left| \theta_j - \frac{a_j}{\ell_0 (k+1)} \cdot \pi \right| < \frac{\pi}{\ell_0}.$$

This implies:

$$\begin{aligned} \left| e^{i\theta_j} - e^{\pi i \frac{a_j}{\ell_0 (k+1)}} \right| &= \left| 1 - e^{i(-\theta_j + \frac{a_j}{\ell_0 (k+1)} \cdot \pi)} \right| \\ &= \left| 1 - e^{i\left|\theta_j - \frac{a_j}{\ell_0 (k+1)} \cdot \pi\right|} \right| \leq \left| 1 - e^{i \frac{\pi}{\ell_0}} \right| < \delta. \end{aligned}$$

Now for each $j = 1, \dots, k$, set $\beta_j := \frac{\pi a_j}{\ell_0 (k+1)} \in (0, \pi)$. Let A_0 be the block diagonal matrix consisting of

$\{R_{\beta_1}, \dots, R_{\beta_k}\}$. Since the eigenvalues of R_{β_j} is within δ to R_{θ_j} , we obtain:

$$\begin{aligned} \|A_0 - S_1^\top U S_1\| &= \left\| \begin{bmatrix} R_{\beta_1} - R_{\theta_1} & & \\ & \ddots & \\ & & R_{\beta_k} - R_{\theta_k} \end{bmatrix} \right\| \\ &= \max\{\|R_{\beta_j} - R_{\theta_j}\|\} < \delta, \end{aligned}$$

where $\|\cdot\|$ denotes the operator norm.

Let $n_1 := 2\ell_0(k+1)$. We have each $\beta_j = \frac{a_j}{n_1} \cdot (2\pi)$, and therefore the rotations $R_{\beta_1}, \dots, R_{\beta_k}$ are all rotations of distinct n_1 -roots of unity. Whilst this is not a *complete* set of rotations of the roots of unity, we can complete the set of rotations by filling in the missing ones. In particular the missing set of rotations is given explicitly by:

$$\mathcal{R}_1 := \left\{ R_{\frac{2\pi a}{n_1}} : a \in \mathbb{Z}, 0 < a < \frac{n_1}{2}, a \neq a_j \right\}$$

Let D denote the $(n_1 - n) \times (n_1 - n)$ block diagonal matrix consisting of:

1. All the missing blocks of rotations described in \mathcal{R}_1 , and
2. Two additional diagonal entries of 1 and -1 .

By construction D is orthogonal since it contains block diagonal matrices of ± 1 and R_θ . Consider the $n_1 \times n_1$ matrix $A := (S_1 A_0 S_1^\top) \oplus D$. Then A is orthogonal by construction and the canonical form of A consists of:

1. A complete set of rotations $R_{\frac{\pi a}{n_1}}$, $a \in \mathbb{Z}, 0 < a < \frac{n_1}{2}$, and
2. An additional diagonal entry of 1 when n_1 is odd and two additional diagonal entries of 1 and -1 when n_1 is even.

This is precisely the canonical form of a $n_1 \times n_1$ full-cycle permutation. Let P be a full-cycle permutation of dimension n_1 , then there exists an orthogonal matrix S such that $A = S^\top P S$. We have,

$$\begin{aligned} \left\| S^\top P S - \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \right\| &= \left\| \begin{bmatrix} S_1 A_0 S_1^\top - U & 0 \\ 0 & 0 \end{bmatrix} \right\| \\ &= \|A_0 - S_1^\top U S_1\| < \delta, \end{aligned}$$

as desired. \square

Remark 8. In practice, the dimension n_1 is usually much smaller than the theoretical upper bound of $2\ell_0(k+1)$. Here, the integer ℓ_0 is chosen to satisfy $|1 - e^{\frac{\pi i}{\ell_0}}| < \delta$, which equivalently means:

$$\frac{\pi}{\ell_0} < \arccos\left(1 - \frac{\delta^2}{2}\right).$$

In practice, a much lower dimension can be achieved. Given a set of angles $\{\theta_i\}$ from an $n \times n$ orthogonal matrix U . For a fixed $n' > n$, we can use maximum matching program in a bipartite graph to check whether: for each θ_i there exists a distinct k_i such that the root-of-unity $\frac{2k_i\pi}{n'}$ approximates θ_i . For fixed n' , define a bipartite graph G with vertex set $A \cup B$ with $A = \{\theta_i\}$ and $B = \{\frac{2a\pi}{n'} : 0 < a < \frac{n'}{2}\}$. An edge e joins $\theta_i \in A$ with $\frac{2a\pi}{n'} \in B$ if $|e^{\theta_i i} - e^{\frac{2a\pi i}{n'}}| < \delta$. One can easily see that we can find distinct k_i to approximate θ_i by $\frac{2k_i\pi}{n'}$ if and only if there exists a matching for this bipartite graph with exactly $|A|$ edges. We let n_C denote the smallest n' such that a desired maximum matching is achieved. We shall demonstrate later, in the numerical experiment section, that n_C is significantly lower than the theoretical upper bound given by Theorem 7.

