

Unifying framework for information processing in stochastically driven dynamical systems

Tomoyuki Kubota,* Hirokazu Takahashi, and Kohei Nakajima

Graduate School of Information Science and Technology, The University of Tokyo, Japan



(Received 26 May 2021; accepted 27 October 2021; published 23 November 2021)

A dynamical system is an information processing apparatus that encodes input streams from the external environment to its state and processes them through state transitions. The information processing capacity (IPC) is an excellent tool that comprehensively evaluates these processed inputs, providing details of unknown information processing in black box systems; however, this measure can be applied only to time-invariant systems. This paper extends the applicable range to time-variant systems and further reveals that the IPC is equivalent to coefficients of polynomial chaos (PC) expansion in more general dynamical systems. To achieve this objective, we tackle three issues. First, we establish a connection between the IPC for time-invariant systems and PC expansion, which is a type of polynomial expansion using orthogonal functions of input history as bases. We prove that the IPC corresponds to the squared norm of the coefficient vector of the basis in the PC expansion. Second, we show that an input following an arbitrary distribution can be used for the IPC, removing previous restrictions to specific input distributions. Third, we extend the conventional orthogonal bases to functions of both time and input history and propose the IPC for time-variant systems. To show the significance of our approach, we demonstrate that our measure can reveal information representations in not only machine learning networks but also a real, cultured neural network. Our generalized measure paves the way for unveiling the information processing capabilities of a wide variety of physical dynamics which have been left behind in nature.

DOI: [10.1103/PhysRevResearch.3.043135](https://doi.org/10.1103/PhysRevResearch.3.043135)

I. INTRODUCTION

Dynamical systems driven by external stimuli can be universally found in nature, especially in biology. The dynamical aspects of information processing found in biology have long been a source of inspiration for researchers who wish to create a high-speed, energy efficient, and robust real-time information processing device, which resolves a von Neumann bottleneck [1]. Reservoir computing (RC) [2–5] is a bioinspired information processing paradigm that capitalizes on this dynamical perspective and has been widely utilized in various fields in recent years [6–13]. It consists of a type of learning framework for recurrent neural networks (RNNs), whose intermediate layer is referred to as the reservoir. Let the number of nodes in a reservoir and its external input be N and u_t , respectively, and the i th ($i = 1, 2, \dots, N$) node state $x_{i,t}$ at the t th time step updates as follows:

$$x_{i,t+1} = f\left(\sum_{j=1}^N w_{ij}x_{j,t} + w_{in,i}u_t\right), \quad (1)$$

where f is the activation function, and w_{ij} and $w_{in,i}$ are the internal and input weights, respectively. To emulate the target

output z_t , we use linear regression to obtain an estimate of z_t , \tilde{z}_t , as follows:

$$\tilde{z}_t = \tilde{\mathbf{w}}_{\text{out}}^{\top} \cdot \mathbf{x}_t, \quad (2)$$

$$\tilde{\mathbf{w}}_{\text{out}} = \arg \min_{\tilde{\mathbf{w}}_{\text{out}}} \sum_{t=1}^T (z_t - \tilde{\mathbf{w}}_{\text{out}}^{\top} \cdot \mathbf{x}_t)^2, \quad (3)$$

where $\mathbf{x}_t = [x_{1,t} \cdots x_{N,t}]^{\top}$ and \mathbf{w}_{out} and $\tilde{\mathbf{w}}_{\text{out}} \in \mathbb{R}^N$ are the weight and solution vector for the target output, respectively. z_t is a function of certain variables, which are expected to include the delayed inputs $\{u_{t-1}, u_{t-2}, \dots\}$. If the state also depends on the input history, the linear combination of states emulates the target well. Therefore this learning method places a constraint on \mathbf{x}_t . RC was proposed by integrating an echo state network (ESN) and a liquid state machine (LSM), which were independently developed by Jaeger [2] and Maass [3], respectively. The prerequisites for \mathbf{x}_t differ between the ESN and the LSM.

The ESN is often represented by using a hyperbolic tangent activation function and fixed internal weights, as in Eq. (1). It requires \mathbf{x}_t to be an echo function, which is a function of only the past input time series $\mathbf{u}_t = \{u_{t-s}\}_{s=1}^t$. This dynamical property is referred to as the echo state property (ESP) [14–16]. We can examine this feature by giving the same input time series to two systems with different initial values and by checking whether the two states coincide after a sufficiently long time. If the input is noise, the resultant phenomenon is called noise-induced synchronization [17,18]. In the case where the input is generated from a deterministic system, the phenomenon, wherein the state is synchronized with the input,

*kubota@ai.u-tokyo.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

is called generalized synchronization [19,20]. Thus the ESP is related to the synchronization phenomenon of nonlinear dynamical systems.

In addition, the LSM uses a spiking neuron model as an activation function and its internal weights are updated by a synaptic plasticity rule (e.g., a leaky integrate-and-fire model with short-term plasticity). It imposes a prerequisite on the power series expansion of states. If the system is time-invariant—the state does not depend on time—and retains decaying inputs, its state can be approximated by the Volterra series [21], which is a series expansion with nonorthogonal bases. Accordingly, the Volterra series operator [22] can be expanded in a nonorthogonal power series basis as follows:

$$\mathbf{x}_t = \sum_{n=1}^{\infty} \prod_{s_1=1}^{\infty} \prod_{s_2=s_1}^{\infty} \cdots \prod_{s_n=s_{n-1}}^{\infty} \mathbf{h}_{s_1 \cdots s_n} u_{t-s_1} \cdots u_{t-s_n},$$

where $\mathbf{h}_{s_1 \cdots s_n} \in \mathbb{R}^N$ is the n th Volterra kernel. Such a memory decay feature is called the fading memory property (FMP) [3,23,24] and has been considered a prerequisite for LSMs.

The above properties suggest that \mathbf{x}_t is limited to states that are independent of time and hold past input time series. As long as the reservoir satisfies these prerequisites, various activation functions can be used for the reservoir node. Furthermore, as the reservoir is not limited to a computer-generated system in Eq. (1), it can be replaced with a real physical system—for example, quantum systems [25–27], analog circuits [6], opto-electronic architectures [28], carbon nanotubes [29], soft robotic systems [30–32], and other dynamical systems [26,33–37]. A reservoir using such a physical system is called a physical reservoir [38] and its scheme is termed physical reservoir computing. These schemes leverage not only RNNs but also physical systems as computational resources for machine learning.

In the literature, a measure called information processing capacity (IPC) [39] has been proposed to quantify the information processing capability of dynamical systems that have ESPs or FMPs. The IPC measures the type and quantity of input history that is handled and held in the system by decomposing the system state into an orthogonal basis [40–42]. Using the N -dimensional state $\mathbf{x}_t \in \mathbb{R}^N$ and the one-dimensional stochastic input $\zeta_t \in \mathbb{R}$ at the t th ($t \in \mathbb{Z}$) time step, the input-driven dynamical system (IDS) determines the next state, as shown below:

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \zeta_t), \tag{4}$$

where \mathbf{f} maps $\mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$. According to the ESP and FMP, \mathbf{x}_t is a function of input history $\{\zeta_{t-1}, \zeta_{t-2}, \dots\}$. If the system holds delayed, processed inputs (i.e., an expanded state \mathbf{x}_t consists of polynomials such as ζ_{t-1} , ζ_{t-1}^2 , and $\zeta_{t-1}\zeta_{t-2}$), we should be able to emulate the polynomials by the linear regression of \mathbf{x}_t . The IPCs evaluate the emulation abilities of the multivariate polynomials of $\{\zeta_{t-1}, \zeta_{t-2}, \dots\}$, which are orthogonal from each other to avoid overlapping abilities.

The multivariate polynomial is obtained from a product of the univariate polynomials. Let the n th-order univariate orthogonal polynomial of ζ be $\mathcal{F}_n(\zeta)$, which is determined by the distribution of ζ —for example, a Legendre polynomial of a uniform random variable and a Hermite polynomial

of a Gaussian random variable are proposed [39]. Using the univariate polynomials, the i th target polynomial $z_t^{(i)}$ ($i = 1, 2, \dots$) is represented by the product of the $n_s^{(i)}$ -th order univariate polynomials of the random variable delayed by s steps, ζ_{t-s} ,

$$z_t^{(i)} = \prod_{s=1}^{\infty} \mathcal{F}_{n_s^{(i)}}(\zeta_{t-s}), \tag{5}$$

where i denotes the index for an infinite family of sets of degrees and delays $\{\{n_s^{(i)}, s\}\}_{s=1}^{\infty}$. As shown in Table I, the infinite set is truncated to numerically compute the IPCs using a maximum delay s_{\max, d_i} for each degree of $z_t^{(i)}$, $d_i = \sum_s n_s^{(i)} (> 0)$. The example set uses the Legendre polynomials with $\{\{d_i, s_{\max, d_i}\}\} = \{\{1, 3\}, \{2, 2\}\}$.

From Eqs. (2) and (3), we obtain an estimate of $z_t^{(i)}$, $\hat{z}_t^{(i)}$. When z_t is an orthogonal function of the independent variables $\{\zeta_{t-s}\}_{s=1}^{\infty}$, the IPC is defined using a normalized emulation error of the reservoir, as follows:

$$\begin{aligned} C(\mathbf{X}, \mathbf{z}^{(i)}) &= 1 - \frac{\min_{\mathbf{w}} \sum_{t=1}^T (z_t^{(i)} - \mathbf{w}^\top \mathbf{x}_t)^2}{\sum_{t=1}^T (z_t^{(i)})^2} \\ &= \frac{\mathbf{z}^{(i)\top} \mathbf{X} \mathbf{X}^+ \mathbf{z}^{(i)}}{\mathbf{z}^{(i)\top} \mathbf{z}^{(i)}}, \end{aligned} \tag{6}$$

where $\mathbf{w} \in \mathbb{R}^N$, $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_T]^\top \in \mathbb{R}^{T \times N}$, and $\mathbf{z}^{(i)} = [z_1^{(i)} \cdots z_T^{(i)}]^\top \in \mathbb{R}^T$ are the weight vector, state, and target output, respectively, while $\mathbf{X}^+ \in \mathbb{R}^{N \times T}$ is the Moore-Penrose inverse of \mathbf{X} . Therefore the IPC is a measure used to evaluate the input information held by the state with the emulation ability of the orthogonal basis.

In this connection, there is a theory about a deterministic dynamical system with a stochastic input ζ_t in a different context. The system can be described as an operator of $\{\zeta_{t-s}\}_{s=1}^{\infty}$ according to the polynomial chaos expansion [43], which is a series expansion using the target outputs of IPC as the bases—i.e., multivariate orthogonal polynomials of the random variables, $z_t^{(i)}$, described in Eq. (5). The polynomial chaos expansion has been frequently utilized to determine the evolution of uncertainty in a dynamical system when there is probabilistic uncertainty in the system parameters [44,45]. These multivariate polynomials are referred to as polynomial chaoses (PCs), and the Hilbert space spanned by PCs is called homogeneous chaos [43,60] expressed as

$$\mathbf{x}_t = \sum_{i=1}^{\infty} \mathbf{c}_i z_t^{(i)}, \tag{7}$$

where $\mathbf{c}_i \in \mathbb{R}^N$ is the i th coefficient vector.

The PCs are obtained from the univariate polynomials and the Cameron-Martin theorem. The univariate polynomials are determined such that they are orthogonal. If a weighting function $w(\zeta_t)$ specific to ζ_t —which is the same as a probability density function of ζ_t for some polynomials—exists, the ensemble average of the inner product in the space

$$\langle \mathcal{F}_n(\zeta_t) \mathcal{F}_m(\zeta_t) \rangle = \sum_{\zeta_t} w(\zeta_t) \mathcal{F}_n(\zeta_t) \mathcal{F}_m(\zeta_t) (n, m = 0, 1, \dots) \tag{8}$$

TABLE I. An example set of orthogonal polynomials, which is truncated with $\{\{d, s_{\max,d}\}\} = \{\{1, 3\}, \{2, 2\}\}$. Note that $\mathcal{F}_n(\zeta)$ denotes the n th-order Legendre polynomial [$\mathcal{F}_0(\zeta) = 1, \mathcal{F}_1(\zeta) = \zeta, \text{ and } \mathcal{F}_2(\zeta) = (3\zeta^2 - 1)/2$].

i	Family of sets of degrees and delays $\{\{n_s^{(i)}, s\}\}$	Degree $d_i = \sum_s n_s^{(i)}$	Maximum delay s_{\max,d_i}	Polynomial $z_t^{(i)} = \prod_s \mathcal{F}_{n_s^{(i)}}(\zeta_{t-s})$
1	$\{\{1, 1\}, \{0, 2\}, \{0, 3\}\}$	1	3	ζ_{t-1}
2	$\{\{0, 1\}, \{1, 2\}, \{0, 3\}\}$	1	3	ζ_{t-2}
3	$\{\{0, 1\}, \{0, 2\}, \{1, 3\}\}$	1	3	ζ_{t-3}
4	$\{\{2, 1\}, \{0, 1\}\}$	2	2	$(3\zeta_{t-1}^2 - 1)/2$
5	$\{\{0, 1\}, \{2, 2\}\}$	2	2	$(3\zeta_{t-2}^2 - 1)/2$
6	$\{\{1, 1\}, \{1, 2\}\}$	2	2	$\zeta_{t-1}\zeta_{t-2}$

satisfies the following relation:

$$\langle \mathcal{F}_n(\zeta_t) \mathcal{F}_m(\zeta_t) \rangle = \begin{cases} \langle \mathcal{F}_n(\zeta_t) \mathcal{F}_n(\zeta_t) \rangle & (n = m) \\ 0 & (n \neq m) \end{cases} \quad (9)$$

in the case of discrete input. If ζ_t is continuous, Eq. (8) is replaced by $\langle \mathcal{F}_n(\zeta_t) \mathcal{F}_m(\zeta_t) \rangle = \int_S w(\zeta_t) \mathcal{F}_n(\zeta_t) \mathcal{F}_m(\zeta_t) d\zeta_t$, where S denotes the support of ζ_t . According to the Cameron-Martin theorem, the multivariate orthogonal polynomial is expected to be the product of univariate orthogonal polynomials as in Eq. (5) [46–48]. Let $\boldsymbol{\zeta}_t = \{\zeta_{t-1}, \zeta_{t-2}, \dots\}$ and represent the i th PC as $z_t^{(i)} = z^{(i)}(\boldsymbol{\zeta}_t)$. In the same manner as the univariate case, if a weighting function specific to $\boldsymbol{\zeta}_t$, $W(\boldsymbol{\zeta}_t)$, exists, the ensemble average of the inner product in the space

$$\langle z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \rangle = \sum_{\boldsymbol{\zeta}_t} W(\boldsymbol{\zeta}_t) z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \quad (10)$$

satisfies the orthogonality relations

$$\langle z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \rangle = \begin{cases} \langle z^{(i)}(\boldsymbol{\zeta}_t) z^{(i)}(\boldsymbol{\zeta}_t) \rangle & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (11)$$

in the case of discrete input of ζ_t . If ζ_t is continuous, Eq. (10) is replaced by $\langle z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \rangle = \int_{S \times S \times \dots} W(\boldsymbol{\zeta}_t) z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) d\zeta_{t-1} d\zeta_{t-2} \dots$, where S denotes the support of ζ_t .

The PC expansion can use various combinations of input and polynomials. Generalized polynomial chaos (gPC) [44] supplies eight combinations of input distribution and univariate polynomial included in the Askey scheme—e.g., the Hermite polynomial for Gaussian distribution, the Legendre polynomial for uniform distribution, and the Charlier polynomial for Poisson distribution are available as $\mathcal{F}_n(\zeta)$ (see Appendix A for details). We distinguish polynomial chaoses by the combination of the univariate polynomial name plus

TABLE II. Generalized polynomial chaoses.

Support of random variable	Random variable ζ_t	Polynomial chaos $z_t^{(i)}$
$(-\infty, \infty)$	Gaussian	Hermite chaos
$[0, \infty)$	Gamma	Laguerre chaos
$[a, b]$	Beta	Jacobi chaos
$[a, b]$	Uniform	Legendre chaos
$\{0, 1, \dots\}$	Poisson	Charlier chaos
$\{0, 1, \dots, N\}$	Binomial	Krawtchouk chaos
$\{0, 1, \dots\}$	Negative binomial	Meixner chaos
$\{0, 1, \dots, N\}$	Hypergeometric	Hahn chaos

chaos—e.g., Hermite, Legendre, and Charlier chaoses. The gPCs are summarized in Table II. Furthermore, arbitrary polynomial chaos (aPC) [45] denotes the PC for random variables following an arbitrary probability distribution. We can compute univariate polynomials by applying the Gram-Schmidt orthogonalization procedure. The n th Gram-Schmidt polynomial $\mathcal{F}_n(\zeta_t)$ is obtained from the following equations:

$$\mathcal{F}_n(\zeta_t) = \zeta_t^n - \sum_{i=0}^{n-1} c_i^{(n)} \mathcal{F}_i(\zeta_t), \quad (12)$$

$$c_i^{(n)} = \frac{\langle \mathcal{F}_i(\zeta_t) \zeta_t^n \rangle}{\langle \mathcal{F}_i(\zeta_t) \mathcal{F}_i(\zeta_t) \rangle}, \quad (13)$$

where $\mathcal{F}_0(\zeta) = 1$; $\mathcal{F}_n(\zeta)$ ($n = 0, 1, \dots$) satisfies Eq. (9). Using these polynomials and Eq. (5), we can calculate the polynomial chaos for an arbitrary input distribution, called Gram-Schmidt polynomial chaos, which is expected to satisfy the orthogonality in Eq. (11).

As previously described, the IPC and PC expansion have a number of similarities. Both schemes use the IDS with stochastic inputs and multivariate orthogonal polynomials, whereas the types of input distributions and orthogonal polynomials are not as limited in the IPC as they are in the PC. One objective of this paper is to establish a connection between the IPC and the PC-expanded IDS. Thus we first aim to reveal this relationship by deriving the IPC from the PC-expanded state. Second, to enlarge the applicable range of the IPC, we extend this relationship for time-variant systems. So far, the IPC has assumed that the system is a function of only the input time series. However, a solution of the dynamical system in Eq. (4) is a function of the input time series and time with a given initial state. By introducing time-dependent orthogonal bases to PCs, we aim to derive the IPC for time-variant systems and illustrate that input information processing is performed by products of a time-varying basis and PC. Finally, to demonstrate the potential of our approach, we apply our theory to three cases—a model frequently used as a benchmark task in the context of temporal machine learning, an artificial neural network, and a real, cultured neural network, which is a typical time-dependent physical reservoir—to reveal their information processing performed in the systems.

