Letter

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

Xing-Yue Duan,[1] Xiong Ying,[2] Si-Yang Leng,[3,4,5] Jürgen Kurths,[3,6] Wei Lin,[2,3,5,7,*] and Huan-Fei Ma[1,†]

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

*wlin@fudan.edu.cn
†hfma@suda.edu.cn

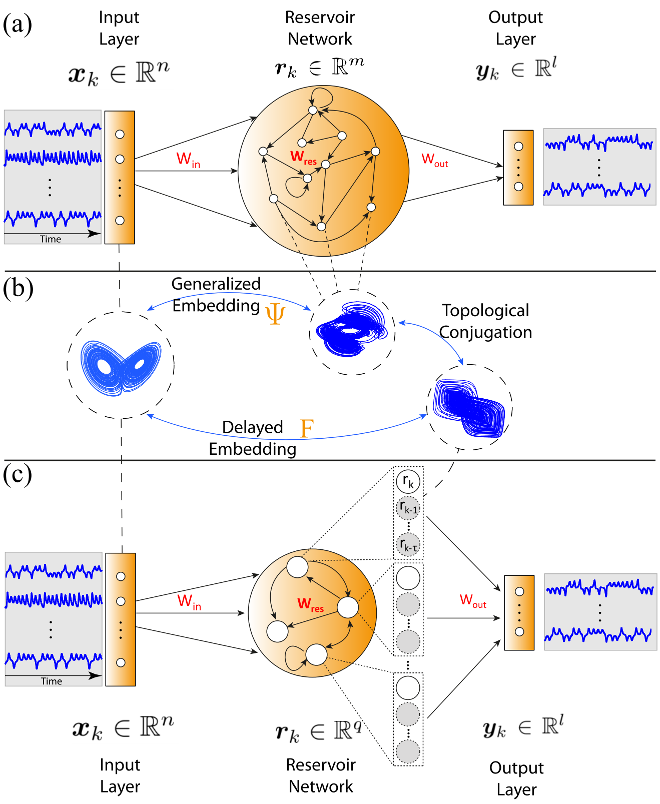

FIG. 1. RCs as different nonlinear dynamical systems and embeddings. (a) A standard RC without time delay. (b) The generalized embedding Ψ from the input dynamics to the standard nondelayed reservoir network and the delayed embedding F from the input dynamics to the delayed reservoir network, which constitute a topological conjugation between the dynamics of the nondelayed reservoir network and the delayed reservoir network. (c) A time-delayed RC with a smaller network size in the reservoir layer.

[21,22]. In the area of photonic neural networks, an architecture of photonic reservoir computing has been developed using a spatiotemporal analogy to translate a delayed differential equation (DDE) into a virtual single-neuron reservoir network [23–25]. Still, despite these significant efforts and achievements, some key questions remain unsolved: how to understand the network dimension of general RC using the theories of nonlinear dynamics and functional analytics, and how to design a small-size network in RC for sustaining its efficacy.

