Learning the dynamics of coupled oscillators from transients

Huawei Fan [1,2] Liang Wang,[1] Yao Du,[1] Yafeng Wang,[3] Jinghua Xiao,[4] and Xingang Wang [1,*]

[1]*School of Physics and Information Technology, Shaanxi Normal University, Xi'an 710119, China*
[2]*School of Science, Xi'an University of Posts and Telecommunications, Xi'an 710121, China*
[3]*Nonlinear Research Institute, Baoji University of Arts and Sciences, Baoji 721016, China*
[4]*School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*

Whereas the importance of transient dynamics to the functionality and management of complex systems has been increasingly recognized, most of the studies are based on models. Yet in realistic situations the mathematical models are often unknown and what is available are only measured time series. Meanwhile, many real-world systems are dynamically stable, in the sense that the systems return to their equilibria in a short time after perturbations. This increases further the difficulty of dynamics analysis, as many information of the system dynamics are lost once the system is settled onto the equilibrium states. The question we ask is: Given the transient time series of a complex dynamical system measured in the stable regime, can we infer from the data some properties of the system dynamics and make predictions, e.g., predicting the critical point where the equilibrium state becomes unstable? We show that for the typical transitions in system of coupled oscillators, including quorum sensing, amplitude death, and complete synchronization, this question can be addressed by the technique of reservoir computing in machine learning. More specifically, by the transient series acquired at several states in the stable regime, we demonstrate that the trained machine is able to predict accurately not only the transient behaviors of the system in the stable regime, but also the critical point where the stable state becomes unstable. Considering the ubiquitous existence of transient activities in natural and man-made systems, the findings may have broad applications.

I. INTRODUCTION

Transient activities are ubiquitous in the real world. When a neuron is stimulated by an external signal, its membrane voltage may raise temporally and then return to the resting state in milliseconds; when a pebble is thrown into a pond, ripples on the surface of the water will be disappeared in minutes; when a traffic congestion is triggered by an accident, the traffic could return to normal in hours; when an epidemic occurs, it might exist for months and finally will pass away. Roughly, transient is the temporal evolution preceding the asymptotic dynamics, which is an intrinsic property of the dynamical systems [1,2]. In traditional studies of system dynamics, attentions have been mainly focusing on the asymptotic dynamics that the system presents after the transient. Yet accumulating evidences show that transient activities might provide a new dimension in our understanding of the system dynamics [3,4], and in many cases are more relevant to the system functions and performance than the asymptotic dynamics, especially for complex nonlinear

systems [5–16]. For instance, in ecological systems [5,6], transient dynamics provides a possible explanation for the occurrence of regime shift under constant environment and the sudden extinction of many species; in neuronal systems [7–9], neurons respond to stimulus mostly during the transients and the information encoded in the transients is more reliable than in the asymptotic dynamics; in climate systems [10,11], transient dynamics provides key information for predicting the extreme events and tipping points; in power-grids [12–14], transient is an important concern for the system stability and security, and provides the time window for responses; in robotics [15,16], transient dynamics plays a crucial role in realizing the autonomous control of robot behaviors. Whereas the importance of transient dynamics has become increasingly recognized in a wide range of disciplines, the study of transient dynamics is challenging, especially for complex systems of spatiotemporal nonlinear dynamics [3]. One reason is that the transient dynamics is different from the asymptotic dynamics. To be specific, the former is governed by the nonattracting sets embedded in the phase space, while the latter is governed by the attracting set. To analyze the transient dynamics in depth, one has to find all the nonattracting sets and their unstable manifolds [17,18], which are extremely difficult for high-dimensional systems [19–21]. Another reason is that in realistic situations the equations governing the system dynamics are normally unknown, and what is available are only the measured time series [4]. This means that any analysis about the transient dynamics must be based on data,

*wangxg@snnu.edu.cn

calling thus for the development of model-free, data-based techniques [22,23].

Model-free prediction of chaotic systems by the technique of reservoir computer (RC) in machine learning has received considerable attention in recent years [24–37]. From the perspective of dynamical systems, RC can be regarded as a complex network of coupled nonlinear units which, driven by the input signals, generates the outputs through a readout function [38,39]. In the training phase, the input signals are provided by the target system, and the purpose of the training is to find the set of coefficients in the readout function for a best fitting of the training data. In the predicting phase, the input signals are replaced by the outputs, and the machine is running as an autonomous system with the fixed parameters. Although structurally simple, RC has shown its super power in many data-oriented applications [38], e.g., speech recognition, channel equalization, robot control, and chaos prediction. In particular, it has been shown that a properly trained RC is able to predict accurately the state evolution of a chaotic system for about half a dozen Lyapunov times [25], which is much longer than the prediction horizon of the traditional methods developed in nonlinear science. Besides predicting the short-term state evolution, RC is also able to replicate faithfully the long-term statistical properties of chaotic systems [27], e.g., the dimension of strange attractors and the Lyapunov exponents. This ability, known as climate replication, has been exploited recently to predict the critical transitions in complex nonlinear systems [32–34]. For instance, by incorporating a parameter-control channel into RC, it is shown that the machine trained by the time series of several states in the oscillatory regime of a dynamical system is able to predict not only the critical point for system collapse, but also the averaged lifetime of the transients in the postcritical regime [32]; training the machine by the time series of coupled oscillators at several states in the desynchronization regime, the machine is able to predict accurately the critical coupling for synchronization [33]. It is noted that in predicting chaotic systems by the technique of RC, the training data are all measured from the asymptotic dynamics that the systems are finally developed to, while the transient behaviors preceding the asymptotic dynamics is normally discarded [24–37].

Suppose that a complex system of coupled dynamical units is operating at a stable state (say, for example, a steady state) and, after each perturbation, restores to the stable state after a short transient, the question we ask is: Based on the transient time series measured at several states in the stable regime, can we predict the critical point where the system becomes unstable? This question is challenging and worth pursuing for several reasons. First, it is not clear whether the machine can be properly trained by the short time series measured on the transient activities. In training RC, a general requirement is that the time series should be sufficiently long (depending on the prediction tasks and the hyperparameters of the RC, the length of the training series may change from hundreds to thousands system oscillations) [38]. Yet in many realistic systems the transient activities sustain for only a short period (normally several system oscillations). Once settled to the asymptotic attractor, e.g., the steady state, many information of the system dynamics will be lost. The limited length of the transient dynamics therefore requires a fast learning of

the machine. At present, it is not clear whether this goal can be achieved by the technique of RC. Second, different from the asymptotic dynamics, the transient dynamics is dependent on the initial conditions. Generally, the larger (smaller) is the perturbation, the longer (shorter) will be the transient [1]. (Statistically, for the fixed system parameters, the lifetime of the transients follows an exponential distribution [4].) At present, it is not clear whether the machine trained by the time series of one transient process is able to predict the evolution of another transient process. Finally, even if the machine is properly trained by the transient time series, it remains unknown whether the trained machine is able to predict the critical point separating the stable and unstable regimes, as the system dynamics in the two regimes are distinctly different. Our main objective in the present paper is to provide an affirmative answer to the above question. Specifically, we are going to demonstrate that for the typical bifurcation scenarios in coupled oscillators, including quorum sensing [40], amplitude death [41], and complete synchronization [42], the machine trained by the transient time series acquired in the stable regime is able to predict not only the system evolution in the stable regime, but also the critical point where the stable state becomes unstable.

The rest of the paper is organized as follows. In Sec. II, we will propose the scheme of parameter-aware RC, and present the details of the training and predicting phases. The results predicted by the machine on the phase transition of three typical models, including an ensemble of indirectly coupled limit-cycle oscillators showing the phenomenon of quorum sensing, two coupled nonidentical limit-cycle oscillators showing the phenomenon of amplitude death, and two coupled chaotic oscillators showing the phenomenon of complete synchronization, will be presented in Sec. III. Discussions and conclusions will be given in Sec. IV.