II. THEORETICAL FRAMEWORK

A. The equivalence of the IPC and coefficient in PC expansion

To show the relationship between the IPC and PC expansion, we derive the IPC from the state expanded in terms of

PCs. First, to prepare for the derivation, we define the inner product and norm for $\mathbf{x} = [x_1 \cdots x_M]^\top$ and $\mathbf{y} = [y_1 \cdots y_M]^\top$ as follows:

$$\begin{aligned} \overline{x_i y_i} &= \sum_{i=1}^M x_i y_i, \\ \|\mathbf{x}\| &= \sqrt{\overline{x_i x_i}}. \end{aligned}$$

Using this inner product, we assume the ergodicity in the homogeneous chaos—i.e., the inner product of the time average

$$\overline{z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t)} = \sum_{t=1}^T z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \quad (14)$$

and the ensemble average in Eq. (10) are equivalent as follows:

$$\langle z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t) \rangle = \overline{z^{(i)}(\boldsymbol{\zeta}_t) z^{(j)}(\boldsymbol{\zeta}_t)}. \quad (15)$$

This assumption allows the PCs in the time direction to be orthogonal bases in the homogeneous chaos.

Next, we transform the IPC into a simpler form using singular value decomposition (SVD), which reduces the N -state time series to $r(\leq N)$ -normalized, linearly independent time-series vectors $\mathbf{p}_j = [p_{j1} \cdots p_{jT}]^\top \in \mathbb{R}^T$ ($j = 1, \dots, r$)—i.e., $\overline{p_{ji} p_{ki}} = 0$ ($j \neq k$) and $\|\mathbf{p}_j\| = 1$. SVD breaks down state $\mathbf{X} \in \mathbb{R}^{T \times N}$ into

$$\mathbf{X} = \mathbf{P} \boldsymbol{\Sigma} \mathbf{Q}^\top, \quad (16)$$

where $\mathbf{P} = [\mathbf{p}_1 \mathbf{p}_2 \cdots \mathbf{p}_r] \in \mathbb{R}^{T \times r}$ and $\mathbf{Q} \in \mathbb{R}^{N \times r}$ are matrices whose column vectors are left singular vectors \mathbf{p}_j and right singular vectors, respectively; $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$ is a diagonal matrix containing the singular values; and $r(\leq N < T)$ is the matrix rank of $\mathbf{X}^\top \mathbf{X}$. Note that, to remove a capacity that originated from a constant term, we subtracted time average $\sum_{t=1}^T \mathbf{x}_t / T$ from \mathbf{x}_t before performing SVD. According to the decomposed state in Eq. (16), $\mathbf{X}^+ = \mathbf{Q} \boldsymbol{\Sigma}^{-1} \mathbf{P}^\top$ and we can rewrite Eq. (6) as

$$\begin{aligned} C(\mathbf{X}, \mathbf{z}^{(i)}) &= \frac{\mathbf{z}^{(i)\top} \mathbf{P} \mathbf{P}^\top \mathbf{z}^{(i)}}{\|\mathbf{z}^{(i)}\| \|\mathbf{z}^{(i)}\|} \\ &= \|\mathbf{P}^\top \boldsymbol{\phi}^{(i)}\|^2, \end{aligned} \quad (17)$$

where $\boldsymbol{\phi}^{(i)} = [\phi_1^{(i)} \cdots \phi_T^{(i)}]^\top = \mathbf{z}^{(i)} / \|\mathbf{z}^{(i)}\| \in \mathbb{R}^T$ is the normalized output ($\|\boldsymbol{\phi}^{(i)}\| = 1$).

Furthermore, assuming that \mathbf{x}_t can be expanded with the normalized PCs, the state \mathbf{x}_t in Eq. (7) is described as

$$\mathbf{x}_t = \sum_{i=1}^{\infty} \phi_t^{(i)} \hat{\mathbf{c}}_i,$$

where $\hat{\mathbf{c}}_i \in \mathbb{R}^N$ is the coefficient vector of the i th normalized basis $\phi_t^{(i)}$. Therefore the state matrix \mathbf{X} is described as

$$\mathbf{X} = \sum_{i=1}^{\infty} \boldsymbol{\phi}^{(i)} \cdot \hat{\mathbf{c}}_i^\top = \lim_{M \rightarrow \infty} \boldsymbol{\Phi} \hat{\mathbf{C}}^\top, \quad (18)$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}^{(1)} \cdots \boldsymbol{\phi}^{(M)}] \in \mathbb{R}^{T \times M}$ and $\hat{\mathbf{C}} = [\hat{\mathbf{c}}_1 \cdots \hat{\mathbf{c}}_M] \in \mathbb{R}^{N \times M}$ are the basis matrix and coefficient matrix, respectively. Comparing Eqs. (16) and (18), we obtain the matrix

$$\mathbf{P} = \lim_{M \rightarrow \infty} \boldsymbol{\Phi} \boldsymbol{\Lambda}^\top, \quad (19)$$

where $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_M] = \boldsymbol{\Sigma}^{-1} \mathbf{Q}^\top \hat{\mathbf{C}} \in \mathbb{R}^{r \times M}$ is a constant matrix, and $\boldsymbol{\lambda}_i \in \mathbb{R}^r$. With Eqs. (17) and (19), the IPC becomes

$$C(\mathbf{X}, \mathbf{z}^{(i)}) = \|\boldsymbol{\lambda}_i\|^2. \quad (20)$$

Let the normalized, linearly independent state vector at the t th step be $\hat{\mathbf{x}}_t \in \mathbb{R}^r$, and $\mathbf{P} = [\hat{\mathbf{x}}_1 \cdots \hat{\mathbf{x}}_T]^\top$. From Eq. (19), $\mathbf{P} = \sum_{i=1}^{\infty} \boldsymbol{\phi}^{(i)} \cdot \boldsymbol{\lambda}_i^\top$, and thus $\hat{\mathbf{x}}_t$ is written as

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{\infty} \phi_t^{(i)} \boldsymbol{\lambda}_i. \quad (21)$$

According to Eq. (21), Eq. (20) illustrates that the computation of the i th IPC is equivalent to expanding the normalized, linearly independent state $\hat{\mathbf{x}}_t$ with the PCs $\{\phi_t^{(i)}\}$ and calculating the squared norm of the coefficient vector of the i th normalized PC $\boldsymbol{\lambda}_i$.

In addition, the IPC has an important property of summation. The total capacity is described as

$$C_{\text{tot}} = \sum_{i=1}^{\infty} C(\mathbf{X}, \mathbf{z}^{(i)}). \quad (22)$$

Using the vector notation $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_r]$, Eq. (17) is rewritten as

$$C(\mathbf{X}, \mathbf{z}^{(i)}) = \sum_{j=1}^r (\mathbf{p}_j^\top \boldsymbol{\phi}^{(i)})^2. \quad (23)$$

From Eq. (23), Eq. (22) becomes

$$\begin{aligned} C_{\text{tot}} &= \sum_{i=1}^{\infty} \sum_{j=1}^r (\mathbf{p}_j^\top \boldsymbol{\phi}^{(i)})^2 \\ &= \lim_{M \rightarrow \infty} \sum_{j=1}^r \|\boldsymbol{\Phi}^\top \mathbf{p}_j\|^2. \end{aligned} \quad (24)$$

Equation (24) yields the sum of the squared norm of \mathbf{p}_j ($j = 1, \dots, r$) projected into an infinite-dimensional space that contains the orthogonal vectors $\boldsymbol{\phi}^{(i)}$ ($i = 1, 2, \dots$); therefore, if every \mathbf{p}_j is a function of only input history—i.e.,

$$\mathbf{p}_j = \mathbf{p}_j(\zeta_{t-1}, \zeta_{t-2}, \dots) \quad (j = 1, \dots, r),$$

then $\lim_{M \rightarrow \infty} \|\boldsymbol{\Phi}^\top \mathbf{p}_j\| = 1$, and Eq. (24) becomes

$$C_{\text{tot}} = r. \quad (25)$$

We call this property a *completeness property*. Under the assumption that the state is a function of only past input history, IPCs hold the completeness property in information processing.

To clearly demonstrate these relations, we show an example of IPC using the Legendre chaoses in Table I. If an input ζ_t follows the uniform distribution in the range of $[-1, 1]$ and produces the solution of a state equation of $x_t = \zeta_{t-1} + \zeta_{t-2}^2$ whose rank is $r = 1$, the expanded state \bar{x}_t is represented as follows:

$$\begin{aligned} \bar{x}_t &= \zeta_{t-1} + \frac{2}{3} \frac{3\zeta_{t-2}^2 - 1}{2} \\ &= z_t^{(1)} + \frac{2}{3} z_t^{(5)} \\ &= \sqrt{\frac{T}{3}} \phi_t^{(1)} + \frac{2}{3} \sqrt{\frac{T}{5}} \phi_t^{(5)}, \end{aligned}$$

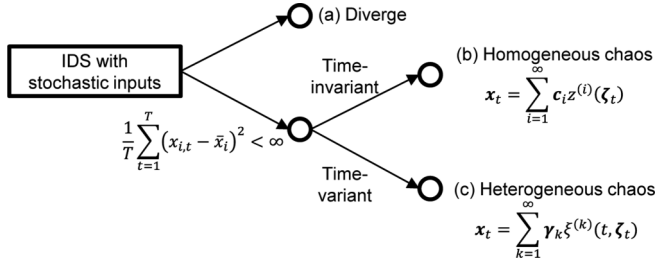


FIG. 1. The classification of IDSs with stochastic inputs based on two conditions: the finiteness of the second moment in Eq. (27) and the time-invariance in Eq. (28). Let $\zeta_t = \{\zeta_{t-s}\}_{s=1}^{\infty}$, and the homogeneous chaos and heterogeneous chaos are the spaces spanned by $z_t^{(i)} = z^{(i)}(\zeta_t)$ and $\xi_t^{(k)} = \xi^{(k)}(t, \zeta_t)$, respectively.

where $\phi_t^{(i)} = z_t^{(i)} / \|z^{(i)}\|$, $\|z^{(1)}\| = \sqrt{T/3}$, and $\|z^{(5)}\| = \sqrt{T/5}$, and a constant term $\sum_{t=1}^T x_t/T = 1/3$ is subtracted from x_t before expansion. Let $\bar{x} = [\bar{x}_1 \cdots \bar{x}_T]^T$, and $\|\bar{x}\| = \sqrt{19T/45}$. Dividing \bar{x}_t by $\|\bar{x}\|$, we obtain the normalized state \hat{x}_t and capacities as follows:

$$\hat{x}_t = \frac{\bar{x}_t}{\|\bar{x}\|} = \sqrt{\frac{15}{19}} \phi_t^{(1)} + \sqrt{\frac{4}{19}} \phi_t^{(5)},$$

$$C(\mathbf{X}, z^{(i)}) = \begin{cases} \frac{15}{19} & (i=1) \\ \frac{4}{19} & (i=5) \\ 0 & (i=2, 3, 4, 6) \end{cases}.$$

Accordingly, $C_{\text{tot}} = 1$, which is equivalent to the rank of $r = 1$.

Under the ergodicity condition, the IPC is equivalent to the square norm of the coefficient vector of the PC. These results provide a new perspective that the coefficient of PC expansion represents the amount of processed input.

B. Extending the completeness in information processing to a time-variant domain

The information processing completeness shown so far holds only if the system is a function of a past input series. Here, using the derived relationship between the IPC and PC expansion, we extend the application range of IPC to time-varying IDSs.

First of all, we must introduce a classification of PC-expanded IDSs based on the conditions imposed on the solution of Eq. (4). From the connection between the IPC and PC, prerequisites for PC expansion are similarly imposed on the time series to compute the IPC. The PC expansion extracts a nondiverging process x_t in the $1 \leq t \leq T$ range from the original sequence, which is obtained according to Eq. (4). By recursively using Eq. (4) from $t = 0$, the solution is clearly determined from the time t , input sequence $\zeta_t = \{\zeta_{t-s}\}_{s=1}^t$, and initial state x_0 , as follows:

$$x_t = g(t, \zeta_t; x_0), \quad (26)$$

where g is determined by f in Eq. (4). However, the state is described as an operator of only $\{\zeta_{t-s}\}$, showing that the PC expansion assumes that x_t is time-invariant. Therefore, as shown in Fig. 1, the PC expansion imposes two conditions on the time series. First, every i th ($i = 1, \dots, N$) state time series

$x_{i,t}$ ($t = 1, \dots, T$) must be a second-order process. We say $x_{i,t}$ to be the second-order process if the second moment of $x_{i,t}$ is finite:

$$\frac{1}{T} \sum_{t=1}^T (x_{i,t} - \bar{x}_i)^2 < \infty \quad (i = 1, \dots, N), \quad (27)$$

where $x_{i,t}$ is the i th element of x_t and $\bar{x}_i = \sum_{t=1}^T x_{i,t}/T$. Finiteness is a prerequisite for expansion convergence in the sense of L_2 due to the Cameron-Martin theorem [46–48], which guarantees that, as shown in Eq. (5), the multivariate orthogonal polynomial is obtained by the product of univariate orthogonal polynomials. Second, the solution g should be time-invariant. The extracted time series must be described only with the input time series ζ_t . Thus $x_t = g(t, \zeta_t; x_0)$ is time-invariant if

$$g(t - \tau, \zeta_{t-\tau}) = g(t, \zeta_{t-\tau}) \quad \text{for } \tau \in \mathbb{Z}, \quad (28)$$

where $\tau \leq t$ and g is independent of x_0 . Therefore the PC expansion assumes the nondivergence and time-invariance of the system.

On the other hand, if the state depends not only on input history but also on time, the homogeneous chaos is no longer a complete orthogonal system. We say x_t is time-variant if the state depends on both time and input history as in Eq. (26)—i.e., $x_t = g(t, \zeta_t; x_0)$ ($t \geq \tau$), which satisfies

$$g(t - \tau, \zeta_{t-\tau}; x_0) \neq g(t, \zeta_{t-\tau}; x_0) \quad \text{for } \tau \in \mathbb{Z} \quad (29)$$

and

$$g(t - \tau, \zeta_{t-\tau}; x_0) \neq g(t - \tau, \zeta_t; x_0) \quad \text{for } \tau \in \mathbb{Z}. \quad (30)$$

To expand the time-variant state with orthogonal bases, we introduce a time-dependent polynomial chaos (TDPC), which is an orthogonal polynomial dependent on both input history and time. Let the i th PC (the degree $d_i = \sum_s n_s^{(i)} \geq 0$) and the j th time-dependent basis be $z_t^{(i)}$ and $\psi_t^{(j)}$, respectively, and assume that both the types of bases satisfy the following orthogonality of time average:

$$\overline{z_t^{(i)} z_t^{(j)}} = \begin{cases} \overline{z_t^{(i)} z_t^{(i)}} & (i = j), \\ 0 & (i \neq j), \end{cases} \quad (31)$$

$$\overline{\psi_t^{(i)} \psi_t^{(j)}} = \begin{cases} \overline{\psi_t^{(i)} \psi_t^{(i)}} & (i = j), \\ 0 & (i \neq j), \end{cases} \quad (32)$$

$$\overline{z_t^{(i)} \psi_t^{(j)}} = 0. \quad (33)$$

For example, the Fourier series expansion uses an orthogonal set of time-dependent bases $\psi_t^{(j)} \in \{1, \cos \Omega t, \sin \Omega t, \cos 2\Omega t, \sin 2\Omega t, \dots, \cos \frac{T}{2}\Omega t, \sin \frac{T}{2}\Omega t\}$ ($\Omega = 2\pi/T$), which are expected to satisfy Eqs. (31)–(33) with an arbitrary type of PCs because $z_t^{(i)}$ and $\psi_t^{(j)}$ are independent of each other, except with the combination of constant terms $z_t^{(1)} = 1$ and $\psi_t^{(1)} = 1$. By removing this combination, the time-variant state x_t is expanded with the bases as follows:

$$x_t = \sum_{\substack{i, j \\ (i, j) \neq (1, 1)}} c_{ij}(x_0) z_t^{(i)} \psi_t^{(j)} \\ = \sum_{k=1}^{\infty} r_k(x_0) \xi_t^{(k)}, \quad (34)$$

TABLE III. An example set of TDPCs $\{z_t^{(i)}\psi_t^{(j)}\}$, which has been truncated with $\{d, s_{\max,d}\} = \{0, 1\}, \{1, 1\}, \{2, 1\}$ and $\psi_t^{(j)} \in \{1, \cos \Omega t, \sin \Omega t\}$. Note that $\mathcal{F}_n(\zeta)$ denotes the n th-order Legendre polynomial [$\mathcal{F}_0(\zeta) = 1, \mathcal{F}_1(\zeta) = \zeta, \text{ and } \mathcal{F}_2(\zeta) = (3\zeta^2 - 1)/2$].

k	Family of sets of degree and delay $\{\{n_s^{(i)}, s\}\}$	Polynomial chaos $z_t^{(i)} = \prod_s \mathcal{F}_{n_s^{(i)}}(\zeta_{t-s})$	Time-dependent basis $\psi_t^{(j)}$
1	$\{\{1, 1\}\}$	ζ_{t-1}	1
2	$\{\{2, 1\}\}$	$(3\zeta_{t-1}^2 - 1)/2$	1
3	$\{\{0, 1\}\}$	1	$\cos \Omega t$
4	$\{\{1, 1\}\}$	ζ_{t-1}	$\cos \Omega t$
5	$\{\{2, 1\}\}$	$(3\zeta_{t-1}^2 - 1)/2$	$\cos \Omega t$
6	$\{\{0, 1\}\}$	1	$\sin \Omega t$
7	$\{\{1, 1\}\}$	ζ_{t-1}	$\sin \Omega t$
8	$\{\{2, 1\}\}$	$(3\zeta_{t-1}^2 - 1)/2$	$\sin \Omega t$

where $c_{ij}(\mathbf{x}_0), \mathbf{y}_k(\mathbf{x}_0) \in \mathbb{R}^N$ are the coefficient vectors depending on \mathbf{x}_0 , and $\xi_t^{(k)}$ denotes the k th TDPC, as long as there is a one-to-one correspondence between $z_t^{(i)}\psi_t^{(j)}$ and $\xi_t^{(k)}$. Table III shows an example of TDPCs in which families of sets of degrees and delays are prepared from $\{d_i, s_{\max,d_i}\} = \{0, 1\}, \{1, 1\}, \{2, 1\}$ and time-varying bases are $\psi_t^{(j)} \in \{1, \cos \Omega t, \sin \Omega t\}$. The TDPCs satisfy the following orthogonality:

$$\overline{\xi_t^{(i)}\xi_t^{(j)}} = \begin{cases} \overline{\xi_t^{(i)}\xi_t^{(i)}} & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (35)$$

In this paper, we call the space spanned by the TDPCs as heterogeneous chaos.