In this Letter, we rigorously study the mechanism of RC from a viewpoint of nonlinear dynamical systems and propose a framework of time-delayed RC. Particularly, combining the delayed embedding theory with the generalized embedding theory, we first prove a general reservoir network rigorously as a high-dimensional embedding of the original input nonlinear dynamical system. Then, we further reveal a trade-off relation between the time delays and the number of neurons by unifying into a universal framework the standard RC without delays and the time-delayed RC where the time delays are introduced into the network's output layer. It therefore allows us to construct a random reservoir network with a significantly reduced physical dimension to achieve the efficacy that the original larger-size RC owns. Surprisingly, we show that a 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 $\boldsymbol{x}_k \in \mathbb{R}^n$ represents the state vector of a dynamical system that is evolving on a compact manifold $\mathcal{M}$ with the evolution operator $\varphi \in \mathrm{Diff}^2(\mathcal{M}) : \boldsymbol{x}_{k+1} = \varphi(\boldsymbol{x}_k)$. The vector $\boldsymbol{r}_k \in \mathbb{R}^m$ represents the state of m reservoir neurons at time step k, the input layer weight matrix W_{in} and the reservoir network matrix W_{res} are, respectively, $m \times n$ and $m \times m$ random matrices generated according to certain distribution laws. The dynamical evolution of the reservoir neurons is governed by (RN): $\boldsymbol{r}_k = (1-\alpha)\boldsymbol{r}_{k-1} + \alpha\phi(W_{\mathrm{res}}\boldsymbol{r}_{k-1} + W_{\mathrm{in}}\boldsymbol{x}_k)$, where α is the leakage factor, and $\phi \in C^2[\mathbb{R}, (-1, 1)]$ is set a sigmoid function (e.g., tanh) in this Letter. The output vector $\boldsymbol{y}_k \in \mathbb{R}^l$ is determined by the output weight matrix $W_{\mathrm{out}} \in \mathbb{R}^{l\times m}$ such that $\boldsymbol{y}_k = W_{\mathrm{out}}\boldsymbol{r}_k$. In the task of nonlinear system reconstruction, given the time series, denoted by $\boldsymbol{x}_k, k = 1, \ldots, N+1$, as training data, the target is to train the output weight matrix W_{out} so as to approximate the one-step dynamics prediction, i.e., $\boldsymbol{y}_k \approx \boldsymbol{x}_{k+1}$. To achieve this, the output weight matrix W_{out} is generally calculated by minimizing the loss function $\mathcal{L} = \sum_{k=1}^{N} \|\boldsymbol{x}_{k+1} - W_{\mathrm{out}}\boldsymbol{r}_k\|^2 + \beta\|W_{\mathrm{out}}\|^2$ 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_{out} and redirect the output $\boldsymbol{y}_k = W_{\mathrm{out}}\boldsymbol{r}_k$ as an approximation of $\boldsymbol{x}_{k+1}$ into the input layer of the network and thus generate the autonomous dynamics for $\boldsymbol{x}_k$ 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

$$\boldsymbol{r}^{\boldsymbol{b}_0}_{k+1,\boldsymbol{x}_0} = \phi\big(W_{\mathrm{res}}\boldsymbol{r}^{\boldsymbol{b}_0}_{k,\boldsymbol{x}_0} + W_{\mathrm{in}}\varphi^{k+1}\boldsymbol{x}_0\big), \quad k = 0, 1, \ldots,$$

and define a map as $\mathfrak{G}^k[\boldsymbol{r}_0, W_{\mathrm{res}}, W_{\mathrm{in}}](\boldsymbol{x}_0) = \boldsymbol{r}^{\boldsymbol{b}_0}_{k,\boldsymbol{x}_0}$. Here, $\boldsymbol{r}^{\boldsymbol{b}_0}_{0,\boldsymbol{x}_0} = \boldsymbol{b}_0$, $\boldsymbol{b}_0 \in \mathbb{I}^m$, and $\mathbb{I} = (-1, 1)$. Thus, we rigorously have the following result.

Theorem 1. Let $m \geqslant 2\dim(\mathcal{M}) + 1$ and $[\boldsymbol{r}_0, W_{\mathrm{res}}, W_{\mathrm{in}}] \in \mathbb{I}^m \times \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n}$ with $\dim(\mathcal{M})$ as the box-counting dimension of the manifold $\mathcal{M}$. Then, there exists a number $k^* > 0$, such that $\mathfrak{G}^k[\boldsymbol{r}_0, W_{\mathrm{res}}, W_{\mathrm{in}}] \in C^1(\mathcal{M}, \mathbb{R}^m)$ is generically an embedding for all $k > k^*$.

