Optimizing input mask for maximum memory performance of time-delay reservoir subjected to state noise

Zichen Kang\textsuperscript{1a)}, Sho Shirasaka\textsuperscript{1}, and Hideyuki Suzuki\textsuperscript{1}

\textsuperscript{1} Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
\textsuperscript{a)} zichen.kang@ist.osaka-u.ac.jp

Received April 19, 2021; Revised June 13, 2021; Published October 1, 2021

Abstract: Reservoir computing is a brain-inspired machine-learning framework that has been successfully used in information processing. A state-of-the-art methodology, called time-delay reservoir (TDR), realizes the reservoir using a single nonlinear physical node with delayed self-feedback. Memory capacity of the TDR is sensitive to the time-multiplexing procedure for input signals with a random mask. The existing memory optimization methods with respect to the masks are limited to the case without state noise, and their computational cost is large. In this article, we optimize the input mask to maximize the memory performance of the TDR with a white Gaussian state noise in a computational time efficient manner within the context of the Fisher memory curve, and then a TDR based on the Mackey-Glass system is used to illustrate our proposed method. The memory-nonlinearity trade-off of the TDRs regarding the input masks is also investigated in view of spectral properties of the spatial Fisher information matrix.

Key Words: reservoir computing, short term memory capacity, time-delay reservoir, fisher memory curve, input mask, optimization

1. Introduction

Reservoir computing (RC) \cite{1–3} is a brain-inspired machine-learning framework that has been successfully used in information processing, such as attractor reconstruction \cite{4} or speech recognition \cite{5}. The basic structure of the RC is composed of an input layer, a reservoir layer and an output layer. The input layer connects the real-world signal to the reservoir layer via fixed weighted connections. In the reservoir layer, an artificial neural network with fixed random internal connections is used to transform the injected signal into a high-dimensional feature space, such as a state space of the echo state network \cite{2}. In training of the RCs, only the connections between the reservoir layer and the output layer are adapted with a simple method such as linear regression so as to generate desired output \cite{2,6}. This design improves the problems of traditional recurrent neural networks, such as local optima and vanishing/exploding gradients \cite{7}.

Recently, a state-of-the-art methodology, called the time-delay reservoir (TDR) \cite{8}, realizes the
reservoir using a single nonlinear physical node with a delay line, as depicted in Fig. 1. The delay line is a chain of points placed equidistantly in time, which are denoted as virtual neurons. In the TDRs, the virtual neurons play a role as the reservoir layer, i.e., they are alternative to the artificial neural network in the conventional RC. The virtual neurons can exhibit high-dimensional transient responses due to the time delay. A major advantage of the TDRs is that they can be easily implemented in hardware. This novel paradigm has attracted several hardware implementations based on analog electronics [8], electromechanical devices [9] and opto-electronic devices [10, 11].

![Fig. 1. Schematic of time-delay reservoirs.](image)

Input preprocessing is essential to improve the TDRs’ performance. A random mask is mostly applied to the input signal in order to generate complex transient responses. Studies on the design of the input mask have been reported, such as a binary mask that maximizes the diversity of the reservoir states for a small set of the virtual neurons [12] and a chaotic mask that improves the performance of an optical TDR for a time-series prediction task [13].

One of the important properties of the TDRs is short-term memory capacity [14], which allows information of past input signal to be stored in the current states of the reservoirs. In [15] Ganguli et al. theoretically studied memory trace of linear echo state networks in the presence of Gaussian state noise using the Fisher information. They showed that the memory capacity is maximized if the input-reservoir connection is chosen as the maximal principal component of the spatial Fisher memory matrix. Grigoryeva et al. optimized the architecture parameters (i.e., input mask, reservoir parameters) of the TDRs by solving structured optimization problems, which have been successfully used in linear and quadratic memory tasks [16]. However, the reservoir performance can be degraded by state noise originating from analog components of the reservoir layer [17–19]. Furthermore, solving the structured optimization problem for determining a high-dimensional input mask is time-consuming. It is desirable to develop a time efficient method to optimize the input mask for the TDRs with considering the impact of the state noise.

In the following, we consider a general model of the TDRs based on a simple nonlinear electronic circuit in the presence of state noise [8]. By taking into account that the input-reservoir connections of the echo state networks correspond to the input mask of the TDRs, we propose a method to optimize their Fisher-information based memory capacity in terms of the mask. In contrast to Grigoryeva et al.’s approach [16], our optimization procedure is independent of the information of driving signals and can easily be carried out with low computational burden.