II. PARAMETER-AWARE RESERVOIR COMPUTER

We generalize the scheme of parameter-aware RC proposed in Refs. [32,33] to predict the phase transition in coupled oscillators. The RC consists of four components: the I/R layer (input-to-reservoir), the parameter-control channel, the reservoir, and the R/O layer (reservoir-to-output). The I/R layer is characterized by the matrix $\mathbf{W}_{in} \in \mathbb{R}^{D_r \times D_{in}}$, which couples the input vector $\mathbf{u}_\beta(t) \in \mathbb{R}^{D_{in}}$ to the reservoir network. Here, $\mathbf{u}_\beta(t)$ denotes the input vector that is acquired from the target system at time t and under the specific bifurcation parameter β. The elements of $\mathbf{W}_{in}$ are randomly drawn from a uniform distribution within the range $[-\sigma, \sigma]$. The parameter-control channel is characterized by the vector $\mathbf{s} = \beta \mathbf{b}$, with β the control parameter and $\mathbf{b} \in \mathbb{R}^{D_r}$ the bias vector. In applications, the control parameter β can be treated as an additional input channel marking the input vector $\mathbf{u}(t)$. The elements of $\mathbf{b}$ are drawn randomly from a uniform distribution within the range $[-\sigma, \sigma]$. The reservoir network contains D_r dynamical nodes, with the initial states of the nodes being randomly chosen from the interval $[-1, 1]$. The states of the nodes in the reservoir network $\mathbf{r}(t) \in \mathbb{R}^{D_r}$ are updated according to the equation

$$\mathbf{r}(t + \Delta t) = \tanh[\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{in}\mathbf{u}_\beta(t) + \beta \mathbf{b}]. \quad (1)$$

Here, Δt is the time step for updating the reservoir network, and $\mathbf{A} \in \mathbb{R}^{D_r \times D_r}$ is the weighted adjacency matrix representing the coupling relationship between nodes in the reservoir. The adjacency matrix $\mathbf{A}$ is constructed as a sparse random Erdös-Rényi matrix: with the probability p, each element of the matrix is arranged a nonzero value drawn randomly from the interval $[-1, 1]$. The matrix $\mathbf{A}$ is rescaled to make its spectral radius equal λ. Before the training process, the reservoir is evolved for a transient period of T_0', so as to avoid the influence induced by the initial states of the nodes. The output layer is characterized by the matrix $\mathbf{W}_{\text{out}} \in \mathbb{R}^{D_{\text{out}} \times D_r}$, which generates the output vector, $\mathbf{v}(t) \in \mathbb{R}^{D_{\text{out}}}$, according to the equation

$$\mathbf{v}(t + \Delta t) = \mathbf{W}_{\text{out}} \tilde{\mathbf{r}}(t + \Delta t), \tag{2}$$

with $\tilde{\mathbf{r}} \in \mathbb{R}^{D_r}$ the new sate vector transformed from the reservoir state (i.e., $\tilde{r}_i = r_i$ for the odd nodes and $\tilde{r}_i = r_i^2$ for the even nodes) [28], and $\mathbf{W}_{\text{out}}$ the output matrix to be obtained by the training process. Except $\mathbf{W}_{\text{out}}$, all other parameters of the RC, e.g., $\mathbf{W}_{\text{in}}$, $\mathbf{A}$, and $\mathbf{b}$, are fixed at the construction. The purpose of the training process is to find a suitable output matrix $\mathbf{W}_{\text{out}}$ so that the output vector $\mathbf{v}(t + \Delta t)$ as calculated by Eq. (2) is as close as possible to the input vector $\mathbf{u}(t + \Delta t)$ for $t = (\tau + 1)\Delta t, \ldots, (\tau + L)\Delta t$, with $T_0 = \tau \Delta t$ the transient period and L the length of the training time series. This can be done by minimizing the following cost function with respect to $\mathbf{W}_{\text{out}}$ [26–28]:

$$\mathbf{W}_{\text{out}} = \mathbf{U}\mathbf{V}^T(\mathbf{V}\mathbf{V}^T + \eta \mathbb{I})^{-1}. \tag{3}$$

Here, $\mathbf{V} \in \mathbb{R}^{D_r \times L}$ is the state matrix whose kth column is $\tilde{\mathbf{r}}[(\tau + k)\Delta t]$, $\mathbf{U} \in \mathbb{R}^{D_{\text{in}} \times L}$ is a matrix whose kth column is $\mathbf{u}[(\tau + k)\Delta t]$, $\mathbb{I}$ is the identity matrix, and η is the ridge regression parameter for avoiding the overfitting. After training, the output matrix $\mathbf{W}_{\text{out}}$ will be fixed, and the RC is ready for prediction. In the predicting phase, first we set the control parameter β to a specific value of interest (not necessarily the parameters used in the training phase), then we evolve the RC as an autonomous dynamical system by taking the output vector $\mathbf{v}(t)$ as the input vector at the next time step $\mathbf{u}_\beta(t + \Delta t)$. The output vector $\mathbf{v}(t)$ gives the prediction. For the sake of simplicity, we set $D_{\text{out}} = D_{\text{in}}$ [26–28].

We note that in the training phase the input data consists of two time series: (1) the input vector $\mathbf{u}_\beta(t)$ representing the state of the target system and (2) the control parameter $\beta(t)$ labeling the condition under which the input vector $\mathbf{u}_\beta(t)$ is acquired. In specific, the input vector $\mathbf{u}_\beta(t)$ is composed of m segments of length L', while each segment is a time series obtained from the target system under the specific control parameter β. As such, $\beta(t)$ is a step-function of time. In the predicting phase, besides replacing $\mathbf{u}_\beta(t)$ with $\mathbf{v}(t)$, we still need to input the control parameter $\beta(t)$ in a step-wise fashion so as to guide the reservoir evolution. The essential difference between the current paper and the previous ones lies in the regime where the training data are acquired. In Ref. [32], the training data are taken from the oscillatory regime of chaotic attractors, and the mission is to predict the critical point leading to the steady states; in Ref. [33], the training data are acquired from the desynchronization regime, and the mission is to predict the critical coupling generating complete synchronization. In both papers, the measured time series

are sufficiently long for training the machine. Different from these papers, in our present paper the training data are taken from the regime of steady states (or complete synchronization state), in which only a short time series, i.e., the transient dynamics, are available for training the machine.

III. RESULTS

A. Quorum sensing in coupled Stuart-Landau oscillators

Quorum sensing refers to the sudden switch of the dynamics of an ensemble elements from the quiescent state to the synchronized oscillatory state as the population density of the elements exceeds a critical value, which is normally observed in systems where a large population of dynamical elements are coupled indirectly through a common medium, e.g., chemical oscillators [40], bacterial systems [43], genetic cellular oscillators [44], and crowd on a footbridge [45]. Below the critical population density, the quiescent state is stable, i.e., once perturbed, the system will return to the quiescent state after a short transient. Assuming that the system is operating in the quiescent regime and only the transient series at several population densities are available, the question here we ask is: Without knowing the equations of the system dynamics, can we predict when will the elements start to oscillate? We are going to demonstrate that for the typical model showing quorum sensing, the question can be addressed by the scheme of parameter-aware RC introduced in Sec. II.

The model for quorum sensing we adopt here is an ensemble of Stuart-Landau oscillators coupled through an external medium [46]. The system dynamics reads

$$\dot{z}_j = (1 + i\omega_j - |z_j|^2)z_j + D(Z - z_j), \tag{4}$$

$$\dot{Z} = -(J + i\omega_0)Z + \frac{\rho D}{N} \sum_{j=1}^{N}(z_j - Z), \tag{5}$$

