Embedding theory of reservoir computing and reducing reservoir network using time delays

Xing-Yue Duan,† Xiong Ying,‡ Si-Yang Leng,§ Jürgen Kurths,∥ Wei Lin,∥ and Huan-Fei Ma

1School of Mathematical Sciences, Soochow University, Suzhou 215006, China
2School of Mathematical Sciences, SCMS, and SCAM, Fudan University, Shanghai 200433, China
3Research Institute of Intelligent Complex Systems and Centre for Computational Systems Biology, Fudan University, Shanghai 200433, China
4Academy for Engineering and Technology, Fudan University, Shanghai 200433, China
5State Key Laboratory of Medical Neurobiology, LCNBI, and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai 200032, China
6Potsdam Institute for Climate Impact Research (PIK), Potsdam 14473, Germany
7Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

(Received 5 November 2021; revised 16 March 2023; accepted 8 May 2023; published 25 May 2023)

Reservoir computing (RC), a particular form of recurrent neural network, is under explosive development due to its exceptional efficacy and high performance in reconstruction and/or prediction of complex physical systems. However, the mechanism triggering such effective applications of RC is still unclear, awaiting deep and systematic exploration. Here, combining the delayed embedding theory with the generalized embedding theory, we rigorously prove that RC is essentially a high-dimensional embedding of the original input nonlinear dynamical system. Thus, using this embedding property, we unify into a universal framework the standard RC and the time-delayed RC where we introduce time delays only into the network’s output layer, and we further find a trade-off relation between the time delays and the number of neurons in RC. Based on these findings, we significantly reduce the RC’s network size and promote its memory capacity in completing systems reconstruction and prediction. More surprisingly, only using a single-neuron reservoir with time delays is sometimes sufficient for achieving reconstruction and prediction tasks, while the standard RC of any large size but without time delay cannot complete them yet.

DOI: 10.1103/PhysRevResearch.5.L022041

The last decades have witnessed the extensive application and development of machine learning technology in data-driven research and in high-technology-oriented industry as well. As a representative leader among many machine learning techniques, the artificial neural network (ANN) has emerged as a powerful approach that is well suited for coping with the supervised learning problems. Among various architectures of ANN, reservoir computing (RC), which is a recently developed framework [1], a special variant of a recurrent neural network, and also known as a generalization of the echo-state network (ESN) [2] or liquid-state machine (LSM) [3], has been reported to have great efficacy in reconstruction and/or prediction of many complex physical systems only based on the observational data of time series [4–7]. The architecture of RC is quite contracted. As shown in Fig. 1(a), only three weight matrices are involved: the input matrix and the reservoir recurrent matrix are randomly generated but fixed, while the output matrix is determined via training. As such, efficient least squares optimization methods rather than the resource-consuming backpropagation algorithm are adopted in the training process [8]. Behind such a contracted architecture, two questions arise naturally: “What is the fundamental mechanism resulting the efficacy of RC?” and “How can the structure be improved using the uncovered mechanism?” These questions have attracted great attention and motivated abundant discussions, including those from the topology and the complexity of random connections [9,10] to the spectral radius of random networks and the edge of chaos [11–13], from the fading memory property [14] to the echo state property [15,16], from the choice of activation functions [17] to the training algorithm of the output layer [18]. Yet, recent understanding of RC is often via heuristic interpretation and it is widely believed that a successful RC should possess high dimensionality, nonlinearity, fading memory, and separation property [6], but barely with rigorous and mathematical demonstrations.

In order to decipher the RC’s capacity of reconstructing and forecasting nonlinear dynamics, several efforts from a viewpoint of dynamical systems have been recently made. For example, the regression model and the dynamical model decomposition method were used to illustrate the usefulness of RC for forecasting chaotic dynamics [19,20], and, to demonstrate the approximation capability of RC, an embedding conjecture was studied and could be partially validated for a specific form of RC under the right technical conditions.
single neuron reservoir, in a standard form without introducing any DDE or time-division multiplexing technique, can sometimes work well for reconstructing and forecasting some representative physical systems. Moreover, we find flexible memory capacity in the time-delayed RC, which makes it possible to accomplish more challenging tasks of dynamics reconstruction that cannot be easily achieved using a standard RC of the same scale.

