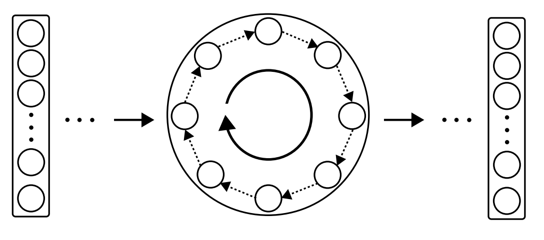
Universality of Simple Cycle Reservoirs

Robert Simon Fong[†], Boyu Li[‡], Peter Tino[†]

[†] University of Birmingham; [‡] New Mexico State University

Overview



Simple Cycle Reservoirs(SCRs) [15] are a type of **highly restricted** neural network architecture where information flows through a **fixed**, **circular** pathway with a **single degree of freedom**.

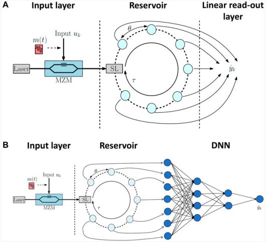
- Efficient processing of sequential data with minimal computational complexity
- Simplicity is particularly advantageous for hardware implementations

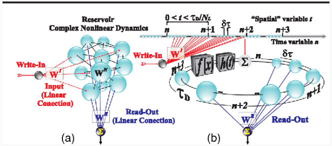
What are Simple Cycle Reservoirs and why do we care?

SCRs are widely adopted in both hardware implementations [2, 13, 4, 1] and excels in time series forecasting tasks [18]. **Yet its theoretical foundations have been lacking.**

This study fills the gap – we demonstrate, **constructively**, the representational power of SCR and show:

SCRs are universal approximators of time-invariant dynamic filters with fading memory in $\mathbb{C}$ and $\mathbb{R}$, respectively.





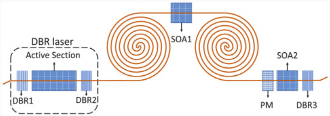


Figure: Photonic implementations of SCR from [3, 11, 9] resp.

Brief overview

- Reservoir Computers are special types of Recurrent Neural Network recently proven, with **existential** proof, to be **universal**.
- **Simple Cycle Reservoirs** is a highly restricted type of Reservoir Computer with **one degree of freedom**, which is important for both digital computing and analogue computers.
- We present a **constructive** proof that **Simple Cycle Reservoirs are universal** in **both** $\mathbb{C}$ and $\mathbb{R}$.
- The universality of SCRs not only advances theoretical understanding but are also pivotal for practical hardware implementations in emerging computing technologies such as photonic integrated circuits.
    - In photonic circuits: Cyclic reservoir is the only thing that is currently implementable. Here we show that SCR is all they need.

Definitions

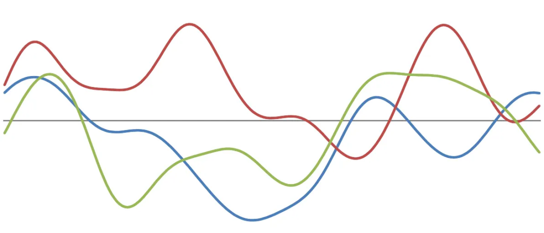
Definitions ... Definitions ... Definitions...

Time series

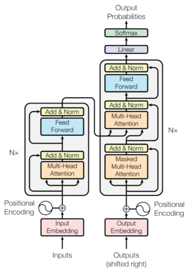
The context of this talk is machine learning on time series data, characterized by their **temporal dependencies**. More formally:

Definition

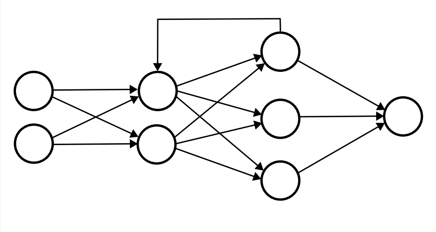
A **Time Series** is a sequence of data points $\{x_t\}_{t \in I}$ indexed by time $t \in I$.



Two Paradigms of learning from Time Series

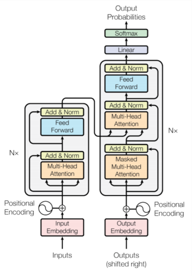


- "Trades time for space" – time series a *static input*
- Temporal correlation is disregarded due to the nature of the static input. [19]
- Example: transformers

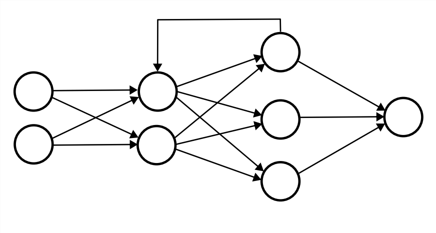


- Captures *temporal dependencies* in the input data stream through *parametric state-space modelling*
- Sequentially encodes input time series in state space
- Examples: Recurrent Neural Networks (RNN), Kalman Filters, etc

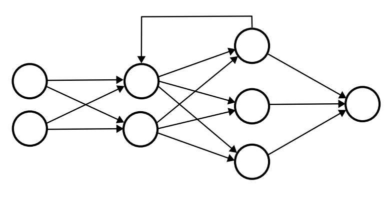
Two Paradigms of learning from Time Series

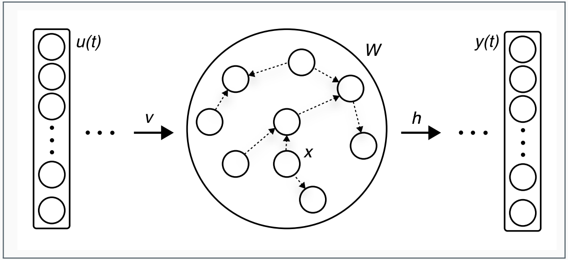


- “Trades time for space" – time series a *static input*
- Temporal correlation is disregarded due to the nature of the static input. [19]
- Example: transformers



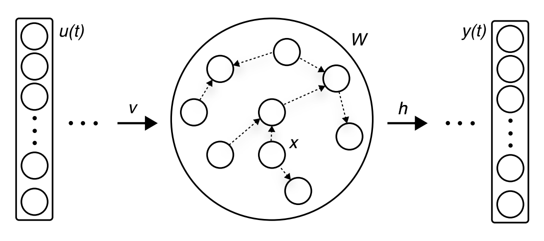
- Captures *temporal dependencies* in the input data stream through *parametric state-space modelling*
- Sequentially encodes input time series in state space
- Examples: Recurrent Neural Networks (RNN), Kalman Filters, etc





- **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.
- The simplest recurrent neural network realization of RC are known as **Echo State Networks** (ESN).

Reservoir Computing as Dynamical System

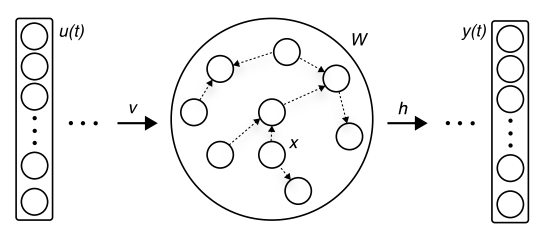


$$\begin{cases} \mathbf{x}_t &= f(W\mathbf{x}_{t-1} + V\mathbf{c}_t) \\ \mathbf{y}_t &= h(\mathbf{x}_t) \end{cases}$$

where:

- $\mathbf{W} \in \mathbb{R}^{N \times N}$ – **fixed** dynamic coupling matrix, spectral radius < 1.
- $V \in \mathbb{R}^N$ – **fixed** input couplings
- $h : \mathbb{R}^N \to \mathbb{R}^d$ – **trainable** readout map
- $f : \mathbb{R}^N \mapsto \mathbb{R}^N$ – **fixed** activation function
- $\{\mathbf{x}(t)\}_t \subset \mathbb{R}^N$ – states
- $\{u(t)\}_t \subset \mathbb{R}$ – inputs
- $\{\mathbf{y}(t)\}_t \subset \mathbb{R}^d$ – outputs

Reservoir Computing as LINEAR Dynamical System



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

where:

- $\mathbf{W} \in \mathbb{R}^{N \times N}$ – **fixed** dynamic coupling matrix, spectral radius < 1.
- $V \in \mathbb{R}^N$ – **fixed** input couplings
- $h : \mathbb{R}^N \to \mathbb{R}^d$ – **trainable** readout map
- $f : \mathbb{R}^N \mapsto \mathbb{R}^N$ – **identity** function
- $\{\mathbf{x}(t)\}_t \subset \mathbb{R}^N$ – states
- $\{u(t)\}_t \subset \mathbb{R}$ – inputs
- $\{\mathbf{y}(t)\}_t \subset \mathbb{R}^d$ – outputs

Definition

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

The corresponding linear dynamical system is given by:

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

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

We make the following assumptions for the system:

1. W is assumed to be strictly **contractive**. In other words, its operator norm $\|W\| < 1$.
2. The input stream is $\{\mathbf{c}_t\}_{t\in\mathbb{Z}_-}$ is **uniformly bounded**. In other words, there exists a constant M such that $\|\mathbf{c}_t\| \leq M$ for all $t \in \mathbb{Z}_-$.

