

Boosting learning ability of overdamped bistable stochastic resonance system based physical reservoir computing model by time-delayed feedback

Zhuozheng Shi^{a,1}, Zhiqiang Liao^{b,*,1}, Hitoshi Tabata^{a,b,*}

^a Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

^b Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:

Received 31 January 2022

Received in revised form 22 May 2022

Accepted 6 June 2022

Available online 17 June 2022

Keywords:

Reservoir computing

Short-term memory ability

Overdamped bistable system

Stochastic resonance

Time-delayed feedback

ABSTRACT

Physical reservoir computing (RC), which can be implemented by various physical systems, is a low-cost neuromorphic framework with a fast learning capability. In the previous studies, an overdamped bistable system-based RC (OBRC) inspired by the FitzHugh-Nagumo neuron model has been proposed to construct an outstanding physical RC system. Benefitting from the stochastic resonance effect, the OBRC requires less power and has stronger noise robustness than many conventional physical RC systems. However, compared with conventional physical RC systems, its learning ability is not superior. To boost the performance of the OBRC, we propose an OBRC with time-delayed feedback (TOBRC). In this work, the TOBRC is implemented in a physical setting with time-multiplexing nodes design and simulated on a conventional computer. Moreover, we adopt a powerful optimization algorithm to automatically determine the optimal hyperparameters for both the OBRC and TOBRC; thus, a more precise quantitative discussion on the upper limit of the system can be made. To compare the TOBRC and OBRC, we conducted short-term memory and parity check tasks to assess the short-term memory ability and nonlinearity, which are the two core abilities of physical RC for learning. The results prove that the short-term memory ability and nonlinearity of the proposed TOBRC are 6.46 and 2.15 times higher than those of the OBRC, respectively. Moreover, the TOBRC outperforms the OBRC under different noise conditions. On the MNIST handwritten digit recognition benchmark, the TOBRC exhibited a lower error rate than the OBRC; it was comparable with that of advanced physical RC systems. Our study confirms that the TOBRC can exhibit excellent learning ability in practical problems.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

In the past decades, artificial neural networks have ushered in a new paradigm in information processing. However, their computing speed on conventional computers is restricted by the separation of the core memory and processing units [1]. To overcome the well-known von Neumann bottleneck, neuromorphic devices based on physical hardware are considered key for the future revolution of artificial intelligence [2]. Recurrent neural networks (RNNs), which are biologically inspired by the recurrent feedback system formed in α -motoneurons [3–5], have continued to occupy an important position in computer science [6–8] and neuromorphic engineering [9–11]. Despite the advantages of RNNs for learning the nonlinear features in data owing to its

characteristics of memory and Turing completeness [12], their performance is limited by the complicated hardware architecture and low convergence rate [13].

To accelerate the learning speed of an RNN, reservoir computing (RC) achieved by a sparsely connected random-fixed network has been proposed [14,15]. The core of an RC is a reservoir with abundant dynamic information, which is responsible for mapping input from low-dimensional space to high-dimensional space to extract nonlinear characteristics [16]. Compared with the RNN, an RC network performs only one readout process and a simple linear regression training [14]; thus, it can significantly reduce the computational burden as well as the local minimum phenomenon. These advantages of the RC not only promote its wide application in software [17–19], but have also attracted the interest of neuromorphic engineers [20,21]. Currently, RC has been widely implemented by various physical systems such as optical [22–24], spintronics [25–28], and electronics systems [29,30]. For physical RC devices, the short-term memory (STM) ability and nonlinearity are the two most important attributes that ensure their learning

* Corresponding authors.

E-mail addresses: liao@bioxide.t.u-tokyo.ac.jp (Z. Liao), tabata@bioeng.t.u-tokyo.ac.jp (H. Tabata).

¹ These authors contributed equally to this work.

ability [25,27]. RC devices can be divided into two categories based on the principle of realizing the STM ability and nonlinearity. The first type is represented by time-delayed RC, whose reservoir part is composed of time-multiplexing physical nodes and a time-delayed feedback loop [22–24,27,28]. The physical nodes of this type of RC typically have strong nonlinearity without the STM ability, which limits the overall learning ability of the RC device. The second type realizes STM ability and nonlinearity using the non-Markov property of the relaxation phenomenon in physical systems [25,26,29]. Although the second type of RC solves the problem of nonlinear nodes not having STM capability, it is generally prone to performance attenuation due to inevitable noise [26].