Here, the generic conclusion in Theorem 1 means that, for all $[\boldsymbol{r}_0, W_{\mathrm{res}}, W_{\mathrm{in}}] \in \mathcal{S}$ where $\mathcal{S} \subset \mathbb{I}^m \times \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n}$ is an open and dense set, $\mathfrak{G}^k[\boldsymbol{r}_0, W_{\mathrm{res}}, W_{\mathrm{in}}]$ 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 $\{W_{\mathrm{in}}, W_{\mathrm{res}}, \phi\}$, the evolutions (RN) of the reservoir neurons, starting from any different initial values $\boldsymbol{r}_0^{(1)}$ and $\boldsymbol{r}_0^{(2)}$, converge to the same dynamics, i.e., $\lim_{k\to\infty} \|\boldsymbol{r}_k^{(1)} - \boldsymbol{r}_k^{(2)}\| = 0$ [15]. Hence, by virtue of Theorem 1, regardless of the choice of the initial value $\boldsymbol{r}_0$, the dynamics of reservoir neurons is determined by the input dynamics, i.e., *there exists a unique embedding* Ψ *such that* $\boldsymbol{r}_k = \Psi(\boldsymbol{x}_k)$ *after a transient phase, while each component* $r_{ik} = \Psi_i(\boldsymbol{x}_k)$ *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 $\{W_{\text{in}}, W_{\text{res}}, \phi\}$ and the input data $\boldsymbol{x}$ 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 $\tilde{\boldsymbol{r}}_k = [r_{1,k}, r_{1,k-\tau}, \ldots, r_{1,k-d_1\tau+\tau}, \ldots, r_{q,k}, \ldots, r_{q,k-d_q\tau+\tau}]^\top$, and, correspondingly, the output matrix W_{out} is calculated by minimizing the L_2 loss function

$$\tilde{\mathcal{L}} = \sum_{k=1}^{N} \|\boldsymbol{x}_{k+1} - W_{\text{out}}\tilde{\boldsymbol{r}}_k\|^2 + \beta\|W_{\text{out}}\|^2,$$

with $W_{\text{out}} \in \mathbb{R}^{l\times d}$ and $d = \sum_{i=1}^{q} d_i$. Here, the new reservoir vector $\tilde{\boldsymbol{r}}_k$ is formed by the lagged dynamics of each neuron, i.e., q neurons with each neuron contributing d_i lagged dynamics $[r_{i,k}, r_{i,k-\tau}, \ldots, r_{i,k-d_i\tau+\tau}]$, where τ 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 $\Psi_1, \ldots, \Psi_q$, $F(\boldsymbol{x}) = \{\Psi_1(\boldsymbol{x}), \Psi_1[\varphi^{-1}(\boldsymbol{x})], \ldots, \Psi_1[\varphi^{-(d_1-1)}(\boldsymbol{x})], \ldots, \Psi_q(\boldsymbol{x}), \Psi_q[\varphi^{-1}(\boldsymbol{x})], \ldots, \Psi_q[\varphi^{d_q-1}(\boldsymbol{x})]\}$ is generically an embedding as long as $\sum_{i=1}^{q} d_i > 2\dim(\mathcal{M})$. Using the above-obtained conclusion that each neuron is generally an observable, we further conclude that the proposed new reservoir vector $\tilde{\boldsymbol{r}}_k$ is also an embedding. Thus, the dynamics of the state vector $\boldsymbol{r}_k$ in the m-neuron reservoir network without time delay is topologically conjugated with the dynamics of the reservoir vector $\tilde{\boldsymbol{r}}_k$ of a q-neuron reservoir network in the sense of embedding as long as $m = d$ with $d = \sum_{i=1}^{q} d_i$, 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