For the complex tasks that require the capacity of nonlinear transformation, such as attractor reconstruction of chaotic dynamical systems from time series data, what kind of input mask can give good performance of the TDRs is also concerned. Previous studies uncovered a trade-off relationship between the short-term memory capacity of the RCs and their degree of capability of nonlinear transformations [20, 21]. As a by-product of our proposed method, input masks of various memory capacity, which is quantitatively characterized by eigenvalues of the spatial Fisher memory matrix, are obtained. Using these masks, the memory-nonlinearity trade-off of the TDRs in terms of the masks can be studied in detail.

This article is organized as follows: In Section 2, we briefly introduce the basic structure of the TDR.
We then derive the Fisher memory curve (FMC) of the linear approximation of the TDR. The input mask in direction of the maximal principal component of the spatial Fisher memory matrix optimizes the memory performance of the TDR for memory tasks. Section 3 reports the numerical results obtained on the TDR based on the Mackey-Glass system. We compare our proposed method and that of Grigoryeva et al. in terms of CPU time and memory task performance in the presence/absence of the state noise. We also illustrate that the optimal performance on benchmark memory tasks involving weakly-nonlinear transformation is obtained when the input mask is chosen as the maximal principal component of the spatial Fisher memory matrix. Furthermore, the memory-nonlinearity trade-off in view of the input mask is investigated using the chaotic attractor learning tasks of two dynamical systems: the Hénon map and the logistic map. Section 4 summarizes the results and discusses their implications.

2. Fisher memory curve of time-delay reservoir

2.1 Reservoir map

The TDR we consider is based on a time delay differential equation as follows:

\[ \dot{x}(t) = -x(t) + f(x(t - \tau), I(t), z(t)), \]  

(1)

where \( x(t) \in \mathbb{R} \) is a dynamical variable, \( f \) is a smooth nonlinear function with a delay feedback term \( x(t - \tau) \), and \( z(t) \in \mathbb{R} \) is state noise. It is assumed that \( z(t) \) is a Gaussian noise with mean \( \mu(z(t)) = 0 \) and covariance \( \langle z(t_1)z(t_2) \rangle = \delta_{t_1,t_2} \) (\( \delta \) is the Dirac delta function). \( z(t) \) can be considered as thermal noise in electronic circuits [18] or spontaneous emission in semiconductor lasers [19]. \( I(t) \in \mathbb{R} \) is time-multiplexed input signal of \( u(t) \in \mathbb{R} \) over the delay period \( \tau \), i.e.,

\[ I(t) = g(t) \cdot u(t). \]  

(2)

The norm \( ||g|| \) determines the scale of the input signals. Generally, the mask function \( g(t) \) is piecewise constant over intervals of \( \theta \) and periodic over the delay period \( \tau \), as \( g(t + \tau) = g(t) \). The states of the virtual neurons are derived from the sampling process of \( x(t) \) within each delay interval of length \( \tau \). The sampling length \( \theta \) is related to the number of virtual neurons \( N \) by \( \theta = \tau/N \). These interconnected virtual neurons emulate a neural network as in the conventional reservoirs. These settings allow the transient response of the simply implemented, single node reservoir to capture the input features without loss of high-dimensional nonlinear characteristics.

The time-discretized form of (1) is expressed as

\[ x(n) = F(x(n-1), I(n), \Delta z(n)). \]  

(3)

Here \( F : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \) is referred to as the reservoir map. It depends on the virtual neuron layer \( x(n-1) := (x_1(n-1), x_2(n-1), ..., x_N(n-1))^\top \in \mathbb{R}^N \) that is parameterized by a discrete time \( n \in \mathbb{Z} \) by setting \( x_i(n) = x(n\tau - (N-i)\theta) \), a time-multiplexed input stream \( I(n) := Mu(n\tau) \), where \( M \in \mathbb{R}^N \) is the mask vector with the elements \( M_i := g(n\tau - (N-i)\theta) \), and the increment \( \Delta z(n) := (\Delta z_1(n), \Delta z_2(n), ..., \Delta z_N(n))^\top \in \mathbb{R}^N \), where \( \Delta z_i(n) = z(n\tau - (N-i)\theta) - z(n\tau - (N-i+1)\theta) \). The increment \( \Delta z(n) \) follows the normal distribution with zero mean and covariance \( \langle \Delta z_i(k)\Delta z_j(l) \rangle = \epsilon\theta\delta_{k,l}\delta_{i,j} \). Note that, \( \Delta z \) in the reservoir map (3) can include some noise sources which do not appear as white in (1) by redefining its strength \( \epsilon \theta \). For example, quantization noise, which crucially affects TDR’s performance [18], can be considered as the discrete-time white Gaussian noise under some conditions [22]. The recursion (3) shows the transient activity of virtual neurons excited by input signals. We can use a discrete-time recursive solution of the system (1) to approximate the reservoir map, as in [16].