To improve the existing physical RC, an overdamped bistable system-based RC model (OBRC) has been recently proposed [31,32]. Although previous studies have focused on the numerical model of the OBRC, the overdamped bistability have been realized in various physical systems, such as optical system [33], condensed matter [34], and electronic circuit [35]. A common feature of these systems is that they can induce stochastic resonance (SR) effect, which is a phenomenon of amplifying effective information in a nonlinear system using noise energy [36–38]. The SR effect has been widely studied in fault diagnosis [39–41], signal processing [42–44], sensory enhancement [45–47], and other fields [48–50]. In previous machine learning studies, SR has also been found to enhance the network performance [51–55]. Compared with conventional RC, the OBRC requires less power and has a stronger noise robustness owing to its SR ability [31,32]. However, due to its simple structure, its learning ability is not superior to those of other physical RC devices [31].

To further enhance the performance of the OBRC, we proposed an improved system called time-delayed OBRC (TOBRC) and numerically set its physical model in the conventional computer. The TOBRC can be constructed by adding a time-delayed feedback structure to the OBRC. Hence, the TOBRC inherits the physical realizability of the OBRC. Although the overdamped bistable SR system with time-delay feedback has been used to enhance the weak signal diagnosis in engineering application [56], there is no reported study regarding it as a learning system. Because the OBRC belongs to the aforementioned second type of RC, the TOBRC can be considered a combination of the two types of RC systems. In addition, in many biological dynamic systems, the noise robustness is enhanced by a time-delayed feedback [57,58]. Therefore, the noise robustness of the TOBRC has the potential to exceed that of the OBRC. Moreover, we combined a powerful optimization algorithm to select the optimal hyperparameters of the TOBRC. Accordingly, compared with the condition of fixed hyperparameters reported in previous studies [31,32], the discussions in this work can more accurately reflect the upper limit of the system performance.

The remainder of this paper is organized as follows. In Section 2, we briefly review the conventional time-multiplexing model and the training process of the physical RC. In Section 3, we introduce the mathematical model of the TOBRC unit and propose a physical RC model based on the TOBRC unit. In Section 4, we first compare the performance of the TOBRC and OBRC on the STM and parity check (PC) benchmarks. The results reveal that the TOBRC has significantly better STM ability and nonlinearity than other physical RC systems, including the OBRC. Besides, the influence of the hyperparameters of the time-delayed feedback loop on the TOBRC performance is explored in detail. Subsequently, we perform the MNIST handwritten digit recognition task. The TOBRC is proven to have a low recognition error rate, comparable to those of advanced physical RC systems. Finally, we conclude the paper in Section 5.

2. Physical reservoir computing model

Fig. 1 displays the generalized framework of the physical RC. The training process of the physical RC can be mathematically described as follows:

$$x(t+1) = F(Wx(t) + W_{in}u(t+1)), \quad (1)$$

$$x(t) = [x(1), x(2), \dots]^T, \quad (2)$$

$$X = \begin{bmatrix} x(1) & \dots & x(m-n+1) \\ \vdots & \ddots & \vdots \\ x(n) & \dots & x(m) \end{bmatrix}, \quad (3)$$

$$W_{out} = Ypinv(X), \quad (4)$$

where $u(t)$ is the input sequence, which is transferred to a physical signal after the masking operation. Masking is equivalent to the input matrix in the software RC, that is, W_{in} in Eq. (1), which improves the learning efficiency of the system [59]. Unlike the software RC, the activation function F and internal connection matrix W cannot be manually set. Instead, they are determined by the internal physical properties of the reservoir. The raw reservoir state $x(t)$ can be obtained by sampling the output signal of the physical reservoir. Because of the single node structure of the physical reservoir, $x(t)$ is a one-dimensional vector that cannot be directly used in the training process. To obtain the reservoir state matrix X that can be used in the training process, for the duration of each input value, we define the sampled n points in X as n time-multiplexing virtual nodes. $x(t)$ can then be reshaped into an n -by- m/n dimensional X matrix, as described in Eq. (3), where m is the total number of sampling points. Because X is not always a square matrix,