The IPCs of \mathbf{x}_t in the heterogeneous chaos can be estimated by replacing the target output $z_t^{(i)}$ in Eq. (6) with the TDPC $\xi_t^{(i)}$. As with the IPC of a time-invariant system, the capacity of a time-variant system is equivalent to expanding the r -normalized, linearly independent state time series $\hat{\mathbf{x}}_t \in \mathbb{R}^r$ with normalized TDPCs and calculating the squared norm of the coefficient vector. If we let the i th TDPC vector and normalized TDPC be $\xi^{(i)} = [\xi_1^{(i)} \dots \xi_T^{(i)}]^\top$ and $\eta_t^{(i)} = \xi_t^{(i)} / \|\xi^{(i)}\|$, respectively, we can describe the normalized, linearly independent state and the capacity as

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{\infty} \eta_t^{(i)} \lambda_i, \quad (36)$$

$$C(\mathbf{X}, \xi^{(i)}) = \|\lambda_i\|^2. \quad (37)$$

We call the capacity with the target TDPCs in Eq. (37) temporal information processing capacity (TIPC). As with the total IPC in Eq. (24), the total TIPC is described by replacing the PC $z^{(i)}$ with the TDPC $\xi^{(i)}$ as follows:

$$\begin{aligned} C_{\text{tot}} &= \sum_{i=1}^{\infty} \sum_{j=1}^r (\mathbf{p}_j^\top \eta^{(i)})^2 \\ &= \lim_{M \rightarrow \infty} \sum_{j=1}^r \|\mathbf{H}^\top \mathbf{p}_j\|^2, \end{aligned} \quad (38)$$

where $\boldsymbol{\eta}^{(i)} = [\eta_1^{(i)} \dots \eta_T^{(i)}]^\top \in \mathbb{R}^T$ and $\mathbf{H} = [\boldsymbol{\eta}^{(1)} \dots \boldsymbol{\eta}^{(M)}] \in \mathbb{R}^{T \times M}$. Equation (38) yields the sum of the squared norm of \mathbf{p}_j ($j = 1, \dots, r$) projected into an infinite-dimensional space that contains the orthogonal vectors $\boldsymbol{\eta}^{(i)}$ ($i = 1, 2, \dots$). Therefore, if every \mathbf{p}_j is a function of time and input history, i.e.,

$$\mathbf{p}_j = \mathbf{p}_j(t, \zeta_{t-1}, \zeta_{t-2}, \dots) \quad (j = 1, \dots, r), \quad (39)$$

then $\lim_{M \rightarrow \infty} \|\mathbf{H}^\top \mathbf{p}_j\| = 1$, and Eq. (38) becomes

$$C_{\text{tot}} = r, \quad (40)$$

indicating that the TIPC provides complete information processing even if the system is time-variant.

The TIPC comprehensively examines information processing in a wide range of input-driven, time-variant dynamical systems because only two assumptions are imposed on the bases and the dynamical system. In the above derivation of Eq. (40), the ergodicity condition is imposed on the polynomial chaos, which is expected to be satisfied if the time series are long enough. The other limitation is that the state is a function of time and input history, as in Eq. (39). If the state includes other independent variables, the completeness property will break down. These conditions are expected to be satisfied by various input-driven time-variant systems running for a long time whether the system is stationary (e.g., a limit cycle with input) or nonstationary (e.g., a drifting state with input). Therefore the extension of the IPC to the time-variant domain can reveal aspects of the information processing performed by various time-variant systems.

C. Demonstration of comprehensive computational capabilities using a general input distribution

To illustrate that various orthogonal polynomials can be used as target outputs for the IPC, we examine the total capacities for eight target gPCs and four target aPCs using a one-dimensional ESN. The ESN is described as

$$x_{t+1} = \tanh(\rho x_t + u_t), \quad (41)$$

$$u_t = \mu + \sigma \zeta_t, \quad (42)$$

where x_t, u_t , and ζ_t are the state, input, and random variable at the t th step, respectively, while ρ, μ , and σ are the internal weight, mean of input, and input intensity, respectively. Based on Fig. 1, we chose $\rho = 0.95$ and $\mu = 0$ to treat the bounded and time-invariant state.

To visualize the capacities, we define the n th-order IPC decomposition as the sum of the n th-order capacities as follows:

$$C_{\text{tot},n} = \sum_{\{i|d_i=n\}} C(\mathbf{X}, z^{(i)}), \quad (43)$$

where $d_i = \sum_s n_s^{(i)}$ and denotes the degree of the i th PC $z^{(i)}$. Figure 2(a) shows the IPC decomposition with eight types of gPCs in the Askey scheme. The selected gPCs are the Hermite, Laguerre, Jacobi, Legendre, Charlier, Krawtchouk, Meixner, and Hahn chaoses. The total IPCs are all 1, suggesting that the gPCs form complete orthogonal systems with any combination of distributions and orthogonal polynomials and thus can be used for the IPC.

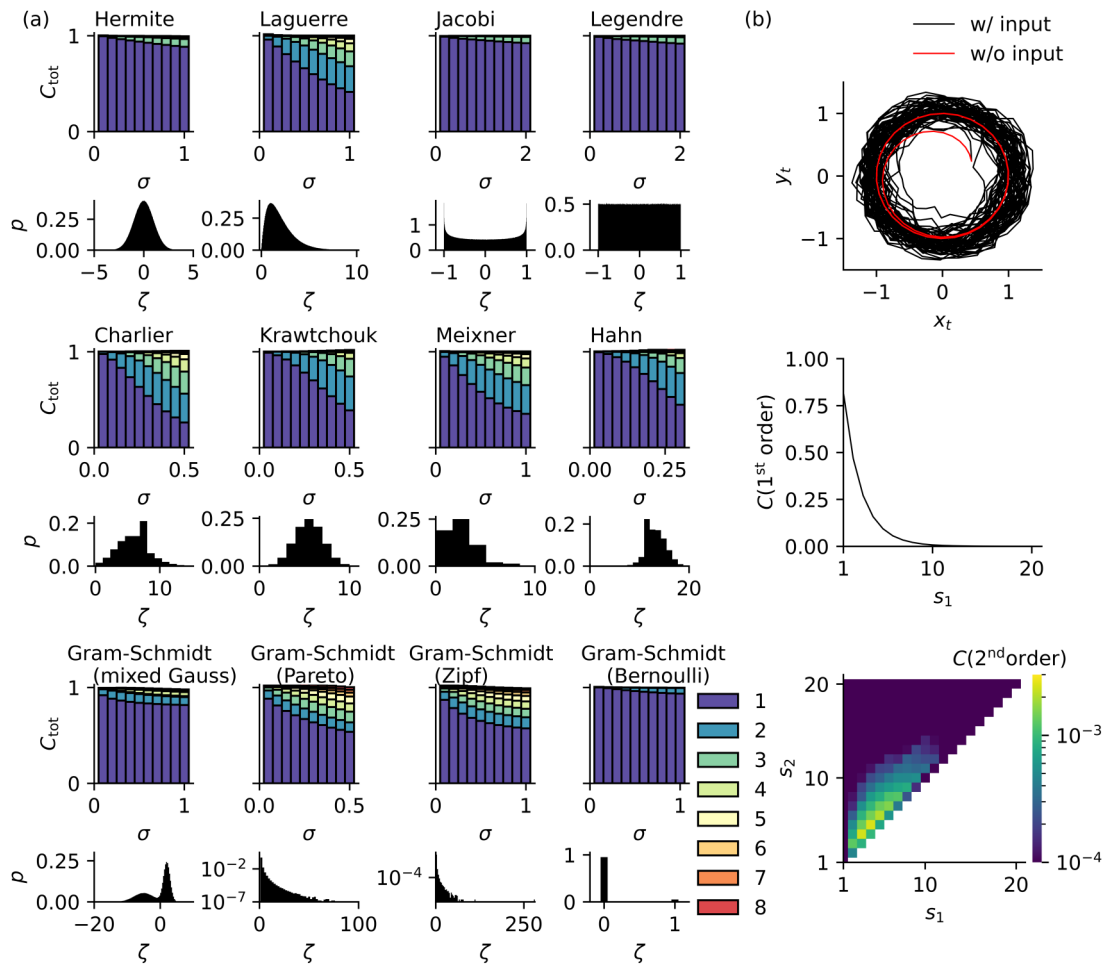


FIG. 2. Demonstration of theories. (a) IPC decomposition of a one-dimensional ESN with eight types of target generalized and four types of target arbitrary PCs of a random variable ζ : Hermite chaos with Gaussian random variables, Laguerre chaos with gamma random variables, Jacobi chaos with beta random variables, Legendre chaos with uniform random variables, Charlier chaos with Poisson random variables, Krawtchouk chaos with binomial random variables, Meixner chaos with negative binomial random variables, Hahn chaos with hypergeometric random variables, and Gram-Schmidt chaos with mixed-Gauss, Pareto, Zipf, and Bernoulli random variables. The n th-order capacities ($n = 1, \dots, 8$) in the total capacity C_{tot} are summarized as in Eq. (43). The variable p represents the probability distribution of ζ . (b) The TIPC of the limit cycle system in Eqs. (44)–(46). First, the phase plane of x_t and y_t in the system (upper panel): the systems without input ($\mu = \sigma = 0$; red line) and with input ($\mu = 0.2, \sigma = 1.5$; black line) are shown. Second, the first-order capacities of $P_1(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{1,s_1})$ with the delay step s_1 (middle panel). Third, the second-order capacities relative to delay steps s_1 and s_2 (lower panel). The dots of the diagonal ($s_1 = s_2$) and upper left triangle ($s_1 < s_2$) show the capacities of $P_2(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{2,s_1})$ and $P_1(\zeta_{t-s_1})P_1(\zeta_{t-s_2}) \cos(\omega\tau t + \alpha_{2,s_1,s_2})$, respectively. In the white area, the TIPC calculation has been omitted.

Furthermore, to demonstrate that aPCs can also be used as the target outputs for the IPC, we estimated the IPCs when the ESN was given four types of inputs that followed a mixed Gaussian, Pareto, Zipf, or Bernoulli distribution, all of which do not follow the Askey scheme. To investigate the IPC of the ESN, we built Gram-Schmidt chaoses. As shown in Fig. 2(a), the total IPCs were 1, indicating that the aPCs formed complete orthogonal systems with the input distributions and Gram-Schmidt chaoses and can be used for the IPC.

These results suggest that both the gPC and aPC are suitable as target outputs of the IPC. The extended IPCs provided complete computational capabilities of systems, which, unfortunately, have not been examined to date due to the limitation of input type.

D. Demonstration of complete computational capabilities in a time-variant system

To demonstrate the extension of the IPC, we offer the TIPC of a two-dimensional limit cycle system. The simple limit cycle system with radius r and azimuth θ in polar coordinates [49] is discretized as follows:

$$r_{t+1} = (1 + \tau)r_t - \tau r_t^3 + \tau u_t, \quad (44)$$

$$\theta_{t+1} = \theta_t + \tau\omega, \quad (45)$$

$$u_t = \mu + \sigma\zeta_t, \quad (46)$$

where $\omega = 2\pi/3$ and $\tau = 0.1$ are the angular velocity and time step width, respectively, and ζ_t is the uniform random

number in the $[-1, 1]$ interval. Therefore the input u_t follows the uniform distribution in the $[\mu - \sigma, \mu + \sigma]$ range ($\mu = 0.2$; $\sigma = 1.5$) and is applied in the radial direction. The Cartesian coordinates are given by $x_t = r_t \cos \theta_t$ and $y_t = r_t \sin \theta_t$.

As shown in Fig. 2(b), input that follows a uniform distribution forces the system to fluctuate around the noninput state. Using this time-variant system, we calculated the TIPC's whose TDPC's were constructed using the PC's $\{z_t^{(i)}\}$ for n_s ($\sum_s n_s < 5$), $s < 10$, and the time-varying bases $\psi_t \in \{\cos \Omega t, \sin \Omega t, \cos 2\Omega t, \sin 2\Omega t, \dots, \cos \frac{T}{2}\Omega t, \sin \frac{T}{2}\Omega t\}$, where $\Omega = 2\pi/T$. Note that, to focus on only the input processing, we removed time-varying terms from the state—i.e., $A_n \cos(2\pi f_n t + \theta_n)$ ($n = 1, 2, \dots$) are subtracted from x_t and y_t , where the amplitude A_n , frequency f_n , and phase θ_n are estimated by the Fourier transform. Since the time-varying bases are orthogonal to the products of the time-varying basis and PC, this process does not affect the existence of the products. From the estimated TIPC's, the state on the Cartesian plane $\mathbf{x}_t = [x_t \ y_t]^\top$ can be expressed as follows:

$$\begin{aligned} \mathbf{x}_t = & \sum_{s_1=1}^{18} \mathbf{p}_{s_1} P_1(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{1,s_1}) \\ & + \sum_{s_1, s_2} \mathbf{q}_{s_1, s_2} P_1(\zeta_{t-s_1}) P_1(\zeta_{t-s_2}) \cos(\omega\tau t + \alpha_{2,s_1, s_2}) \\ & + \sum_{s_1=2}^8 \mathbf{q}_{s_1} P_2(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{2,s_1}), \end{aligned} \quad (47)$$

where $P_n(\zeta)$ represents the n th-order Legendre polynomial and the phases α_1 and α_2 depend on the initial values. Note that although TIPC's depend on the initial values of the system in general, the final outcome of the TIPC's is the same in this case because the choice of the initial values affects not the coefficient vectors, but the phases, in Eq. (47). The coefficient vectors for the first and second-order terms are \mathbf{p} and $\mathbf{q} \in \mathbb{R}^2$, respectively, indicating that the TIPC's of the system are composed of the capacities of the product of the time-varying basis $\psi_t \in \{\cos \omega\tau t, \sin \omega\tau t\}$ and PC $\{\{P_1(\zeta_{t-s})\}, \{P_1(\zeta_{t-s})P_1(\zeta_{t-s_2})\}, \{P_2(\zeta_{t-s})\}\}$.

Moreover, in the case of using a conventional IPC, the total IPC saturates at the rank of the state matrix, r , only with the state that is a function of input history [39] characterized by the ESP. In the limit cycle system we adopted, the Lyapunov exponents of Eqs. (44)–(46) are zero in the azimuthal and negative in the radial directions; therefore, the maximum Lyapunov exponent is zero (see Appendix C for further details) and the system does not satisfy the ESP. However, even if the system does not satisfy the conventional condition, the total TIPC becomes $C_{\text{tot}} = 1.987$, saturating at the rank $r = 2$, as shown in Fig. 2(b). Note that the total TIPC is composed of $C_{\text{tot},1} = 1.952$ and $C_{\text{tot},2} = 3.45 \times 10^{-2}$, which are computed from Eq. (43) by replacing the PC $z^{(i)}$ to the TDPC $\xi^{(i)}$.

These results suggest that the information processing that was lacking with conventional IPC can be measured by adding time-varying bases, and the total TIPC can reach the rank even if the system does not hold the ESP.

III. APPLICATIONS

To illustrate the usefulness of our theory, we demonstrate the information processing capabilities of three systems.

A. The benchmark task

First, we analyze the computational capabilities required to predict a simple model for a time-series benchmark test, which is a well-known nonlinear autoregressive moving average (NARMA) model called the NARMA10 [50]. The NARMA10 task is widely utilized to evaluate the computational capabilities of dynamical systems [4,6,25–37,51–57], but the meaning of predicting this model is unknown. As mentioned earlier, a target output is expected to be a function of the input history, which can be examined by the IPC.

Let the state and input at the t th step be y_t and u_t , respectively. The NARMA10 model is expressed by

$$y_{t+1} = \alpha y_t + \beta y_t \sum_{s=0}^9 y_{t-s} + \gamma u_t u_{t-9} + \delta, \quad (48)$$

$$u_t = \mu + \kappa \zeta_t, \quad (49)$$

where the default constant parameters ($\alpha, \beta, \gamma, \delta$) are set to (0.3, 0.05, 1.5, 0.1); ζ_t is the random variable at the t th step and follows a uniform distribution in the $[-1, 1]$ interval; and μ and κ are the mean of ζ_t and the input intensity parameter, respectively. This paper uses two ranges: $u_t \in [0, \sigma]$ ($\mu = \kappa = \sigma/2$) and $u_t \in [-\sigma, \sigma]$ ($\mu = 0, \kappa = \sigma$).

According to Fig. 1, we classified the model with certain parameter regions as time-invariant (see Appendix D). In Fig. 3, p expresses the probability of not diverging as a function of σ for different random series $\{\zeta_t\}$, and the model is stable with certain parameters. Using the nondivergent model, we estimated its IPC's for the target Legendre chaos with delayed time step $s (< 16)$ and degree n_s ($\sum_s n_s < 9$).

Since the NARMA10 model is a one-dimensional system, the total capacity is 1, but the decomposition of the IPC changes with some parameters. Figure 3(a) and 3(b) show that the capacities using the uniform random variable in an asymmetric range $u_t \in [0, \sigma]$ differ significantly from those using the input in a symmetric range $u_t \in [-\sigma, \sigma]$. The capacity decomposition with the symmetric input includes only the second-order capacities ($\sum_s n_s = 2$). In contrast, the capacities with the asymmetric input contain those of the first-order because the input term emerges as $u_{t-9}u_t = \sigma^2/4(P_1(\zeta_t)P_1(\zeta_{t-9}) + P_1(\zeta_t) + P_1(\zeta_{t-9}) + 1)$ in Eq. (48), including the first-order terms of ζ . Hence, in the case of using an asymmetric input, we can regard the model as the system receiving the first-order inputs, which are retained for a few steps. These results suggest that the input should be changed according to the dynamical system when one uses the NARMA10 task; for example, as the nodes of an ESN are represented by an odd function, such as a hyperbolic tangent, the ESN has only odd capacities [39]. From this property and our results, the ESN with $u_t \in [0, \sigma]$ emulates the NARMA10 model, whereas the one with $u_t \in [-\sigma, \sigma]$ does not predict it at all.

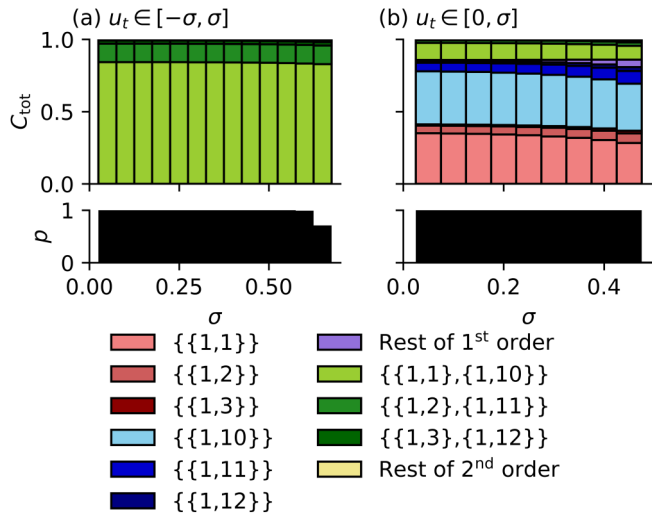


FIG. 3. The IPC decomposition of the NARMA10 model as a function of the input intensity σ and the probability ρ that y_t does not diverge. The labels indicate representative combinations of $\{\{n_s, s\}\}$, where n_s is the degree of the polynomial, s is the delayed time step of the input, and the desired output is $\prod_s P_{n_s}(\zeta_{t-s})$. The labels for other combinations have been omitted. The probability ρ represents the proportion of y_t values that do not diverge to infinity and is calculated using 100 input time series ζ_t generated from 100 random seeds. In all the figures, the capacities are not stacked if the output diverged or if $\sigma = 0$. Here, u_t follows the uniform distribution in the ranges (a) $[-\sigma, \sigma]$ and (b) $[0, \sigma]$.