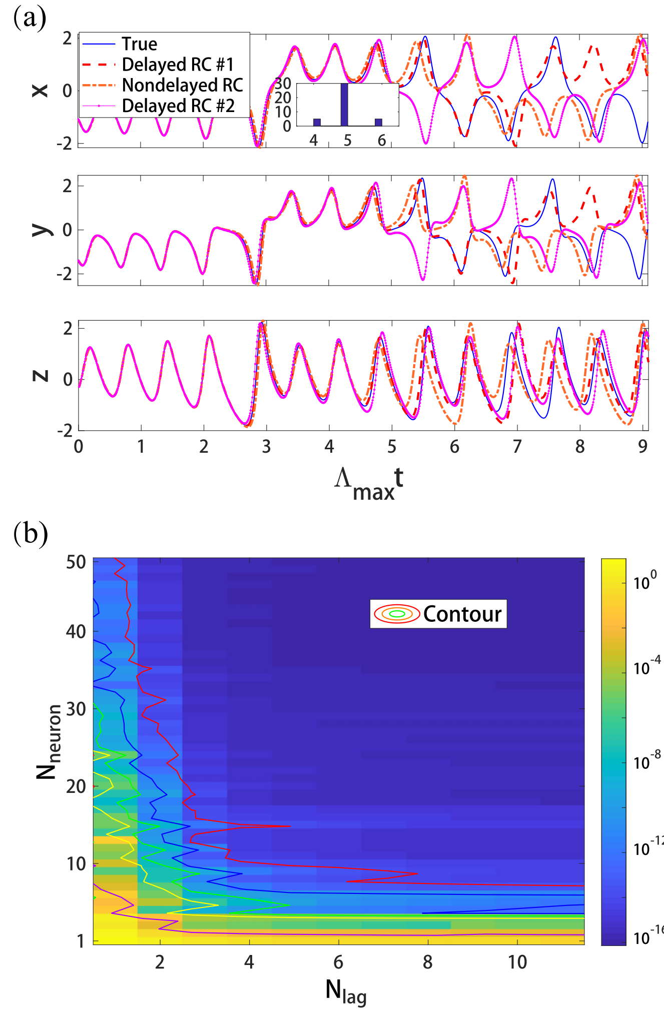

FIG. 2. (a) Reconstructed dynamics of the Lorenz system by a nondelayed RC including 200 neurons, a delayed RC #1 including 40 neurons with uniformly five lags for each neuron, and a delayed RC #2 including 40 neurons with random lags for each neuron. Here, the time unit is expressed in the Lyapunov time, and the random lags are generated by a distribution centered at 5 as shown in the inset. (b) System reconstruction test for the Lorenz system with different combinations of N_{neuron} and N_{lag}, where the training MSE in a log-scale and the contour curves are, respectively, highlighted. Here, $\tau = 5$ and the sampling step size is $\Delta t = 0.01$. All the other parameter settings are introduced in Ref. [26].

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 results are also obtained for the other benchmark systems, as presented in Ref. [26] (see Fig. S5). All these further reinforce the above conclusion that, whenever $d \approx m$ with moderate N_{lag} and N_{neuron}, the time-delayed and the nondelayed RCs generally share the same ability in system representation, and the number of neurons and of time lags can be traded off mutually in these frameworks.

Such a trade-off relationship further puts the nondelayed and time-delayed RCs into a unified framework where the output dimension d becomes the effective reservoir dimension that finally decides the ability of the system representation. The standard nondelayed RC is actually a degenerated form in this unified framework where all the neurons have zero lag. More surprisingly, we find that it is even possible to reduce the number of neurons into one and realize a single-neuron reservoir in the proposed framework. To see this, we consider a gene regulation model with multiple delays: $\dot{x}(t) = -kx(t) + gf_1[x(t-\tau_1)]f_2[x(t-\tau_2)]$, which describes self-inhibition and self-activation with distinct delays τ_1 and τ_2, 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 multiple neurons. Frankly, the single-neuron RC in this numerical illustration is only a special case that is not universally suitable for any 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 neuron. As for the task in Fig. 2(a), in order to get a successful prediction for the three components of the Lorenz system, a single-neuron reservoir is not adequate even with multiple time delays. In addition to the equivalent representation 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, using the definition in Ref. [8] and with different combinations of N_{neuron} and N_{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

$$\mathrm{MC}_k = \frac{\operatorname{cov}\,[x(t-k), \hat{y}_k(t)]^2}{\operatorname{var}[x(t)] \cdot \operatorname{var}\,[\hat{y}_k(t)]},$$

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 $\operatorname{cov}(\cdot)$ and $\operatorname{var}(\cdot)$, respectively, represent covariance and variance.