Theorem 9. Let U be an $n \times n$ orthogonal matrix and $W = \lambda U$ with $\lambda \in (0, 1)$. Let $R = (W, V, h)$ be a reservoir system with state coupling W . For any $\epsilon > 0$, there exists a reservoir system $R_c = (W_c, V_c, h_c)$ that is ϵ -close to R such that:

1. W_c is a contractive full-cycle permutation with $\|W_c\| = \|W\| = \lambda \in (0, 1)$, and
2. h_c is h with linearly transformed domain.

Proof. The proof follows that of an analogous statement in the complex domain in (Li, Fong, and Tino 2024, Theorem 14). The arguments follow through by replacing unitary matrices by orthogonal matrices and conjugate transpose by regular transpose. \square

From Cyclic Permutation to SCR

We have now ready to prove the main result (Theorem 12). So far, we have proved that any linear reservoir system $R = (W, V, h)$ is ϵ -close to another reservoir system $R' = (W', V', h')$ where W' is a permutation or a full-cycle permutation. It remains to show that one can make the entries in the input-to-state coupling matrix V to be all ± 1 .

We first recall the following useful Lemmas.

Lemma 10 ((Li, Fong, and Tino 2024, Lemma 16)). Let n, k be two natural numbers such that $\gcd(n, k) = 1$. Let P be an $n \times n$ full-cycle permutation. Consider the $nk \times nk$ matrix:

$$P_1 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & P \\ P & 0 & 0 & \dots & 0 & 0 \\ 0 & P & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & \dots & & & P & 0 \end{bmatrix}.$$

Then P_1 is a full-cycle permutation.

Lemma 11 ((Li, Fong, and Tino 2024, Lemma 17)). For any $n \times m$ real matrix V and $\delta > 0$, there exists k matrices $\{F_1, \dots, F_k\} \subset \mathbb{M}_{n \times m}(\{-1, 1\})$ and a constant integer $N > 0$ such that:

$$\left\| V - \frac{1}{N} \sum_{j=1}^k F_j \right\| < \delta$$

Moreover, k can be chosen such that $\gcd(k, n) = 1$.

We now obtain our main theorem on the universality of SCR over \mathbb{R} . In comparison to (Li, Fong, and Tino 2024, Theorem 20), the coupling matrix V in a SCR over \mathbb{R} contains only ± 1 instead of $\{\pm 1, \pm i\}$.

Theorem 12. For any reservoir system $R = (W, V, h)$ of dimensions (n, m, d) and any $\epsilon > 0$, there exists a SCR $R' = (W', V', h')$ of dimension (n', m, d) that is ϵ -close to R . Moreover, $\|W\| = \|W'\|$ and h' is h with linearly transformed domain.

Proof. One may refer to the proof of (Li, Fong, and Tino 2024, Theorem 20). Crucially, because the dynamic coupling matrix V in the intermediate steps are all over \mathbb{R} instead of \mathbb{C} , the resulting matrix V' only have ± 1 . \square

(Grigoryeva and Ortega 2018b)(Corollary 11) shows that linear reservoir systems with polynomial readouts are universal. They can approximate to arbitrary precision time-invariant causal fading memory filters. This result, together with Theorem 12, establish universality of SCRs in the real domain. Indeed, given a time-invariant fading memory filter, one can find an approximating linear reservoir system with polynomial readout h approximating the filter to the desired precision. By Theorem 12, we can in turn *constructively* approximate this reservoir system with a SCR, again to arbitrary precision in the sense of Definition 2. Moreover, the SCR readout is a polynomial of the same degree as h , since it is h with linearly transformed domain. We summarize our results in the main theorem below:

Theorem 13. Any time-invariant fading memory casual filter over uniformly bounded inputs can be approximated to arbitrary precision by a SCR.