Filters

Definition

Filters are transformations of discrete time signals of infinite length. More formally a filter $\mathcal{F}$ for discrete-time system is represented as:

$$y_t = \mathcal{F}(x_t).$$

Both Kalman filter and RNN are examples of filters

Properties of Filters

Definition

- **A** filter $\mathcal{F}$ is **time-invariant** if time-invariant if its output does not change when the input is shifted in time:

$$y_{t-\tau} = \mathcal{F}(x_{t-\tau}).$$

- A filter $\mathcal{F}$ has **fading memory property** if the influence of past inputs on the current output diminishes over time.

Contractive dynamical coupling W guarantees the so-called **Echo State Property** (ESP), which in turn implies the filter corresponding to a (linear) reservoir system is time-invariant and has fading memory property (FMP).

Universality of Echo State Networks

Given the simplicity of ESN, it is natural to ask what is their representational power, compared with the general class of time-invariant fading memory dynamic filters. ESN is shown to be **universal** in recent work in various settings. Most relevant to us is the following:

Theorem (Grigoryeva and Ortega [8](Corollary 11), paraphrased)

Linear reservoir systems with polynomial readouts are universal, in the sense that any time-invariant fading memory filter can be approximated by to arbitrary precision by a linear reservoir system.

Other settings include:

- Universal approximation capability was first established in the L^∞ sense for deterministic, as well as almost surely uniformly bounded stochastic inputs [8].
- This was later extended in [7] to L^p, $1 \leq p < \infty$ and not necessarily almost surely uniformly bounded stochastic inputs.

BUT These results are **existential** arguments. Choice of the fixed reservoir and input mapping remains an open problem such as in neural architecture search

Random connection is not optimal [14, 17]

Definitions in Simple Cycle Reservoirs

Definition

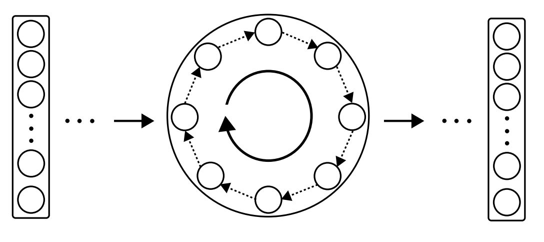
Let $P = [p_{ij}]$ be an $n \times n$ matrix.

1. We say P is a **permutation matrix** if there exists a permutation σ in the symmetric group S_n such that
$$p_{ij} = \begin{cases} 1, & \text{if } \sigma(i) = j, \\ 0, & \text{if otherwise.} \end{cases}$$
2. We say a permutation matrix P is a **full-cycle permutation** if its corresponding permutation $\sigma \in S_n$ is a cycle permutation of length n.
   - Also called **left circular shift** or **cyclic permutation**.

A matrix W is **a contractive permutation** (resp. **a contractive full-cycle permutation** if $W = aP$ for some scalar $a \in (0, 1) \subset \mathbb{R}$ and P is a permutation (resp. full-cycle permutation).

$$W = a \cdot \begin{bmatrix} 0 & 0 & 0 & & \dots & 0 & 1 \\ 1 & 0 & 0 & & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & & 0 & 0 \\ \vdots & & \ddots & & & \vdots & \vdots \\ 0 & \dots & & & & 1 & 0 \end{bmatrix}$$

Simple Cycle Reservoirs

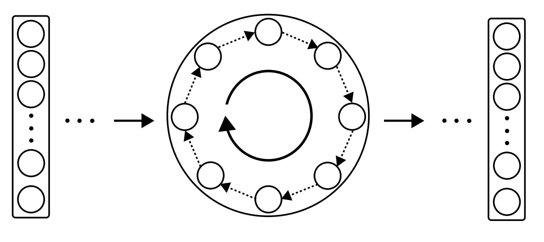


Simple Cycle Reservoirs (SCRs) is a type of highly restricted Reservoir Computer defined by a fixed **cyclic** dynamical coupling and fixed aperiodic input coupling of ± 1 (or $\{\pm 1, \pm i\}$ in the complex case).

Given the wide applicability of SCR, its natural to ask whether SCR is also **universal** in the same sense.

In this talk we answer the question affirmatively in $\mathbb{C}$ and $\mathbb{R}$ respectively (same domain applies for input and output throughout).

Simple Cycle Reservoirs: formal definition

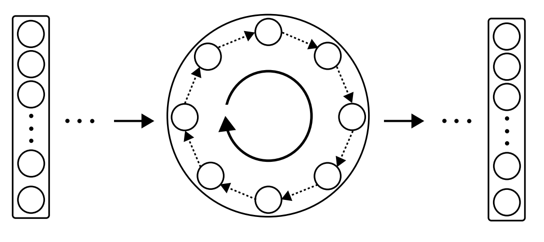


Definition

A linear reservoir system $R = (W, V, h)$ with dimensions (n, m, d) is called:

- A **Simple Cycle Reservoir (SCR)** if:
  1. W is a contractive full-cycle permutation, and
  2. $V \in \mathbb{M}_{n \times m}(\{-1, 1\})$.
- A **Complex Simple Cycle Reservoir ($\mathbb{C}$-SCR)** if:
  1. W is a contractive full-cycle permutation, and
  2. $V \in \mathbb{M}_{n \times m}$ and all entries of V are either ± 1 or $\pm i$.

Simple Cycle Reservoirs: Detail definition



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

where:

- $\mathbf{W} \in \mathbb{R}^{N\times N}$ – **fixed** contractive full-cycle permutation
- $V \in \mathbb{R}^N$ – **fixed** input coupling, which is either:
  - For SCR: $V \in \mathbb{M}_{n\times m}(\{-1,1\})$, or
  - For $\mathbb{C}$-SCR: $V \in \mathbb{M}_{n\times m}(\{\pm 1, \pm i\})$.
- $h : \mathbb{R}^N \to \mathbb{R}^d$ – **trainable** readout map

- $\{\mathbf{x}(t)\}_t \subset \mathbb{R}^N$ – states
- $\{u(t)\}_t \subset \mathbb{R}$ – inputs
- $\{\mathbf{y}(t)\}_t \subset \mathbb{R}^d$ – outputs

Definition

For $k > 1$, a linear reservoir system $R = (W, V, h)$ with dimensions (n, m, d) is called a **Multi-Cycle Reservoir of order** k if:

1. W is block-diagonal with k (not necessarily identical) blocks of contractive full-cycle permutation couplings W_1, ..., W_k, of dimensions $n_i \times n_i$, $i = 1, 2, ..., k$,

$$W := \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_k \end{bmatrix}, \quad \sum_{i=1}^{k} n_i = n, \text{and}$$

2. $V \in \mathbb{M}_{n \times m}(\{-1, 1\})$.