where $j = 1, 2, ..., N$ is the oscillator index, $z_j = x_j + iy_j$ is the complex variable denoting the state of oscillator j, and $Z = X + iY$ is the complex variable denoting the state of the external medium. In Eq. (4), ω_j is the natural frequency of oscillator j, which is randomly chosen within the range $[-1, 1]$. The oscillators are isolated from each other, but are all coupled to the external medium through the diffusions, with the coupling strength (diffusing coefficient) being represented by D. In Eq. (5), ω_0 denotes the natural frequency of the medium, J represents the relaxation parameter, and ρ denotes the population density, which plays as the bifurcation parameter for generating quorum sensing. (As the number of oscillators is fixed, the increase of the population density corresponds to the decrease of the system volume in experiments [40,43–47].) In simulations, we set the system parameters as $N = 100$, $D = 1.5$, $\omega_0 = 1$, and $J = 0.2$. The initial states of the oscillators and the external medium (including the real and imaginary parts) are randomly chosen from the range $(-\delta, \delta)$. The equations are solved numerically by the fourth-order Runge-Kutta method with the time step $\Delta t = 0.1$. Following Refs. [40,43–47], we characterize the asymptotic behaviors of the oscillators and the external medium by the the order parameters $R_z = \langle |\frac{1}{N} \sum_{j=1}^{N} z_j| \rangle_T$ and $R = \langle |Z| \rangle_T$, respectively. Here, $\langle ... \rangle_T$ denotes that the results
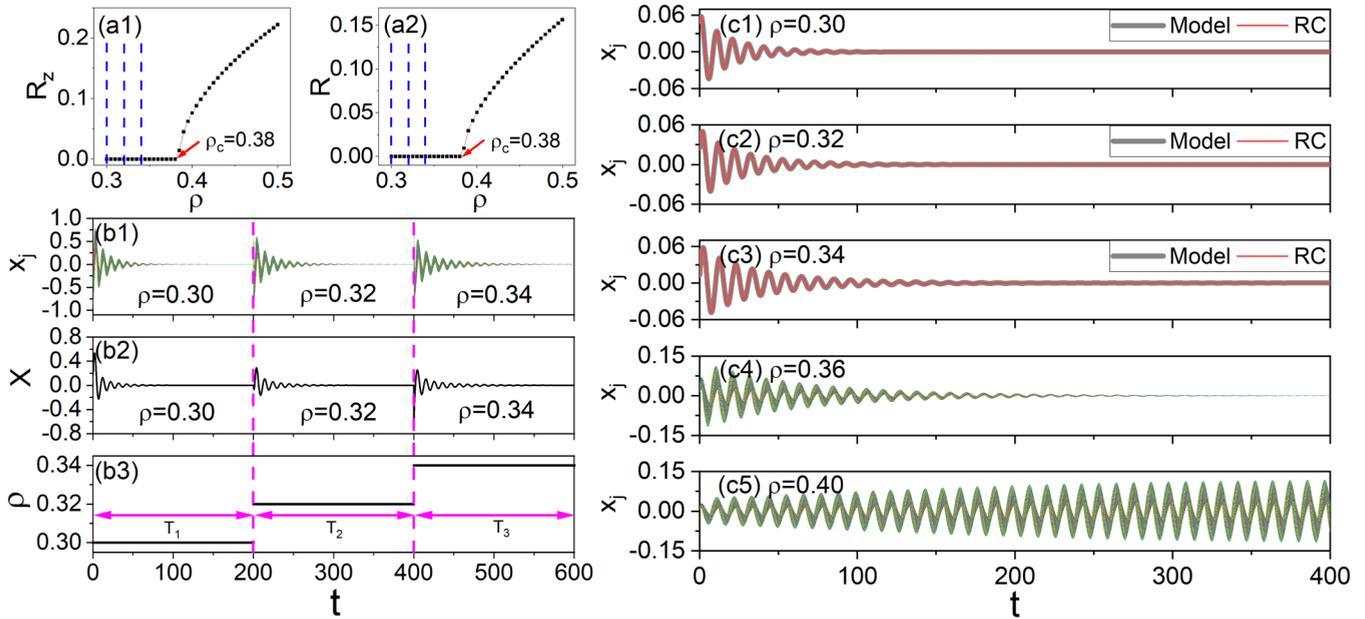
FIG. 1. Predicting the dynamics of an ensemble of Stuart-Landau oscillators coupled through an external medium. [(a1),(a2)] The variations of the system order parameters R_z and R with respect to the population density ρ obtained from model simulations. The transition from the quiescent to oscillatory states occurs at $\rho_c \approx 0.38$. Vertical lines denote the sampling states where the training data are acquired. [(b1),(b2)] The time evolution of the oscillators and the medium under different values of ρ adopted in the quiescent regime. (b3) The time series of the control parameter. $T_1 = T_2 = T_3 = 200$ is the time length of the sampling time series. [(c1)–(c5)] The time evolutions of the system predicted by the machine for different values of ρ. Black curves are the results obtained from model simulations. Red curves are the results predicted by the machine.

are averaged over a time period of T. To focus on the asymptotic behaviors of the system, a transient period of T' has been discarded in calculating the order parameters.

Setting $\delta = 1$, $T' = 1 \times 10^4$ and $T = 2 \times 10^4$, we plot in Fig. 1(a) the variations of R_z and R with respect to ρ. It is seen that $R_z = R = 0$ for $\rho < \rho_c \approx 0.38$ and, as ρ exceeds ρ_c, the values of R_z and R are gradually increased. That is, the asymptotic dynamics of the oscillators and the medium is quiescent for $\rho < \rho_c$ and is oscillatory for $\rho > \rho_c$. To show the transient dynamics of the system in the quiescent regime, we plot in Figs. 1(b1) and 1(b2) the time evolution of the oscillators and the medium for $m = 3$ sampling population densities ($\rho = 0.30$, 0.32, and 0.34), each of time duration $T = 200$. It is seen that in all three cases, the oscillators and the medium are damped to the origin after a short transient ($t < 100$). Here our mission is to infer from the transient time series measured at the three sampling states the transition point ρ_c.

To implement the parameter-aware RC, we first generate the training data by combining the time series of the three sampling states into a long time sequence. The training series are plotted in Figs. 1(b1) and 1(b2). As the time series of each sampling state contains $L' = T/\Delta t = 2 \times 10^3$ data points, the length of the combined sequence therefore is $L = mL' = 6 \times 10^3$. The corresponding time series of the population density, as shown in Fig. 1(b3), is used as the inputs of the parameter-control channel, i.e., replacing β with ρ in Eq. (1). Then, using the combined times series and the time series of $\rho(t)$ as the inputs, we calculate the output matrix according to Eq. (3). This completes the training phase.

In obtaining the output matrix, a short segment of $\tau = 100$ data points in each sampling series is used to drive the reservoir out of the transient. The hyperparameters of RC are chosen as $(D_r, p, \sigma, \lambda, \eta) = (5 \times 10^3, 0.1, 1, 0.1, 1 \times 10^{-9})$, which are obtained by the optimizer "optimoptions" in MATLAB. In addition, to reduce the complexity of the input matrix $\mathbf{W}_{\text{in}}$, we set only one nonzero element in each row of the input matrix. That is, each node in the reservoir network receives only one component from the input vector $\boldsymbol{u} = [x_1, y_1, ..., x_N, y_N, X, Y]^T$.

Before predicting the transition point of quorum sensing, we check first the performance of the trained machine in predicting the transient dynamics of the sampling parameters, including $\rho = 0.30$, 0.32, and 0.34. In the predicting phase, we replace the input vector $\mathbf{u}$ in Eq. (1) with the output vector $\mathbf{v}$ calculated by Eq. (2), while setting the control parameter β in Eq. (1) to be one of the sampling parameters. The reservoir network is started from the random initial conditions, and is driven by the testing data for $\tau' = 10$ time steps [48]. (This strategy of "cold start" is necessary for predicting the time evolution of the system state, but is not when the mission is to predict the transition point [32,33,36].) The predicted results are plotted in Figs. 1(c1)–1(c3). We see that in all the cases, the machine predicts accurately not only the transient evolution of the system state, but also the quiescent state the system is finally settled to.

We check further the capability of the trained machine in predicting the time evolution of the system state under a new parameter that is not included in the sampling set. To demonstrate, we set $\rho = 0.36$ as the control parameter, which
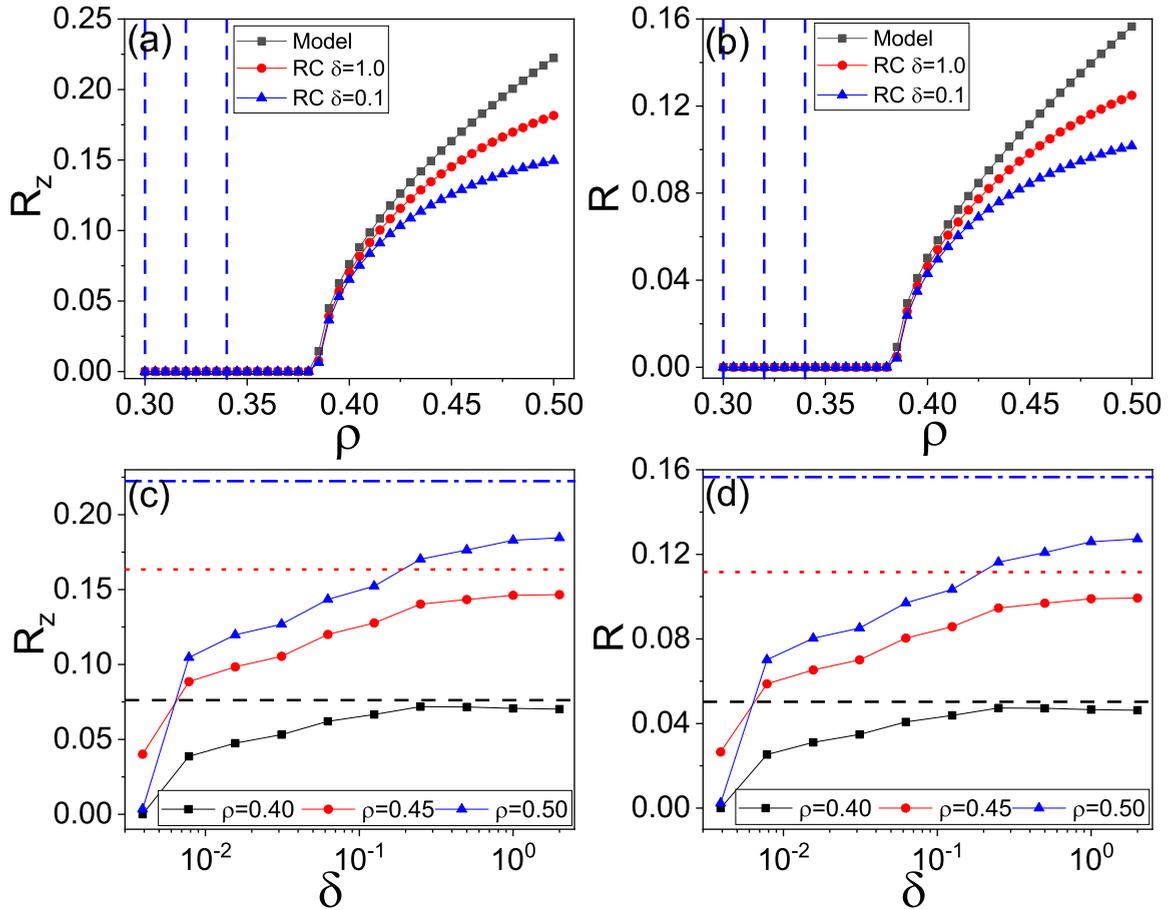
FIG. 2. Predicting quorum-sensing transition in an ensemble of indirectly coupled Stuart-Landau oscillators. [(a),(b)] The variation of the order parameters R_z and R with respect to ρ predicted by machine. The initial conditions of the system are randomly chosen from the range $(-\delta, \delta)$. Black squares: the results obtained by model simulations. Red discs: the predicted results for $\delta = 1.0$. Blue triangles: the predicted results for $\delta = 0.1$. Vertical lines denote the sampling states. [(c),(d)] The variation of R_z and R with respect to δ for different values of ρ adopted in the oscillatory regime. Horizontal lines are the corresponding order parameters obtained from model simulations. As δ increases, the predicted order parameters approach the actual values.