B. The machine learning network

Second, we analyzed the performance of a machine learning network that solved the NARMA10 task using ESNs. We used three types of ESNs with different activation functions. Let the i th ($i = 1, 2, \dots, N$) state of the ESN at the t th step be $x_{i,t}$. The state equations using linear, hyperbolic tangent, and analog integrator functions are as follows:

$$x_{i,t+1} = \rho \sum_{j=1}^N w_{ij} x_{j,t} + \iota w_{in,i} \zeta_t, \quad (50)$$

$$x_{i,t+1} = \tanh \left(\rho \sum_{j=1}^N w_{ij} x_{j,t} + \iota w_{in,i} \zeta_t \right), \quad (51)$$

$$x_{i,t+1} = \left(1 - \frac{1}{\tau} \right) x_{i,t} + \frac{1}{\tau} \tanh \left(\rho \sum_{j=1}^N w_{ij} x_{j,t} + \iota w_{in,i} \zeta_t \right), \quad (52)$$

where w_{ij} was initialized with a uniform random number in the range of $[-1, 1]$ and multiplied by a constant so that the maximum eigenvalue of the matrix $\mathbf{W} = [w_{ij}]$ was 1. The input weight $w_{in,i}$ was also set to a uniform random number in the range of $[-1, 1]$; and ρ , $\iota (= 0.1)$, and $N (= 50)$ represent the spectral radius of $\rho \mathbf{W}$, input intensity, and the number of nodes, respectively. In Eq. (52), τ was set to 1.25.

As shown in Fig. 4(a), we also emulated the target NARMA10 model using 50-node ESNs with three activation functions and compared the output of the ESN \hat{y}_t

and the target y_t with the normalized root-mean-square errors (NRMSEs). For all the functions, the NRMSEs decreased with an increase in ρ and increased when $\rho \geq 1$. To analyze the outputs of the ESNs training with the NARMA10 model, the IPCs of the output \hat{y}_t were estimated. Figure 4(b)–4(d) shows the change in IPC decomposition of the output y_t with an increase in the spectral radius ρ of the ESN with the linear, hyperbolic tangent, and analog integral functions, respectively. To emulate the NARMA10 model, whose IPC decomposition is shown in Fig. 4(e), the nine types of Legendre chaoses $\{P_1(\zeta_{t-s})\}_{s=1,2,3,10,11,12}$ and $\{P_2(\zeta_{t-s})\}_{s=1,2,3}$ are mainly required in a certain ratio. According to Fig. 4(b)–4(d), for any activation function, as $\rho (< 1)$ increases, $\{P_1(\zeta_{t-s})\}_{s=1,2,3,10,11,12}$ approaches the required rate. On the other hand, the three types of second-order IPCs $\{P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})\}_{s=1,2,3}$ are almost zero in ESNs with an activation function. Therefore, in the NARMA10 task with ESNs, performance is compared using only the first-order IPCs.

To investigate why the second-order IPCs did not appear in the decomposition, we estimated the IPCs from ESN states. Figure 4(f)–4(h) shows the change in IPC decomposition with the increase in the spectral radius ρ in the ESNs with linear, hyperbolic tangent, and analog integral functions, respectively. The rank r of the ESN state increases with $\rho (\leq 1)$ and corresponds to C_{tot} . As these three systems have small amounts of target, second-order IPCs (each is less than 0.2), the ESNs cannot emulate the target, second-order Legendre chaoses.

The above results demonstrate that our method is capable of decomposing the computational capability of the machine learning network before and after training, as well as clarifying whether the computational components required for the task have been extracted through learning and exist in the original network.

C. The real neural network

Finally, to demonstrate the broad applicability of our theory, we prepared a dissociated culture of neurons for a physical reservoir, which is an open, nonequilibrium system that fluctuates due to external inputs and has parameters that can be considered time-dependent. Real neurons extracted from the cortices of rat embryos were pharmacologically and mechanically isolated and then seeded on an electrode array. After the culture matured [Fig. 5(a)], we constructed a physical reservoir using electrodes with active neurons [Fig. 5(b); see Appendix F]. We repeatedly applied bipolar waves with 10, 20, or 30 ms interpulse intervals (IPIs) to 29 stimulus electrodes: their amplitudes ζ_t followed a Gaussian distribution with mean $\mu = 200, 300,$ or 400 mV and standard deviation $\sigma = 50$ mV [Fig. 5(c)]. Furthermore, we computed the number of spikes in an IPI-width bin from $N = 792$ measurement electrodes as the reservoir states [Fig. 5(d)]. As a result, we obtained a long, single trajectory of the activation of the electrodes according to the input stream, which consisted of 20,000 time steps and was used for our TIPC analysis. Note that in the following analysis, we assumed that the initial value for the time-variant system was given and fixed.

Using these data, we computed the TIPC of the physical system. Figure 5(a) shows the first-order TIPC of the delay

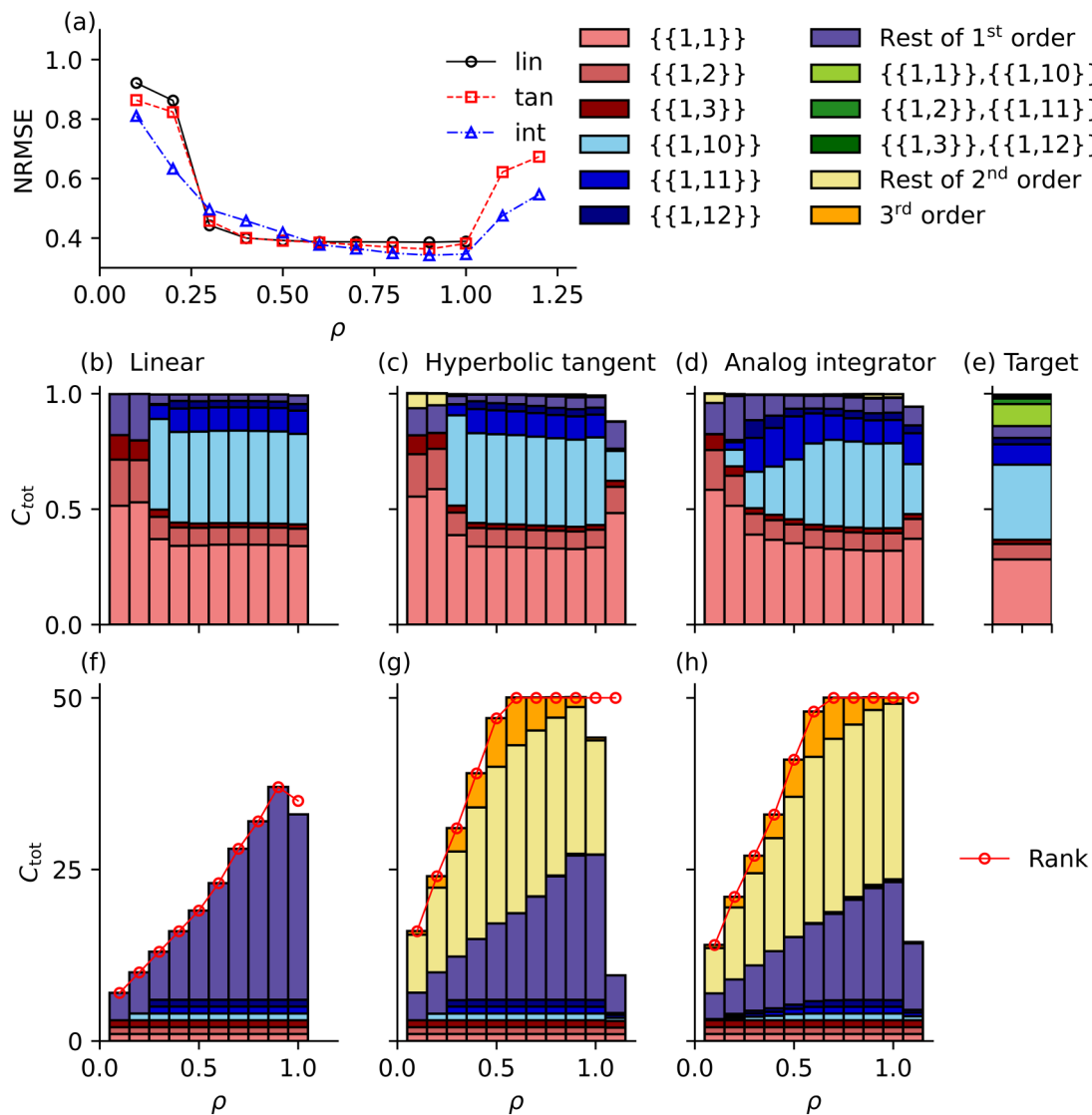


FIG. 4. The performance decomposition of NARMA10-trained ESNs with the IPC. (a) shows the NRMSEs of the NARMA10 task relative to the spectral radius ρ of ESNs composed of linear, hyperbolic tangent, or analog integrator functions. (b), (c), and (d) display the relationships between ρ and the IPC decomposition in the outputs from the NARMA10-trained ESNs with linear, hyperbolic tangent, and analog integrator functions, respectively. The labels indicate representative combinations of $\{\{n_s, s\}\}$, where n_s is the degree of the polynomial, s is the delayed time step of the input, and the desired output is $\prod_s P_{n_s}(\zeta_{t-s})$. In the case of the first-order IPCs, the representative combinations— $\{\{1,1\}\}$, $\{\{1,2\}\}$, $\{\{1,3\}\}$, $\{\{1,10\}\}$, $\{\{1,11\}\}$, and $\{\{1,12\}\}$ —are shown; the other combinations are summarized as “Rest of the first order.” The representative combinations of the second-order IPCs— $\{\{1,1\}, \{1,10\}\}$, $\{\{1,2\}, \{1,11\}\}$, and $\{\{1,3\}, \{1,12\}\}$ —are barely held by the ESN; thus, other combinations are summarized as “Rest of the second order.” (e) shows the IPC decomposition of the target NARMA10 model with input in the $[0, 0.45]$ range. (f), (g), and (h) represent the relationships between ρ and the IPC decomposition of the ESN state using linear, hyperbolic tangent, and analog integrator functions, respectively. The red line denotes the rank of state r , which corresponds to the total IPC in the time-invariant domain.

step s ; they contain the memory function (MF) [58,59] and the four temporal memory functions (TMFs) whose targets were the products of the time-varying basis and PC $z_t = P_1(\zeta_{t-s}) \cos(n\omega t)$ or $z_t = P_1(\zeta_{t-s}) \sin(n\omega t)$ ($n = 1, 2$). Note that $n\omega = 2\pi n/T$ ($n = 1, \dots, T/2$) denotes the frequency. The MFs and TMFs monotonically decay with an increase in s . Since the PCs and the products are orthogonal, the dissociated culture held the old inputs not only by delaying them but also by multiplying them by time-varying terms. In addition, the time-variant IPC in the dissociated culture had a fading-memory-like property that the input history encoded

in the product terms and it decayed as an increase in the delay.

Next, to investigate the frequency characteristics of the TIPC, we plotted it with $s = 1$. Figure 5(b) and 5(c) show the relationships between frequency and TIPC with time-varying cosine and sine targets, respectively. We term such a graph the TIPC spectrum. Both spectra have larger TIPC at lower frequencies. Therefore the input was processed by the products of the low frequency sinusoidal wave and the past inputs, suggesting that the superposition of these waves represents a gradual trend which may be caused by nonstationary changes

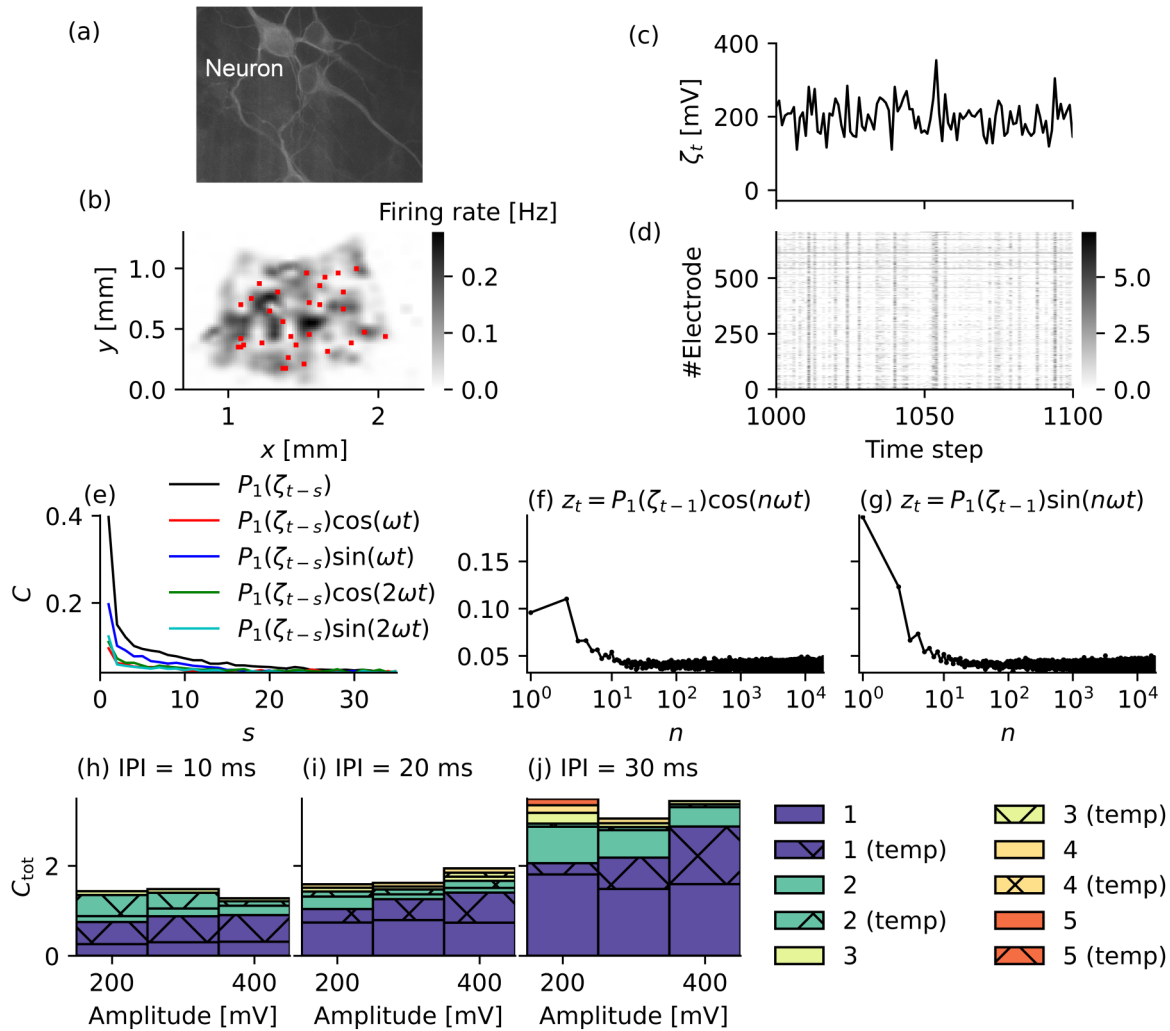


FIG. 5. TIPCs of the dissociated culture of neurons. (a) shows the dissociated culture of neurons on the electrode array. (b) shows the spontaneous firing rate of the culture on the electrode array, where (x, y) denotes the position of the electrode. The rate is filtered by the Gaussian kernel. The red dots represent the 29 stimulation electrodes. (c) and (d) represent the amplitude of stimulus ζ_t , which follows a Gaussian distribution ($\mu = 200$ mV), and the spike count of each electrode x_t ($\mu = 200$ mV and IPI = 10 ms), respectively. (e) illustrates the temporal memory functions (TMFs) of delay step s , and (e)–(g) show the TMFs when the IPI = 30 ms and $\mu = 400$ mV. TMFs with the time-invariant target $z_t = P_1(\zeta_{t-1})$ and four time-variant targets—i.e., $z_t = P_1(\zeta_{t-1}) \cos(n\omega t)$ and $P_1(\zeta_{t-1}) \sin(n\omega t)$ ($n = 1, 2$)—are plotted. (f) and (g) show the first-order TIPC spectrum with the cosine $\zeta_t = P_1(\zeta_{t-1}) \cos(n\omega t)$ and sine $\zeta_t = P_1(\zeta_{t-1}) \sin(n\omega t)$ targets, respectively. (h), (i), and (j) depict the total capacity of the culture when the IPI = 10, 20, and 30 ms, respectively, and the horizontal axis is the mean of amplitude μ . The hatched bar represents the total capacity using time-varying n th-order polynomials.

in the neural activities. For example, as the stimuli were repeatedly applied, the spike rates decreased (neural adaptation) and the intensity of the interaction between neurons changed (synaptic plasticity), which further changed the activities. Therefore the obtained time series was nonstationary, which embedded the time-dependent information processing of the dissociated culture of neurons.

Figure 5(d)–5(f) illustrates the total capacities C_{tot} with different μ and IPIs. Every C_{tot} contained time-varying IPCs, and as the IPIs decreased, the ratio of time-varying IPCs to C_{tot} increased. As with the time-invariant case, the time-variant IPC increased as the degree decreased. Therefore the total capacity of the dissociated culture of neurons contained time-varying IPCs in all cases.

In summary, the above results reveal that the dissociated culture of neurons had not only the time-invariant IPC but also

the time-variant IPC, indicating that this system processed inputs through the products of the time-varying basis and PC in addition to the functions of past input series. Thus the TIPC demonstrated time-dependent information processing that could not be handled to date.

IV. DISCUSSION

A. The relationship between an attractor and information processing

In the present paper, random variables were used as input in the time-invariant systems whose states were represented by echo functions, which depend, not on the time t , but on the past input time series. In a situation where the system with no input converges into a fixed point (e.g., the intersection point of the NARMA10 system), this function illustrates that

the system stays a fixed-point attractor and fluctuates around the fixed point due to the input. A conventional IPC targets time-invariant systems and quantifies the input processing of a state that depends only on the input time series. Therefore a conventional IPC sometimes represents the input information processing performed around a certain fixed-point attractor.

In addition, the IPC is equal to the squared norm of the coefficient vector of the temporal basis vectors expanded with PCs. Since these coefficient vectors obviously change depending on the fixed point, different types of information processing are performed at different fixed points. However, an IPC extended for time-variant systems was applied to the limit cycle system, which does not satisfy the prerequisites for RC. From the TIPC estimates, the expanded solution contains the products of the time-varying basis and PC, $P_1(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{1,s_1})$, $P_2(\zeta_{t-s_1}) \cos(\omega\tau t + \alpha_{2,s_1})$, and $P_1(\zeta_{t-s_1})P_1(\zeta_{t-s_2}) \cos(\omega\tau t + \alpha_{2,s_1,s_2})$, revealing that the processed input represents the amplitude scale of the sinusoidal. Therefore the fluctuation of a periodic attractor determined by input can be processed around the limit cycle.