The state $\mathbf{x} \in \mathbb{C}^n$ of such a multi-cycle system is composed of the k component states $\mathbf{x}^{(i)} \in \mathbb{C}^{n_i}$, $i = 1, 2, ..., k$, $\mathbf{x} = (\mathbf{x}^{(1)}, ..., \mathbf{x}^{(k)})$. In our case, the readout will act on a linear combination of the component states,

$$h(\mathbf{x}) = h\left(\sum_{i=1}^{k} a_i \cdot \mathbf{x}^{(i)}\right), \quad a_i \in \mathbb{C} \text{ are mixing coefficients.}$$

SMRC, Twin SCR, and transformation of readout

Definition

A linear reservoir is called a **Simple Multi-Cycle Reservoir (SMCR) of order** k if it is a Multi-Cycle Reservoir of order k with k **identical** (contractive full-cycle permutation) blocks.

Definition

A linear reservoir is called a **Twin Simple Cycle Reservoir (Twin SCR)** if it is a Multi-Cycle Reservoir of order 2.

It will often be the case that when we transform one reservoir system into another the corresponding readout mappings will be closely related to each other. In particular:

Definition

Given two functions h, g sharing the same domain $D \subset \mathbb{K}^n$, where $\mathbb{K}^n$ is a field, we say that g **is** h **with linearly transformed domain** if there exists a linear transformation over $\mathbb{K}^n$ with the corresponding matrix A such that $g(\mathbf{x}) = h(A\mathbf{x})$ for all $\mathbf{x} \in D$.

Equivalent Reservoir Systems

Definition

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, the two systems generate the same output stream. More precisely, for any input $c = \{\mathbf{c}_t\}_{t\in\mathbb{Z}_-}$, the solutions $\{(\mathbf{x}_t, \mathbf{y}_t)\}_t$ and $\{(\mathbf{x}'_t, \mathbf{y}'_t)\}_t$ for systems R and R', given by:

$$\mathbf{y}_t = h\left(\mathbf{x}_t(c)\right) = h\left(\sum_{j\geq 0} W^j V \mathbf{c}_{t-j}\right) \quad \text{and}$$

$$\mathbf{y}'_t = h'\left(\mathbf{x}'_t(c)\right) = h'\left(\sum_{j\geq 0} \left(W'\right)^j V' \mathbf{c}_{t-j}\right),$$

   respectively, satisfy $\mathbf{y}_t = \mathbf{y}'_t$ for all t.

2. For $\epsilon > 0$, we say the **two systems are ϵ-close** if the outputs of the two systems, given any input stream, are ϵ-close. That is (under the notation above), $\left\|\mathbf{y}_t - \mathbf{y}'_t\right\|_2 < \epsilon$ for all t.

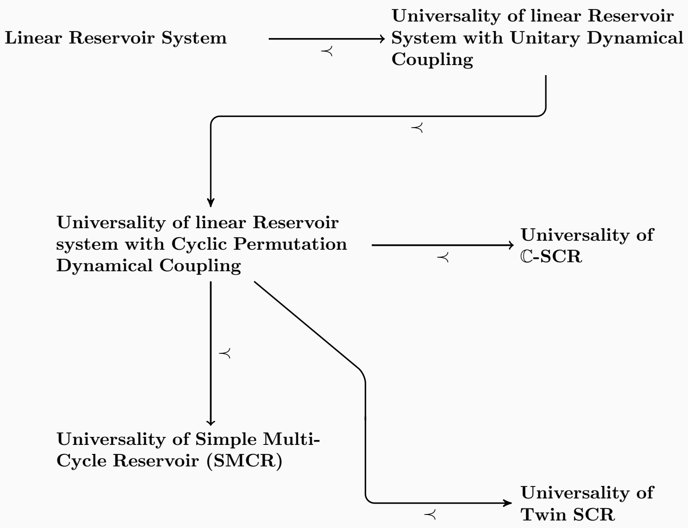
Essentially we view the system as a whole on the input-to-output level.

Part I: Simple Cycle Reservoirs (and friends) are universal over $\mathbb{C}$.

Theorem (Li, F., Tino [12])

Any time-invariant fading memory filter over uniformly bounded inputs can be approximated to arbitrary precision by a Simple Multi-Cycle Reservoir, a $\mathbb{C}$-SCR, or a Twin SCR, each endowed with a polynomial readout.

Summary of main results



Unitary Dilation of Linear Reservoirs – an intermediate step towards constructing $\mathbb{C}$-SCR approximators.

Throughout the proofs we will jiggle between the coupling matrix and input map.

We begin by transforming the arbitrary coupling dynamics to a unitary one with dilation theory.

Definition

A (bounded linear) **operator** is a linear transformation $T : \mathcal{H} \to \mathcal{H}$ on a Hilbert space $\mathcal{H}$.

- Example: $T(x) = ax$ for some scalar a in the space.

Definition

An operator U is:

- **Unitary** if:

$$U^* U = UU^* = I.$$

- **Orthogonal** if:

$$U^\top U = UU^\top = I.$$

- **Dilation Concept:**
  - Dilation theory studies how an operator on a smaller space can be extended to a larger space, where the operator takes a more manageable form.
  - Given an operator T on a Hilbert space $\mathcal{H}$, dilation theory seeks to find a larger Hilbert space $\mathcal{K} \supseteq \mathcal{H}$ and a unitary operator U on $\mathcal{K}$ such that T is a "compression" of U.
  - This means $T = P_{\mathcal{H}} U|_{\mathcal{H}}$, where $P_{\mathcal{H}}$ is the orthogonal projection onto $\mathcal{H}$.
- **Why Dilation?**
  - Unitary operators are easier to study because they have well-understood spectral properties.
  - Dilation transforms the problem of analyzing T into the simpler problem of analyzing U.

Theorem (Sz.-Nagy [16])

Given a contraction $W \in \mathbb{M}_{n\times n}$ over $\mathbb{C}$ ($\|W\| \leq 1$), there exists a unitary operator on an infinite-dimensional Hilbert space $\mathcal{H}$ and an isometric embedding $J : \mathbb{C}^n \to \mathcal{H}$ such that $W^k = J^* U^k J$ for all $k \in \mathbb{Z}$.

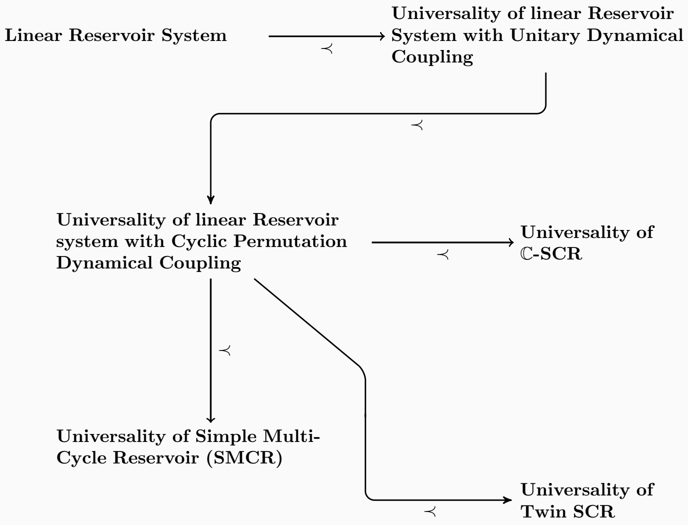
If we only require $W^k = J^* U^k J$ for all $1 \leq k \leq N$, Egerváry showed the following:

Theorem (Egerváry [5])

Given a contraction $W \in \mathbb{M}_{n\times n}$ over $\mathbb{C}$ ($\|W\| \leq 1$), there exists a unitary operator on an $(N+1)\cdot n$ dimensional Hilbert space $\mathcal{H}$ and an isometric embedding $(N+1)\cdot n \times n$ matrix J over $\mathbb{C}$ such that $W^k = J^* U^k J$ for all $k \in \mathbb{Z}$, where:

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

Summary of main results



Unitary Dilation of Linear Reservoirs

Using the dilation technique we get a ϵ-close approximating reservoir systems with a unitary coupling.