Note that most of previous analysis via the FMC is limited to linear reservoirs. In order to make application of the FMC to the TDR (1) tractable, we will linearize the reservoir map (3) about a stable equilibrium. Actually, it is well-known that the optimal dynamical regime of the TDR (1) is around a stable equilibrium. This condition implies parameter settings in system (1) must ensure the existence of a stable equilibrium when both \( u(t) \) and \( z(t) \) are set to zero. The global asymptotic stability of the system (1) is relevant to the fading memory property [23].
2.2 Jacobian linearization

Suppose \( \bar{x} \in \mathbb{R} \) is the stable equilibrium of the continuous time model (1) with \( u(t) = 0 \) and \( z(t) = 0 \). The fixed point of the reservoir map (3) is \( \bar{x} := (\bar{x}, \bar{x}, ..., \bar{x})^\top \in \mathbb{R}^N \). We know that if we start the system (3) at \( x(0) = \bar{x} \) and apply \( I(n) = 0, \Delta z(n) = 0 \), then the state of the system (3) will remain fixed at \( x(n) = \bar{x} \) for all \( n \). We consider the Jacobian linearization of the reservoir map (3) about the equilibrium point \( (\bar{x}, 0, 0) \). By performing the Taylor expansion of the right hand side and neglecting all higher (than the 1st) order terms, we obtain an expression of the form:

\[
\begin{align*}
\dot{x}(n) &= F(\bar{x}, 0, 0) + \frac{\partial F}{\partial x}(x, I, \Delta z) = (\bar{x}, 0, 0) (x(n - 1) - \bar{x}) + \frac{\partial F}{\partial I}(x, I, \Delta z) = (\bar{x}, 0, 0) I(n) \\
& \quad + \frac{\partial F}{\partial \Delta z}(x, I, \Delta z) = (\bar{x}, 0, 0) \Delta z(n).
\end{align*}
\]

For convenience, we define these three \( N \times N \) Jacobian matrices as

\[
A := \frac{\partial F}{\partial x}(x, I, \Delta z) = (\bar{x}, 0, 0), \quad B := \frac{\partial F}{\partial I}(x, I, \Delta z) = (\bar{x}, 0, 0), \quad C := \frac{\partial F}{\partial \Delta z}(x, I, \Delta z) = (\bar{x}, 0, 0).
\]

Note that the time-multiplexed signal \( I(n) \) is linearly dependent on the sampled input signal \( u(n \tau) \) with the mask vector \( M \). Then the reservoir state update equation can be expressed as follows

\[
x(n) = \bar{x} + A(x(n - 1) - \bar{x}) + BM u(n \tau) + C \Delta z(n).
\]

Here \( A \) is an \( N \times N \) constant matrix that is referred to as the connectivity matrix. The feedforward connections from the input signal into the reservoir layer are represented as the product of the matrix \( B \) and the mask \( M \).

The linearized system (4) is an approximation of the nonlinear system (1) that is valid in a small region around the equilibrium \( (\bar{x}, 0, 0) \). The asymptotic stability ensures that the small perturbations away from the equilibrium decay with time. The TDR (1) exhibits this asymptotic behavior when the spectral radius \( \rho(A) \) of the matrix \( A \) is smaller than one. For an arbitrary number of virtual neurons \( N, \rho(A) \leq \|A\|_{\infty} < 1 \) if and only if \( |\frac{\partial F}{\partial x}(\bar{x}, 0, 0)| < 1 \) [16].