Based on these findings, we can conclude that a conventional IPC can evaluate computational capabilities around a fixed point, while the TIPC can also evaluate capabilities around a periodic attractor. Since the recent RC framework exploits not only fixed points or periodic attractors but also deterministically chaotic ones [60–64], future work should examine the relationship between various attractors and information processing.

B. Methods to utilize a time-variant system as a computational resource

In calculating the TIPCs for a limit cycle system, we demonstrated that the state of the system can include products of a time-varying basis and PC. Since the product is represented by the time-varying basis ψ_t and PC ϕ_t , the conventional IPC for time-invariant terms cannot quantify their amount of information. The TIPC indicates that even in a time-variant system, information can be processed by the PC ϕ_t in the product. In addition, to date, RC has trained a static readout weight by linear regression, assuming that the ESP or FMP is satisfied; the state in the reservoir is a function of the input time series. Since the target output is described by the input history, performance decreases when a time-dependent reservoir (e.g., an ESN with a spectral radius of $\rho \geq 1$) is used. As the number of input history terms in the expanded state decreases, the performance drops, while the number of product terms increases. The input information in a product term can be used for the task and extracted by giving a readout weight that cancels out ϕ_t in the product term (e.g., a time-varying weight). Therefore, even in a reservoir where the ESP or FMP does not hold, the input time series may be processed by TDPCs and the task can be successfully solved by using new types of readout.

In this way, a method already exists for exploiting periodic systems as computational resources. As shown in Fig. 6(a) and 6(b), the time-multiplexing technique [6] switches the input $u(t)$ with the time width τ and extracts N nodes $\mathbf{x}(t) = [x(t), x(t + \frac{\tau}{N}), x(t + \frac{2\tau}{N}), \dots, x(t + \frac{(N-1)\tau}{N})]^T$ from a single node. This method virtually increases the number of nodes

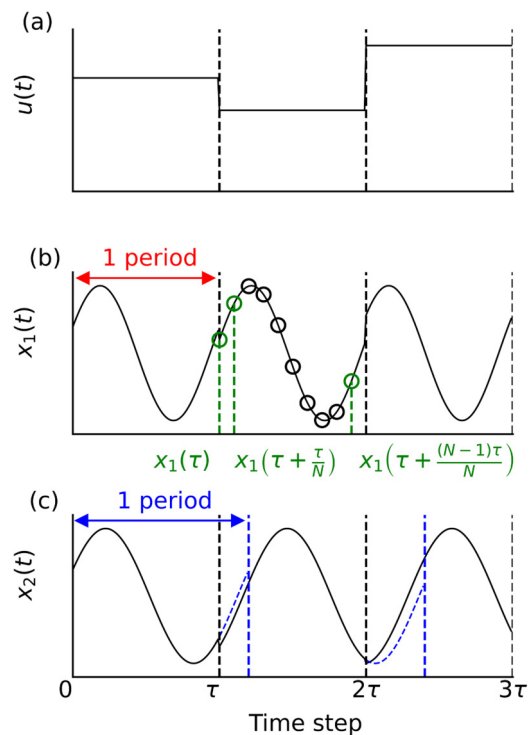


FIG. 6. The time-multiplexing method transforms a periodic system into a time-invariant one. (a) The input $u(t)$ is switched with time width τ . (b) and (c) show a periodic state time series whose frequency is and is not consistent with that of the input, respectively. In (b), the circles represent the virtual nodes $\mathbf{x}_1(t) = [x_1(t), x_1(t + \frac{\tau}{N}), \dots, x_1(t + \frac{(N-1)\tau}{N})]^T$.

in the reservoir to improve its computational capabilities. If $x(t)$ is a periodic function and oscillates with a period specific to $u(t)$ [Fig. 6(b)], applying an input with the same width τ as the period can extract time-invariant virtual nodes because the i th virtual node always corresponds to $x(t)$ at a certain phase and is not affected by the periodic fluctuation. However, if the period does not match τ [Fig. 6(c)], the phase shifts, and the scheme cannot exploit the computational capabilities of the periodic system. Therefore time multiplexing can be interpreted as a method capable of extracting a computational resource by transforming a time-varying system into a time-invariant system.

From the above discussion, a time-variant system can process input information through product terms, and rich input information can be extracted from time-variant systems by designing readouts for the systems. In the future, information processing using a time-varying reservoir and novel design methods for readouts will be investigated.

C. Extension to TIPC with multiple input variables

In the present paper, the TIPC assumes that the input dimension is one, but this can easily be extended to a multiple input version. Let M independent stochastic inputs that follow multiple arbitrary distributions be $\zeta_t^{(1)}, \dots, \zeta_t^{(M)} \in \mathbb{R}$, and the state equation in Eq. (4) is modified to

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \zeta_t^{(1)}, \dots, \zeta_t^{(M)}),$$

TABLE IV. An example set of PCs $\{z_t^{(i)}\}$ with two inputs, which has been truncated with $\{\{d, s_{\max, d}\}\} = \{\{1, 2\}, \{2, 2\}\}$. Note that $\mathcal{F}_n^{(m)}(\zeta^{(m)})$ denotes the n th-order Legendre polynomial for the m th input [$\mathcal{F}_1^{(m)}(\zeta^{(m)}) = \zeta^{(m)}$ and $\mathcal{F}_2^{(m)}(\zeta^{(m)}) = \{3(\zeta^{(m)})^2 - 1\}/2$].

i	Family of sets of index, degree, and delay $\{\{m_{ij}, n_{ij}, s_{ij}\}\}$	Degree $d_i = \sum_j n_{ij}$	Polynomial chaos $z_t^{(i)}$
1	$\{\{1, 1, 1\}\}$	1	$\zeta_{t-1}^{(1)}$
2	$\{\{2, 1, 1\}\}$	1	$\zeta_{t-1}^{(2)}$
3	$\{\{1, 1, 2\}\}$	1	$\zeta_{t-2}^{(1)}$
4	$\{\{2, 1, 2\}\}$	1	$\zeta_{t-2}^{(2)}$
5	$\{\{1, 2, 1\}\}$	2	$\{3(\zeta_{t-1}^{(1)})^2 - 1\}/2$
6	$\{\{2, 2, 1\}\}$	2	$\{3(\zeta_{t-1}^{(2)})^2 - 1\}/2$
7	$\{\{1, 1, 1\}, \{2, 1, 1\}\}$	2	$\zeta_{t-1}^{(1)}\zeta_{t-1}^{(2)}$
8	$\{\{1, 1, 1\}, \{2, 1, 2\}\}$	2	$\zeta_{t-1}^{(1)}\zeta_{t-2}^{(2)}$
9	$\{\{2, 1, 1\}, \{1, 1, 2\}\}$	2	$\zeta_{t-1}^{(2)}\zeta_{t-2}^{(1)}$
10	$\{\{1, 2, 2\}\}$	2	$\{3(\zeta_{t-2}^{(1)})^2 - 1\}/2$
11	$\{\{2, 2, 2\}\}$	2	$\{3(\zeta_{t-2}^{(2)})^2 - 1\}/2$
12	$\{\{1, 1, 2\}, \{2, 1, 2\}\}$	2	$\zeta_{t-2}^{(1)}\zeta_{t-2}^{(2)}$

whose solution is described as

$$\mathbf{x}_t = \mathbf{g}(t, \zeta_t^{(1)}, \dots, \zeta_t^{(M)}),$$

where $\zeta_t^{(m)} = \{\zeta_{t-s}^{(m)}\}_{s=1}^t$ ($m = 1, \dots, M$). Under the assumption that the state is a second-order process in Eq. (27) and time-invariant for a version of multiple inputs in Eq. (28)—i.e.,

$$\mathbf{g}(t - \tau, \zeta_{t-\tau}^{(1)}, \dots, \zeta_{t-\tau}^{(M)}) = \mathbf{g}(t, \zeta_{t-\tau}^{(1)}, \dots, \zeta_{t-\tau}^{(M)})$$

for $\tau \in \mathbb{Z}$,

\mathbf{x}_t can be expanded by aPCs with multiple input variables [65]. Let the n th univariate polynomial for the m th input be $\mathcal{F}_n^{(m)}(\zeta_t^{(m)})$, and the i th PC is described by

$$z_t^{(i)} = \prod_j \mathcal{F}_{n_{ij}}^{(m_{ij})}(\zeta_{t-s_{ij}}^{(m_{ij})}), \quad (53)$$

where m_{ij} , n_{ij} , and s_{ij} denote the index of input, degree, and delay, respectively, of the j th element for the i th PC. Table IV provides an example of PCs with two inputs, which are truncated with $\{\{d_i, s_{\max, d_i}\}\} = \{\{1, 2\}, \{2, 2\}\}$. As all of these aPCs are orthogonal in the meaning of ensemble average, their orthogonality is defined by the same inner product as Eqs. (10) and (11). Under the ergodicity assumption in Eq. (15), the PCs in the time direction compose a complete orthogonal system of multiple input histories $\{\zeta_t^{(1)}, \dots, \zeta_t^{(M)}\}$.

On the other hand, if the state is time-variant, \mathbf{x}_t can be expanded by TDPCs of multiple inputs. As with the TDPC with a single input, the TDPC with multiple inputs is expected to be orthogonal, as in Eq. (35), using the PCs $z_t^{(i)}$ in Eq. (53) and time-varying bases $\psi_t^{(j)}$ that meet Eqs. (31)–(33). Using the TDPCs, we can expand the normalized, linearly independent state in Eq. (36) and define its TIPC in Eq. (37). The total TIPC is described by Eq. (38). If all of the normalized states

are functions of time and multiple input histories—i.e.,

$$\mathbf{p}_j = \mathbf{p}_j(t, \zeta_t^{(1)}, \dots, \zeta_t^{(M)}) \quad (j = 1, \dots, r),$$

the total TIPC becomes $C_{\text{tot}} = r$.

Physical systems often receive various inputs from the external environment, resulting in such multiple input-driven systems. For example, in the dissociated culture of neurons, the ranks of the state are 724–792, but the total capacities is less than 3.5 [Fig. 5(c)], which are much smaller than their ranks. One possible speculation regarding this issue is that in our scheme, the state may be expressed as a function of multiple stochastic inputs, including the one we gave (e.g., electrical noise, synaptic noise, thermal noise, and shot noise [66]). Of all the inputs, the ones we could observe were limited, and the total capacity computed only from the inputs did not reach the rank. Therefore physical systems can receive unobservable inputs, which disturb the examination of all capacities.

D. A complete orthogonal system with PCs for a finite data set

This paper demonstrated that the total IPCs with various types of PCs reach the ranks in the time-invariant systems. According to the interconvertibility between the IPC and the PC expansion, these results imply that the PC expansion approximates the original time series with high accuracy, which is achieved by a complete orthogonal system of the PCs. In this paper, we adopted the following settings to satisfy the orthogonality and the completeness of the PCs.

From the perspective of the numerical computation, the orthogonality imposed on the PCs is decomposed into two parts: (i) the independence between different order polynomials $\mathcal{F}_n(\zeta)$, $\mathcal{F}_m(\zeta)$ ($n \neq m$) in Eq. (9), and (ii) the independence between inputs with different delays ζ_{t-s_1} , ζ_{t-s_2} ($s_1 \neq s_2$). The first condition is expected to be met by the sufficiently long length of time series in the case of gPC. Since $\mathcal{F}_n(\zeta)$ is given based on the theoretical distribution of ζ , the orthogonality will be satisfied if the distribution of sampled ζ approaches the ideal one. In addition, the first condition for aPC is satisfied through the Gram-Schmidt procedure in Eqs. (12) and (13) regardless of the length of time. The first condition is not sufficient to guarantee all of the orthogonal relations—for example, the orthogonality of univariate polynomials with different delays $\mathcal{F}_n(\zeta_{t-s_1})$, $\mathcal{F}_n(\zeta_{t-s_2})$ ($s_1 \neq s_2$) and the orthogonality of products of univariate polynomials $\mathcal{F}_1(\zeta_{t-s_1})\mathcal{F}_1(\zeta_{t-s_1-1})$, $\mathcal{F}_1(\zeta_{t-s_2})\mathcal{F}_1(\zeta_{t-s_2-1})$ ($s_1 \neq s_2$) are not satisfied. These orthogonalities are achieved by the independence of the inputs with different delays ζ_{t-s_1} , ζ_{t-s_2} . Their inner products approach zero as the lengths of time series become long enough. In summary, in both the cases of gPC and aPC, the orthogonality depends on the length of time series T , which was set to 10^6 – 10^7 in this paper.

Furthermore, the search range of the PCs is important to construct a complete orthogonal system of a target dynamical system. Some systems have first-order capacities with over 1000 delays, and others have a first-order capacity with only one delayed step. Additionally, as the order of capacity in a system becomes high, the maximum delay tends to be shorter. For example, if the first-order capacities are up to 50 delay steps, the second-order ones are within 10 steps; thus, the

third-order ones are within five steps. By leveraging these properties and examining a sufficiently wide range of the family of sets of degrees and delays, the completeness can be achieved.

Therefore, to numerically construct the complete orthogonal system, we adopted the long enough lengths of time series and the large enough families of sets of degrees and delays. By building a sufficiently large model with a sufficient amount of data, we can build a highly accurate system.

V. CONCLUSION

This paper attempted to clarify the unknown relationship between the PC expansion and IPC by deriving the IPC from the PC-expanded system. To illustrate this relation, we showed that the IPC can be measured using gPCs and aPCs. In addition, using the NARMA10 model and ESNs, we concretely described the relationship and showed the usefulness of our theory. Next, taking into account the characteristics of the general solution of the input-driven dynamical system, we proposed the IPC for time-variant systems—the TIPC. To demonstrate that a time-variant system has such TIPCs, we investigated the TIPC decomposition of a limit cycle system and a cultured neural network. The primary results are summarized as follows.

(1) Using SVD, we can obtain the orthogonal temporal basis vectors from the state time series. These vectors can be expanded with PCs to obtain the coefficient vector of each basis. The IPC is equivalent to the squared norm of the coefficient of the PC used as the target output. Therefore the expanded basis coefficients represent the amount of input processing information.

(2) Using eight types of polynomials in the Askey scheme and Gram-Schmidt PCs, we estimated the IPCs of a one-dimensional ESN, whose total IPCs were equal to one. These results indicate that various types of PCs within the Askey scheme and Gram-Schmidt orthogonalization are suitable for the target output of the IPC.

(3) We calculated the TIPC of a time-variant system using a simple limit cycle and revealed that the input information processing is performed by the products of the time-varying basis and PC.

(4) IPC analysis revealed that the NARMA10 model is mainly composed of $P_1(\zeta_{t-s})$ ($s = 1, 2, 3, 10, 11, 12$) and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$ ($s = 1, 2, 3$). The NARMA10 benchmark task can be solved by holding the nine types of input information in a reservoir. Consequently, combining the IPC and PC expansion provides a clear and concise picture of information processing.

(5) The dissociated culture of neurons had not only the time-invariant IPC but also the time-variant IPC. These results suggest that the neural system processes inputs with a trended state which varies over time (due to the synaptic plasticity and neural adaptation). To date, such nonstationary information processing has been left behind; however, we revealed that this type of processing is performed in the real neural network.

The above results suggest that the connection between the IPC and PC expansion allows for a simpler description of information processing in dynamical systems. In the future, information processing using time-variant systems should

also be researched. This scheme can be applied to nonstationary systems and thus may be suitable for elucidating information processing in neural circuits, a topic which has been overlooked so far. In addition, it can be applied not only to neural systems but also to other physical systems that can be time-variant, such as quantum, spintronics, optical systems, and fluid. For example, it is reported recently that some types of vortex have been generated when a fluid flows passed a bluff body, which can be used as an information processing device [67]. In the analysis, it was found that near the critical Reynolds number where the flow exhibits a twin vortex before the onset of the Kármán vortex shedding associated with the Hopf bifurcation (i.e., transition from a fixed point to a limit cycle), the information processing capability was maximized. This was also characterized by the breakdown of the ESP due to the periodic attractor. According to our results, it may be possible to evaluate the type and amount of information processing even in the Kármán vortex shedding, which would be a topic for future work.

ACKNOWLEDGMENTS

This paper is based on results obtained from the Exploration of Neuromorphic Dynamics towards Future Symbiotic Society project commissioned by the NEDO. H.T. was partly supported by JSPS KAKENHI (20H04252), AMED (JP20dm0307009), the Naito Foundation, and the Asahi Glass Foundation. K.N. was supported by JSPS KAKENHI (JP18H05472), by JST CREST (JPMJCR2014), and MEXT Quantum Leap Flagship Program (MEXT Q-LEAP; JPMXS0118067394 and JPMXS0120319794).

APPENDIX A: GENERALIZED POLYNOMIAL CHAOS

The generalized polynomial chaos (gPC) [44] is PC composed of the univariate polynomial $\mathcal{F}_n(\zeta)$ included in the Askey scheme [68] tree (Fig. 7). The Askey scheme represents various orthogonal polynomials (e.g., Hermite, Jacobi, Laguerre, and Charlier) using the hypergeometric series ${}_rF_s$ of x , along with parameters a_1, \dots, a_r and b_1, \dots, b_s :

$${}_rF_s\left(x; \begin{matrix} a_1, \dots, a_r \\ b_1, \dots, b_s \end{matrix}\right) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_r)_k x^k}{(b_1)_k \cdots (b_s)_k k!},$$

$$(a)_n = \begin{cases} 1 & (n = 0) \\ a(a+1) \cdots (a+n-1) & (n = 1, 2, \dots) \end{cases},$$

where $(a)_n$ is the Pochhammer symbol. For example, the Laguerre polynomial with a parameter α , $L_n^{(\alpha)}(\zeta)$ can be written as

$$L_n^{(\alpha)}(\zeta) = \frac{(\alpha+1)_n}{n!} {}_1F_1\left(\zeta; \begin{matrix} -n \\ \alpha+1 \end{matrix}\right).$$

Note that in Fig. 7, an upper polynomial with the limit of a certain parameter or parameters becomes a lower polynomial connected with a line; for example, the Laguerre polynomial becomes the Hermite polynomial $H_n(\zeta)$ by taking the following limit of α :

$$\lim_{\alpha \rightarrow \infty} \left(\frac{2}{\alpha}\right)^{n/2} L_n^{(\alpha)}((2\alpha)^{1/2}\zeta + \alpha) = \frac{(-1)^n}{n!} H_n(\zeta).$$

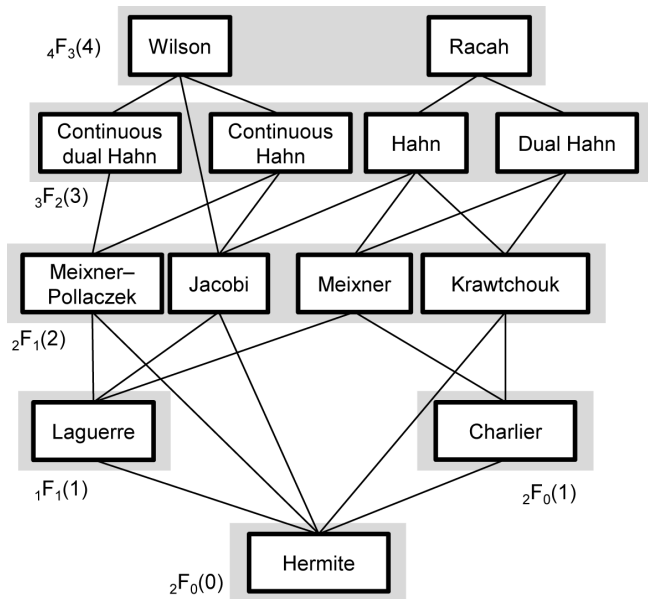


FIG. 7. The relationship between polynomials in the Askey scheme. Each polynomial is represented with the hypergeometric series ${}_rF_s(p)$, where p represents the number of parameters substituted in a_i ($i = 1, 2, \dots, r$), b_i ($i = 1, 2, \dots, s$), or z . An upper polynomial with the limit of a certain parameter or parameters becomes a lower polynomial connected with a line.