is also within the quiescent regime. The predicted results for this new parameter are plotted in Fig. 1(c4). It is seen that the transient dynamics and the quiescent state are well predicted by the machine too. Setting $\rho = 0.4$, we plot in Fig. 1(c5) the time evolution predicted by the machine, which suggests that, instead of reaching the quiescent state, the system will finally settle to an oscillatory state. That is, the machine predicts that quorum sensing occurs before the population density $\rho = 0.4$. This prediction is consistent with the results obtained from model simulations, as depicted in Fig. 1(a).

Having justified the capability of RC in predicting the dynamics of the individual states in both the quiescent and oscillatory regimes, we proceed to predict the transition point ρ_c by the scheme of parameter-aware RC introduced in Sec. II. In doing this, we increase the control parameter ρ gradually from 0.3 to 0.5 by the step $\Delta\rho = 0.05$. For each value of ρ, we first run the machine (starting from the random initial conditions) for a transient period of $T_0' = 1 \times 10^3$, and then calculate from the outputs the time-averaged order parameters R_z and R over a period of $T = 1 \times 10^4$. The results are plotted in Figs. 2(a) and 2(b) (red discs). We see that the machine predicts accurately the critical population density, $\rho_c \approx 0.38$,

and also the progressive increase of R_z and R as ρ increases from ρ_c. Figures 2(a) and 2(b) also show that in the oscillatory regime ($\rho > \rho_c$), the values of R_z and R predicted by the machine are smaller to the values obtained from model simulations (black squares). More specifically, as ρ increases from ρ_c, the error between the predicted and actual results is gradually enlarged.

As the training data are collected from states in the quiescent regime, the useful information for machine training is thus contained in the transient activities only. It is natural to conjecture that the longer the transient, the more information will the machine infer from the data, and the more accurate will be the prediction. To justify this conjecture, we enlarge the range $(-\delta, \delta)$ over which the initial conditions of the oscillators and the medium are chosen, and check again the prediction performance by training a new machine. Statistically, with the increase (decrease) of δ, the lifetime of the transient will be extended (shortened) [4]. Decreasing δ to 0.1, we generate the training data by the same sampling states in the quiescent regime ($\rho = 0.30$, 0.32, and 0.34), and train the RC again. The length of the sampling series and the hyperparameters of the RC are the same to the ones used in

Fig. 1. The results predicted by the new machine are plotted in Figs. 2(a) and 2(b) (blue triangles). We see that the transition point ρ_c is also accurately predicted, but, comparing with the results of $\delta = 1.0$, the prediction is worse in the oscillatory regime. To explore further the impact of δ on the prediction performance, we plot in Figs. 2(c) and 2(d) the variation of the order parameters with respect to δ for different values of ρ in the oscillatory regime. It is seen that with the increase of δ, the order parameters predicted by the machine approach gradually the actual order parameters calculated from model simulations. Numerical evidences thus suggest that it is the transient activities of the quiescent states that provide the useful information for machine training, and, with the increase of the transient lifetime, the prediction performance is gradually improved.

Additional simulations have been conducted to check the impacts of the sampling states on the prediction performance, including the number of the sampling states and the locations of the sampling states. The general findings are the following (not shown): (1) when the number of sampling parameters is fixed, the closer the sampling parameters to the transition point, the better is the prediction; (2) the prediction performance is improved by adopting more sampling states in the quiescent regime. The additional results are understandable, as: (1) in the quorum-sensing transition, as the bifurcation parameter approaches the transition point from below, the lifetime of the transient will be gradually increased [43–47]; (2) by combining the time series of more sampling states, the total number of transient data in the training series will be increased. These additional results suggest again that the useful information for training the machine comes from the transient dynamics, but not the asymptotic dynamics associated with the quiescent state.

B. Amplitude death in coupled Stuart-Landau oscillators

Amplitude death refers to the cessation of oscillation in coupled oscillators as the system bifurcation parameter passes through a critical value [41,49]. For its important implications to the functioning of many real-world systems, the phenomenon of amplitude death has been extensively studied by researchers in different fields in the past decades [50]. Similar to the phenomenon of quorum sensing, in amplitude death the system dynamics is also transited from the quiescent to oscillatory states at a critical point. To generate amplitude death, a general approach is introducing a parameter mismatch among the oscillators and using the coupling strength of the oscillators as the bifurcation parameter [51]. This phenomenon can also be observed in systems of identical oscillators by adopting new coupling schemes, such as introducing time-delay to the couplings [51], using the conjugate coupling functions or dynamical couplings [52,53]. Recently, stimulated by the blooming of network science, amplitude death in networked oscillators has been explored, in which the important role of network structure on the transition has been revealed [54–56]. Depending on the system functions, amplitude death might be desired or undesired. In neuronal systems, an ensemble of neural cells may stop pulsating under strong couplings, leading to the pathological conditions [57]. In physiological context, the quenching of oscillations means

the loss of rhythms, resulting in diseases such as sudden cardiac death [58]. In cases like this, amplitude death is undesired and the mission is to prevent its occurrence. However, in physical systems such as coupled lasers [59], quenched oscillations are necessary for generating a stable output; in ecological systems, large-amplitude oscillations increase the risk of species extinctions [60]. In cases like this, amplitude death is desired and the mission becomes maintaining the quiescent state.

As in realistic situations the system equations are generally unknown, the prediction of amplitude death therefore relies on the development of model-free techniques. By the scheme of parameter-aware RC, in Ref. [34] the authors are able to predict successfully the transition to amplitude death in coupled chaotic and periodic oscillators. Different from our present paper, in Ref. [34] the training data are collected in the oscillatory regime, in which the length of the measured time series at each sampling state is sufficiently long (about a few hundreds oscillations.) The question here we ask is: If the time series are collected in the quiescent regime in which the transient activities sustain for only a short period (e.g., a few dozens oscillations), can we predict the critical point where the systems is survived from amplitude death? We are going to show next that for the model of coupled limit cycles, the prediction is feasible.