We start with a standard RC as sketched in Fig. 1(a). Here, the input data $x_k \in \mathbb{R}^n$ represents the state vector of a dynamical system that is evolving on a compact manifold $\mathcal{M}$ with the evolution operator $\varphi \in \text{Diff}^2(\mathcal{M}) : x_{k+1} = \varphi(x_k)$. The vector $r_k \in \mathbb{R}^m$ represents the state of $m$ reservoir neurons at time step $k$, the input layer weight matrix $W_\text{in}$ and the reservoir network matrix $W_\text{res}$ are, respectively, $m \times n$ and $m \times m$ matrices generated according to certain distribution laws. The dynamical evolution of the reservoir neurons is governed by (RN): $r_k = (1 - \alpha) r_{k-1} + \alpha \varphi(W_\text{in} x_{k-1} + W_\text{res} r_{k-1})$, where $\alpha$ is the leakage factor, and $\varphi \in C^2[\mathbb{R}, (-1, 1)]$ is set a sigmoid function (e.g., tanh) in this Letter. The output vector $y_k \in \mathbb{R}^l$ is determined by the output weight matrix $W_\text{out}$ such that $y_k = W_\text{out} r_k$. In the task of nonlinear system reconstruction, given the time series, denoted by $x_k$, $k = 1, \ldots, N+1$, as training data, the target is to train the output weight matrix $W_\text{out}$ so as to approximate the one-step dynamics prediction, i.e., $y_k \approx x_{k+1}$. To achieve this, the output weight matrix $W_\text{out}$ is generally calculated by minimizing the loss function $L = \sum_{k=1}^{N} \| x_{k+1} - W_\text{out} r_k \|^2 + \beta \| W_\text{out} \|^2$ over the training data set, where $\beta > 0$, the $L_2$-regularization coefficient, is introduced to make optimization robust. After training, one can fix the output weight matrix $W_\text{out}$ and redirect the output $y_k = W_\text{out} r_k$ as an approximation of $x_{k+1}$ into the input layer of the network and thus generate the autonomous dynamics for $x_k$ with $k > N$.

To rigorously establish an embedding theory for RC, we consider directly the evolution (RN) of the reservoir neurons with the leakage factor $\alpha = 1$ as

$$\rho_{k+1,x_0}^0 = \varphi(W_\text{res} \rho_{k,x_0}^0 + W_\text{in} x_0), \quad k = 0, 1, \ldots,$$

and define a map as $\Theta^n[r_0, W_\text{res}, W_\text{in}] : x_0 \mapsto \rho_{k,x_0}^0$. Here, $\rho_{0,x_0} = b_0$, $b_0 \in \mathbb{R}^m$, and $\mathbb{R} = (-1, 1)$. Thus, we rigorously have the following result.

**Theorem 1.** Let $m \geq 2 \dim(\mathcal{M}) + 1$ and $[r_0, W_\text{res}, W_\text{in}] \in \mathbb{R}^m \times \mathbb{R}^{m \times m} \times \mathbb{R}^{m \times n}$ with $\dim(\mathcal{M})$ as the box-counting dimension of the manifold $\mathcal{M}$. Then, there exists a number $k^* > 0$, such that $\Theta^n[r_0, W_\text{res}, W_\text{in}] \in C^1(\mathbb{M}, \mathbb{R}^m)$ is generically an embedding for all $k > k^*$.