Numerical Experiments

We conclude the paper with numerical experiments illustrating our contributions. For reproducibility of the experiments, all experiments are CPU-based and are performed on Apple M3 Max with 128GB of RAM. The source code and data of the numerical analysis is openly available at:

Code — <https://github.com/Lampertos/RSCR>

Dilation of Linear Reservoirs on Time Series Forecasting

This section illustrates the structural approximation properties of linear reservoir systems when dilating the reservoir coupling matrix, when applied to univariate time series forecasting. The readout function will be assumed to be linear

throughout the numerical analysis in this section, as the primary objective of this paper is to examine the structural approximation properties of the state-space representation of linear reservoirs as determined by the coupling matrix.

Initially, a linear reservoir system featuring a randomly generated coupling matrix W is constructed. We then approximate this system by linear reservoir systems under two types of dilated coupling matrices: (1) U – Orthogonal dilation of W (Theorem 6), and (2) C – Cyclic dilation of U (Theorem 9).

The results demonstrate that the prediction loss approximation error diminishes progressively as the dimension of dilation increases. For demonstration purposes, we keep dimensionality of the original reservoir system to be approximated by SCR low ($n = 5$). The elements of W are independently sampled from the uniform distribution $U(0, 1)$. The elements of input-to-state coupling V is generated by scaling the binary expansion of the digits of π by 0.05 (Rodan and Tiño 2010). Univariate forecasting performance of the initial and the approximating reservoir systems are compared on two popular datasets used in recent time series forecasting studies (e.g. (Zhou et al. 2020) (adopting their training/validation/test data splits)):

ETT The Electricity Transformer Temperature dataset consists of measurements of oil temperature and six external power-load features from transformers in two regions of China. The data was recorded for two years, and measurements are provided either hourly (indicated by 'h') or every 15 minutes (indicated by 'm'). In this paper we used oil temperature of the ETTm2 dataset for univariate prediction with train/validation/test split being 12/4/4 months.

Datasets — <https://github.com/zhouhaoyi/ETDataset>

ECL The Electricity Consuming Load consists of hourly measurements of electricity consumption in kWh for 321 Portuguese clients during two years. In this paper we used client MT 320 for univariate prediction. The train/validation/test split is 15/3/4 months.

Datasets — <https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014>

The readout h of the original reservoir system is trained using ridge regression with a ridge coefficient of 10^{-9} . Note that modified versions of the input-to-state map V and the readout map h will be employed in all subsequent dilated systems. Specifically, the readout map will not be subject to further training in these systems. In all simulations, we maintain a spectral radius $\lambda = 0.9$ and prediction horizon is set to be 300.⁴

The initial system $R = (W, V, h)$ is dilated over a set of pre-defined dilation dimensions $\mathcal{D} := \{2, 6, 10, 15, 19, 24, 28, 33, 37, 42\}$ ⁵. For each $N \in \mathcal{D}$, by

⁴These two parameters are primarily chosen not for accuracy of the forecasting capacities but to demonstrate the structural approximation properties proven in this paper.

⁵Recall that the corresponding orthogonal dilation will have dimensions $(N + 1) \cdot 5 \times (N + 1) \cdot 5$ for each $N \in \mathcal{D}$.

Theorem 6, we construct a linear reservoir system R_U with an orthogonal dynamic coupling W_U of dimension $n_U = (N + 1)n$. Then by Theorem 9, we dilate R_U into an ϵ -close linear reservoir system R_C with contractive cyclic-permutation dynamic coupling.

For the orthogonal dilation, the linear reservoir system $R_U := (W_U, V_U, h_U)$ is defined by:

$W_U := \lambda \cdot U$, where

$$U := \begin{bmatrix} W & & & & D_{W^\top} \\ D_W & & & & -W^\top \\ & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix} \in \mathbb{M}_{n_U \times n_U}(\mathbb{R})$$

$$V_U := \begin{bmatrix} V \\ 0 \end{bmatrix}, \quad h_U(x) = h(P_n(x)),$$

where $n_U := 5 \cdot (N + 1)$, and $P_n : \mathbb{R}^{n_U} \hookrightarrow \mathbb{R}^5$ denote the projection onto the first $n = 5$ coordinates.

Since U is orthogonal, it's canonical form T_U can be obtained via the real version of Schur's decomposition:

$$U = J_U T_U J_U^\top,$$

where we note that both U and T_U have unit spectral radius.

By Remark 8, given $\epsilon > 0$ and the canonical form T_U of U , the maximum matching program in bipartite graphs allows us to find the canonical form of the $n_C \times n_C$ -dimensional root-completed-matrix A (along with the corresponding dimension n_C), given by $T := A_0 \oplus D$, described in the proof of Theorem 7. By construction T is ϵ -close to $T_U \oplus D$ in terms of operator norm.