Theorem (Li, F., Tino [12])

Let $R = (W, V, h)$ be a reservoir system defined by contraction W with $\|W\| =: \lambda \in (0, 1)$ and satisfying the assumptions of Definition 2. Given $\epsilon > 0$, there exists a reservoir system $R' = (W', V', h')$ that is ϵ-close to R with $W' = \lambda U$ for a unitary U. Moreover, h' is h with linearly transformed domain.

Unitary Dilation of Linear Reservoirs: Proof

Sketch of Proof.

$$
\begin{cases}
R := (W, V, h) \\
W \in \mathbb{C}_{n \times n} \\
V \in \mathbb{C}_{m \times n} \\
h : \mathbb{C}^n \to \mathbb{C}^d \\
\lambda := \|W\|
\end{cases}
\xrightarrow[\prec]{\qquad}
\begin{cases}
\textbf{Unitary universal} \\
R_U := (W_U, V_U, h_U) \\
W_U := \lambda \cdot U \in \mathbb{C}_{(N+1)n \times (N+1)n} \\
U := \begin{bmatrix} W & & & & D_{W^*} \\ D_W & & & & -W^* \\ & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}, \quad V_U := \begin{bmatrix} V \\ 0 \end{bmatrix} \\
h_U(x) = h\left(P_n(x)\right) \\
n_U := (N+1)\, n
\end{cases}
$$

□

From Unitary to Full-Cycle Permutation State Coupling

We first show that matrix similarity of dynamical coupling implies reservoir equivalence.

Theorem (Li, F., Tino [12])

Let W be a contraction and let $R = (W, V, h)$ denote the corresponding reservoir system. Suppose S is an invertible matrix such that $W' = S^{-1}WS$ and $\|W'\| < 1$. Then there exists a reservoir system $R' = (W', V', h')$ that is equivalent to R.

The proof is technical and will be omitted in the talk.

From Unitary to Full-Cycle Permutation State Coupling

With the lemma we first show the equivalence on the matrix level.

In particular, we show that any for given unitary state coupling we can always find a full-cycle permutation that is close to it to arbitrary precision.

This is done by perturbing a given unitary matrix to one that is unitarily equivalent to a cyclic permutation.

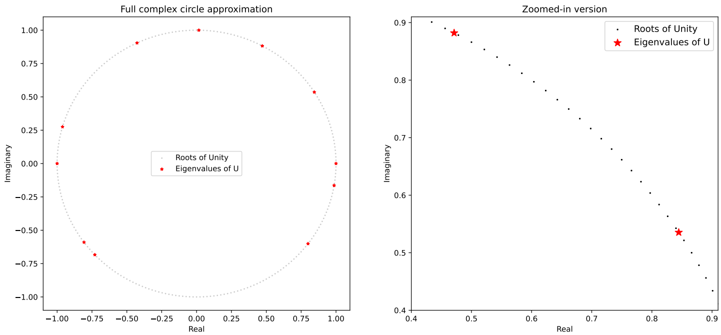
Theorem (Li, F., Tino [12])

Let U be an $n \times n$ unitary matrix and $\delta > 0$ be an arbitrarily small positive number. There exists an $n_1 \times n_1$ matrix A with $n_1 > n$ that is unitarily equivalent to a full-cycle permutation, and an $(n_1 - n) \times (n_1 - n)$ diagonal matrix D such that:

$$\left\| A - \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \right\| < \delta.$$

Sketch of Proof.

It is well known that the eigenvalues of unitary matrices lie on the unit circle $\mathbb{T}$ in $\mathbb{C}$, and the eigenvalues of cyclic permutations are the roots of unities in $\mathbb{C}$.
Given a unitary matrix U, we can therefore first perturb its eigenvalues to a subset of eigenvalues of a cyclic permutation matrix. The remaining roots of unity (eigenvalues of the cyclic permutation) not covered by the previous operation are then filled in using direct sum with the diagonal matrix consisting of the missing eigenvalues.



From Unitary to Full-Cycle Permutation State Coupling

Combining the two results, we obtain the universality of linear reservoirs with full-cycle permutation coupling:

Theorem (Li, F., Tino [12])

Let U be an $n \times n$ unitary matrix and $W = \lambda U$ with $\lambda \in (0, 1)$. Let $R = (W, V, h)$ be a reservoir system that satisfies the assumptions of Definition 2 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.*

Sketch of Proof.

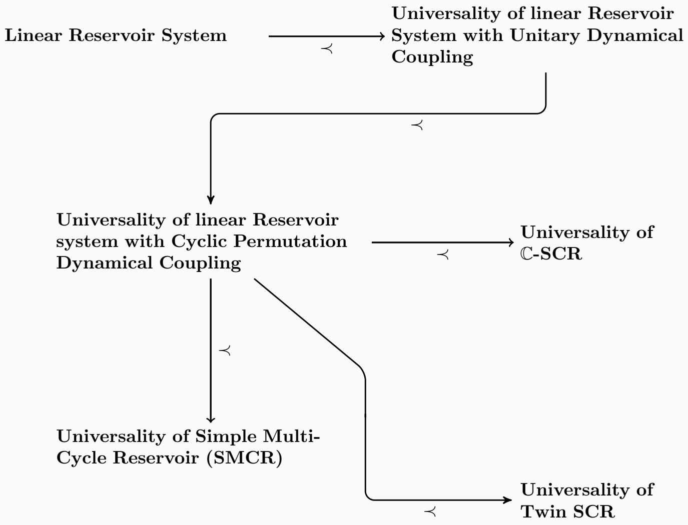
$$
\begin{cases}
R := (W, V, h) \\
W \in \mathbb{C}_{n \times n} \\
V \in \mathbb{C}_{m \times n} \\
h : \mathbb{C}^n \to \mathbb{C}^d \\
\lambda := \|W\|
\end{cases}
\xrightarrow[\prec]{}
\begin{cases}
\textbf{Unitary universal} \\
R_U := (W_U, V_U, h_U) \\
W_U := \lambda \cdot U \in \mathbb{C}_{(N+1)n \times (N+1)n} \\
U := \begin{bmatrix} W & & & & D_{W^*} \\ D_W & & & & -W^* \\ & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}, \quad V_U := \begin{bmatrix} V \\ 0 \end{bmatrix} \\
h_U(x) = h\left(P_n(x)\right) \\
n_U := (N+1)\, n
\end{cases}
$$

$$
\xrightarrow[\prec]{}
\begin{cases}
\textbf{Cyclic Permutation universal} \\
R'_U := (W'_U, V'_U, h'_U) \\
W'_U = \lambda \cdot \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \cong \lambda \cdot P, \\
P - \text{cyclic permutation, } P \in \mathbb{C}_{n'_U \times n'_U} \\
V'_U := S \begin{bmatrix} V_U \\ 0 \end{bmatrix}, \quad h'_U(\mathbf{x}) = h_U\left(P_{n'_U}\left(S^* \mathbf{x}\right)\right) \\
n'_U > n_U, \; S\text{– unitary transform}
\end{cases}
$$

□

We showed that the dynamical coupling W of the reservoir system can be made into a full-cycle permutation. It remains to show that the input-coupling V can be made into $\mathbb{M}_{n\times m}(\{\pm 1\})$ or $\mathbb{M}_{n\times m}(\{\pm 1, \pm i\})$.

Summary of main results



Universality of SMCR

Universality of Simple Multi-Cycle Reservoir

We begin by enforcing all entries of the input-coupling to be in ± 1.

Theorem (Li, F., Tino [12])

For any reservoir system $R = (W, V, h)$ that satisfies the assumptions of Definition 2 and any $\epsilon > 0$, there exists a Simple Multi-Cycle Reservoir $R' = (W', V', h')$ that is ϵ-close to R. Moreover, $\|W\| = \|W'\|$ and h' is h with linearly transformed domain.

It is important to note that

- Here W' is a contractive **permutation**. That is, it is no longer a full-cycle permutation.
- We managed to get $V' \in \mathbb{M}_{n \times m}(\{\pm 1\})$, as per definition of SMCR.