Here, the generic conclusion in Theorem 1 means that, for all $[r_0, W_\text{res}, W_\text{in}] \in \mathcal{S}$ where $\mathcal{S} \subset \mathbb{R}^m \times \mathbb{R}^{m \times m} \times \mathbb{R}^{m \times n}$ is an open and dense set, $\Theta^n[r_0, W_\text{res}, W_\text{in}]$ is an embedding for any sufficiently large $k$. The detailed and rigorous proof with respect to the $C^1$ topology is provided in the Supplemental Material (SM) [26]. Moreover, the echo state property, a necessary condition for constructing an RC, requires that, with the general configuration $[W_\text{in}, W_\text{res}, \varphi]$, the evolutions (RN) of the reservoir neurons, starting from any different initial values $r_k^{(1)}$ and $r_k^{(2)}$, converge to the same dynamics, i.e., $\lim_{k \to \infty} \| r_k^{(1)} - r_k^{(2)} \| = 0$ [15]. Hence, by virtue of Theorem 1, regardless of the choice of the initial value $r_0$,
the dynamics of reservoir neurons is determined by the input dynamics, i.e., there exists a unique embedding \( \Psi \) such that \( r_s = \Psi(x_s) \) after a transient phase, while each component \( r_s = \Psi_i(x_s) \) implies that the dynamics of each neuron is an observable of the original dynamics.

In the standard RC investigated above, \( m \), the number of reservoir neurons and also known as the reservoir dimension, is often required to be huge \([6,8]\). To design a different RC framework, significantly reducing \( m \), we introduce time delays into the output layer, as sketched in Fig. 1(c). While all the configuration \( \{W_{in}, W_{res}, \phi\} \) and the input data \( x \) are set in the same manner, the reservoir network is assumed to include \( q \ (< m) \) neurons only. Thus, a new reservoir vector before the output layer is designated as \( \tilde{r}_k = [r_1, k, r_1-k, \ldots, r_1, k-d_1, k+\tau, \ldots, r_q, k, r_q, k-d_1, k+\tau] \), and, correspondingly, the output matrix \( W_{out} \) is calculated by minimizing the \( L_2 \) loss function

\[
\tilde{L} = \sum_{k=1}^{N} \| x_{k+1} - W_{out} \tilde{r}_k \|^2 + \beta \| W_{out} \|^2,
\]

with \( W_{out} \in \mathbb{R}^{L \times d} \) and \( d = \sum_{i=1}^{d_1} d_i \). Here, the new reservoir vector \( \tilde{r}_k \) is formed by the lagged dynamics of each neuron, i.e., \( q \) neurons with each neuron contributing \( d_i \) lagged dynamics \([r_1, k, r_1-k, \ldots, r_1, k-d_1, k+\tau, \ldots, r_q, k, r_q, k-d_1, k+\tau]\), where \( \tau \) is a time delay, and \( d \) is assigned as the output dimension of this delayed RC.

Now, we are in a position to demonstrate that the time-delayed RC with the above-assigned \( d \) has the same representation and computation ability as the standard RC involving \( m \) neurons without time delay under the same parameter settings, as long as \( d \approx m \). Actually, based on the delayed embedding theory and its applications \([27–30]\), an approximate combination of the lagged observable can also generically form an embedding, i.e., for smooth observational functions \( \Psi_1, \ldots, \Psi_q \), \( F(x) = [\Psi_1(x), \Psi_1(\phi^{-1}(x)), \ldots, \Psi_q(\phi^{-1}(x))] \), \( \Psi_q(\phi^{-1}(x)) \) is generically an embedding as long as \( \sum_{i=1}^{q} d_i > 2d \text{dim}(M) \). Using the above-obtained conclusion that each neuron is generally an observable, we further conclude that the proposed new reservoir vector \( \tilde{r}_k \) is also an embedding. Thus, the dynamics of the state vector \( r_s \) in the \( m \)-neuron reservoir network without time delay is topologically conjugated with the dynamics of the reservoir vector \( \tilde{r}_k \) of a \( q \)-neuron reservoir network in the sense of embedding as long as \( m = d \) with \( d = \sum_{i=1}^{d_1} d_i \), as sketched in Fig. 1(b). Consequently, we come to the conclusion that the delayed observables of the RC state, seen as additional nonlinear observables, have the same computational power in the system reconstruction.