Let C be the $n_C \times n_C$ -dimensional full cycle permutation matrix of unit spectral radius. Since C is orthogonal, we can once again apply Schur's decomposition to obtain it's canonical form:

$$C = J_C T_C J_C^\top.$$

By construction, T contains rotational matrices with angles at the roots of unity that are rearrangements of the eigenvalues of cyclic permutation of the same size. Therefore there exists permutation matrix \tilde{P} such that:

$$\tilde{P} T \tilde{P}^\top = T_C.$$

Therefore by the proof of Theorem 9 (following that of (Li, Fong, and Tino 2024, Theorem 14)), the linear reservoir system $R_C := (W_C, V_C, h_C)$ for the cyclic dilation is defined by:

$$W_C := \lambda \cdot C, \quad V_C := P \begin{bmatrix} V_U \\ 0 \end{bmatrix},$$

$$h_C(\mathbf{x}) = h(P_n(P^\top \mathbf{x})), \quad P := J_C \tilde{P} \tilde{J}_U^\top, \quad (2)$$

where $C \in \mathbb{M}_{n_C \times n_C}(\mathbb{R})$ is a cyclic permutation of dimension $n_C > 5 \cdot (N + 1)$, and \tilde{J}_U is a $n_C - 5 \cdot (N + 1) \times n_C - 5 \cdot (N + 1)$ matrix with an J_U on the upper left hand corner and zero everywhere else.

By the uniform continuity of the readout h , it suffices to evaluate the closeness of the state trajectories of the original and the approximating cyclic dilation systems, as they are driven by the same input time series ETTm2 and ECL. The two state activation sequences are not directly comparable, but they become comparable if the states of the approximating system are transformed by the orthogonal matrix P (Equation (2)) and projected into the first n coordinates. Figure 2 shows the mean square differences between the states of the two systems as a function of the dilation dimension N . As expected, MSE between the states decays to zero exponentially as the dilation dimension increases.

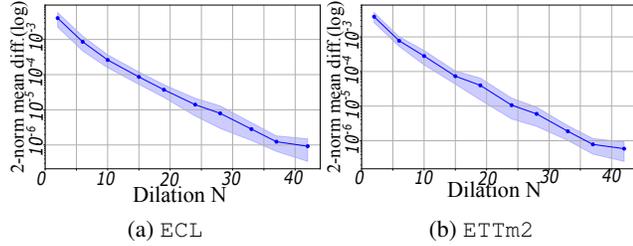


Figure 2: Mean and 95% confidence intervals of the mean square differences of the states of the original reservoir and the approximating cyclic dilation systems over 15 randomized generations of the original system. The data used is labelled in the sub-caption.

Reduction of Dilation Dimension With Maximum Matching in Bipartite Graph

In this section we illustrate how the dimension n_C of the cyclic dilation obtained from the maximum matching program in bipartite graphs discussed in Remark 8 can yield reservoir sizes drastically lower than the theoretical upper bound given by the approximating full-cycle permutation P :

$$n_1 = 2 \cdot \ell_0 \cdot (k + 1) > \left[2 \cdot \frac{\pi}{\arccos\left(1 - \frac{\delta^2}{2}\right)} \cdot (k + 1) \right].$$

We generate 10 orthogonal matrices U uniformly randomly for each initial dimension $n \in \{20, 40, \dots, 140, 160\}$. We perform cyclic dilation as described in the previous section and compare it against the theoretical upper bound. Notice that the y -axis is in log scale. The dimension n_C is significantly lower than the theoretical upper bound, reaching ≈ 300 – 400 units for initial reservoirs of size 80 – 160 , which is well within possibilities of hardware implementations of such reservoirs.

Conclusion

In this paper, we rigorously demonstrated the universality of Simple Cycle Reservoir (SCR) in the real domain, adopting the strategy from (Li, Fong, and Tino 2024). Specifically, we proved that SCRs are universal approximations for any real-valued unrestricted linear reservoir system and any real-valued time-invariant fading memory filter over uniformly bounded input streams.

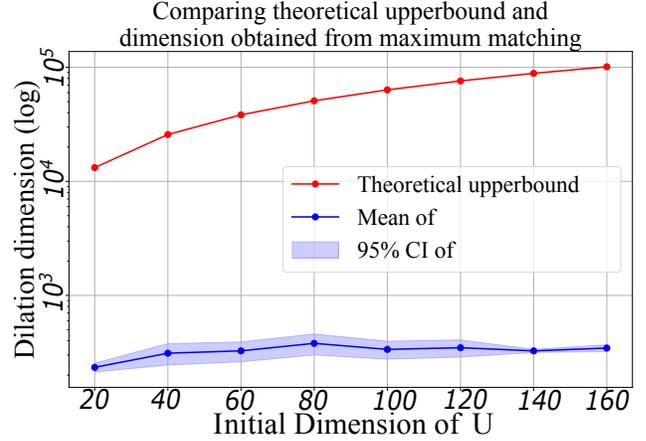


Figure 3: Theoretical upper bound v.s. dimension from maximum matching of bipartite graph.