Sketch of Proof.

$$\begin{cases} \textbf{Cyclic Permutation universal} \\ R'_U := (W'_U, V'_U, h'_U) \\ W'_U = \lambda \cdot \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \cong \lambda \cdot P, \\ P - \text{cyclic permutation}, \ P \in \mathbb{C}_{n'_U \times n'_U} \\ V'_U := S \begin{bmatrix} V_U \\ 0 \end{bmatrix}, \quad h'_U(\mathbf{x}) = h_U \left(P_{n'_U} \left(S^* \mathbf{x} \right) \right) \\ n'_U > n_U, \ S - \text{unitary transform} \end{cases}$$

$$\Big\downarrow \prec$$

$$\begin{cases} \textbf{SMCR universal} \\ R_P := (W_P, V_P, h_P) \\ W_P \in \mathbb{C}_{(n'_U)^2 \cdot m \times (n'_U)^2 \cdot m} \\ W_P - \text{contractive permutation.} \\ W_P := \begin{bmatrix} \lambda \cdot P & & \\ & \ddots & \\ & & \lambda \cdot P \end{bmatrix} \\ V_P \in \mathbb{M}_{m \times (n'_U)^2 \cdot m} \left(\{-1, 1\} \right) \\ h_P(\mathbf{x}_1, \cdots, \mathbf{x}_{n'_U m}) := h'_U \left(\sum_{i=1}^{n'_U m} a_i \mathbf{x}_i \right) \\ n_p = n'_U \cdot (n'_U m) \end{cases}$$

Universality of $\mathbb{C}$-SCR

Moving back to cyclic permutation coupling

In order to regain the **full-cycle permutation** dynamical coupling matrix. This requires a more careful construction.

We first show that an arrangement of full-cycle block of individual full-cycle permutation blocks can be under some conditions rearranged into a larger full-cycle permutation matrix.

Lemma (Li, F., Tino [12])

Let n, k be two natural numbers such that $\mathrm{gcd}(n,k) = 1$. 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.

On the gcd condition

We emphasis that the condition $\gcd(n, k) = 1$ is crucial.

Consider a simple example where $n = 2$ and $k = 3$. Let P be the matrix for cyclic permutation $(1, 2)$,

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

From our construction, the matrix P_1 is

$$P_1 = \begin{bmatrix} & & & & & 1 \\ & & & & 1 & \\ & 1 & & & & \\ 1 & & & & & \\ & & & 1 & & \\ & & 1 & & & \end{bmatrix}$$

One can check that P_1 corresponds to the cyclic permutation $(1, 4, 5, 2, 3, 6)$. If we picked $k = 2$, then

$$P_1 = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix},$$

which is not a full-cycle permutation.

Universality of $\mathbb{C}$-SCR and Twin SCRs

The second machinery we require is a way to break any input-coupling to $\mathbb{M}(\pm 1)$ and $\mathbb{M}(\pm 1, \pm i)$. The lemma is technical and we refer to the paper for detailed exposition. It is sort of like finding the closest rational number to each real (or complex) entry.

Lemma (Li, F., Tino [12])

*For any $n \times m$ **real** matrix V and $\delta > 0$, there exists k matrices $\{F_1, \cdots, F_k\} \subset \mathbb{M}_{n \times m}(\{-1, 1\})$ 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 $\gcd(k, n) = 1$.

Corollary (Li, F., Tino [12])

*For any $n \times m$ **complex** matrix V and $\delta > 0$, there exists k matrices $\{F_1, \cdots, F_k\}$ where each $F_i \in \mathbb{M}_{n \times m}(\pm 1)$ or $\mathbb{M}_{n \times m}(\pm i)$ 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 $\gcd(k, n) = 1$.

Finally, we are ready to show our main theorem:

Theorem (Li, F., Tino [12])

For any reservoir system $R = (W, V, h)$ of dimensions (n, m, d) that satisfies the assumptions of Definition 2 and any $\epsilon > 0$, there exists a $\mathbb{C}$-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.

By construction, W' is a contractive full-cycle permutation and entries of V' are either all ± 1 or $\pm i$. This comes at a price of dimension increase.

Sketch of Proof.

$$\begin{cases} \textbf{Cyclic Permutation universal} \\ R'_U := (W'_U, V'_U, h'_U) \\ W'_U = \lambda \cdot \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \cong \lambda \cdot P, \\ P - \text{cyclic permutation, } P \in \mathbb{C}_{n'_U \times n'_U} \\ V'_U := S \begin{bmatrix} V_U \\ 0 \end{bmatrix}, \quad h'_U(\mathbf{x}) = h_U\left(P_{n'_U}\left(S^*\mathbf{x}\right)\right) \\ n'_U > n_U,\ S\text{– unitary transform} \end{cases} \xrightarrow[\prec]{} \begin{cases} \mathbb{C}\textbf{-SCR universal} \\ R_{\mathbb{C}} := (W_{\mathbb{C}}, V_{\mathbb{C}}, h_{\mathbb{C}}) \\ W_{\mathbb{C}} := \lambda \cdot P_1 \in \mathbb{C}_{n'_U \cdot k \times n'_U \cdot k} \\ P_1 := \begin{bmatrix} 0 & 0 & 0 & & \dots & 0 & P \\ P & 0 & 0 & & \dots & 0 & 0 \\ 0 & P & 0 & \dots & & 0 & 0 \\ \vdots & & \ddots & & & \vdots & \vdots \\ 0 & \dots & & & & P & 0 \end{bmatrix} \cdots (\dagger) \\ V_{\mathbb{C}} \in \mathbb{M}_{m \times n'_U \cdot k}\left(\{-1, 1\}\right) \text{ OR } V_{\mathbb{C}} \in \mathbb{M}_{m \times n'_U \cdot k}\left(\{-i, i\}\right) \\ h_{\mathbb{C}}(\mathbf{x}_1, \dots, \mathbf{x}_k) = h'_U\left(\frac{1}{N^{\mathbb{C}}} \sum_{j=1}^{k} \mathbf{x}_j\right) \\ n_{\mathbb{C}} = n'_U \cdot k;\ k \text{ satisfies } \gcd(k, n'_U) = 1 \end{cases}$$

□

Universality of Twin SCR

We now wish to push the input-coupling V back to $\mathbb{M}(\pm 1)$ **while maintaining the cyclic-permutation dynamical coupling** W. To do this we require the “direct sum" of exactly **two** SCRs:

Theorem (Li, F., Tino [12])

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

Note that:

- W' is a direct sum of two full-cycle permutation matrices.
- $V' \in \mathbb{M}_{n\times m}(\{\pm 1\})$.
- This is **NOT** a stronger version of SMRC. The dynamical coupling of the Twin SCR described here require is **two big blocks of SCR**. Whereas in the previous theorem SMRC has small blocks of full-cycle permutation.

Universality of Twin SCR : Proof

Sketch of Proof.