To demonstrate the capability of our time-delayed RC, we first consider the benchmark Lorenz system. After a training phase including \( N = 6000 \) samples, the autonomously generated dynamics by the RC are shown in Fig. 2(a). Particularly used are a standard RC, a time-delayed RC containing fewer neurons with uniformly lagged dynamics for each neuron, and a time-delayed RC containing the same number of neurons but with random lags for each neuron. Clearly, the time-delayed RC has almost the same performance of system reconstruction as the nondelayed one, no matter whether the lags are uniformly or randomly generated. Actually, this coincides with the above-performed arguments from a viewpoint of embedding that the dynamics of this nondelayed RC is a generalized embedding to the input dynamics with generically \( 200 \) observables, while the dynamics of the time-delayed RC forms an embedding of dimension \( 200 \) when the sum of lags equals \( 200 \) for either uniform or random lags. Such a trade-off relation is further clearly illustrated in Fig. 2(b), where a different neuron number with a different lag number for each neuron is combined, and, for each combination, a training error is calculated as the mean squared error (MSE) on the training data set based on over 20 independent runs. As depicted in Fig. 2(b), for a fixed moderate number of neurons, the training error decreases monotonically with the lag number for each neuron, and, for a fixed moderate lag number, the training error also decreases monotonically with the neuron number. Analogous...
the memory capacity (MC) for different RC frameworks, us-

job quite well. To understand this phenomenon, we calculate

one reservoir network usually requires more than one single

the task of system reconstruction for multiple variables using

for any system reconstruction. Due to the multiscale property,

multiple neurons. Frankly, the single-neuron RC in this numerical

same reconstruction ability of the time-delayed RC with mul-

However, with all the same reservoir environment, the time-

delayed RC [both RC#1 and RC#2 in Fig.3(a)] can fulfill the

is an essential measure for RC’s reconstruction ability for

a single-neuron reservoir is not adequate even with multiple

Frankly, the single-neuron RC in this numerical illustration is only a special case that is not universally suitable

for system reconstruction. Due to the multiscale property, the

task of system reconstruction for multiple variables using

one reservoir network usually requires more than one single

As for the task in Fig.2(a), in order to get a successful

one reservoir in the proposed framework. To see this, we con-

side a gene regulation model with multiple delays: \( \dot{x}(t) =

\) describes self-

and with specific parameters the one-dimensional model has

chaotic dynamics \([26,31,32]\). Specifically, a time-delayed RC

including only one neuron with 600 lags is used to reconstruct

the dynamics, and the autonomously generated dynamics after

training are shown in Fig.3(a). The results confirm that the

single-neuron, time-delayed RC performs well, achieving the

same reconstruction ability of the time-delayed RC with mul-

tiple neurons. Frankly, the single-neuron RC in this numerical

ability in the sense of embedding, we further discover that the

time-delayed RC has a more flexible memory capacity, which

is an essential measure for RC’s reconstruction ability for delayed systems. In the dynamics reconstruction job for the

above gene regulation model in Fig.3(a), the chaotic dynamics

cannot be reconstructed by a standard RC, no matter how large the reservoir is, according to the dimension test \([26]\).

However, with all the same reservoir environment, the time-

delayed RC [both RC#1 and RC#2 in Fig.3(a)] can fulfill the

job quite well. To understand this phenomenon, we calculate

the memory capacity (MC) for different RC frameworks, us-

ing the definition in Ref. \([8]\) and with different combinations of

\( N_{\text{neuron}} \) and \( N_{\text{lag}} \) but satisfying the same output dimension, i.e., \( N_{\text{neuron}} \cdot N_{\text{lag}} = 600 \). Specifically, MC of a reservoir refers to its ability to retain information from previous time steps and it is defined in Ref. \([8]\) as

\[
MC_k = \frac{\text{cov}[x(t-k), \hat{y}_k(t)]^2}{\text{var}[x(t)] \cdot \text{var} [\hat{y}(t)]},
\]

where a random sequence of input values \( x(t) \) is presented to the reservoir, and the reservoir output \( \hat{y}_k(t) \) is trained to predict a previous input value \( x(t-k) \), and here cov(·) and var(·), respectively, represent covariance and variance.

Figure 3(b) clearly shows that, as \( N_{\text{lag}} \) increases, the reservoir computer with different delay settings has stronger

memory capacity though still keeping a fading memory fash-

ion. This is essential for the dynamics reconstruction job, particularly for time-delayed physical or biological systems such as the gene regulation model above. Thus, the proposed time-delayed RC framework has a more flexible capability to deal with dynamics reconstruction jobs requiring tunable MC.