The i th PC $z_t^{(i)}$ is represented by the product of the $n_s^{(i)}$ -th order polynomial of ζ_t delayed by s steps, $\mathcal{F}_{n_s^{(i)}}(\zeta_{t-s})$. When the sets of degree n_s and delay step s are given by the i th family of sets $\mathcal{N}_i = \{\{n_s^{(i)}, s\}\}$, $z_t^{(i)}$ is represented as

$$z_t^{(i)} = \prod_s \mathcal{F}_{n_s^{(i)}}(\zeta_{t-s}).$$

Using the hypergeometric series ${}_rF_s(\zeta)$, $\mathcal{F}_n(\zeta_{t-k})$ can be applied to the following eight types of orthogonal polynomials.

1. The Hermite polynomial and a Gaussian distribution

The n th ($n = 1, 2, \dots$)-order Hermite polynomial $H_n(\zeta)$ is given by

$$H_{n+1} = \zeta H_n - nH_{n-1},$$

where $H_0 = 1$, and $H_{-1} = 0$. ζ follows the standard normal distribution

$$f(\zeta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta^2}{2}\right).$$

2. The Laguerre polynomial and a gamma distribution

The n th ($n = 1, 2, \dots$)-order Laguerre polynomial $L_n^{(\alpha)}(\zeta)$ is given by

$$L_n^{(\alpha)}(\zeta) = \sum_{i=0}^n (-1)^i \binom{n+\alpha}{n-i} \frac{\zeta^i}{i!},$$

where ζ follows the gamma distribution

$$f(\zeta) = \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} \zeta^\alpha e^{-\zeta/\beta}.$$

The parameter $\alpha > -1$, $\beta = 1$, and $\Gamma(\cdot)$ is the gamma function. In Fig. 2(a), the parameter α was set to 1.

3. The Jacobi polynomial and a beta distribution

The n th ($n = 1, 2, \dots$)-order Jacobi polynomial $P_n^{(\alpha,\beta)}(\zeta) := P_n$ is given by

$$P_{n+1} = \frac{(\gamma_n + 1)(\gamma_n(\gamma_n + 2)\zeta + \alpha^2 - \beta^2)P_n}{2(n+1)(\gamma_n - n + 1)\gamma_n} - \frac{2(n+\alpha)(n+\beta)(\gamma_n + 2)}{2(n+1)(\gamma_n - n + 1)\gamma_n} P_{n-1},$$

where $\gamma = 2n + \alpha + \beta$ and ζ follows a beta distribution in the range of $[-1, 1]$:

$$f(\zeta) = \frac{1}{B(\alpha, \beta)} \zeta^{\alpha-1} (1-\zeta)^{\beta-1},$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

In Fig. 2(a), the parameters (α, β) were set to $(-0.25, -0.25)$.

4. The Legendre polynomial and a uniform distribution

The n th ($n = 1, 2, \dots$)-order Legendre polynomial $P_n(\zeta)$ is given by

$$P_n(\zeta) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{2^n} \binom{n}{k} \binom{2n-2k}{n} \zeta^{n-2k},$$

where $\lfloor \cdot \rfloor$ represents the floor function and ζ follows a uniform distribution in the range of $[-1, 1]$:

$$f(\zeta) = \frac{1}{2}.$$

5. The Charlier polynomial and a Poisson distribution

The n th ($n = 1, 2, \dots$)-order Charlier polynomial $C_n(\zeta; a)$ is given by

$$C_n(\zeta; a) = \sum_{i=0}^n \binom{n}{i} \binom{\zeta}{i} i! (-a)^{n-i},$$

where ζ follows the Poisson distribution

$$f(\zeta) = \frac{e^{-a} a^\zeta}{\zeta!}.$$

In Fig. 2(a), the parameter a was set to 6.

6. The Krawtchouk polynomial and a binomial distribution

The n th ($n = 1, 2, \dots$)-order Krawtchouk polynomial $K_n(\zeta; p, N)$ is given by

$$K_n(\zeta; p, N) = \sum_{i=0}^n (-1)^{n-i} \binom{N-\zeta}{n-i} \binom{\zeta}{i} p^{n-i} (1-p)^i,$$

where ζ follows the binomial distribution

$$f(\zeta) = \binom{N}{\zeta} p^\zeta (1-p)^{N-\zeta}.$$

In Fig. 2(a), the parameters (p, N) were set to $(0.5, 10)$.

7. The Meixner polynomial and a negative binomial distribution

The n th ($n = 1, 2, \dots$)-order Meixner polynomial $M_n(\zeta; \beta, c) := M_n$ is given by

$$M_{n+1} = \frac{\{(c-1)\zeta + n + c(n+\beta)\}M_n - nM_{n-1}}{c(n+\beta)},$$

where $M_0 = 1$ and $M_{-1} = 0$. ζ follows the negative binomial distribution

$$f(\zeta; \beta, c) = \binom{\zeta-1}{\beta-1} (1-c)^\beta c^{\zeta-\beta}.$$

In Fig. 2(a), the parameters (c, β) were set to $(0.2, 10)$.

8. The Hahn polynomial and a hypergeometric distribution

The n th-order Hahn polynomial $Q_n(\zeta; \alpha, \beta, N) := Q_n$ is given by

$$Q_{n+1} = \frac{A_n + C_n - \zeta}{A_n} Q_n - \frac{C_n}{A_n} Q_{n-1},$$

$$A_n = \frac{(n + \alpha + \beta + 1)(n + \alpha + 1)(N - n)}{(2n + \alpha + \beta + 1)(2n + \alpha + \beta + 2)},$$

$$C_n = \frac{n(n + \alpha + \beta + N + 1)(n + \beta)}{(2n + \alpha + \beta)(2n + \alpha + \beta + 1)},$$

where $Q_0 = 1$ and $Q_{-1} = 0$. ζ follows a hypergeometric distribution with $m = -\alpha - 1$ and $n = -\beta - 1$:

$$f(\zeta; m, n, N) = \binom{m}{\zeta} \binom{n}{N-\zeta} p^\zeta (1-p)^{N-\zeta}.$$

In Fig. 2(a), the parameters (m, n, N) were set to $(100, 50, 20)$.

APPENDIX B: THRESHOLD OF IPC

To remove the estimation error, $C(X, z)$ was set to zero if it was smaller than the threshold ϵ . The modified capacity $C_\epsilon(X, z)$ is represented by

$$C_\epsilon(X, z) = \begin{cases} C(X, z) & (\text{if } C(X, z) \geq \epsilon) \\ 0 & (\text{otherwise}) \end{cases}.$$

The threshold is determined using random shuffle surrogates. We prepared $N (= 200)$ input time series that were shuffled in the time direction and calculated IPCs using the shuffled inputs to obtain N capacities. Furthermore, we let the significance level be $\alpha (= 1)\%$ and chose the original IPC, which exceeded 1.2–3 times the value in the top $\alpha/2\%$ of N capacities. The above operation was performed for each z_t , and significant IPCs were obtained.

APPENDIX C: LYAPUNOV EXPONENTS OF THE LIMIT CYCLE SYSTEM

According to Eqs. (44)–(46), the Jacobian matrix J_t of $\mathbf{r} = [r \ \theta]^\top$ is represented by

$$J_t = \frac{\partial \mathbf{r}_{t+1}}{\partial \mathbf{r}_t} = \begin{bmatrix} 1 + \tau - 3\tau r_t^2 & 0 \\ 0 & 1 \end{bmatrix},$$

where $\tau = 0.1$. Since J_t is the diagonal matrix, the eigenvalues of $J_t^\top J_t$ for r and θ are $\rho_r = (1 + \tau - 3\tau r_t^2)^2$ and

$\rho_\theta = 1$, respectively. Using these eigenvalues, the Lyapunov exponents for r and θ , λ_r and λ_θ , are obtained as follows:

$$\lambda_r = \frac{1}{2T} \sum_{t=1}^T \ln \rho_r, \tag{C1}$$

$$\lambda_\theta = \frac{1}{2T} \sum_{t=1}^T \ln \rho_\theta, \tag{C2}$$

where $T (= 1.5 \times 10^6)$ denotes the total time. Using Eq. (C1) and r_t ($t = 1, \dots, T$) computed from Eqs. (44)–(46), we obtained $\lambda_r = -0.577$. In addition, according to $\rho_\theta = 1$ and Eq. (C2), we obtained $\lambda_\theta = 0$. From the above results, we concluded that the maximum Lyapunov exponent of the limit cycle system was zero.

APPENDIX D: CLASSIFICATION OF THE NARMA10 MODEL

1. Attractor analysis

To classify the NARMA10 model based on Fig. 1, we studied the divergence and time-dependence of the NARMA10 model. First, we analyzed an attractor of the model without input. We defined new variables for a time-delay system as $z_t^{(s)} \equiv y_{t+1-s}$ ($s = 1, 2, \dots, 10$), and the model without input ($\mu = \kappa = 0$) was rewritten as follows:

$$z_{t+1}^{(1)} = \alpha z_t^{(1)} + \beta z_t^{(1)} \sum_{i=1}^{10} z_t^{(i)} + \delta, \tag{D1}$$

$$z_{t+1}^{(s)} = z_t^{(s-1)} \quad (s = 2, 3, \dots, 10). \tag{D2}$$

Equation (D1) makes use of $z_t^{(1)}$ and $\sum_{i=1}^{10} z_t^{(i)}$; thus, we defined $w_t^{(1)} = z_t^{(1)}$ and $w_t^{(2)} = \sum_{i=1}^{10} z_t^{(i)}$, and the discrete derivatives $\Delta w_t^{(1)} (= w_{t+1}^{(1)} - w_t^{(1)})$ and $\Delta w_t^{(2)} (= w_{t+1}^{(2)} - w_t^{(2)})$ were derived from Eqs. (D1) and (D2) as follows:

$$\Delta w_t^{(1)} = (\alpha - 1)w_t^{(1)} + \beta w_t^{(1)} w_t^{(2)} + \delta, \tag{D3}$$

$$\Delta w_t^{(2)} = \alpha w_t^{(1)} + \beta w_t^{(1)} w_t^{(2)} + \delta - z_t^{(10)}. \tag{D4}$$

From Eqs. (D3) and (D4), the nullclines can be obtained:

$$w_t^{(2)} = \frac{1 - \alpha}{\beta} - \frac{\delta}{\beta w_t^{(1)}},$$

$$w_t^{(2)} = -\frac{\alpha}{\beta} + \frac{z_t^{(10)} - \delta}{\beta w_t^{(1)}}.$$

Note that the intersection point of the nullclines is a saddle point $(w_t^{(1)}, w_t^{(2)}) = (z_t^{(10)}, (1 - \alpha)/\beta - \delta/\beta z_t^{(10)})$. Furthermore, we plotted the nullclines on the $w_t^{(1)}-w_t^{(2)}$ plane to examine the increase and decrease in $w_t^{(1)}$ and $w_t^{(2)}$ [Fig. 8(b) and Table V]. Since $z_t^{(10)}$ depends on the time step t , the nullcline changes over time and three regimes can be distinguished: $z_t^{(10)} < \delta$, $z_t^{(10)} = \delta$, and $z_t^{(10)} > \delta$ [Figs. 8(b) and 9]. The signs of $\Delta w^{(1)}$ and $\Delta w^{(2)}$ are shown in Tables V–VII. From the above results, we found that the model with no input has a fixed-point attractor.

2. Divergence

Next, we examined the conditions under which the NARMA10 model diverges. Since the nullclines

TABLE V. The derivative table of $\Delta w_t^{(1)}$ and $\Delta w_t^{(2)}$ in the stability diagram [$z_t^{(10)} > \delta$; Fig. 8(b)].

# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$
(i)	-	-	(iii)	+	+	(v)	-	+	(vii)	+	+
(ii)	+	-	(iv)	-	-	(vi)	+	-			

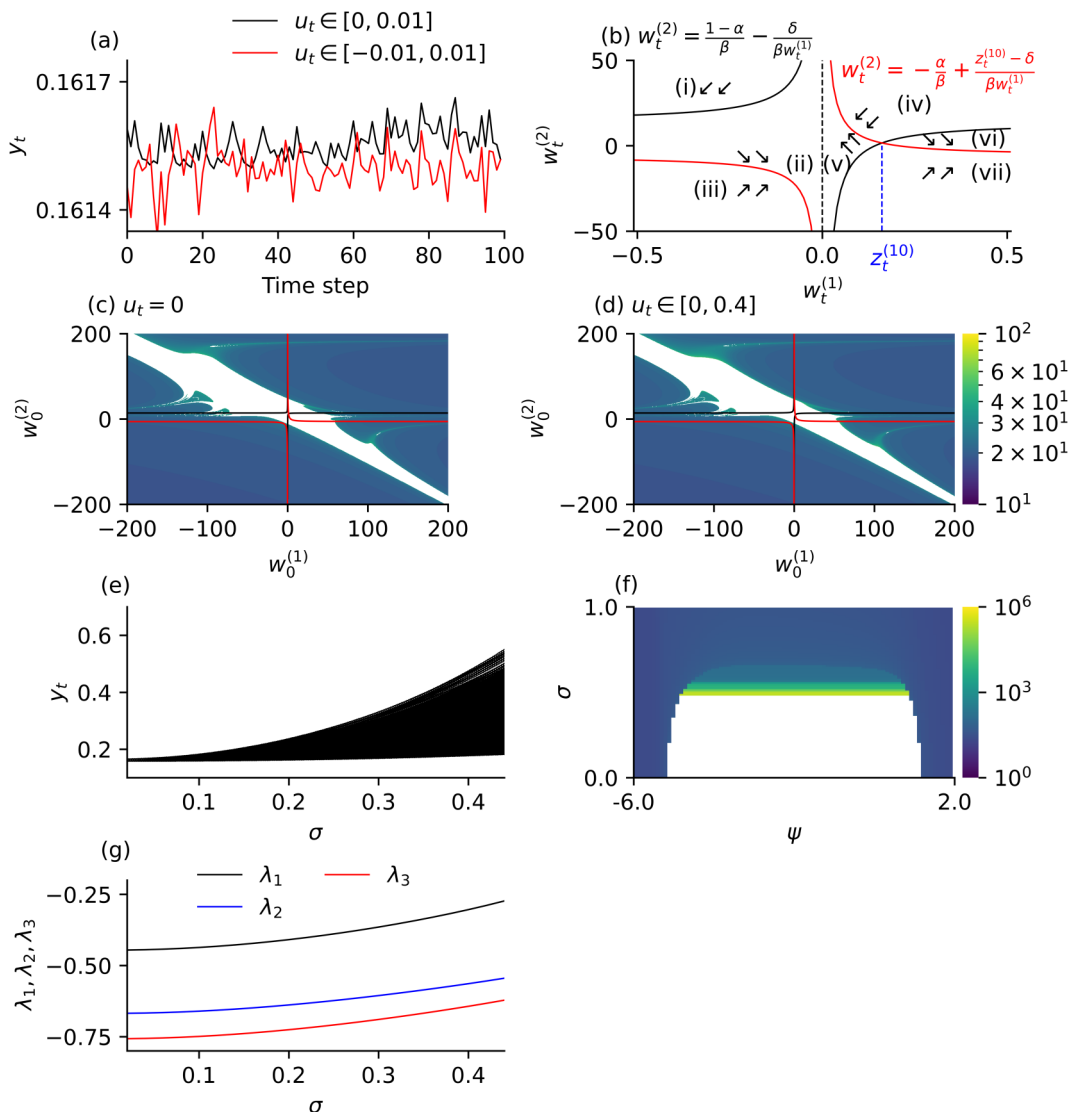


FIG. 8. Properties of the NARMA10 dynamical system with $(\alpha, \beta, \gamma, \delta) = (0.3, 0.05, 1.5, 0.1)$. (a) shows the time series of Eq. (48) with $u_t \in [0, 0.01]$ (black) and $u_t \in [-0.01, 0.01]$ (red), and (b) the stability diagram of $w_t^{(1)} = z_t^{(1)}$ and $w_t^{(2)} = \sum_{i=1}^{10} z_t^{(i)}$ as $z_t^{(10)} > \delta$. The two nullclines, $w_t^{(2)} = \frac{1-\alpha}{\beta} - \frac{\delta}{\beta w_t^{(1)}}$ (solid black) and $w_t^{(2)} = -\frac{\alpha}{\beta} + \frac{z_t^{(10)} - \delta}{\beta w_t^{(1)}}$ (solid red), are displayed. The signs of $\Delta w_t^{(1)}$ and $\Delta w_t^{(2)}$ are shown in Table V. Note that the point $(z_t^{(10)}, (1-\alpha)/\beta - \delta/\beta z_t^{(10)})$ is a saddle point. (c) and (d) show the basins of attraction of Eq. (48) relative to $w_0^{(1)}$ and $w_0^{(2)}$ with $u_t = 0$ and $u_t \in [0, 0.4]$, respectively, as well as the two nullclines from (b). In (c), (d), and (f), each dot shows the time step at which y_t diverges to infinity, and the white dot represents the initial values for which y_t does not do so. (e) illustrates the bifurcation diagram of Eq. (48), while (f) shows the basin of attraction of Eq. (48) relative to σ and ψ , where the initial values were $y_i = \psi$ ($i = 0, 1, \dots, 9$). The three largest Lyapunov spectra λ_i ($i = 1, 2, 3$) relative to σ are shown in (g).

TABLE VI. The derivative table of $\Delta w_t^{(1)}$ and $\Delta w_t^{(2)}$ in the stability diagram [$z_t^{(10)} = \delta$; Fig. 9(a)].

# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$
(i)	-	-	(iii)	+	+	(v)	-	+	(vii)	+	+
(ii)	+	-	(iv)	-	-	(vi)	+	-			

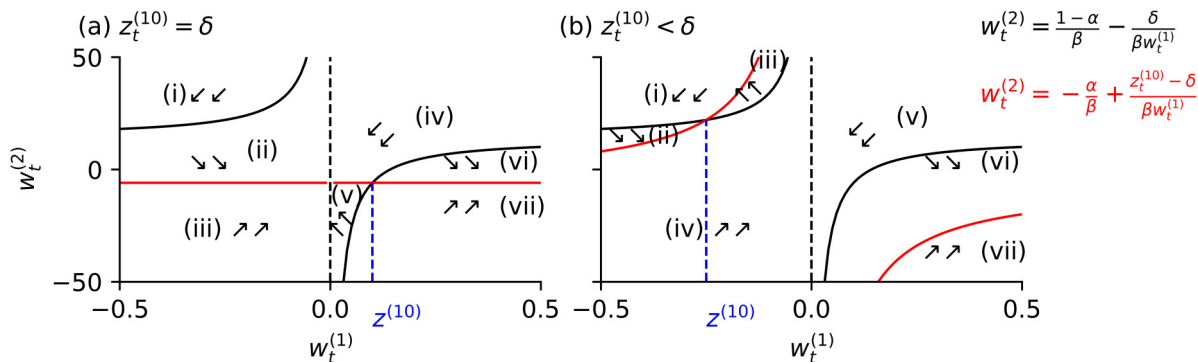


FIG. 9. Stability diagrams. The solid black and solid red lines represent $\Delta w_t^{(1)} = 0$ and $\Delta w_t^{(2)} = 0$, respectively. The signs of $\Delta w_t^{(1)}$ and $\Delta w_t^{(2)}$ in (a) and (b) are shown in Tables VI and VII, respectively. In (a) and (b), the stability diagrams of $w_t^{(1)}$ and $w_t^{(2)}$ when $z_t^{(10)} = \delta$ and $z_t^{(10)} > \delta$, respectively, are shown. Note that the intersection point of the nullclines is $(w_t^{(1)}, w_t^{(2)}) = (z_t^{(10)}, (1 - \alpha)/\beta - \delta/\beta z_t^{(10)})$.

are time-varying due to $z_t^{(10)}$, the increase or decrease of $w_t^{(1)}$ and $w_t^{(2)}$ in each area indicated by Fig. 8(b) and Table V can change. Thus, in the case of no input ($\mu = \kappa = 0$), we examined the time step at which y_t diverges with the initial values of $w_0^{(1)}$ and $w_0^{(2)}$. The initial values of y_t were set as $y_0 = w_0^{(1)}$ and $y_1 = y_2 = \dots = y_9 = (w_0^{(2)} - w_0^{(1)})/9$. As shown in Fig. 8(c), the combination of initial values $(w_0^{(1)}, w_0^{(2)})$ where y_t does not diverge is distributed in a complex manner. We attribute the difference between the theoretical and numerical basins of attraction to the change in $z_t^{(10)}$, which produced a complicated distribution of the nullclines [see Figs. 8(b) and 9, and Tables V–VII].

In the presence of input, y_t diverges due to the initial condition. As shown in Fig. 8(d), the model with input $u_t \in [0, 0.4]$ ($\mu = \kappa = 0.2$) diverges at an initial value similar to the one with no input. Figure 8(e) shows the bifurcation diagram of y_t given an initial value at which y_t converges to the fixed point. Since the fixed point around which y_t fluctuates is the saddle point, y_t can diverge with the given input. To investigate the divergence conditions caused by the input, we ran the model over 10^6 time steps and examined the time step at which y_t diverges by altering the initial values $y_0 = \dots = y_9 = \psi$ and input intensity σ . Figure 8(f) shows that y_t diverges to infinity when ψ or σ exceeds three thresholds: (i) $\psi < -5.15$, (ii) $1.239 < \psi$, and (iii) $\sigma > 0.45$. In the (i) and (ii) cases, y_t diverges at a shorter time step ($t < 100$) than (iii) because the divergence is caused by the initial value ψ . However, in case of (iii), y_t diverges according to the input. After converging to the fixed point, y_t can diverge when successively receiving large positive inputs. According to the time steps in Fig. 8(f), when y_t diverges as the runtime becomes longer, y_t diverges with a smaller σ . Consequently, the threshold (iii) is the boundary that depends on the input time series. These results suggest that y_t , in the vicinity of the fixed point, can diverge, depending on the input.

In Fig. 3, p expresses the probability of not diverging as a function of σ for different random series $\{\zeta_t\}$. As previously shown, once y_t converges to the fixed point, the model stochastically diverges due to the input, and the probability p depends on σ ; thus, even though p is high, y_t can potentially diverge. For example, two typical ranges of input, $u_t \in [0, 0.5]$ and $[0, 1]$, have been used for the NARMA10 task; however, y_t diverges in both cases [Fig. 3(b), $\sigma = 0.5, 1.0$] because σ is relevant to the average time until divergence, and the runtime, 10^6 time steps, is much longer than the time used for the benchmark task. Consequently, the divergence probability of y_t , p , changes depending on the parameter σ . Therefore, although the NARMA10 model produces the fixed-point attractor, it can potentially diverge depending on the initial values, input time series, and parameter settings.

3. Time-variance analysis

Finally, we investigated the time dependence of the NARMA10 model. The state of a dynamical system receiving noise input can transit from order to stochastic chaos [60], an effect referred to as noise-induced chaos [69]. As chaotic behavior is exhibited by a time-variant system, we investigated whether the system is chaotic or ordered by calculating the maximum Lyapunov exponent λ_1 . Thus we derived the Lyapunov spectrum of the model λ_i ($i = 1, 2, \dots, 10$) based on the ten-dimensional time-delay system in Eqs. (D1) and (D2). When we expressed $z_t^{(s)}$ ($s = 1, \dots, 10$) as a vector $\mathbf{z}_t = [z_t^{(1)} \dots z_t^{(10)}]^T$, the Jacobian matrix of Eq. (48) $\mathbf{J}_t \in \mathbb{R}^{10 \times 10}$ with respect to \mathbf{z}_t can be written as follows:

$$\mathbf{J}_t = \frac{\partial \mathbf{z}_{t+1}}{\partial \mathbf{z}_t} = \begin{bmatrix} X_t & Y_t & \dots & Y_t & Y_t \\ 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \end{bmatrix}, \quad (\text{D5})$$

TABLE VII. The derivative table of $\Delta w_t^{(1)}$ and $\Delta w_t^{(2)}$ in the stability diagram [$z_t^{(10)} < \delta$; Fig. 9(b)].

# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$	# area	$\Delta w_t^{(1)}$	$\Delta w_t^{(2)}$
(i)	−	−	(iii)	−	+	(v)	−	−	(vii)	+	+
(ii)	+	−	(iv)	+	+	(vi)	+	−			

where $X_t = \alpha + 2\beta z_t^{(1)} + \beta \sum_{i=2}^{10} z_t^{(i)}$ and $Y_t = \beta z_t^{(1)}$. Using Eq. (D5), the Lyapunov spectrum was computed as follows:

$$\lambda_i = \frac{1}{2T} \sum_{t=1}^T \ln \rho_i(\mathbf{J}_{t+M-1} \mathbf{J}_{t+M-2} \cdots \mathbf{J}_t)$$

$$(i = 1, 2, \dots, 10),$$

where $\rho_i(\mathbf{J}_{t+M-1} \mathbf{J}_{t+M-2} \cdots \mathbf{J}_t)$ is the i th singular value of matrix $\mathbf{J}_{t+M-1} \mathbf{J}_{t+M-2} \cdots \mathbf{J}_t$, while T and M were set to 6000 and 40, respectively. Figure 8(g) shows the three largest Lyapunov spectra, λ_1 , λ_2 , and λ_3 , all of which are negative relative to σ , indicating that the system does not demonstrate stochastic chaos. Therefore the NARMA10 model is neither a stochastically chaotic nor a time-variant system.

The divergence and time-invariance analysis results revealed that the NARMA10 model converges to the fixed point and varies in the vicinity of the point. We considered the y_t fluctuating around the fixed point to be time-invariant.

APPENDIX E: INTERCONVERTIBILITY OF THE PC EXPANSION AND IPC

To clearly demonstrate that the PC expansion and IPC are interconvertible, we derived an approximate model that has nearly the same decomposition of the IPC as the original decomposition using Legendre chaoses. From the above capacity analysis, we narrowed the polynomial terms to $P_1(\zeta_{t-s})$ ($s = 1, 2, \dots$) and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$ ($s = 1, 2, \dots$), which yielded significantly greater capacities. The expanded state of y_t in Eqs. (48) and (49) is expressed as follows:

$$y_t = p + \sum_{s \in \mathcal{N}_1} q_s P_1(\zeta_{t-s}) + \sum_{s \in \mathcal{N}_2} r_s P_1(\zeta_{t-s}) P_1(\zeta_{t-s-9}), \quad (E1)$$

where p , q_s , and r_s are coefficients for the Legendre chaoses $P_0 = 1$, $P_1(\zeta_{t-s})$, and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$, respectively, while $P_1(\zeta) = \zeta$. ζ_t follows a uniform distribution in $[-1, 1]$, and \mathcal{N}_1 and \mathcal{N}_2 represent the sets of delayed time steps s for $P_1(\zeta_{t-s})$ and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$, respectively. Let the normalized Legendre chaoses be $\phi_0 = \frac{1}{\sqrt{T}}$, $\phi_{1,t}^{(s)} = \frac{P_1(\zeta_{t-s})}{\sqrt{\sum_{t=1}^T P(\zeta_{t-s})^2}}$, and $\phi_{2,t}^{(s)} = \frac{P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})}{\sqrt{\sum_{t=1}^T \{P(\zeta_{t-s})P(\zeta_{t-s-9})\}^2}}$, and the state in Eq. (E1) is represented as follows:

$$y_t = \hat{p}\phi_0 + \sum_{s \in \mathcal{N}_1} \hat{q}_s \phi_{1,t}^{(s)} + \sum_{s \in \mathcal{N}_2} \hat{r}_s \phi_{2,t}^{(s)}, \quad (E2)$$

where $\hat{p} = p\sqrt{T}$, $\hat{q}_s = q_s \sqrt{\sum_{t=1}^T P(\zeta_{t-s})^2}$, and $\hat{r}_s = r_s \sqrt{\sum_{t=1}^T \{P(\zeta_{t-s})P(\zeta_{t-s-9})\}^2}$ are the modified coefficients for p , q_s , and r_s , respectively.

Debiasing the state in Eq. (E2) and using Eq. (6), the IPCs for $\phi_{1,t}^{(s)}$ and $\phi_{2,t}^{(s)}$ become

$$C = \frac{q_s^2}{\sum_{s \in \mathcal{N}_1} q_s^2 + \sum_{s \in \mathcal{N}_2} r_s^2} \quad (E3)$$

and

$$C = \frac{r_s^2}{\sum_{s \in \mathcal{N}_1} q_s^2 + \sum_{s \in \mathcal{N}_2} r_s^2}, \quad (E4)$$

respectively. Therefore Eqs. (E3) and (E4) show that each IPC is the normalized squared coefficient in the polynomial chaos expansion. To demonstrate this model, we employ $\mathcal{N}_1 = \{1, 2, 3, 10, 11, 12\}$ and $\mathcal{N}_2 = \{1, 2, 3\}$. As shown in Fig. 10(a) and 10(b), for $u_t \in [-\sigma, \sigma]$ and $[0, \sigma]$, respectively, the approximate model successfully reproduced the original NARMA10 model. Furthermore, as in Fig. 10(c) and 10(d), we confirmed that the IPC decomposition of the approximate model reproduced the original decomposition.

1. Derivation

Here, we derive the following equations:

$$y_t = p + \sum_{s \in \mathcal{N}_1} q_s P_1(\zeta_{t-s}) + \sum_{s \in \mathcal{N}_2} r_s P_1(\zeta_{t-s}) P_1(\zeta_{t-s-9}), \quad (E5)$$

$$p = \frac{1-\alpha}{20\beta} - \sqrt{\left(\frac{1-\alpha}{20\beta}\right)^2 - \frac{\gamma\mu^2 + \delta}{10\beta}}, \quad (E6)$$

$$q_s = \begin{cases} \gamma\mu\kappa (s = 1) \\ (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^{s-2} \beta p q_{s-j-1} \\ (s = 2, 3, \dots, 9) \\ \gamma\mu\kappa + (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^{s-2} \beta p q_{s-j-1} \\ (s = 10) \\ (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^9 \beta p q_{s-j-1} \\ (s = 11, 12, \dots) \end{cases}, \quad (E7)$$

$$r_s = \begin{cases} \gamma\kappa^2 (s = 1) \\ (\alpha + 10\beta p)r_{s-1} + \beta q_{s-1} \sum_{j=0}^9 q_{s+8-j} \\ + \beta \sum_{j=0}^{s-2} (pr_{s-j-1} + q_{s+8}q_{s-j-1}) \\ (s = 2, 3, \dots, 10) \\ (\alpha + 10\beta p)r_{s-1} + \beta q_{s-1} \sum_{j=0}^9 q_{s+8-j} \\ + \beta \sum_{j=0}^9 (pr_{s-j-1} + q_{s+8}q_{s-j-1}) \\ (s = 11, 12, \dots) \end{cases}, \quad (E8)$$

where \mathcal{N}_1 and \mathcal{N}_2 represent the sets of delayed time steps $s \in \mathbb{N}$ for $P_1(\zeta_{t-s})$ and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$, respectively.

First, y_t in Eq. (48) is expanded using the Legendre chaoses of input time series ζ_{t-s} ($s = 1, 2, \dots, t$) with time-varying coefficients, as follows:

$$y_t = p_t + \sum_{s=1}^t q_{t,s} P_1(\zeta_{t-s}) + \sum_{s=1}^t r_{t,s} P_1(\zeta_{t-s}) P_1(\zeta_{t-s-9}) + \dots, \quad (E9)$$

where p_t denotes a time-varying term independent of ζ_{t-s} ($s = 1, 2, \dots, t$), and $q_{t,s}$ and $r_{t,s}$ are the s th coefficients of $P_1(\zeta_{t-s})$ and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$, respectively. The NARMA10 model with ζ_t can be expressed as

$$y_{t+1} = \alpha y_t + \beta y_t \sum_{j=0}^9 y_{t-j} + \gamma(\mu + \sigma \zeta_t)(\mu + \sigma \zeta_{t-9}) + \delta. \quad (E10)$$

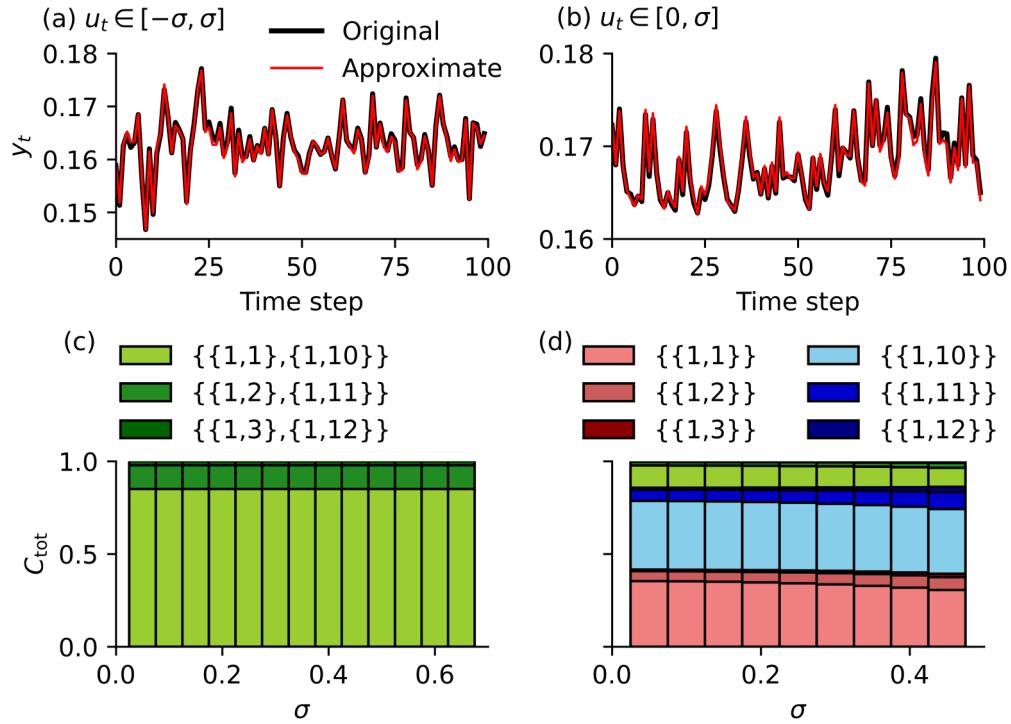


FIG. 10. The proposed model for the benchmark task. (a) and (b) show the time-series data with $u_t \in [-\sigma, \sigma]$ and $[0, \sigma]$, respectively. The solid black and solid red lines represent the output of the NARMA10 and our models, respectively. (c) and (d) show the IPC with $u_t \in [-\sigma, \sigma]$ and $[0, \sigma]$, respectively. The labels represent combinations of $\{\{n_s, s\}\}$, where n_s is the degree of the polynomial and s is the delayed time step of the input. Here, the desired output is $\prod_s P_{n_s}(\zeta_{t-s})$.