Cyclic Permutation universal

$$\begin{cases} R'_U := (W'_U, V'_U, h'_U) \\ W'_U = \lambda \cdot \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \cong \lambda \cdot P, \\ P - \text{cyclic permutation}, \ P \in \mathbb{C}_{n'_U \times n'_U} \\ V'_U := S \begin{bmatrix} V_U \\ 0 \end{bmatrix}, \quad h'_U(\mathbf{x}) = h_U\left(P_{n'_U}\left(S^* \mathbf{x}\right)\right) \\ n'_U > n_U, \ S - \text{unitary transform} \end{cases}$$

$\prec$

$\mathbb{C}$-SCR universal

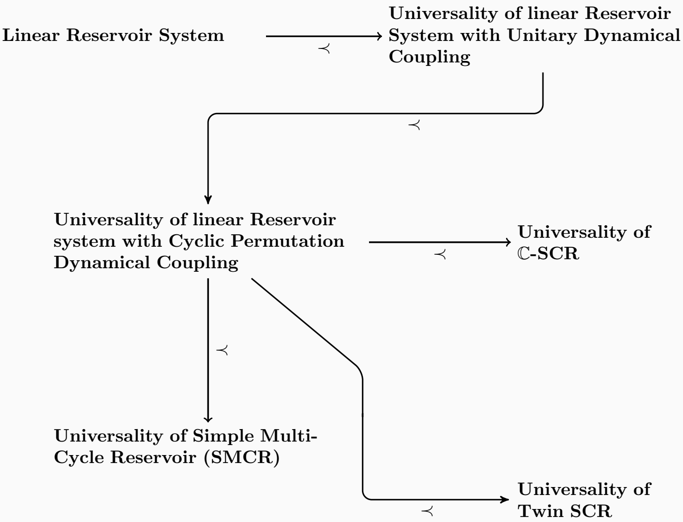
$$\begin{cases} R_{\mathbb{C}} := (W_{\mathbb{C}}, V_{\mathbb{C}}, h_{\mathbb{C}}) \\ W_{\mathbb{C}} := \lambda \cdot P_1 \in \mathbb{C}_{n'_U \cdot k \times n'_U \cdot k} \\ P_1 := \begin{bmatrix} 0 & 0 & 0 & & \dots & 0 & P \\ P & 0 & 0 & & \dots & 0 & 0 \\ 0 & P & 0 & \dots & & 0 & 0 \\ \vdots & & \ddots & & & \vdots & \vdots \\ 0 & \dots & & & & P & 0 \end{bmatrix} \cdots (\dagger) \\ V_{\mathbb{C}} \in \mathbb{M}_{m \times n'_U \cdot k}(\{-1, 1\}) \text{ OR } V_{\mathbb{C}} \in \mathbb{M}_{m \times n'_U \cdot k}(\{-i, i\}) \\ h_{\mathbb{C}}(\mathbf{x}_1, \dots, \mathbf{x}_k) = h'_U\left(\frac{1}{N^{\mathbb{C}}} \sum_{j=1}^{k} \mathbf{x}_j\right) \\ n_{\mathbb{C}} = n'_U \cdot k; \ k \text{ satisfies } \gcd(k, n'_U) = 1 \end{cases}$$

$\prec$

Twin SCR universal

$$\begin{cases} R_{\mathbb{R}} := (W_{\mathbb{R}}, V_{\mathbb{R}}, h_{\mathbb{R}}) \\ W_{\mathbb{R}} := \lambda (P_r \oplus P_i) = \lambda \cdot \begin{bmatrix} P_r & 0 \\ 0 & P_i \end{bmatrix}, \\ \text{where both } P_r, P_i, \text{ has form } (\dagger). \\ W_{\mathbb{R}} \in \mathbb{C}_{n'_U \cdot (k_r + k_i) \times n'_U \cdot (k_r + k_i)} \\ V_{\mathbb{R}} \in \mathbb{M}_{m \times n'_U \cdot (k_r + k_i)}(\{-1, 1\}) \\ h_{\mathbb{R}}(\mathbf{x}_1, \dots, \mathbf{x}_{k_r}, \mathbf{x}'_1, \dots, \mathbf{x}'_{k_i}) \\ \quad = h'_U\left(\frac{1}{N^{\mathbb{R}}_r} \sum_{j=1}^{k_r} \mathbf{x}_j + \frac{i}{N^{\mathbb{R}}_i} \sum_{j=1}^{k_i} \mathbf{x}'_j\right) \\ n_{\mathbb{R}} = n'_U \cdot (k_r + k_i) \\ k_r, k_i \text{ satisfies } \gcd(k_r, n'_U) = \gcd(k_r, n'_U) = 1 \end{cases}$$

Summary of main results



Summary of main results: Detailed

$$\begin{cases} R := (W,V,h) \\ W \in \mathbb{C}_{n\times n} \\ V \in \mathbb{C}_{m\times n} \\ h : \mathbb{C}^n \to \mathbb{C}^d \\ \lambda := \|W\| \end{cases} \quad \xrightarrow{\prec}$$

$$\begin{cases} \textbf{Unitary universal} \\ R_U := (W_U, V_U, h_U) \\ W_U := \lambda \cdot U \in \mathbb{C}_{(N+1)n\times(N+1)n} \\ U := \begin{bmatrix} W & & & & D_{W^*} \\ D_W & & & & -W^* \\ & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}, \quad V_U := \begin{bmatrix} V \\ 0 \end{bmatrix} \\ h_U(x) = h\left(P_n(x)\right) \\ n_U := (N+1)\, n \end{cases}$$

$\downarrow \prec$

$$\begin{cases} \textbf{Cyclic Permutation universal} \\ R'_U := (W'_U, V'_U, h'_U) \\ W'_U = \lambda \cdot \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \cong \lambda \cdot P, \\ P - \text{cyclic permutation, } P \in \mathbb{C}_{n'_U \times n'_U} \\ V'_U := S \begin{bmatrix} V_U \\ 0 \end{bmatrix}, \quad h'_U(\mathbf{x}) = h_U\left(P_{n_U}(S^*\mathbf{x})\right) \\ n'_U > n_U,\ S\text{– unitary transform} \end{cases}$$

$\xrightarrow{\prec}$

$$\begin{cases} \mathbb{C}\textbf{-SCR universal} \\ R_{\mathbb{C}} := (W_{\mathbb{C}}, V_{\mathbb{C}}, h_{\mathbb{C}}) \\ W_{\mathbb{C}} := \lambda \cdot P_1 \in \mathbb{C}_{n'_U \cdot k \times n'_U \cdot k} \\ P_1 := \begin{bmatrix} 0 & 0 & 0 & & \dots & 0 & P \\ P & 0 & 0 & & \dots & 0 & 0 \\ 0 & P & 0 & \dots & & 0 & 0 \\ \vdots & & \ddots & & & \vdots & \vdots \\ 0 & \dots & & & & P & 0 \end{bmatrix} \dots (\dagger) \\ V_{\mathbb{C}} \in \mathbb{M}_{m\times n'_U \cdot k}(\{-1,1\}) \text{ OR } V_{\mathbb{C}} \in \mathbb{M}_{m\times n'_U \cdot k}(\{-i,i\}) \\ h_{\mathbb{C}}(\mathbf{x}_1,\dots,\mathbf{x}_k) = h'_U\left(\frac{1}{N^{\mathbb{C}}}\sum_{j=1}^{k}\mathbf{x}_j\right) \\ n_{\mathbb{C}} = n'_U \cdot k;\ k \text{ satisfies } \gcd(k, n'_U) = 1 \end{cases}$$

Cyclic Permutation universal $\downarrow \prec$

$$\begin{cases} \textbf{SMCR universal} \\ R_P := (W_P, V_P, h_P) \\ W_P \in \mathbb{C}_{(n'_U)^2 \cdot m \times (n'_U)^2 \cdot m} \\ W_P - \text{contractive permutation.} \\ W_P := \begin{bmatrix} \lambda \cdot P & & \\ & \ddots & \\ & & \lambda \cdot P \end{bmatrix} \\ V_P \in \mathbb{M}_{m \times (n'_U)^2 \cdot m}(\{-1,1\}) \\ h_P(\mathbf{x}_1,\cdots,\mathbf{x}_{n'_U m}) := h'_U\left(\sum_{i=1}^{n'_U m} a_i \mathbf{x}_i\right) \\ n_p = n'_U \cdot (n'_U m) \end{cases}$$

Cyclic Permutation universal $\xrightarrow{\prec}$

$$\begin{cases} \textbf{Twin SCR universal} \\ R_{\mathbb{R}} := (W_{\mathbb{R}}, V_{\mathbb{R}}, h_{\mathbb{R}}) \\ W_{\mathbb{R}} := \lambda(P_r \oplus P_i) = \lambda \cdot \begin{bmatrix} P_r & 0 \\ 0 & P_i \end{bmatrix}, \\ \text{where both } P_r, P_i, \text{ has form } (\dagger). \\ W_{\mathbb{R}} \in \mathbb{C}_{n'_U\cdot(k_r+k_i)\times n'_U\cdot(k_r+k_i)} \\ V_{\mathbb{R}} \in \mathbb{M}_{m\times n'_U\cdot(k_r+k_i)}(\{-1,1\}) \\ h_{\mathbb{R}}(\mathbf{x}_1,\dots,\mathbf{x}_{k_r},\mathbf{x}'_1,\dots,\mathbf{x}'_{k_i}) \\ \quad = h'_U\left(\frac{1}{N^{\mathbb{R}}_r}\sum_{j=1}^{k_r}\mathbf{x}_j + \frac{i}{N^{\mathbb{R}}_i}\sum_{j=1}^{k_i}\mathbf{x}'_j\right) \\ n_{\mathbb{R}} = n'_U \cdot (k_r + k_i) \\ k_r, k_i \text{ satisfies } \gcd(k_r, n'_U) = \gcd(k_r, n'_U) = 1 \end{cases}$$

Universality over $\mathbb{C}$

We now showed that: Given a linear reservoir system R with a polynomial readout h, we can *construct* a $\mathbb{C}$-SCR (also SMCR or Twin SCR) R' that is ϵ-close to R in the space of linear reservoir systems.