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

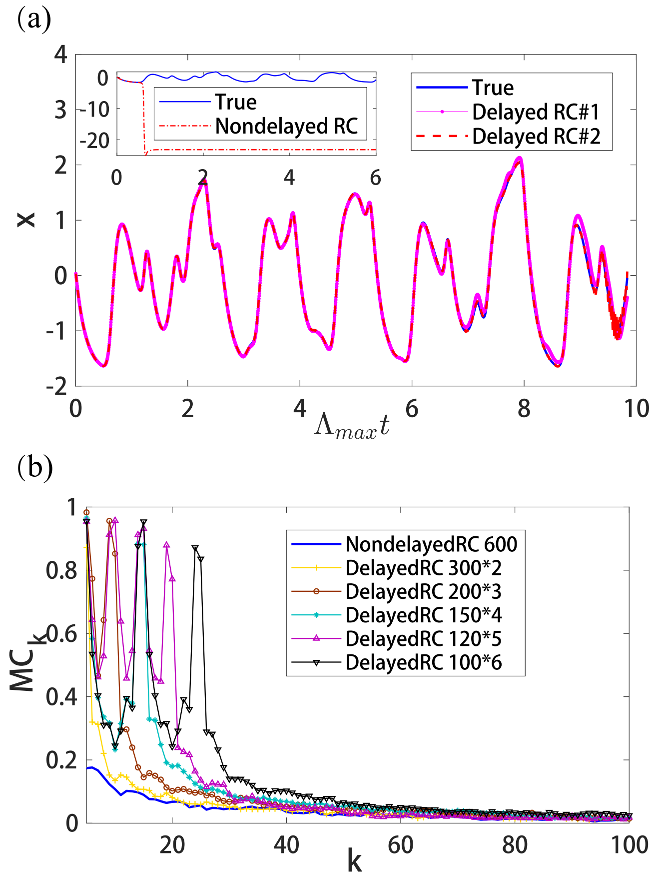

FIG. 3. (a) Reconstructed dynamics of the chaotic gene regulation model by the standard RC including 600 neurons (see the inset), the single-neuron, time-delayed RC#1 with 600 lags, and the time-delayed RC#2 including six neurons and 100 lags for each neuron. (b) MC test with different combinations of $N_{\text{neuron}} \cdot N_{\text{lag}}$. Here, $\tau = 5$ and the sampling step size is $\Delta t = 0.1$. All the other parameter settings are introduced in Ref. [26].

memory capacity though still keeping a fading memory fashion. 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 system, we consider the ideal storage cellular automation model (ISCAM) simulating heterocatalytic reaction-diffusion processes at metal surfaces [33,34]. Considering the extremely high dimension (the 100×100 grids yields 10 000 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 τ, the time delay, which definitely affect RC's efficacy in system reconstruction. In fact, the existing literature included some criteria for selecting such parameters in system reconstruction using delayed embedding theory. We thus implement these criteria, the dimension test and the delayed mutual information (DMI), to