The model of amplitude death we consider here consists of two diffusively coupled nonidentical Stuart-Landau oscillators [41,49],

$$\dot{z}_1 = (1 + i\omega_1 - |z_1|^2)z_1 + \varepsilon(z_2 - z_1),$$
$$\dot{z}_2 = (1 + i\omega_2 - |z_2|^2)z_2 + \varepsilon(z_1 - z_2), \quad (6)$$

where $z_{1,2} = x_{1,2} + iy_{1,2}$ are the complex variables denoting the oscillator states, and ε is the coupling strength. As analyzed in Ref. [49], the two oscillators are ceased to the origin ($z_1 = z_2 = 0$) when $\varepsilon > \varepsilon_c = 1$ and $\Delta\omega > 2\sqrt{2\varepsilon - 1}$, with $\Delta\omega \equiv |\omega_1 - \omega_2|$ the frequency mismatch between the oscillators. In simulations, we fix the natural frequencies of the two oscillators as $\omega_1 = 2.0$ and $\omega_2 = 7.0$, and tune the coupling strength ε to investigate the transition. The initial states of the oscillators, i.e., the values of $x_{1,2}$ and $y_{1,2}$ at $t = 0$, are chosen randomly from the range $(-1, 1)$. The time step used in simulating Eq. (6) is $\Delta t = 0.02$. Setting $\varepsilon = 0.8 < \varepsilon_c$, we plot in Fig. 3(a) the evolutions of x_1 and x_2 with time. We see that the system shows the feature of permanent oscillation. An example of the quenched state is plotted in Fig. 3(b), in which the coupling strength is set as $\varepsilon = 1.15 > \varepsilon_c$. We see that in this case the oscillation is ceased after a transient period about $T = 20$. To have a global picture on the bifurcation diagram, we plot in Fig. 3(c) the variation of the local extrema of x_1 with respect to ε, which confirms the theoretical prediction that amplitude death occurs at $\varepsilon_c = 1$. Here our mission is to reproduce the bifurcation diagram based on the transient time series acquired at several states in the death regime.

In predicting the bifurcation diagram, we first generate the training data by combining the time series of several states in the death regime. As illustration, we choose the sampling states at three coupling parameters: $\varepsilon = 1.05, 1.1,$ and 1.15. The length of the time series for each state is $T = 40$, and the training data contains in total $L = mT/\Delta t = $
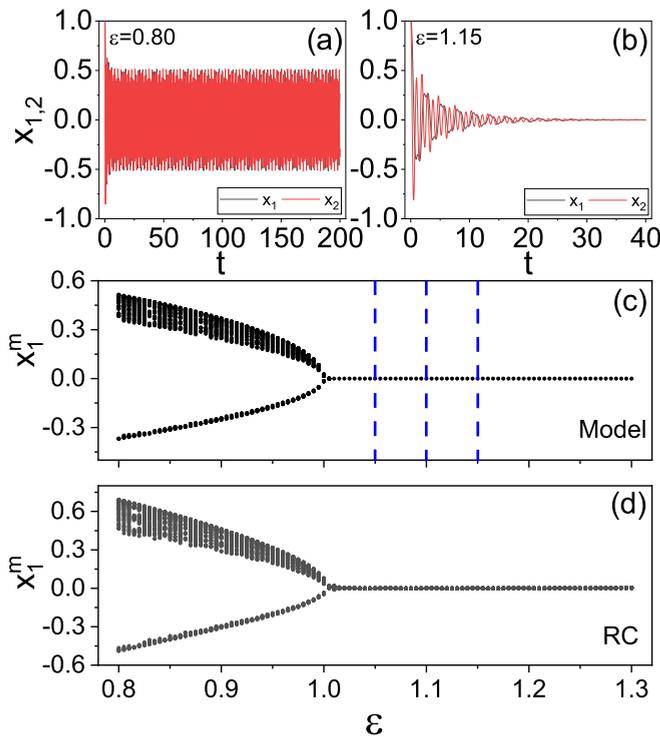
FIG. 3. Predicting the transition from oscillation to amplitude death in two coupled nonidentical Stuart-Landau oscillators. (a) The time evolutions of x_1 and x_2 for the coupling strength $\varepsilon = 0.80$. (b) The time evolutions of x_1 and x_2 for the coupling strength $\varepsilon = 1.15$. (c) The bifurcation diagram obtained from model simulations. x_1^m is the local maximum or minimum of the variable x_1 during the course of system evolution. Vertical lines denote the locations of the sampling states. (d) The bifurcation diagram predicted by the machine.

6×10^3 points. We then input the training data and the corresponding time series of the control parameter, $\varepsilon(t)$, to the machine, and estimate the output matrix according to Eq. (3). In this application, the optimized hyperparameters of the RC are $(D_r, p, \sigma, \lambda, \eta) = (600, 0.1, 1, 0.1, 1 \times 10^{-7})$. Finally, we tune the control parameter ε to different values, and calculate from the machine outputs the variation of the system dynamics with respect to ε. The predicted results are plotted in Fig. 3(d), in which x_1^m denotes the local maximum or minimum of the variable x_1 during the course of system evolution. Comparing with the results of model simulation shown in Fig. 3(c), we see that both the location of the transition point ε_c and the oscillating amplitudes in the oscillatory regime are well predicted by the machine.

C. Complete synchronization in coupled chaotic oscillators

We finally employ the scheme of parameter-aware RC to predict the transition to complete synchronization in coupled chaotic oscillators. As a universal concept in nonlinear science, synchronization has been extensively studied in literature for decades [61]. Briefly, synchronization refers to the coherent motion of coupled oscillators, which occurs normally when the interacting strength of the oscillators is larger to a critical value. Depending on the degree of the

correlation, the oscillators could be synchronized in different forms, such as complete synchronization, phase synchronization, and generalized synchronization [62]. Among them, complete synchronization has the strongest correlation, as in complete synchronization the states of the oscillators are identical during the time course of system evolution [42]. Recently, stimulated by the discoveries of the small-world and scale-free features in many realistic systems, synchronization behaviors in complex networks have received considerable attention [63].

In the study of complete synchronization, a typical model employed in literature is an ensemble of identical chaotic oscillators coupled through a linear function (i.e., the diffusive coupling), and one of the central tasks is to predict, theoretically or numerically, the critical coupling strength for generating synchronization [61]. When the dynamical equations of the system are available, the critical coupling in many cases can be predicted theoretically, e.g., by the method of master-stability function [64,65]. Yet in realistic situations the dynamical equations are mostly unknown, which means that the prediction should be made based on solely the time series measured from the system evolutions. One approach to address this question is to reconstruct first the system dynamics by techniques such as compressive sensing, and then estimate the critical coupling by model simulations [66]. A different approach proposed recently to address this question is exploiting RC, in which the synchronization error between the oscillators under arbitrary coupling strength can be calculated directly from the machine outputs [33]. In these studies, a common requirement is that the time series must be measured from the desynchronization states. This requirement, however, is not met in many physical and biological systems, e.g., the power grids and the circadian clock, whose normal functions rely on the synchronized motion of the dynamical units. Assuming that these systems are already operating on the stable synchronous states, i.e., the systems will restore to the synchronous states quickly after perturbations, the question we ask here is: Given the transient time series and the coupling parameter of several states in the synchronization regime, can we predict whether the system will be desynchronized if a small drift is introduced to the coupling parameter? We are going to demonstrate that for the typical systems of coupled chaotic oscillators, the prediction can be accomplished by the scheme of parameter-aware RC introduced in Sec. II.

We consider first the model of two coupled chaotic Logistic maps. The system dynamics reads

$$x(n+1)_{1,2} = F[x(n)_{1,2}] + \varepsilon[H(x(n)_{2,1}) - H(x(n)_{1,2})], \quad (7)$$

where $x(n)$ denotes the state of the map at the nth iteration, $F(x) = 4x(1-x)$ is the mapping function describing the local dynamics, $H(x) = F(x)$ is the coupling function, and ε is the coupling strength. The initial states of the maps are chosen randomly from the interval (0,1). When the system equations are available, the critical coupling for synchronization can be analyzed by the method of master stability function [64,65], which shows that complete synchronization occurs in a bounded region in the parameter space, $\varepsilon \in (\varepsilon_1, \varepsilon_2)$, with $\varepsilon_1 = 0.25$ and $\varepsilon_2 = 0.75$. To generate the training data from the synchronization regime, we choose $\varepsilon = 0.26, 0.28$, and 0.30 as the sampling parameters.
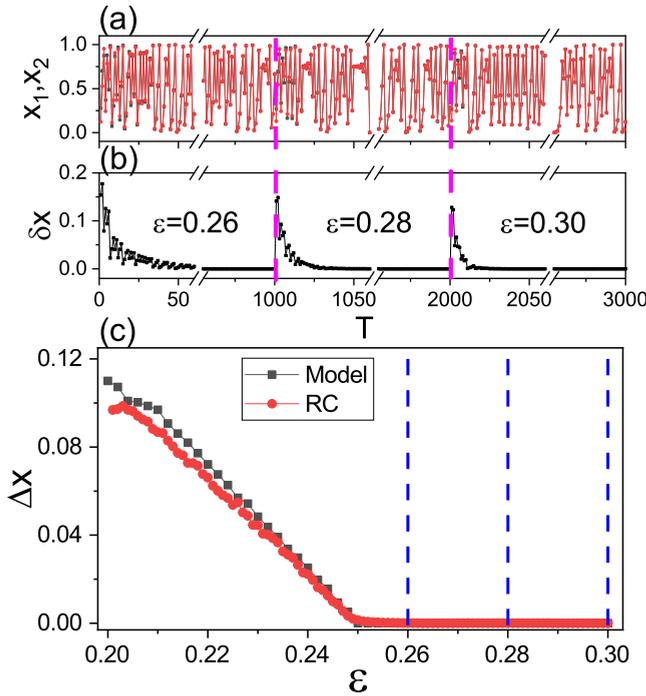
FIG. 4. Predicting synchronization transition in two coupled chaotic Logistic maps. (a) The training time series, which is generated by combining the time series of three sampling states in the synchronization regime, $\varepsilon = 0.26$, 0.28, and 0.30. (b) The time evolution of the synchronization error $\delta x = |x_1 - x_2|$ for the sampling states. (c) The variation of the time-averaged synchronization error Δx with respect to ε around the critical point $\varepsilon_1 = 0.25$. Black squares are the results obtained by model simulations. Red discs are the results predicted by the machine. Vertical dashed lines denote the locations of the sampling states.