Combining this with the follow result:

Theorem (Grigoryeva and Ortega [8](Corollary 11), paraphrased)

Linear reservoir systems with polynomial readouts are universal, in the sense that any time-invariant fading memory filter can be approximated by to arbitrary precision by a linear reservoir system.

We obtain our main theorem of Part I:

Theorem (Li, F., Tino [12])

Any time-invariant fading memory filter over uniformly bounded inputs can be approximated to arbitrary precision by a Simple Multi-Cycle Reservoir, a $\mathbb{C}$-SCR, or a Twin SCR, each endowed with a polynomial readout.

It is worth noting that our results are not restricted to polynomial readouts, as long as they are continuous.

Part II: Simple Cycle Reservoirs are universal over $\mathbb{R}$.

Theorem (F., Li, Tino [6])

Any time-invariant fading memory filter can be approximated to arbitrary precision by Simple Cycle Reservoirs (Full stop!)

From $\mathbb{C}$ to $\mathbb{R}$

Reducing from $\mathbb{C}$ to $\mathbb{R}$ is far from straightforward! In particular we have the following open questions:

- In $\mathbb{C}$ attempts to even partially restrict SCR in from $\mathbb{C}$ to $\mathbb{R}$ would result in more complex multi-reservoir structures.
- We want all components to be in $\mathbb{R}$ while **maintaining**:
  1. Dynamical coupling W – **full-cycle permutation**
  2. Input coupling V – **all entries** in ± 1.
- Moreover, how to we **restrict the dimension expansion** from the two-step dilation?

In Part II we will address **all** these questions by:

- We prove that SCRs operating in real domain are universal approximators of time-invariant dynamic filters with fading memory.
- 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.

Dilation Theorem of Egerváry

On the real domain, Egerváry's dilation theorem can be extended to:

Theorem (Egerváry [5])

*Given a contraction $W \in \mathbb{M}_{n\times n}$ over $\mathbb{R}$ ($\|W\| \leq 1$), there exists an **orthogonal** operator on an $(N+1)\cdot n$ dimensional Hilbert space $\mathcal{H}$ and an isometric embedding $(N+1)\cdot n \times n$ matrix J over $\mathbb{C}$ such that $W^k = J^T U^k J$ for all $k \in \mathbb{Z}$, where:*

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

Egerváry's dilation theorem allows us to get a ϵ-close approximating reservoir system with an **orthogonal** dynamic coupling matrix given an arbitrary linear reservoir.

Theorem (F., Li, Tino [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.

It 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 the $\mathbb{C}$ case because it would inevitably involve a unitary matrix over $\mathbb{C}$ during the diagonalization process. **To keep things in** $\mathbb{R}$, we convert an orthogonal matrix to its **canonical form** via a (real) orthogonal matrix.

Canonical form of Orthogonal Matrix

For $\theta \in [0, 2\pi)$, consider 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 $e^{\pm i\theta}$ Note that in the $\mathbb{R}$ case, eigenvalues comes in pairs.. For any orthogonal matrix $C \in O(n)$, there exists an orthogonal matrix S such that the product $S^\top CS$ has the following form:

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

where $\theta_i \in (0, \pi)$, and $\Upsilon := \text{diag}\{a_1, a_2, ..., a_q\}$, $a_i \in \{-1, +1\}$, $i = 1, 2, ..., q$, is a diagonal matrix with q entries of ± 1's. For simplicity, **we will assume for the rest of the talk an even dimension** n.

Canonical form of Orthogonal Matrix

Observing:

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

Υ can therefore be written as a block diagonal matrix with blocks R_0 and R_π.

Hence $S^\top CS$ is a block diagonal matrix consisting of $\{R_{\theta_1}, \ldots, R_{\theta_m}\}$, $\theta_i \in [0, \pi]$, $i = 1, 2, ..., m$, 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.

Permutation matrices are orthogonal, therefore we can compute their corresponding canonical form. In particular, given an $\ell \times \ell$ full-cycle permutation P (ℓ even), we can find an orthogonal matrix Q such that $Q^\top PQ$ is a block diagonal matrix of $\{1, -1, R_{\frac{2\pi j}{\ell}} : 1 \leq j < \frac{\ell}{2}\}$.

Here, note that eigenvalues comes in pairs: 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.

We construct such an orthogonal matrix X and show:

Theorem (F., Li, Tino [6])

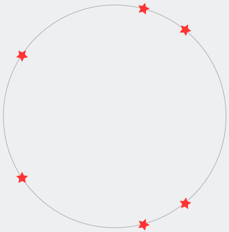
Let U be an $n \times n$ orthogonal matrix and $\delta > 0$ be an arbitrarily small positive number. There exists $n_1 \geq n$, an $n_1 \times n_1$ orthogonal matrix S, an $n_1 \times n_1$ full-cycle permutation P and an $(n_1 - n) \times (n_1 - n)$ orthogonal matrix D, such that

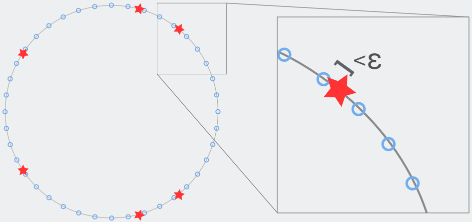
$$\left\| S^\top PS - \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix} \right\| < \delta.$$

From Orthogonal to Full-Cycle Permutation Coupling

Sketch of Proof.

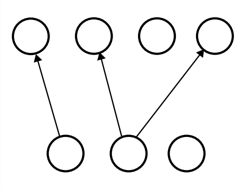
Given an orthogonal matrix U, we find its canonical form. The corresponding angles once again lies on the unit circle $\mathbb{T}$ in $\mathbb{C}$. We then perturb the pairs of angles to the closest subset of roots of unity. The remaining roots of unity not covered by the previous operator are filled in as a direct sum of blocks of R_θ. The resulting matrix constitutes the canonical form of the approximating system.

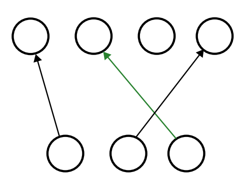




Given a reservoir system with dynamic coupling W and its equivalent with orthogonal coupling U, rotational angles in the canonical form of U are shown as red dots. Roots of unity corresponding to cyclic dilation approximate the rotational angles to a prescribed precision ϵ. □

Maximum Bipartite Matching





- **Problem Definition:** Given a bipartite graph $G = (A \cup B, E)$, where A and B are disjoint sets of vertices, find a matching $M \subseteq A \times B$ such that the number of edges in M is maximized, and no two edges in M share a vertex.
- Hopcroft-Karp Algorithm [10]:

$$\text{Time Complexity: } O(\sqrt{|U \cup V|} \cdot |E|)$$

$$\text{Space Complexity: } O(|U \cup V| + |E|)$$

Dilation Dimension Reduction via Maximum Bipartite Matching

In practice, the dimension $n_1 = \dim\left(S^\top PS\right)$ 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:

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

With Maximum Bipartite Matching, a lower dimension can be achieved:

- Let $\{\theta_i\}$ denote angles in the canonical form of $n \times n$ orthogonal matrix U.
- Construct bipartite graph $G = (A \cup B, E)$ with
  - Vertex set $A \cup B$ where: $A = \{\theta_i\}$ and $B = \{\frac{2a\pi}{n'} : 0 < a < \frac{n'}{2}\}$.
  - $e \in E$ joins $\theta_i \in A$ with $\frac{2a\pi}{n'} \in B$ if and only if $\left|e^{\theta_i i} - e^{\frac{2a\pi i}{n'}}\right| < \delta$.
- By construction we can find distinct k_i roots of unity $\frac{2k_i\pi}{n'}$ to approximate θ_i if and only if there exists a matching for this bipartite graph with exactly $|A|$ edges.