According to Eqs. (E9) and (E10), y_{t+1} is rewritten as

$$\begin{aligned}
 y_{t+1} = & \left(\alpha p_t + \beta p_t \sum_{j=0}^9 p_{t-j} + \gamma \mu^2 + \delta \right) \\
 & + \gamma \mu \kappa (P_1(\zeta_t) + P_1(\zeta_{t-9})) \\
 & + \alpha \sum_{s=1}^t q_{t,s} P_1(\zeta_{t-s}) \\
 & + \beta p_t \sum_{j=0}^9 \sum_{s=1}^{t-j} q_{t-j,s} P_1(\zeta_{t-s-j}) \\
 & + \beta \sum_{j=0}^9 p_{t-j} \sum_{s=1}^t q_{t,s} P_1(\zeta_{t-s}) \\
 & + \gamma \kappa^2 \zeta_t \zeta_{t-9} \\
 & + \alpha \sum_{s=1}^t r_{t,s} P_1(\zeta_{t-s}) P_1(\zeta_{t-s-9}) \\
 & + \beta p_t \sum_{j=0}^9 \sum_{s=1}^{t-j} r_{t-j,s} P_1(\zeta_{t-s-j}) P_1(\zeta_{t-s-j-9}) \\
 & + \beta \sum_{j=0}^9 p_{t-j} \sum_{s=1}^t r_{t,s} P_1(\zeta_{t-s}) P_1(\zeta_{t-s-9})
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{s=1}^t q_{t,s} P_1(\zeta_{t-s}) P_1(\zeta_{t-s}) \sum_{j=0}^9 \sum_{s=1}^{t-j} q_{t-j,s} P_1(\zeta_{t-s-j}) \\
 & + \dots, \tag{E11}
 \end{aligned}$$

where $P_1(\zeta) = \zeta$. When increasing t by one in Eq. (E9), the following equation is obtained:

$$\begin{aligned}
 y_{t+1} = & p_{t+1} + \sum_{s=1}^{t+1} q_{t+1,s} P_1(\zeta_{t+1-s}) \\
 & + \sum_{s=1}^{t+1} r_{t+1,s} P_1(\zeta_{t+1-s}) P_1(\zeta_{t+1-s-9}) + \dots \tag{E12}
 \end{aligned}$$

Equating the coefficients in Eqs. (E11) and (E12) yields

$$p_{t+1} = \alpha p_t + \sum_{j=0}^9 \beta p_t p_{t-j} + \gamma \mu^2 + \delta, \tag{E13}$$

$$q_{t+1,s} = \begin{cases} \gamma \mu \kappa (s = 1) \\ (\alpha + \beta \sum_{j=0}^9 p_{t-j}) q_{t,s-1} \\ + \sum_{j=0}^{s-2} \beta p_t q_{t-j,s-j-1} (s = 2, 3, \dots, 9) \\ \gamma \mu \kappa + (\alpha + \beta \sum_{j=0}^9 p_{t-j}) q_{t,s-1} \\ + \sum_{j=0}^{s-2} \beta p_t q_{t-j,s-j-1} (s = 10) \\ (\alpha + \beta \sum_{j=0}^9 p_{t-j}) q_{t,s-1} \\ + \sum_{j=0}^9 \beta p_t q_{t-j,s-j-1} (s = 11, 12, \dots) \end{cases}, \tag{E14}$$

$$r_{t+1,s} = \begin{cases} \gamma\kappa^2 (s=1) \\ (\alpha + \beta \sum_{j=0}^9 p_{t-j})r_{t,s-1} \\ + \beta q_{t,s-1} \sum_{j=0}^9 q_{t-j,s+8-j} \\ + \beta \sum_{j=0}^{s-2} (p_t r_{t-j,s-j-1} + q_{t,s+8} q_{t-j,s-j-1}) \\ (s=2, 3, \dots, 10) \\ (\alpha + \beta \sum_{j=0}^9 p_{t-j})r_{t,s-1} \\ + \beta q_{t,s-1} \sum_{j=0}^9 q_{t-j,s+8-j} \\ + \beta \sum_{j=0}^9 (p_t r_{t-j,s-j-1} + q_{t,s+8} q_{t-j,s-j-1}) \\ (s=11, 12, \dots) \end{cases} \quad (E15)$$

According to Eq. (E13), p_t has both a stable and an unstable equilibrium point. If $p_t < (1 - \alpha)/20\beta + \sqrt{((1 - \alpha)/20\beta)^2 - (\gamma\mu^2 + \delta)/10\beta}$, it converges to the stable point. When t is large enough, p_t converges to

$$p = \lim_{t \rightarrow \infty} p_t = \frac{1 - \alpha}{20\beta} - \sqrt{\left(\frac{1 - \alpha}{20\beta}\right)^2 - \frac{\gamma\mu^2 + \delta}{10\beta}}.$$

According to Eq. (E14), $q_{t,s}$ also converges to

$$q_s = \lim_{t \rightarrow \infty} q_{t+1,s} = \begin{cases} \gamma\mu\kappa (s=1) \\ (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^{s-2} \beta p q_{s-j-1} \\ (s=2, 3, \dots, 9) \\ \gamma\mu\kappa + (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^{s-2} \beta p q_{s-j-1} \\ (s=10) \\ (\alpha + 10\beta p)q_{s-1} + \sum_{j=0}^9 \beta p q_{s-j-1} \\ (s=11, 12, \dots) \end{cases}.$$

In the same manner, $r_{t,s}$ converges to

$$r_s = \lim_{t \rightarrow \infty} r_{t+1,s} = \begin{cases} \gamma\kappa^2 (s=1) \\ (\alpha + 10\beta p)r_{s-1} + \beta q_{s-1} \sum_{j=0}^9 q_{s+8-j} \\ + \beta \sum_{j=0}^{s-2} (p r_{s-j-1} + q_{s+8} q_{s-j-1}) \\ (s=2, 3, \dots, 10) \\ (\alpha + 10\beta p)r_{s-1} + \beta q_{s-1} \sum_{j=0}^9 q_{s+8-j} \\ + \beta \sum_{j=0}^9 (p r_{s-j-1} + q_{s+8} q_{s-j-1}) \\ (s=11, 12, \dots) \end{cases}.$$

Therefore, when t is large enough, and Eq. (E9) is approximated with the constant term p and the Legendre chaoses of

$P_1(\zeta_{t-s})$ and $P_1(\zeta_{t-s})P_1(\zeta_{t-s-9})$, whose delayed time steps s are limited to sets \mathcal{N}_1 and \mathcal{N}_2 , respectively, Eqs. (E5)–(E8) are obtained.

APPENDIX F: HOW TO COMPOSE THE DISSOCIATED CULTURE RESERVOIR

All experiments were approved by the ethical committee of the University of Tokyo and followed the ‘‘Guiding Principles for the Care and Use of Animals in the Field of Physiological Science’’ established by the Physiological Society of Japan. Embryonic rat cortices were dissected from E18 rats and used for cortical cell cultures. The cortices were dissociated in 2 mL of 0.25% trypsin-ethylenediaminetetraacetic acid (Trypsin-EDTA, Life Technologies), from which cells were isolated by trituration, and 38,000 cells were seeded on each microelectrode array (MEA; MaxWell Biosystems). For cell adhesion, 5 mL of 0.05% Polyethyleneimine (PEI; Sigma-Aldrich) and 5 μ l of 0.02 mg/ml Laminin (Sigma-Aldrich) were used before plating the cells. Then, after 24 hours, the plating media [70] were changed to growth media [71]. The plating media were composed of Neurobasal 850 μ l (Life Technologies), 10% horse serum (HyClone), 0.5 mM GlutaMAX (Life Technologies), and 2% B27 (Life Technologies). The growth media were composed of DMEM 850 μ l (Life Technologies), 10% horse serum (HyClone), 0.5 mM GlutaMAX (Life Technologies), and 1 mM sodium pyruvate (Life Technologies). All experiments were conducted in an incubator at 37 $^{\circ}$ C with 5% CO₂. The MEAs were sealed with a lid to prevent water evaporation and invasion of bacteria and fungus.

The MEA had 26,400 electrodes, which were placed 17.5 μ m apart and arranged in a 120 \times 220 grid. The MEA can simultaneously use up to 1,024 of 26,400 electrodes. We selected electrodes with a high firing rate as measurement electrodes and, for simulation electrodes, electrodes on which the axon places. We applied bipolar pulse stimuli that had an amplitude of ζ_t , which followed a normal distribution with mean μ and standard deviation σ , and had an inter-pulse interval (IPI) of 10, 20, and 30 ms to the stimulation electrodes. Furthermore, a 6th-order Butterworth bandpass filter and zero-phase IIR filter were applied to the voltage traces observed from the measurement electrodes to extract 300–3000 Hz components. At all electrodes, stimulus-induced artifacts were removed by eliminating traces ± 2 ms from the stimulus times. The standard deviation of extracted signals was calculated as follows [72]:

$$\sigma = \text{median} \left\{ \frac{|x|}{0.6745} \right\}.$$

If the amplitude of an extracted signal exceeded 4σ , the value of the spike train was set to one; otherwise, it was set to zero. As the measurement frequency was 20 kHz, the above spike train was separated by a 1-ms time bin, and if one or more spikes appeared in one bin, the modified spike train was set to one; otherwise, it was set to zero. The train was divided into bins by IPI width, and the number of spikes in the bin was used for the state x_t .

- [1] A. Z. Stieg, A. V. Avizienis, H. O. Sillin, C. Martin-Olmos, M. Aono, and J. K. Gimzewski, Emergent criticality in complex turing b-type atomic switch networks, *Adv. Mater.* **24**, 286 (2012).
- [2] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [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] D. Verstraeten, B. Schrauwen, M. d’Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, *Neural Netw.* **20**, 391 (2007).
- [5] K. Nakajima and I. Fischer, *Reservoir Computing* (Springer, Singapore, 2021).
- [6] 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).
- [7] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states, *Nat. Commun.* **4**, 1364 (2013).
- [8] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Experimental demonstration of reservoir computing on a silicon photonics chip, *Nat. Commun.* **5**, 3541 (2014).
- [9] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Reservoir computing using dynamic memristors for temporal information processing, *Nat. Commun.* **8**, 2204 (2017).
- [10] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima *et al.*, Neuromorphic computing with nanoscale spintronic oscillators, *Nature (London)* **547**, 428 (2017).
- [11] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, Temporal data classification and forecasting using a memristor-based reservoir computing system, *Nat. Electr.* **2**, 480 (2019).
- [12] M. J. Marinella and S. Agarwal, Efficient reservoir computing with memristors, *Nat. Electr.* **2**, 437 (2019).
- [13] K. Lüdge and A. Röhm, Computing with a camera, *Nat. Mach. Intell.* **1**, 551 (2019).
- [14] 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.
- [15] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, Re-visiting the echo state property, *Neural Netw.* **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] A. Maritan and J. R. Banavar, Chaos, Noise, and Synchronization, *Phys. Rev. Lett.* **72**, 1451 (1994).
- [18] R. Toral, C. R. Mirasso, E. Hernández-García, and O. Piro, Analytical and numerical studies of noise-induced synchronization of chaotic systems, *Chaos: An Interdiscip. J. Nonlin. Sci.* **11**, 665 (2001).
- [19] Z. Lu, B. R. Hunt, and E. Ott, Attractor reconstruction by machine learning, *Chaos: An Interdiscip. J. Nonlin. Sci.* **28**, 061104 (2018).
- [20] M. Inubushi, K. Yoshimura, Y. Ikeda, and Y. Nagasawa, On the characteristics and structures of dynamical systems suitable for reservoir computing, in *Reservoir Computing* (Springer, Berlin, 2021), pp. 97–116.
- [21] S. Boyd and L. Chua, Fading memory and the problem of approximating nonlinear operators with volterra series, *IEEE Trans. Circuit Syst.* **32**, 1150 (1985).
- [22] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover, Mineola, NY, 1959).
- [23] W. Maass and H. Markram, On the computational power of circuits of spiking neurons, *J. Comput. Syst. Sci.* **69**, 593 (2004).
- [24] W. Maass, Liquid state machines: motivation, theory, and applications, in *Computability in Context: Computation and Logic in the Real World* (World Scientific, Singapore, 2011), pp. 275–296.
- [25] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Boosting Computational Power Through Spatial Multiplexing in Quantum Reservoir Computing, *Phys. Rev. Appl.* **11**, 034021 (2019).
- [26] K. Fujii and K. Nakajima, Harnessing Disordered-Ensemble Quantum Dynamics for Machine Learning, *Phys. Rev. Appl.* **8**, 024030 (2017).
- [27] Q. H. Tran and K. Nakajima, Higher-order quantum reservoir computing, [arXiv:2006.08999](https://arxiv.org/abs/2006.08999).
- [28] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, *Sci. Rep.* **2**, 287 (2012).
- [29] M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer, Evolving carbon nanotube reservoir computers, in *International Conference on Unconventional Computation and Natural Computation* (Springer, Berlin, 2016), pp. 49–61.
- [30] K. Nakajima, H. Hauser, R. Kang, E. Guglielmino, D. G. Caldwell, and R. Pfeifer, Computing with a muscular-hydrostat system, in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation* (IEEE, New York, 2013), pp. 1504–1511.
- [31] K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, Exploiting the dynamics of soft materials for machine learning, *Soft Robot.* **5**, 339 (2018).
- [32] E. Torres, K. Nakajima, and I. S. Godage, Information processing capability of soft continuum arms, in *Proceedings of the 2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)* (IEEE, New York, 2019), pp. 441–447.
- [33] J. Bürger, A. Goudarzi, D. Stefanovic, and C. Teuscher, Hierarchical composition of memristive networks for real-time computing, in *Proceedings of the 2015 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH’15)* (IEEE, New York, 2015), pp. 33–38.
- [34] M. Hermans, P. Antonik, M. Haelterman, and S. Massar, Embodiment of Learning in Electro-Optical Signal Processors, *Phys. Rev. Lett.* **117**, 128301 (2016).
- [35] J. Yin, Y. Meng, and Y. Jin, A developmental approach to structural self-organization in reservoir computing, *IEEE Trans. Autonom. Mental Develop.* **4**, 273 (2012).
- [36] B. Barazani, G. Dion, J.-F. Morissette, L. Beaudoin, and J. Sylvestre, Microfabricated neuroaccelerometer: Integrating sensing and reservoir computing in mems, *J. Microelectromech. Syst.* **29**, 338 (2020).
- [37] T. Okumura, M. Tai, and M. Ando, Experimental study on parallel and analog optical reservoir computing with delayed

- feedback system for physical implementation, *Nonlin. Theory Appl. IEICE* **10**, 236 (2019).
- [38] K. Nakajima, Physical reservoir computing—an introductory perspective, *Jpn. J. Appl. Phys.* **59**, 060501 (2020).
- [39] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, *Sci. Rep.* **2**, 514 (2012).
- [40] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, Information processing capacity of spin-based quantum reservoir computing systems, *Cognitive Computation* (2020).
- [41] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, Gaussian states of continuous-variable quantum systems provide universal and versatile reservoir computing, *Commun. Phys.* **4**, 53 (2021).
- [42] N. Akashi, T. Yamaguchi, S. Tsunegi, T. Taniguchi, M. Nishida, R. Sakurai, Y. Wakao, and K. Nakajima, Input-driven bifurcations and information processing capacity in spintronic reservoirs, *Phys. Rev. Res.* **2**, 043303 (2020).
- [43] N. Wiener, The homogeneous chaos, *Am. J. Math.* **60**, 897 (1938).
- [44] D. Xiu and G. E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.* **24**, 619 (2002).
- [45] S. Oladyshkin and W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, *Reliab. Eng. Syst. Safety* **106**, 179 (2012).
- [46] R. H. Cameron and W. T. Martin, The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals, *Ann. Math.* **48**, 385 (1947).
- [47] K. Itô, Multiple wiener integral, *J. Math. Soc. Jpn.* **3**, 157 (1951).
- [48] H. Ogura, Orthogonal functionals of the poisson process, *IEEE Trans. Inf. Theory* **18**, 473 (1972).
- [49] S. Strogatz, *Nonlinear Dynamics and Chaos* (Perseus, Cambridge, MA, 2003).
- [50] A. F. Atiya and A. G. Parlos, New results on recurrent network training: Unifying the algorithms and accelerating convergence, *IEEE Trans. Neural Netw.* **11**, 697 (2000).
- [51] H. Jaeger, Adaptive nonlinear system identification with echo state networks, *Adv. Neural Inf. Proc. Syst.* **15**, 593 (2003).
- [52] A. Rodan and P. Tino, Minimum complexity echo state network, *IEEE Trans. Neural Netw.* **22**, 131 (2010).
- [53] F. M. Bianchi, L. Livi, C. Alippi, and R. Jenssen, Multiplex visibility graphs to investigate recurrent neural network dynamics, *Sci. Rep.* **7**, 44037 (2017).
- [54] F. Duport, A. Smerieri, A. Akrouf, M. Haelterman, and S. Massar, Fully analogue photonic reservoir computer, *Sci. Rep.* **6**, 22381 (2016).
- [55] M. Hermans and B. Schrauwen, Recurrent kernel machines: Computing with infinite echo state networks, *Neural Comput.* **24**, 104 (2012).
- [56] M. Inubushi and K. Yoshimura, Reservoir computing beyond memory-nonlinearity trade-off, *Sci. Rep.* **7**, 10199 (2017).
- [57] S. Kan, K. Nakajima, Y. Takeshima, T. Asai, Y. Kuwahara, and M. Akai-Kasaya, Simple Reservoir Computing Capitalizing on the Nonlinear Response of Materials: Theory and Physical Implementations, *Phys. Rev. Appl.* **15**, 024030 (2021).
- [58] H. Jaeger, *Short Term Memory in Echo State Networks* (GMD-Forschungszentrum Informationstechnik, 2001), Vol. 5.
- [59] O. L. White, D. D. Lee, and H. Sompolinsky, Short-Term Memory in Orthogonal Neural Networks, *Phys. Rev. Lett.* **92**, 148102 (2004).
- [60] This paper uses three terms that include “chaos”: homogeneous chaos, deterministic chaos, and stochastic chaos. First, homogeneous chaos refers to the stationary stochastic process, which was introduced to explain a random behavior in physical phenomena [43]. Second, deterministic chaos represents unpredictable behavior occurring from a deterministic dynamical system. Third, stochastic chaos is a process generated from a stochastically driven deterministic dynamical system [73]. Their meanings are different but all the “chaoses” represent unpredictable processes.
- [61] D. Sussillo and L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, *Neuron* **63**, 544 (2009).
- [62] R. Laje and D. V. Buonomano, Robust timing and motor patterns by taming chaos in recurrent neural networks, *Nat. Neurosci.* **16**, 925 (2013).
- [63] W. Nicola and C. Clopath, Supervised learning in spiking neural networks with force training, *Nat. Commun.* **8**, 2208 (2017).
- [64] K. Inoue, K. Nakajima, and Y. Kuniyoshi, Designing spontaneous behavioral switching via chaotic itinerancy, *Sci. Adv.* **6**, eabb3989 (2020).
- [65] R. Ahlfeld, B. Belkouchi, and F. Montomoli, Samba: Sparse approximation of moment-based arbitrary polynomial chaos, *J. Comput. Phys.* **320**, 1 (2016).
- [66] A. A. Faisal, L. P. Selen, and D. M. Wolpert, Noise in the nervous system, *Nat. Rev. Neurosci.* **9**, 292 (2008).
- [67] K. Goto, K. Nakajima, and H. Notsu, Twin vortex computer in fluid flow, *New J. Phys.* **23**, 063051 (2021).
- [68] R. Askey and J. A. Wilson, *Some Basic Hypergeometric Orthogonal Polynomials that Generalize Jacobi Polynomials* (American Mathematical Society, Providence, Rhode Island, 1985), Vol. 319.
- [69] J. P. Crutchfield, J. D. Farmer, and B. A. Huberman, Fluctuations and simple chaotic dynamics, *Phys. Rep.* **92**, 45 (1982).
- [70] G. J. Brewer, J. Torricelli, E. Ege, and P. Price, Optimized survival of hippocampal neurons in b27-supplemented neurobasal, a new serum-free medium combination, *J. Neurosci. Res.* **35**, 567 (1993).
- [71] S. M. Potter and T. B. DeMarse, A new approach to neural cell culture for long-term studies, *J. Neurosci. Methods* **110**, 17 (2001).
- [72] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering, *Neural Comput.* **16**, 1661 (2004).
- [73] S. Lovejoy and D. Schertzer, Stochastic chaos and multifractal geophysics, in *Chaos, Fractals and Models*, edited by F. Guindani and G. Salvadori (Italian University Press, 1998), Vol. 96, pp. 38–52.