For each sampling parameter, we record the system state, $\mathbf{u}(t) = [x_1, x_2]^T$, for a period of $T = L' = 1 \times 10^3$ iterations. The combination of the three time series forms the training series, which is plotted in Fig. 4(a). Figure 4(b) plots the time evolution of the synchronization error, $\delta x = |x_1 - x_2|$, for different parameters. We see that in all three cases the value of δx is decreased to 0 after a short transient ($T < 50$). Figure 4(b) shows also that with the increase of ε, the lifetime of the transient is shortened. The shortened transient at larger coupling strengths is attributed to the decreased conditional Lyapunov exponent. Specifically, as ε increases from ε_c in the synchronization regime, the Lyapunov exponent characterizing the stability of the dynamics in the phase space transverse to the synchronous manifold will be decreased from 0 gradually, while the more negative is the exponent, the quicker will be the system returning to the synchronous state and the shorter will be the transient lifetime.

The training series, together with the corresponding time series of the coupling strength $\varepsilon(n)$, are then fed into the machine for obtaining the output matrix. In this application, the optimized hyperparameters of the RC are $(D_r, p, \sigma, \lambda, \eta) = (100, 0.1, 1, 1 \times 10^{-5}, 1 \times 10^{-7})$. In obtaining the output matrix, a short segment of $\tau = 5$ data points in each sampling series is used to drive the reservoir out of the transient states. In the predicting phase, we set ε as the control parameter

and tune it from 0.30 to 0.20 by the step $\Delta\varepsilon = 2 \times 10^{-3}$. For each value of ε, we first run the machine for a transient period of $T_0' = 1 \times 10^3$ iterations in order of removing the impact induced by the initial conditions, and then calculate from the machine outputs the time-averaged synchronization error $\Delta x = \langle |x_1 - x_2| \rangle_T$, with $T = 1 \times 10^4$. The variation of Δx with respect to ε is plotted in Fig. 4(c), which shows that the two maps are desynchronized at the critical coupling $\varepsilon_c \approx$ 0.25 and, as ε decreases from ε_c, the value of Δx is gradually increased. To evaluate the prediction performance, we plot in Fig. 4(c) also the results obtained from model simulations. We see that the predicted results are in good agreement with the results obtained from model simulations.

We consider next the model of two coupled chaotic Lorenz oscillators. The system dynamics reads

$$\dot{x}_{1,2} = \alpha(y_{1,2} - x_{1,2}) + \varepsilon(x_{2,1} - x_{1,2}),$$
$$\dot{y}_{1,2} = x_{1,2}(\beta - z_{1,2}) - y_{1,2} + \varepsilon(y_{2,1} - y_{1,2}),$$
$$\dot{z}_{1,2} = x_{1,2}y_{1,2} - \gamma z_{1,2} + \varepsilon(z_{2,1} - z_{1,2}). \quad (8)$$

The parameters of the oscillators are $(\alpha, \beta, \gamma) = (10, 28, 2)$, by which the motion of isolated oscillator is chaotic. Analysis based on the method of master stability function shows that the two oscillators are synchronized when $\varepsilon > \varepsilon_c \approx$ 0.42 [64,65]. Still, our mission here is to predict the critical coupling ε_c based on the time series acquired at several states in the synchronization regime.

As the illustration, we choose $\varepsilon = 0.5$, 0.55, and 0.6 as the sampling parameters (all within the synchronization regime). In model simulations, the initial conditions of the oscillators are randomly chosen within the range $(-1, 1)$, and the system is evolved according to Eq. (8) by the time step $\Delta t = 0.02$. For each sampling parameter, we record the system state, $\mathbf{u}(t) = [x_1, y_1, z_1, x_2, y_2, z_2]^T$, for a period of $T = 80$. The combination of the recorded time series forms the training series, as depicted in Fig. 5(a) for the variables x_1 and x_2. The time evolutions of the synchronization error $\delta x = |x_1 - x_2|$ under different parameters are plotted in Fig. 5(b). We see that in each case the synchronization error is damped to 0 after a short transient ($T < 60$). The training series and the corresponding time series of the coupling strength are then fed into the machine for estimating the output matrix. Here, a short segment of $\tau = 50$ data points at the beginning of each sampling series is used to drive the reservoir out of the transient states. In this application, the hyperparameters of the RC are $(D_r, p, \sigma, \lambda, \eta) = (2 \times 10^3, 0.2, 0.05, 0.1, 1 \times 10^{-7})$.

To predict the critical coupling by the scheme of parameter-aware RC, we set ε as the control parameter and decrease it from 0.6 to 0.3 by the step $\Delta\varepsilon = 0.1$. For each value of ε, we first run the machine for a transient period of $T_0' = 1 \times 10^3$, and then calculate from the machine outputs the time averaged synchronization error $\Delta x = \langle |x_1 - x_2| \rangle_T$, with $T = 1 \times 10^3$. The variation of Δx with respect of ε is plotted in Fig. 5(c), which shows that the system is desynchronized at the critical coupling $\varepsilon_c \approx 0.43$. To check the performance of the predictions, we plot in Fig. 5(c) also the results obtained from model simulations. We see that the predicted results agree with the results of model simulations reasonably well.

For the case of synchronization transition, additional simulations have been also conducted to check the impacts of the
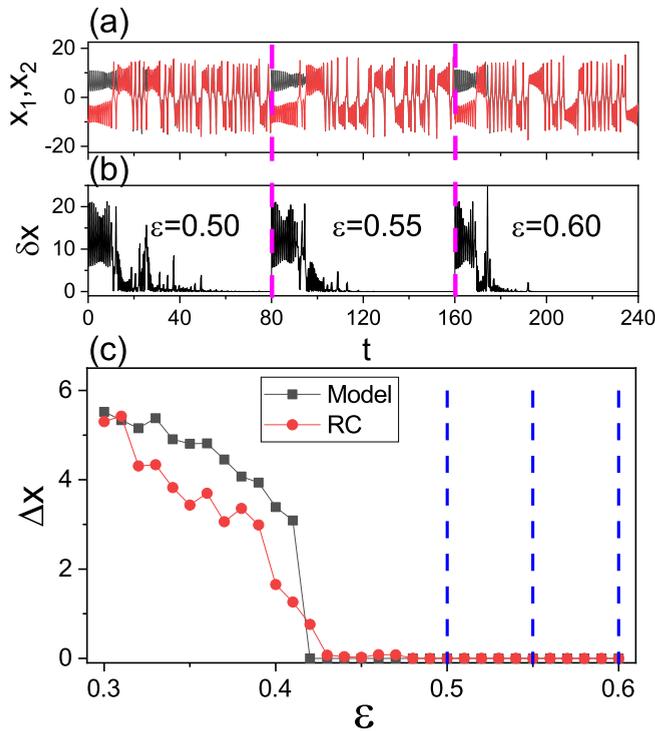
FIG. 5. Predicting synchronization transition in two coupled chaotic Lorenz oscillators. (a) The training series of the variables x_1 and x_2, which is generated by combining the time series of three sampling states in the synchronization regime, $\varepsilon = 0.5, 0.55$, and 0.6. (b) The time evolution of the synchronization error $\delta x = |x_1 - x_2|$ for the sampling states. (c) The variation of the time-averaged synchronization error Δx with respect to ε around the critical point $\varepsilon_c \approx 0.42$. Black squares are the results obtained from model simulations. Red discs are the results predicted by the machine. Vertical dashed lines denote the sampling parameters.

sampling states on the prediction performance (not shown). It is found that for the fixed number of sampling states, the prediction performance becomes worse when choosing the larger sampling parameters. In particular, it is found that for the model of coupled Logistic maps, the machine fails to predict the second transition point at $\varepsilon_2 \approx 0.75$, despite the number and location of the sampling parameters. A close look to the transient behaviors of the system nearby ε_2 shows that the failure is due to the ultrashort transient in approaching synchronization. In specific, starting from the random initial conditions, the synchronization error between the two coupled maps is decreased to 0 after just several iterations (~ 5). (Numerically, it is found that for the machine to be properly trained and is able to predict the transition point, the lifetime of the transient dynamics should be not less than $T \approx 15$ iterations.) For such a short transient, the machine is not able to get from the training data sufficient information about the system dynamics, and therefore is not able to predict the transition. This finding is consistent with the findings in predicting the quorum-sensing transition, where it is shown that the prediction performance is deteriorated gradually as the lifetime of the transient is shortened [Figs. 2(c) and 2(d)].