2.3 Fisher memory curve at the equilibrium

In [15] Ganguli et al. used the Fisher Information to quantify the memory of past input signals in linear reservoirs in the presence of Gaussian state noise. In the Fisher memory framework, the input sequence \( u := (\ldots, u((n - 2)\tau), u((n - 1)\tau), u(n\tau), \ldots) \) is considered as the parameter on which the probability of \( x \) depends. This means that the conditional distribution \( p(x(n)|u) \) varies with the input sequence \( u \). The average rate of change of the log-likelihood function \( \log(p(x(n)|u)) \) with respect to small perturbation in the input history is measured by the Fisher memory matrix whose elements are defined as

\[
\mathcal{I}_{l,k}(u) = -E_{p(x(n)|u)} \left[ \frac{\partial^2}{\partial u_l((n - l)\tau) \partial u_k((n - k)\tau)} \log(p(x(n)|u)) \right].
\]

The diagonal elements \( J_k = \mathcal{I}_{k,k} \) quantify how much information about a slight change in the input signal at \( k \) time steps in the past can be held by \( x(n) \). The collection of \( \{J_k\}_{k=0}^{\infty} \) is referred to as the Fisher memory curve. It captures the decay of information that can be preserved in the reservoir states about the past input signals. Specially, Ganguli et al. have shown that if \( x(n) \) follows the Gaussian distribution with mean \( \mu \) and covariance \( C \), the Fisher memory matrix can be obtained as

\[
\mathcal{I}_{l,k}(u) = \frac{\partial^2}{\partial u_l((n - l)\tau) \partial u_k((n - k)\tau)} C^{-1} \frac{\partial u_l}{\partial u_{(n - l)\tau}} \frac{\partial u_k}{\partial u_{(n - k)\tau}} [15].
\]

For the purpose of studying the linear memory capacity of the TDR, we apply the Fisher memory to the linearized system (4). Suppose that the initial condition of the reservoir map is prepared at sufficiently far past. Then its effect disappears and using (4) the state of the virtual neurons at time \( n \) can be expressed as follows

\[
x(n) = \bar{x} + \sum_{k=0}^{\infty} [A^k BM u((n - k)\tau) + A^k C \Delta z(n - k)].
\]
Recalling that the increment $\Delta z(t)$ follows the Gaussian distribution, the conditional distribution $p(x(n)|u)$ is also Gaussian with mean $\mu(x(n)) = \bar{x} + \sum_{k=0}^{\infty} A^k B M u ((n-k)\tau)$ and covariance matrix $C_n = \sigma^2 \sum_{k=0}^{\infty} A^k C A^\top$. The mean is only dependent on the input signal history linearly, and the covariance matrix is independent of the input signal. The Fisher memory matrix at the equilibrium $(\bar{x}, 0, 0)$ can be written as

$$\mathcal{I}_{l,k} = M^\top B^\top A^\top C^{-1} A^k B M$$

and is independent of the input signal.

From (7) the total memory can be expressed as

$$J_{\text{tot}} = \sum_{k=0}^{\infty} J_k = M^\top J^s M.$$  

Here $J^s = \sum_{k=0}^{\infty} B^\top A^k C^{-1} A^k B$ is called the spatial Fisher memory matrix. Because $J^s$ is a real symmetric matrix, the eigenvalues are real and the eigenvectors associated with the different eigenvalues are orthogonal. The total memory $J_{\text{tot}}$ can be characterized by the eigenvalues of $J^s$ when the associated eigenvectors are chosen to be the input mask $M$. We can see that $J_{\text{tot}}$ can be maximized by the eigenvector corresponding to the largest eigenvalue of $J^s$. In the following, the total memory always refers to the largest eigenvalue of $J^s$. The above description does not mean that the short-term memory capacity can be changed by the input mask. Actually, the short-term memory capacity is immutable when the reservoir layer of the TDR is designed. What the input mask affects is the performance of the short-term memory capacity.

### 3. Simulation experiments

#### 3.1 The memory capacity of the TDR based on the Mackey-Glass system

In the following, we focus on the TDR based on the Mackey-Glass system

$$\dot{x}(t) = -x(t) + \eta [x(t-\tau) + \gamma I(t) + z(t)] \frac{1}{1 + [x(t-\tau) + \gamma I(t) + z(t)]^p},$$

$$I(t) = g(t) \cdot u(t).$$

Here the parameter $\eta > 0$ and $\gamma > 0$ are referred to as the feedback gain and the input gain respectively. The parameter $p > 0$ controls the ‘strength’ of the nonlinearity.