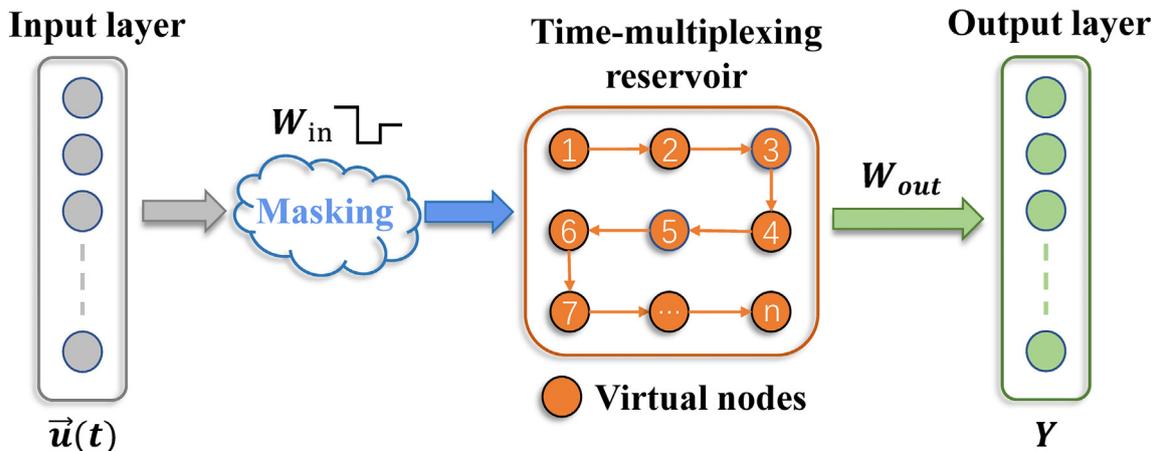


Fig. 1. General physical RC framework with a time-multiplexing structure.

$W_{out} = YX^{-1}$ cannot be directly solved. Therefore, when optimizing W_{out} for physical RC, the pseudo inverse matrix of X is widely used [25–27], which is represented by $\text{pinv}()$ in Eq. (3).

In this work, the masking matrix W_{in} was randomly determined at the beginning of the training process and remained unchanged during training, which is similar to the software RC. In addition, to ensure the reliability of the results, all the calculated results were averaged over ten independent experiments.

3. Time-delayed bistable stochastic resonance system based RC

3.1. Mathematical model of TOBRC

The TOBRC unit used in this work describes the dynamics of the Brownian particles driven by an external driving force in a time-delayed overdamped bistable system, which can be mathematically expressed as:

$$\frac{dx(t)}{dt} = ax(t) - bx(t)^3 + cx(t - \tau) + U(t) + \xi(t), \quad (5)$$

where a and b are the coefficients of the potential function. ϵ and τ are defined as the learning rate and delayed distance, respectively. It is noteworthy that when $\epsilon = 0$ in Eq. (5), the TOBRC unit degenerates into the OBRC unit, which is similar to the FitzHugh-Nagumo neuron model [32]. $U(t)$ is the input signal, and $\xi(t)$ is the noise term which is satisfied:

$$\langle \xi_n[t] \rangle = 0, \quad (6)$$

$$\langle \xi_n[t] \cdot \xi_n[0] \rangle = 2D\delta(t), \quad (7)$$

where D and $\delta(t)$ are the noise intensity and Dirac delta function, respectively. For simplifying the discussion, we temporarily normalize a and b to 1 in this section and assume that $U(t)$ is the cosine signal with a small amplitude A . Subsequently, under the condition of a weak periodic signal, we can simplify the non-Markov process of Eq. (5) into a Markov process described by the following formula using the probability density approach [60]:

$$\frac{dx(t)}{dt} = (1 + \epsilon\tau)(x(t) - x(t)^3 + cx(t) + U(t)) + \xi(t). \quad (8)$$

Accordingly, the effective potential function in Eq. (7) can be obtained as [61]:

$$V(x) = -(1 + \epsilon\tau)(0.5x^2 - 0.25x^4) - 0.5\epsilon(1 + \epsilon\tau)x^2 + (1 + \epsilon\tau)xU(t). \quad (9)$$

Fig. 2(a) depicts $V(x)$ with different ϵ and τ when there is no external driving force. Compared with the conventional overdamped bistable potential well, the additional parameter ϵ can adjust the potential well depth with the determined a and b . Specifically, a positive ϵ can deepen the potential well, whereas a negative ϵ can decrease the potential well depth. In addition, the sign of ϵ determines the modulation effect of τ on the potential well. When $\epsilon < 0$, the potential barrier decreases with the increase in τ , which is not conducive to the noise robustness of the system. In particular, in the case of $\epsilon = -0.5$ and $\tau = 2$, $V(x)$ almost loses the bistable nonlinearity, as shown in Fig. 2(a). When $\epsilon > 0$, the potential barrier increases with the increase in τ , which reduces the probability of state switching caused by noise disturbance. Hence, under the modulation of ϵ and τ , the dynamics of the TOBRC can be adjusted more flexibly under different external input conditions.