IV. DISCUSSION AND CONCLUSION

In exploiting RC for predicting chaotic systems, two general requirements are that (1) the measured time series should reflect the actual dynamics of the target system and (2) the measured time series should be sufficiently long. For complex systems of coupled oscillators, when the system dynamics is degenerated to a low-dimensional manifold in the phase space, e.g., a fixed point (the cases of quorum sensing and amplitude death) or the synchronous state (the case of complete synchronization), many information of the system dynamics are missed in the asymptotic dynamics. For instance, when two oscillators are quenched to the origin, we get no information about the system dynamics from the constant time series. This is one of the reasons why in previous papers of phase-transition prediction the training data are all collected from the oscillatory regime of high-dimensional dynamics [32–34]. Different from the previous papers, in the current paper we predict the transition in the opposite direction, e.g., predicting the transition by the data collected from the regime of degenerated, low-dimensional dynamics. The key lies in the transient activities preceding the asymptotic dynamics. We have demonstrated in different models that the transient activities, although of short lifetimes, reflect the actual dynamics of the target system and, by the scheme of parameter-aware RC, can be exploited to predict the critical point of phase transition. In systems of coupled-limit cycles, we have shown that by the transient activities acquired in the quiescent regime in which the asymptotic dynamics is represented by a fixed point, the trained machine is able to predict not only the transition point, but also the system dynamics in the oscillatory regime (e.g., the variation of the order parameters); in system of coupled chaotic oscillators, we have shown that the transient dynamics in the synchronization regime can be exploited to predict not only the critical point for synchronization, but also the behavior of the synchronization error in the desynchronization regime. In both cases, it is found that the prediction performance is strongly affected by the transients, i.e., by increasing the lifetime of the transients, the prediction performance will be gradually improved. Considering the facts that many real-world systems are operating at stable, low-dimensional states and transient activities are ubiquitously observed in these systems, the findings may have broad applications, e.g., predicting the tipping points of climate systems [11] and identifying the complex structure of ecological networks [63].

Transient activities could be also exploited for the inference of multiple coexisting attractors. An important feature of many nonlinear system is that the asymptotic dynamics of the system is dependent on its initial state, i.e., the problem of attracting basins. This poses a new challenge in machine learning of chaotic dynamics, as the machine trained by the time series of one attractor is not able to infer the dynamics of other attractors in the phase space. To cope with this problem, in Ref. [30] the authors propose the idea of training the machine by a long time series composed of many short series acquired from different attractors, and then in the predicting phase use a short transient sequence to drive the machine to the desired attractor. Another approach for learning multiple attractors is exploiting noisy time series [68]. In this approach,

the training data is acquired from the noise-perturbed trajectory of a specific attractor. It is shown that the machine, if properly trained, is able to infer the dynamics of other coexisting attractors in the phase space that are not exposed to the machine in the training phase, namely the "unseen" attractors. As did in Ref. [30], to infer the "unseen" attractors in the predicting phase, a short transient sequence is needed to "washout" the reservoir. In the approach of noisy trajectory, the role of noise perturbations is preventing the system from settling onto a specific attractor, i.e., keeping the system at transient permanently, so as to explore the global information of the system dynamics. Our current paper is different from these papers in the following aspects. First, the length of the training data is different. In Refs. [30,68], the training data is hundreds or thousands system oscillations; in our present paper the training data sustains for only several system oscillations. Second, the driving scheme is different. In Refs. [30,68], the inference of the new attractors is guided by a transient sequence, whereas in our present paper the reservoir dynamics is guided by a constant control parameter. Lastly, the mission is different. The task in Refs. [30,68] is to infer the coexisting attractors in the phase space, whereas in our present paper the mission is to infer the critical point for phase transition. Despite the differences, all studies point to the same fact: transient activities provide global information about the system dynamics. This understanding is believed to be helpful to the machine learning of multiple attractors.

While our studies show preliminarily the feasibility of predicting phase transition based on the time series of transients, many questions remain open. One is about the generality of the findings. For illustrative purposes, we have employed the model of coupled oscillators to demonstrate the predictions, it is not clear whether the similar results can be found in other type of models, e.g., systems showing the phenomenon of supertransient [67]. Another question is about the connection between the transient lifetime and the prediction performance. Our studies suggest that, to predict the transition point accurately, the sampling parameters should be close to the critical point. Normally, as the sampling parameter leaves away from the critical point, the lifetime of the transient will be decreased by an algebraic scaling law [4]. When the sampling parameter is too far away from the critical point, the lifetime of the transient will be too short to train the machine, leading to the failed prediction. This phenomenon has been observed in predicting the second critical coupling ε_2 in the synchronization of coupled chaotic Logistic maps. In this case, the change of the initial condition of the maps does not extend the lifetime, and new methods must be proposed for training the machine. One approach to tackle this question could be generating a longer sequence by combing the transients of several system realizations; another approach could be adopting noisy trajectory (i.e., generating the training series by a stochastic dynamics) [68]. A full exploration on the dependence of machine prediction performance on transient lifetime is out the scope of the current paper. It is our hope that this question can be addressed by further studies.

To summarize, exploiting the technique of RC in machine leaning, we have shown that the critical point of phase transition in coupled oscillators can be predicted by the time series of transient activities in the stable regime. The predictions are demonstrated in three different models, including an ensemble of indirectly coupled periodic oscillators showing quorum-sensing transition, two coupled nonidentical oscillators showing amplitude-death transition, and two coupled chaotic oscillators showing complete-synchronization transition. In all cases, it is found that the machine trained by the transient dynamics at several states in the stable regime is able to predict not only the transient evolution in the stable regime, but also the critical point where the system becomes unstable. Our paper shows the feasibility of training RC by transient dynamics, which provides an alternative approach for the model-free prediction of phase transition in complex dynamical system, and shed lights on the learning of coexisting attractors and attracting basins by machine.

[1] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 2002).

[2] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, Boca Raton, FL, 2015).

[3] T. Tél and Y.-C. Lai, Chaotic transients in spatially extended systems, Phys. Rep. **460**, 425 (2008).

[4] Y.-C. Lai and T. Tél, *Transient Chaos - Complex Dynamics on Finite Time Scales* (Springer, New York, 2011).

[5] A. Hastings, K. C. Abbott, K. Cuddington, T. Francis, G. Gellner, Y.-C. Lai, A. Morozov, S. Petrivskii, K. Scranton, and M. L. Zeeman, Transient phenomena in ecology, Science **361**, 6406 (2018).

[6] H. Fan, L.-W. Kong, X. Wang, A. Hastings, and Y.-C. Lai, Synchronization within synchronization: Transients and intermittency in ecological networks, Nati. Sci. Rev. **8**, nwaa269 (2020).

[7] S. M. Courtney, L. G. Ungerleider, K. Keil, and J. V. Haxby, Transient and sustained activity in a distributed neural system for human working memory, Nature (London) **386**, 608 (1997).

[8] O. Mazor and G. Laurent, Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons, Neuron **48**, 661 (2005).

[9] M. Rabinovich, R. Huerta, and G. Laurent, Transient dynamics for neural processing, Science **321**, 48 (2008).

[10] V. V. Kharin and F. W. Zwiers, Estimating extremes in transient climate change simulations, J. Clim. **18**, 1156 (2005).

[11] J. Fan, J. Meng, J. Ludescher, X. Chen, Y. Ashkenazy, J. Kurths, S. Havlin, and H. J. Schellnhuber, Statistical physics approaches to the complex earth system, Phys. Rep. **896**, 1 (2021).

[12] A. E. Motter, Cascade Control and Defense in Complex Networks, Phys. Rev. Lett. **93**, 098701 (2004).

[13] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing, Transient Dynamics Increasing Network Vulnerability to Cascading Failures, Phys. Rev. Lett. **100**, 218701 (2008).

[14] L. Halekotte, A. Vanselow, and U. Feudel, Transient chaos enforces uncertainty in the British power grid, J. Phys. Complex. **2**, 035015 (2021).