How the number of equilibria of the autonomous form ($I(t) = z(t) = 0$) of (10) depends on the parameters is summarized as follows [24]. In the case $p = 1$, both $x_0 = 0$ and $x_0 = \eta - 1$ are the equilibria. When $\eta \in (0, 1)$, the equilibrium $x_0 = 0$ is stable and the other $x_0 = \eta - 1$ is unstable. The transcritical bifurcation occurs when $\eta$ crosses 1. For $p \geq 2$, the number of the equilibria is more than 2. We recall that the RC functions well for broad class of input only when it has a unique stable equilibrium. If the number of the stable equilibria is more than 1, the transient response of the reservoir near a stable equilibrium can be driven closer to the other due to input or noise. Thus the multistability makes the RC’s output dependent not only on input history $u$ but on its initial condition, that is, it violates the echo state property (ESP) [2], which is commonly assumed in order to guarantee an RC well-functioning. Thus in this study we set $p = 1$, $\eta \in (0, 1)$. Moreover, even when the reservoir in its autonomous form is monostable, large input gain can break the ESP. Hence we set the input gain low enough to guarantee the ESP. For our numerical simulations we set $\gamma = 0.1$, unless noted otherwise.

For each value of the feedback gain, the input mask is always taken as the eigenvector corresponding to the largest eigenvalue of $J^s$. Figure 2a shows the FMC as a function of the feedback gain $\eta$. The high-level feedback can enhance the strength of memory trace. Furthermore, increasing the number of virtual neurons in the reservoir layer can prevent the memory from decaying. In Fig. 2b, the total memory of the TDR based on the Mackey-Glass system is shown. For the TDR, there is a limit to improving total memory by increasing the number of virtual neurons. As mentioned earlier the virtual neurons corresponding to the points on the delay line placed equidistantly. As the number of
the virtual neurons increases, the adjacent virtual neurons are too close to be independent; namely, the increase in the effective degrees of freedom is limited.

Equation (9) shows that the amount of memory performance gained from the reservoir layer is dependent on the input mask. The FMCs for some input masks are shown in Fig. 3. These input masks are taken as the eigenvectors, which are arranged in descending order according to the eigenvalues of $J^s$. The FMC colored blue is obtained by the eigenvector corresponding to the largest eigenvalue of $J^s$. It depicts the fading of the memory trace of the TDR whose total memory $J_{\text{tot}}$ is maximized.

### 3.2 Performance evaluation

In [16], the input mask for the memory tasks is determined by solving structured optimization problems, which have been successfully used in linear and quadratic memory tasks. In these tasks, a linear or quadratic function is used to generate a one-dimensional target signal from the time-lagged input signal. However, the reservoir performance can be degraded by state noise originating from analog components of the reservoir layer [17–19]. In the input layer, time-multiplexing is used to inject the input signal into temporally separated virtual neurons, then the dimension of the input mask is determined by the size of the reservoir layer. For a large set of the virtual neurons used to improve the reservoir performance, the Grigoryeva et al.’s method gives rise to a high-dimensional nonlinear optimization problem of high computational cost. On the other hand, our method only involves a linear problem, i.e., finding the maximal principal component of the spatial Fisher memory matrix.

Considering the 5-lag linear memory task, we compare the time performance of the Grigoryeva et al.’s method with the proposed method in Table I. Here, we solve the optimization problem obtained from the Grigoryeva et al.’s method by using the particle swarm optimization (PSO) algorithm [25].
Table I. CPU-time performance of the Grigoryeva et al.’s method and the proposed method.

| Number of virtual neurons | 10   | 15   | 20   | 25   | 30   | 35   | 40   |
|---------------------------|------|------|------|------|------|------|------|
| Grigoryeva et al.’s method | 11.75| 29.58| 47.66| 82.30| 117.16| 233.16| 823.55|
| Proposed method            | 0.22 | 0.25 | 0.31 | 0.48 | 0.64 | 4.41 | 5.23 |