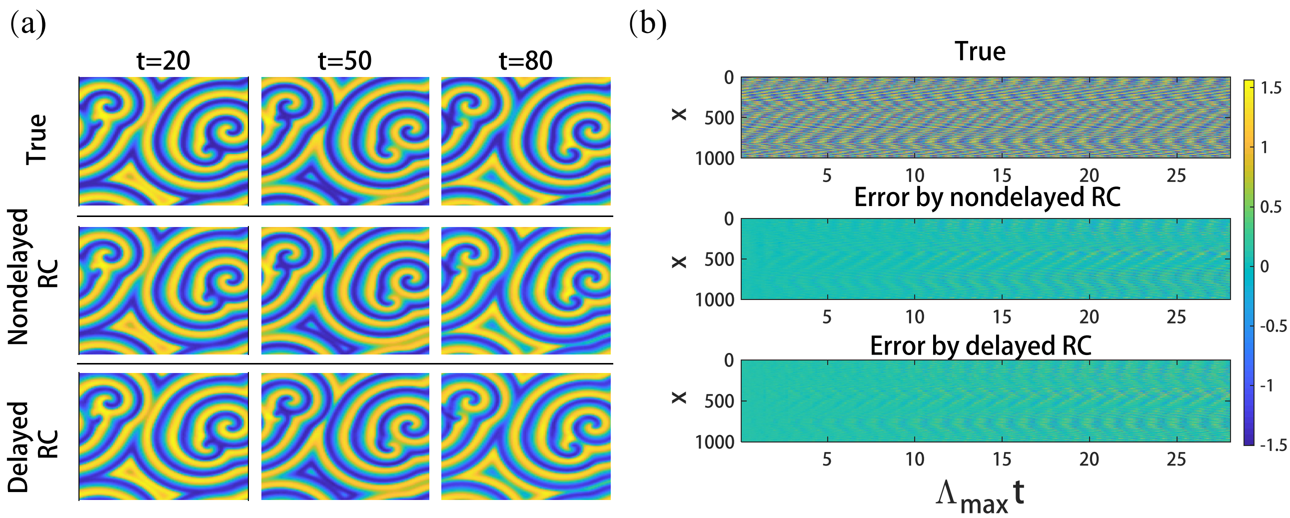

FIG. 4. Reconstructed chaotic patterns for the ISCAM model by a standard nondelayed RC including 5000 neurons and a time-delayed RC including 1000 neurons and five lags for each neuron. (a) Selected dynamical pattern using different evolution rules. (b) Reconstruction errors deviating from true dynamics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

determine d and τ. From a perspective of embedding, d is only required to be larger than $2 \cdot \dim(\mathcal{M})$ while practically the box-counting dimension of the manifold $\mathcal{M}$ is usually very small, i.e., $\dim(\mathcal{M})$ 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 τ, 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 $\tau \cdot \Delta t \cdot N_{\text{lag}} < \Lambda_{\max}$ for the choice of τ and N_{lag} in practice, where Δt 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).

[1] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[2] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks – with an erratum note, GMD Technical Report No. 148, German National Research Center for Information Technology, Bonn, 2001.

[3] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Comput. **14**, 2531 (2002).

[4] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach, Phys. Rev. Lett. **120**, 024102 (2018).

[5] Q. Zhu, H. Ma, and W. Lin, Detecting unstable periodic orbits based only on time series: When adaptive delayed feedback control meets reservoir computing, Chaos **29**, 093125 (2019).

[6] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, Neural Networks **115**, 100 (2019).

[7] Y. Tang, J. Kurths, W. Lin, E. Ott, and L. Kocarev, Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics, Chaos **30**, 063151 (2020).

[8] H. Jaeger, *Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the 'Echo State Network" Approach* (GMD-Forschungszentrum Informationstechnik, Bonn, 2002), Vol. 5.

[9] A. Haluszczynski and C. Räth, Good and bad predictions: Assessing and improving the replication of chaotic attractors by means of reservoir computing, Chaos **29**, 103143 (2019).

[10] A. Griffith, A. Pomerance, and D. J. Gauthier, Forecasting chaotic systems with very low connectivity reservoir computers, Chaos **29**, 123108 (2019).

[11] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. **3**, 127 (2009).