[15] K. Inoue, K. Nakajima, and Y. Kuniyoshi, Designing spontaneous behavioral switching via chaotic itinerancy, Sci. Adv. **6**, eabb3989 (2020).

[16] S. Kitsunai, W. Cho, C. Sano, S. Saetia, Z. Qin, Y. Koike, M. Frasca, N. Yoshimura, and L. Minati, Generation of diverse insect-like gait patterns using networks of coupled Rössler systems, Chaos **30**, 123132 (2020).

[17] C. Grebogi, E. Ott, and J. Yorke, Crises, sudden changes in chaotic attractors, and transient chaos, Physica D **7**, 181 (1983).

[18] K. Kaneko and I. Tsuda, Chaotic itinerancy, Chaos. **13**, 926 (2003).

[19] Y.-C. Lai and R. L. Winslow, Geometric Properties of the Chaotic Saddle Responsible for Supertransients in Spatiotemporal Chaotic Systems, Phys. Rev. Lett. **74**, 5208 (1995).

[20] E. L. Rempel and A. C.-L. Chian, Origin of Transient and Intermittent Dynamics in Spatiotemporal Chaotic Systems, Phys. Rev. Lett. **98**, 014101 (2007).

[21] A. Morozov, K. Abbott, K. Cuddington, T. Francis, G. Gellner, A. Hastings, Y.-C. Lai, S. Petrovskii, K. Scranton, and M. L. Zeeman, Long transients in ecology: Theory and applications, Phys. Life Rev. **32**, 1 (2020).

[22] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Predicting Catastrophes in Nonlinear Dynamical Systems by Compressive Sensing, Phys. Rev. Lett. **106**, 154101 (2011).

[23] W.-X. Wang, Y.-C. Lai, and C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, Phys. Rep. **644**, 1 (2016).

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

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

[26] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, Reservoir observers: Model-free inference of unmeasured variables in chaotic systems, Chaos **27**, 041102 (2017).

[27] J. Pathak, Z. Lu, B. Hunt, M. Girvan, and E. Ott, Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data, Chaos **27**, 121102 (2017).

[28] 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).

[29] H. Fan, J. Jiang, C. Zhang, X. G. Wang, and Y.-C. Lai, Long-term prediction of chaotic systems with machine learning, Phys. Rev. Research **2**, 012080(R) (2020).

[30] Z. Lu and D. S. Bassett, Invertible generalized synchronization: A putative mechanism for implicit learning in neural systems, Chaos **30**, 063133 (2020).

[31] A. Flynn, V. A. Tsachouridis, and A. Amann, Multifunctionality in a reservoir computer, Chaos **31**, 013125 (2021).

[32] L.-W. Kong, H. Fan, C. Grebogi, and Y.-C. Lai, Machine learning prediction of critical transition and system collapse, Phys. Rev. Research **3**, 013090 (2021).

[33] H. Fan, L.-W. Kong, Y.-C. Lai, and X. G. Wang, Anticipating synchronization with machine learning, Phys. Rev. Research **3**, 023237 (2021).

[34] R. Xiao, L.-W. Kong, Z.-K. Sun, and Y.-C. Lai, Predicting amplitude death with machine learning, Phys. Rev. E **104**, 014205 (2021).

[35] Y. L. Guo, H. Zhang, L. Wang, H. Fan, J. H. Xiao, and X. G. Wang, Transfer learning of chaotic systems, Chaos **31**, 011104 (2021).

[36] H. Zhang, H. Fan, L. Wang, and X. G. Wang, Learning Hamiltonian dynamics with reservoir computing, Phys. Rev. E **104**, 024205 (2021).

[37] D. Patel, D. Canaday, M. Girvan, A. Pomerance, and E. Ott, Using machine learning to predict statistical properties of nonstationary dynamical processes: System climate, regime transitions, and the effect of stochasticity, Chaos **31**, 033149 (2021).

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

[39] K. Nakajima and I. Fischer, *Reservoir Computing: Theory, Physical Implementations, and Applications* (Springer, Singapore, 2021).

[40] A. F. Taylor, M. R. Tinsley, F. Wang, Z. Huang, and K. Showalter, Dynamical quorum sensing and synchronization in large populations of chemical oscillators, Science **323**, 614 (2009).

[41] R. E. Mirollo and S. H. Strogatz, Amplitude death in an array of limit-cycle oscillators, J. Stat. Phys. **60**, 245 (1990).

[42] L. M. Pecora and T. L. Carroll, Synchronization in Chaotic Systems, Phys. Rev. Lett. **64**, 821 (1990).

[43] M. B. Miller, and B. L. Bassler, Quorum sensing in bacteria, Annu. Rev. Microbiol. **55**, 165 (2001).

[44] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing, Proc. Natl. Acad. Sci. U.S.A. **101**, 10955 (2004).

[45] S. H. Strogatz, D. M. Abrams, A. McRobie, B. Eckhardt, and E. Ott, Crowd synchrony on the millennium bridge, Nature (London) **438**, 43 (2005).

[46] D. J. Schwab, A. Baetica, and P. Mehta, Dynamical quorum-sensing in oscillators coupled through an external medium, Physica D **241**, 1782 (2012).

[47] B.-W. Li, C. Fu, H. Zhang, and X. Wang, Synchronization and quorum sensing in an ensemble of indirectly coupled chaotic oscillators, Phys. Rev. E **86**, 046207 (2012).

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

[49] D. G. Aronson, G. B. Ermentrout, and N. Kopell, Amplitude response of coupled oscillators, Physica D **41**, 403 (1990).

[50] W. Zou, D. V. Senthilkumar, M. Zhan, and J. Kurths, Quenching, aging, and reviving in coupled dynamical networks, Phys. Rep. **931**, 1 (2021).

[51] D. V. Ramana Reddy, A. Sen, and G. L. Johnston, Time Delay Induced Death in Coupled Limit Cycle Oscillators, Phys. Rev. Lett. **80**, 5109 (1998).

[52] T. Singla, N. Pawar, and P. Parmananda, Exploring the dynamics of conjugate coupled Chua circuits: Simulations and experiments, Phys. Rev. E **83**, 026210 (2011).

[53] K. Konishi, Amplitude death induced by dynamic coupling, Phys. Rev. E **68**, 067202 (2003).

[54] Z. Hou and H. Xin, Oscillator death on small-world networks, Phys. Rev. E **68**, 055103(R) (2003).

[55] J. Yang, Transitions to amplitude death in a regular array of nonlinear oscillators, Phys. Rev. E **76**, 016204 (2007).

[56] W. Q. Liu, X. G. Wang, S. G. Guan, and C. H. Lai, Transition to amplitude death in scale-free networks, New J. Phys. **11**, 093016 (2009).

[57] A. Babloyantz, Self-organization phenomena resulting from cell-cell contact, J. Theoret. Biol. **68**, 551 (1977).

[58] D. P. Zipes and H. J. Wellens, Sudden cardiac death, Circulation **98**, 2334 (1998).

[59] A. Prasad, Y.-C. Lai, A. Gavrielides, and V. Kovanis, Amplitude modulation in a pair of time-delay coupled external-cavity semiconductor lasers, Phys. Lett. A **318**, 71 (2003).

[60] R. M. May, *Stability and Complexity in Model Ecosystems* (Princeton University Press, Princeton, 1974).

[61] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Science* (Cambridge University Press, Cambridge, 2003).

[62] S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. S. Zhou, The synchronization of chaotic systems, Phys. Rep. **366**, 1 (2002).

[63] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. S. Zhou, Synchronization in complex networks, Phys. Rep. **469**, 93 (2008).

[64] L. M. Pecora and T. L. Carroll, Master Stability Functions for Synchronized Coupled Systems, Phys. Rev. Lett. **80**, 2109 (1998).

[65] L. Huang, Q. Chen, Y.-C. Lai, and L. M. Pecora, Generic behavior of master-stability functions in coupled nonlinear dynamical systems, Phys. Rev. E **80**, 036204 (2009).

[66] R.-Q. Su, X. Ni, W.-X. Wang, and Y.-C. Lai, Forecasting synchronizability of complex networks from data, Phys. Rev. E **85**, 056220 (2012).

[67] C. Grebogi, E. Ott, and J. A. Yorke, Fractal Basin Boundaries, Long-Lived Chaotic Transients, and Unstable-Unstable Pair Bifurcation, Phys. Rev. Lett. **50**, 935 (1983).

[68] A. Röhm, D. J. Gauthier, and I. Fischer, Model-free inference of unseen attractors: Reconstructing phase space features from a single noisy trajectory using reservoir computing, Chaos **31**, 103127 (2021).