In Fig. 4, the reservoir performance optimized by using the Grigoryeva et al.’s method and the proposed method are compared in the 5-lag linear memory task and the 3-lag quadratic memory task. The reservoir performance is expressed by the normalized root mean square error (NRMSE). In this comparison, the termination condition for the PSO algorithm is set as if the relative change of the objective function over the last 20 iterations is less than $10^{-6}$. The simulation results of the TDR in the absence of state noise are shown in Fig. 4a and Fig. 4c. In this case, Grigoryeva et al.’s method outperforms the proposed method, as in the partial enlarged view of Fig. 4a and Fig. 4c. This is somewhat natural since the Grigoryeva et al.’s method is a task-specific optimization while our approach is task-independent. Figure 4b and Fig. 4d illustrate the impact of the state noise on the memory tasks. We set the state noise strength $\epsilon = 10^{-4}$ in these simulation experiments. It is seen that through our method the TDR achieves good memory performance. These results show that...
the proposed method is helpful for improving the noise robustness of memory of the TDRs.

3.3 The performance of the short-term memory affected by the input mask

In this section, we use the eigenvectors of $J^s$ as the input mask for the TDRs and compare their test errors for a memory task. We consider an i.i.d. random signal $u(t)$ from a uniform distribution over $[-1, 1]$. The TDRs are trained to produce an output $w(u) = \sin(u(t - \tau'))$ driven by the input signal of $\tau'$ step before as closely as possible. For the task of $\sin(u(t - \tau'))$, the memory performance is necessary but little performance of the nonlinear transformation is required. It allows us to focus on the studying of the short-term memory capacity with the case of the different input masks. The task parameter $\tau'$ controls the requirement of memory. In our experiments, the TDRs are implemented by the Mackey-Glass node with the parameter setting: $\tau = 80$. The separation distance of the virtual neurons is set at $\theta = 0.2$ that offers a suitable short-term memory capacity as in Fig. 2b. Then the number of virtual neurons is $N = 400$. The state noise of strength $\epsilon = 10^{-4}$ is considered. The readout weights were trained by an online algorithm [6].

Concretely, we let each TDR run for 5000 steps, starting from a zero initial state, and discard the initial 100 steps during which the effect of the initial state dies out. Figure 5 depicts the test error of the task $\sin(u(t - \tau'))$ as a function of the feedback gain and the input mask. The indices of the input masks are arranged in descending order according to the eigenvalues of $J^s$. It can be seen that a decrease in NRMSE corresponds to an increase in feedback gain $\eta$. In Fig. 5a and Fig. 5c, the optimal performance is shown for the eigenvector of the largest eigenvalue of $J^s$ regardless of the value the feedback gain $\eta$. The NRMSEs for $\eta = 0.1, 0.5, 0.9$ are shown in Fig. 5b and Fig. 5d from top to bottom. We confirmed that similar results are obtained by employing other tasks, such as

![Simulation results for sin(u(t - 2))](image)

**Fig. 5.** Simulation results for the learning task of $\sin(u(t - \tau'))$. (a) Error surface for $\sin(u(t - 2))$, as a function of the input mask and the feedback gain. (b) The NRMSE versus the input masks for $\sin(u(t - 2))$ with parameter setting: $\eta = 0.1, 0.5, 0.9$ from top to bottom. (c) Error surface for $\sin(u(t - 3))$, as a function of the input masks and the feedback gain. (d) The NRMSE versus the input masks for $\sin(u(t - 3))$ with parameter setting: $\eta = 0.1, 0.5, 0.9$ from top to bottom. The indices of the input masks are arranged in descending order according to the eigenvalues of $J^s$. 

669
\( f(u) = u(t - \tau') \) or \( f(u) = u^2(t - \tau') \). We note that the masks with indices not less than 27 seem not approximate eigenvectors of the exact spatial Fisher memory matrix but numerical artifacts (see Appendix A). Thus the "local minima" of the NRMSEs around the 200th indices for \( \eta = 0.9 \) seen in Fig. 5b and Fig. 5d might be numerical phenomena which have no relevance to the dynamical properties of the TDR.

These results show that the input mask in direction of the maximal principal component of the spatial Fisher memory matrix optimizes the reservoir performance of the memory tasks.

### 3.4 The performance of the nonlinear transformation affected by the input mask

Nonlinear transformation is also an essential property for the TDRs. The input signal is transformed in various nonlinear ways via the high-dimensional nonlinear dynamics of the reservoirs. In this section, we test the TDRs on the chaotic attractor learning tasks of two dynamical systems: the Hénon map and the logistic map. For these learning tasks, the performance of the nonlinear transformation is necessary.