[12] J. Jiang and Y.-C. Lai, Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius, Phys. Rev. Res. **1**, 033056 (2019).

[13] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, Information processing in echo state networks at the edge of chaos, Theory Biosci. **131**, 205 (2012).

[14] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, Sci. Rep. **2**, 514 (2012).

[15] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, Re-visiting the echo state property, Neural Networks **35**, 1 (2012).

[16] G. Manjunath and H. Jaeger, Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks, Neural Comput. **25**, 671 (2013).

[17] H. Chang, S. Nakaoka, and H. Ando, Effect of shapes of activation functions on predictability in the echo state network, arXiv:1905.09419.

[18] J. J. Steil, Online reservoir adaptation by intrinsic plasticity for backpropagation–decorrelation and echo state learning, Neural Networks **20**, 353 (2007).

[19] E. Bollt, On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD, Chaos **31**, 013108 (2021).

[20] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, Next generation reservoir computing, Nat. Commun. **12**, 5564 (2021).

[21] A. Hart, J. Hook, and J. Dawes, Embedding and approximation theorems for echo state networks, Neural Networks **128**, 234 (2020).

[22] A. Hart, J. Hook, and J. Dawes, Echo state networks trained by Tikhonov least squares are l2 (μ) approximators of ergodic dynamical systems, Physica D **421**, 132882 (2021).

[23] L. Appeltant, M. C. Soriano, G. Van 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, Nat. Commun. **2**, 468 (2011).

[24] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger, Tutorial: Photonic neural networks in delay systems, J. Appl. Phys. **124**, 152004 (2018).

[25] 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, Phys. Rev. X **7**, 011015 (2017).

[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.L022041 for details of the proof of the embedding theory, the system models, the parameters settings, and the comparison study with the photonic reservoir.

[27] F. Takens, Detecting strange attractors in turbulence, Lect. Notes Math. **898**, 366 (1981).

[28] T. Sauer, J. Yorke, and M. Casdagli, Embedology, J. Stat. Phys. **65**, 579 (1991).

[29] H.-F. Ma, S. Leng, K. Aihara, W. Lin, and L. Chen, Randomly distributed embedding making short-term high-dimensional data predictable, Proc. Natl. Acad. Sci. **115**, E9994 (2018).

[30] J.-W. Hou, H.-F. Ma, D. He, J. Sun, Q. Nie, and W. Lin, Harvesting random embedding for high-frequency change-point detection in temporal complex systems, Natl. Sci. Rev. **9**, nwab228 (2022).

[31] Y. Suzuki, M. Lu, E. Ben-Jacob, and J. N. Onuchic, Periodic, quasi-periodic and chaotic dynamics in simple gene elements with time delays, Sci. Rep. **6**, 21037 (2016).

[32] H. Wernecke, B. Sándor, and C. Gros, Chaos in time delay systems, an educational review, Phys. Rep. **824**, 1 (2019).

[33] A. W. M. Dress, W. Hordijk, W. Lin, and P. Serocka, The ideal storage cellular automaton model, in *Structure Discovery in Biology: Motifs, Networks & Phylogenies in Dagstuhl Seminar Proceedings* (Schloss Dagstuhl-Leibniz-Zentrum für Informatik, Wadern, Germany, 2010).

[34] A. W. Dress and W. Lin, Dynamics of a discrete-time model of an "ideal-storage" system describing hetero-catalytic processes on metal surfaces, I. J. Bifurcation Chaos **21**, 1331 (2011).

[35] M. J. McGuinness, The fractal dimension of the Lorenz attractor, Phys. Lett. A **99**, 5 (1983).

[36] D. Viswanath, The fractal property of the Lorenz attractor, Physica D **190**, 115 (2004).

[37] Y. Sakemi, K. Morino, T. Leleu, and K. Aihara, Model-size reduction for reservoir computing by concatenating internal states through time, Sci. Rep. **10**, 21794 (2020).