Finally, to further validate the efficacy of the time-delayed RC in reconstructing a high-dimensional spatial-temporal sys-

tem, we consider the ideal storage cellular automation model (ISCAM) simulating heterocatalytic reaction-diffusion pro-

cesses at metal surfaces \([33,34]\). Considering the extremely high dimension (the \( 100 \times 100 \) grids yields \( 10000 \) input dimension), it is a challenging job to reconstruct the chaotic spatial-temporal patterns. As shown in Fig.4, with the same reservoir output dimension, the time-delayed RC has almost the same reconstruction ability as the nondelay one.

Our framework uses a few hyperparameters, such as \( d \), the
effective reservoir dimension, and \( \tau \), the time delay, which
definitely affect RC’s efficacy in system reconstruction. In

fact, the existing literature included some criteria for select-
such parameters in system reconstruction using delayed
embossing theory. We thus implement these criteria, the di-
mension test and the delayed mutual information (DMI), to
determine $d$ and $\tau$. From a perspective of embedding, $d$ is only required to be larger than $2 \cdot \dim(M)$ while practically the box-counting dimension of the manifold $M$ is usually very small, i.e., $\dim(M)$ is between 2 and 3 [35,36] for the chaotic Lorenz attractor. However, to design an effective RC, $d$ is required to be moderately large (see all the examples above). This is probably because, although the generic property in the embedding theory means open and dense in a topological sense, there are still degenerated situations in practice, particularly for randomly generated networks (see Fig. S1 in Ref. [26]). Moreover, to reveal the mechanism from representation to computation, the recent efforts used the universal approximation theory [21] and the dynamic mode decomposition (DMD) [19] framework, which further demonstrate the necessity of a large network size of RC in achieving good approximations. Thus, the dimension tests are used to seek a suitable $d$ for each computation. As for the delay $\tau$, either a too-small or too-large value renders computation problematic in system reconstruction, which naturally prompts us to introduce a modified DMI test taking into account the intrinsic time scales of the neuronal dynamics in RC. Finally, it is noted that, for chaotic systems, the lagged observables earlier than the Lyapunov time have diminishing predictive power for the current time step, so we suggest the constraint $\tau \cdot \Delta t \cdot N_{\text{lag}} < \Lambda_{\text{max}}$ for the choice of $\tau$ and $N_{\text{lag}}$ in practice, where $\Delta t$ is the sampling step size. The details for the choice of these hyperparameters are given in Ref. [26].

In conclusion, we have provided deep and rigorous insight into the mechanism of RC from a viewpoint of embedding theory and nonlinear dynamical systems. Based on our analytical findings, we have studied the role of time delay in the reservoir network and proposed a framework of time-delayed RC. This framework can significantly reduce the network size and promote the memory capacity, making its ability attain or even transcend the ability owned by the standard RC. Considering the computational costs, which are crucially dependent on the network size in the dynamical evolution of RC, and the hardware costs related to the circuit size in those overwhelmingly developed physical RCs [6], a smaller-size reservoir is always expected to promote its real and extensive applications. Moreover, we notice a recently published and independent work [37] where a method, different from the perspective of embedding theory and memory capacity presented here, was proposed to concatenating internal states through time in RC and realize model-size reduction. Lastly, any contributions to designing RC frameworks of low resource consumption are believed to advance the direction of machine learning and thus be of broad applicability in solving data-driven science and engineering problems.

This work is supported by the NSF of China (Grants No. 11925103 and No. 12171350), by the STCSM (Grants No. 18DZ1201000 and No. 2021SHZDZX0103), and by the SMEC (Grant No. 2023ZKZD04).


FIG. 4. Reconstructed chaotic patterns for the ISCAM model by a standard nondelayed RC including 5000 neurons and a time-delayed RC including 1000 neurons and five lags for each neuron. (a) Selected dynamical pattern using different evolution rules. (b) Reconstruction errors deviating from true dynamics from different RC frameworks.
[10] A. Griffith, A. Pomerance, and D. J. Gauthier, 