- The training data set of the Hénon map is generated by

\[
\begin{align*}
  x(t+1) &= 1 - ax(t)^2 + by(t), \\
  y(t+1) &= x(t-1),
\end{align*}
\]

with parameter setting: \( a = 1.4 \) and \( b = 0.3 \). For this parameter value the Hénon map is chaotic.
• The logistic map is described by the equation

\[ x(t + 1) = rx(t)(1 - x(t)). \]  

The time series \( x(t) \) is chaotic with parameter setting: \( r = 4 \).

We choose \( x(t) \) as the input signal, and the TDRs are trained to produce an output of \( x(t + 1) \). The TDRs are trained from a run of 5000 steps and the first 100 steps were discarded. In these simulation experiments, the variance \( \epsilon \) of the white noise is equal to \( 10^{-5} \).

Figure 6a and Fig. 6c depict the approximation errors for the learning tasks of the Hénon map and the logistic map, respectively, as a function of the feedback gain and the input mask. We can see that a broad range of the input mask can provide good performance of the nonlinear transformation. It is interesting to note that the performance is degraded when the input mask takes the eigenvector corresponding to the largest eigenvalue of \( J^s \). Specially, we compare the test errors for the tasks of the Hénon map and the logistic map with the test errors for the tasks \( w(u) = \sin(u(t - \tau')) \) in different cases of \( \eta = 0.1, 0.5, 0.9 \), as in Fig. 6b and Fig. 6d. We can see that the test errors show opposite trends with respect to the input mask. This phenomenon can be interpreted as the memory-nonlinearity trade-off [20, 21] in the TDRs.

4. Conclusions

In this work, we developed a method to optimize memory performance of the TDRs subjected to state noise with respect to input masks in a task-independent and computationally-efficient manner. In comparison with the existing optimization method, the proposed one is running-time-efficient since it only involves linear problem in the optimization procedure and improves memory performance in a great extent in the presence of state noise. Through numerical experiments, we confirmed that the input mask in direction of the maximal principal component of the spatial Fisher memory matrix certainly provides the best performance on benchmark memory tasks requiring only weak nonlinear transformation capacity. Also, we illustrated a memory-nonlinearity trade-off in terms of the input masks, whose associated memory performance can be quantitatively characterized in detail through spectral properties of the spatial Fisher memory matrix, using chaotic times series prediction tasks. Our proposed method facilitates the design of the input masks for TDRs with desired computational properties.

![Fig. A-1. Heatmap of the eigenvectors of \( J^s \) with parameter setting: \( \gamma = 0.1 \), \( p = 1 \), \( \eta = 0.9 \), \( \tau = 80 \) and \( \theta = 0.2 \). The color of each pixel is determined by the absolute value of components of the input mask. The indices of the input masks are arranged in descending order according to the eigenvalues of \( J^s \).](image-url)
Acknowledgments
This study is supported by JST CREST Grant Number JPMJCR18K2 and by JSPS KAKENHI Grant Number 20H00596.

Appendix

A. Eigenvectors of the spatial Fisher memory matrix
In Fig. A-1, the heatmap is used to visualize the eigenvectors of the spatial Fisher memory matrix $J^s$. The eigenvectors of indices ranging from 27th to 399th seem spurious. The reason is that the eigenvalues with indices of 27th or more are located extremely close to unity.