To achieve this, we constrained our approach to the real domain throughout the approximation pipeline. We performed cyclic dilation of orthogonally dilated coupling matrices by completing the set of roots of unity in the canonical form of orthogonal matrices, rather than using the eigendecomposition method involving unitary matrices. This ensured that all approximant systems remained in the real domain under orthogonal similarity.

We facilitated the completion of roots of unity by utilizing a maximum matching program in bipartite graphs, enabling a tighter dimension expansion of the approximation system. This method ensured efficient and effective expansion to achieve the desired approximation accuracy.

The fully constructive nature of our results is a crucial step towards the physical implementations of reservoir computing (Appeltant et al. 2011; Nakajima, Tanaka, and Hashimoto 2021; Coarer et al. 2018; Abe et al. 2024).

Acknowledgements

B. Li is supported in part by NSF grant (DMS-2350543). P. Tiño was supported by the EPSRC Prosperity Partnerships grant ARCAN, EP/X025454/1.

References

- Abe, Y.; Nakada, K.; Hagiwara, N.; Suzuki, E.; Suda, K.; Mochizuki, S.-i.; Terasaki, Y.; Sasaki, T.; and Asai, T. 2024. Highly-integrable analogue reservoir circuits based on a simple cycle architecture. *Sci. Rep.*, 14(10966): 1–10.
- Appeltant, L.; Soriano, M. C.; der Sande, G. V.; Danckaert, J.; Massar, S.; Dambre, J.; Schrauwen, B.; Mirasso, C. R.; and Fischer, I. 2011. Information processing using a single dynamical node as complex system. *Nature Communications*, 2.
- Coarer, F. D.-L.; Sciamanna, M.; Katumba, A.; Freiberger, M.; Dambre, J.; Bienstman, P.; and Rontani, D. 2018. All-Optical Reservoir Computing on a Photonic Chip Using Silicon-Based Ring Resonators. *IEEE Journal of Selected Topics in Quantum Electronics*, 24(6): 1 – 8.

Gonon, L.; and Ortega, J.-P. 2019. Reservoir computing universality with stochastic inputs. *IEEE transactions on neural networks and learning systems*, 31(1): 100–112.

Grigoryeva, L.; and Ortega, J.-P. 2018a. Echo state networks are universal. *Neural Networks*, 108: 495–508.

Grigoryeva, L.; and Ortega, J.-P. 2018b. Universal Discrete-time Reservoir Computers with Stochastic Inputs and Linear Readouts Using Non-homogeneous State-affine Systems. *J. Mach. Learn. Res.*, 19(1): 892–931.

Halmos, P. R. 1950. Normal dilations and extensions of operators. *Summa Brasil. Math.*, 2: 125–134.

Jaeger, H. 2001. The "echo state" approach to analysing and training recurrent neural networks. Technical report gmd report 148, German National Research Center for Information Technology.

Jaeger, H. 2002a. Short term memory in echo state networks. Technical report gmd report 152, German National Research Center for Information Technology.

Jaeger, H. 2002b. A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. Technical report gmd report 159, German National Research Center for Information Technology.

Jaeger, H.; and Haas, H. 2004. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless telecommunication. *Science*, 304: 78–80.

Li, B.; Fong, R. S.; and Tino, P. 2024. Simple Cycle Reservoirs are Universal. *Journal of Machine Learning Research*, 25(158): 1–28.

Lukosevicius, M.; and Jaeger, H. 2009. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3): 127–149.

Maass, W.; Natschlager, T.; and Markram, H. 2002. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, 14(11): 2531–2560.

Nakajima, M.; Tanaka, K.; and Hashimoto, T. 2021. Scalable reservoir computing on coherent linear photonic processor. *Communications Physics*, 4(1): 20.

Paulsen, V. 2002. *Completely bounded maps and operator algebras*, volume 78 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. ISBN 0-521-81669-6.

Rodan, A.; and Tiño, P. 2010. Minimum complexity echo state network. *IEEE transactions on neural networks*, 22(1): 131–144.

Tiño, P.; and Dorffner, G. 2001. Predicting the future of discrete sequences from fractal representations of the past. *Machine Learning*, 45(2): 187–218.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2020. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *AAAI Conference on Artificial Intelligence*.