We will see that the dimension obtained is significantly lower than that of the theoretical upperbound.

With the previous results, we can now show

Theorem (F., Li, Tino [6])

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.*

The proof follows that of an analogous statement in $\mathbb{C}$. The arguments follow through by replacing unitary matrices by orthogonal matrices and conjugate transpose by regular transpose.

To show the universality of SCR in $\mathbb{R}$, we recall two useful lemmas from $\mathbb{C}$ which carries naturally onto $\mathbb{R}$.

Lemma (Li, F., Tino [12])

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:

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

Then P_1 is a full-cycle permutation.

Lemma (Li, F., Tino [12])

*For any $n \times m$ **real** matrix V and $\delta > 0$, there exists k matrices $\{F_1, \cdots, F_k\} \subset \mathbb{M}_{n\times m}(\{-1, 1\})$ 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 $\gcd(k,n) = 1$.

From Full-Cycle Permutation Coupling to SCR

Finally, we now obtain our main theorem showing the universality of SCR over $\mathbb{R}$.

Theorem (F., Li, Tino [6])

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.

The proof follows the same flow with one major difference: Since 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.

Once again, combining this with the follow result:

Theorem (Grigoryeva and Ortega [8](Corollary 11), paraphrased)

Linear reservoir systems with polynomial readouts are universal, in the sense that any time-invariant fading memory filter can be approximated by to arbitrary precision by a linear reservoir system.

We obtain our main theorem of Part II:

Theorem (F., Li, Tino [6])

Any time-invariant fading memory filter can be approximated to arbitrary precision by Simple Cycle Reservoirs.

Part III: Numerical Analysis

For each trail of the experiment we initialize a linear reservoir system with $\dim(W) = 5$, where:

- The entries 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

The numerical analysis goes as follows:

- Each initial system $R = (W, V, h)$ is dilated over a set of pre-defined dilation dimensions $\mathcal{D} := \{2, 6, 10, 15, 19, 24, 28, 33, 37, 42\}$.
- For each $N \in \mathcal{D}$, we construct a linear reservoir system R_U with an orthogonal dynamic coupling W_U of dimension $n_U = (N + 1)n$.
- Finally we dilate R_U into an ϵ-close linear reservoir system R_C with contractive cyclic-permutation dynamic coupling.

Numerical Results

Mean and 95% confidence intervals of the MSE of the states of the original reservoir and the approximating cyclic dilation systems over 15 randomized generations of the original system.

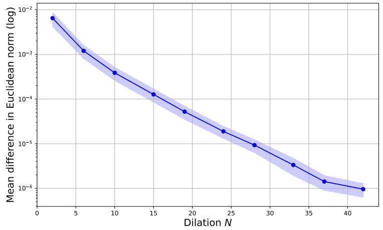


Figure: ECL

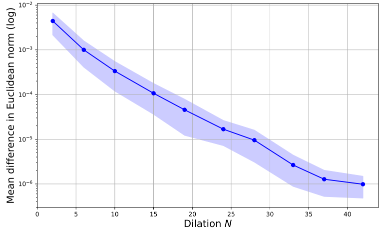


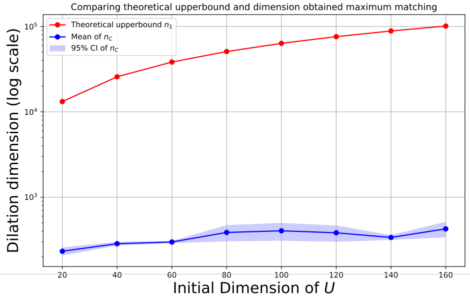
Figure: Ettm2

Reduction of Dilation Dimension with Maximum Bipartite Matching

Finally, we illustrate how the dimension n_C of the cyclic dilation obtained from the maximum matching program in bipartite graphs can yield reservoir sizes drastically lower than the **theoretical upper bound**:

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

Theoretical upper bound v.s. dimension obtained from maximum matching from cyclic dilations of 10 uniformly generated orthogonal matrices U for each initial dimension $n \in \{20, 40, \ldots, 140, 160\}$



Concluding Remarks

- In $\mathbb{C}$, we showed **constructively** the **universal approximation properties** of Complex - Simple Cycle Reservoirs (and friends) in the space of time-invariant fading-memory filters.
- In $\mathbb{R}$, we proved **constructively** that SCRs are **universal** approximations for any real-valued time-invariant fading memory filter over uniformly bounded input streams.
- 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.
- The fully constructive nature of our results is a crucial step towards understanding the **intrinsic properties of state-space models** and the **physical implementations** of reservoir computing in analogue computers such as photonic integrated circuits.

References I

[1] Y. Abe, K. Nakada, N. Hagiwara, E. Suzuki, K. Suda, S.-i. Mochizuki, Y. Terasaki, T. Sasaki, and T. Asai.
Highly-integrable analogue reservoir circuits based on a simple cycle architecture.
Sci. Rep., 14(10966):1–10, May 2024.

[2] L. Appeltant, M. C. Soriano, G. V. der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer.
Information processing using a single dynamical node as complex system.
Nature Communications, 2, 2011.

[3] I. Bauwens, G. Van der Sande, P. Bienstman, and G. Verschaffelt.
Using photonic reservoirs as preprocessors for deep neural networks.
Frontiers in Physics, 10:1051941, 2022.

[4] F. D.-L. Coarer, M. Sciamanna, A. Katumba, M. Freiberger, J. Dambre, P. Bienstman, and D. Rontani.
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, Nov. 2018.

[5] E. Egerváry.
On the contractive linear transformations of n-dimensional vector space.
Acta Sci. Math. (Szeged), 15:178–182, 1954.

References II

[6] R. S. Fong, B. Li, and P. Tiňo.
Universality of real minimal complexity reservoir.
arXiv preprint arXiv:2408.08071, 2024.

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

[8] L. Grigoryeva and J.-P. Ortega.
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, Jan. 2018.

[9] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. Van der Sande.
Demonstrating delay-based reservoir computing using a compact photonic integrated chip.
Optics express, 28(3):3086–3096, 2020.

[10] J. E. Hopcroft and R. M. Karp.
An n^5/2 algorithm for maximum matchings in bipartite graphs.
SIAM Journal on computing, 2(4):225–231, 1973.

[11] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot.
High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification.
Physical Review X, 7(1):011015, 2017.

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

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

[14] M. C. Ozturk, D. Xu, and J. Principe.
Analysis and design of echo state network.
Neural Computation, 19(1):111–138, 2007.

[15] A. Rodan and P. Tino.
Minimum complexity echo state network.
IEEE transactions on neural networks, 22(1):131–144, 2010.

References IV

[16] B. Sz.-Nagy.
Sur les contractions de l'espace de Hilbert.
Acta Sci. Math. (Szeged), 15:87–92, 1953.

[17] P. Tino.
Dynamical systems as temporal feature spaces.
Journal of Machine Learning Research, 21(44):1–42, 2020.

[18] P. Tino, R. S. Fong, and R. F. Leonarduzzi.
Predictive modeling in the reservoir kernel motif space.
arXiv preprint arXiv:2405.07045, 2024.

[19] A. Zeng, M. Chen, L. Zhang, and Q. Xu.
Are transformers effective for time series forecasting?
In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.