References
[1] H. Jaeger and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication,” Science, vol. 304, no. 5667, pp. 78–80, 2004. DOI: 10.1126/science.1091277
[2] H. Jaeger, “The “echo state” approach to analysing and training recurrent neural networks,” GMD Report, vol. 148, no. 34, p. 13, January 2001.
[3] W. Maass, T. Natschläger, and H. Markram, “Real-time computing without stable states: a new framework for neural computation based on perturbations,” Neural Comput., vol. 14, no. 11, pp. 2531–2560, November 2002. DOI: 10.1162/089976602760407955
[4] Z. Lu, B.R. Hunt, and E. Ott, “Attractor reconstruction by machine learning,” Chaos, vol. 28, no. 6, 061104, June 2018. DOI: 10.1063/1.5039508
[5] D. Verstraeten, B. Schrauwen, and D. Stroobandt, “Reservoir-based techniques for speech recognition,” The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 1050–1053, October 2006.
[6] D. Sussillo and L.F. Abbott, “Generating coherent patterns of activity from chaotic neural networks,” Neuron, vol. 63, no. 4, pp. 544–557, August 2009. DOI: 10.1016/j.neuron.2009.07.018
[7] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning, The MIT Press, 2016.
[8] L. Appeltant, M. Soriano, and G. Sande, et al., “Information processing using a single dynamical node as complex system,” Nat Commun., vol. 13, no. 2, 468, September 2011. DOI: 10.1038/ncomms1476
[9] G. Dion, S. Mejaouri, and J. Sylvestre, “Reservoir computing with a single delay-coupled nonlinear mechanical oscillator,” Journal of Applied Physics, vol. 124, no. 15, 152132, October 2018. DOI: 10.1063/1.5038038
[10] L. Larger, M.C. Soriano, D. Brunner, L. Appeltant, J.M. Gutierrez, L. Pesquera, C.R. Mirasso, and I. Fischer, “Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing,” Opt. Express, vol. 20, no. 3, pp. 3241–3249, January 2012. DOI: 10.1364/OE.20.003241
[11] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V.S. Udaltsov, Y.K. Chembo, and M. Jacquot, “High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification,” Phys. Rev. X, vol. 7, no. 1, 011015, February 2017. DOI: 10.1103/PhysRevX.7.011015
[12] L. Appeltant, G. Van der Sande, J. Danckaert, and F. Ingo, “Constructing optimized binary masks for reservoir computing with delay systems,” Scientific Reports, vol. 4, 3629, 2014. DOI: 10.1038/srep03629
[13] J. Nakayama, K. Kanno, and A. Uchida, “Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal,” Opt. Express, vol. 24, no. 8, pp. 8679–8692, April 2016. DOI: 10.1364/OE.24.008679
[14] H. Jaeger, “Short term memory in echo state networks,” GMD Report, vol. 152, 2002.
[15] S. Ganguli, D. Huh, and H. Sompolinsky, “Memory traces in dynamical systems,” Proceedings of the National Academy of Sciences, vol. 105, no. 48, pp. 18970–18975, October 2008. DOI: 10.1073/pnas.0804451105
[16] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega “Optimal nonlinear information processing capacity in delay-based reservoir computers,” *Scientific Reports*, vol. 5, 12858, 2015. DOI: 10.1038/srep12858

[17] M.C. Soriano, S. Ortín, D. Brunner, L. Larger, C. Mirasso, I. Fischer, and L. Pesquera, “Optoelectronic reservoir computing: tackling noise-induced performance degradation,” *Opt. Express*, vol. 21, no. 1, pp. 12–20, January 2013. DOI: 10.1364/OE.21.000012

[18] M.C. Soriano, S. Ortín, L. Keuninckx, L. Appeltant, J. Danckaert, L. Pesquera, and G. van der Sande, “Delay-based reservoir computing: Noise effects in a combined analog and digital implementation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 388–393, February 2015. DOI: 10.1109/TNNLS.2014.2311855

[19] R.M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. Van der Sande, “Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics,” *Opt. Express*, vol. 22, no. 7, pp. 8672–8686, 2014. DOI: 10.1364/OE.22.008672

[20] J. Dambre, D. Verstraeten, B. Schrauwen, and M. Serge “Information Processing Capacity of Dynamical Systems,” *Scientific Reports*, vol. 2, 514, 2012. DOI: 10.1038/srep00514

[21] M. Imbishi and K. Yoshimura, “Reservoir computing beyond memory-nonlinearity trade-off,” *Scientific Reports*, vol. 7, 10199, 2017. DOI: 10.1038/s41598-017-10257-6

[22] B. Widrow and I. Kollár, *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, Cambridge University Press, 2010. DOI: doi:10.1017/CBO9780511754661

[23] S. Boyd and L. Chua, “Fading memory and the problem of approximating nonlinear operators with Volterra series,” *IEEE Transactions on Circuits and Systems*, vol. 32, no. 11, pp. 1150–1161, November 1985. DOI: 10.1109/TCS.1985.1085649

[24] J. Wei and D. Fan, “Hopf bifurcation analysis in a Mackey-Glass system,” *International Journal of Bifurcation and Chaos*, vol. 17, no. 6, pp. 2149–2157, 2007. DOI: 10.1142/S0218127407018282

[25] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1945, 1995. DOI: 10.1109/ICNN.1995.488968