From Eqs. (8) and (9), the two stable-state potential positions of TOBRC model can be calculated as $x_{\pm} = \pm\sqrt{1 + \epsilon}$, which are not affected by τ . When a particle transitions between these two minimum potentials, its escape rate can be described by Kramer's dead escape rate [38]:

$$\gamma = \frac{(1 + \epsilon)(1 + \epsilon\tau)}{\sqrt{2\pi}} \exp\left[-\frac{(1 + \epsilon)^2(1 + \epsilon\tau)}{4D}\right]. \quad (10)$$

Referring to the methodology in the previous research [62], the output signal-to-noise ratio (SNR) of the TOBRC can be obtained as:

$$\text{SNR} = \frac{\pi A^2}{2D^2}(1 + \epsilon)(1 + \epsilon\tau)^2\gamma. \quad (11)$$

Fig. 2(b) illustrates the relationship between SNR and $D/\Delta V$ under the normalized amplitude $A = 1$, where $\Delta V = \frac{1}{4}(1 + \epsilon)^2(1 + \epsilon\tau)$ is the potential barrier height. In a determined potential well, one can expect that SNR first increase and then decrease with increasing D , which symbolizes the SR effect. More importantly, when τ is positive, the value of SNR- $D/\Delta V$ curve is higher than that when $\tau = 0$, which indicates that the TOBRC model has stronger SR effect than the OBRC model. According to our previous research, SR effect enable the OBRC naturally weaken the input noise under regardless of working parameters [31], which indirectly makes the learning performance of the OBRC decline more slowly with increasing D than the RC with traditional activation function [32]. Hence, owing to the stronger SR effect in the TOBRC, it can be expected that the noise robustness in the TOBRC is also stronger than that of the OBRC when performing specific tasks.

Fig. 2(b) shows the TOBRC model with the existence of a feedback loop. In the simulation environment, the output of the TOBRC unit cannot be obtained by an analytical method; thereby, we adopt the fourth-order Runge-Kutta method [63]. Using the fourth-order Runge-Kutta method, the raw reservoir state $x(t)$ can be calculated by the following equations:

$$S(n) = Wx(n) + W_{in}u(n + 1), \quad (12)$$

$$k_1 = h(ax(n) - bx(n)^3 + cx(n - \tau) + S(n)), \quad (13)$$

$$k_2 = h\left(a\left(x(n) + \frac{k_1}{2}\right) - b\left(x(n) + \frac{k_1}{2}\right)^3 + \epsilon\left(x(n - \tau) + \frac{k_1}{2}\right) + S(n)\right), \quad (14)$$

$$k_3 = h\left(a\left(x(n) + \frac{k_2}{2}\right) - b\left(x(n) + \frac{k_2}{2}\right)^3 + \epsilon\left(x(n - \tau) + \frac{k_2}{2}\right) + S(n + 1)\right), \quad (15)$$

$$k_4 = h\left(a\left(x(n) + k_3\right) - b\left(x(n) + k_3\right)^3 + \epsilon\left(x(n - \tau) + k_3\right) + S(n + 1)\right), \quad (16)$$

$$x(n + 1) = x(n) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \quad (17)$$

where $x(n)$ and $u(n)$ denote the discrete forms of $x(t)$ and $u(t)$, respectively. h represents the calculation step size, which is fixed at 0.1 in this work. After obtaining $x(n)$, Eqs. (3) and (4) were implemented to complete the training process. When applying the benchmarks, as described in Section 4, to eliminate the randomness of the testing, we repeated the test 10 times independently and considered the average value of the evaluation indices by changing the training and testing sets.

3.2. Optimization of hyperparameters

Due to the lack of recognized parameter selection guidance, the overdamped bistable system generally depends on the optimization algorithm to automatically obtain the optimal parameters [43]. In this study, the Optuna optimization algorithm was employed to determine the hyperparameters of the OBRC and TOBRC [64].

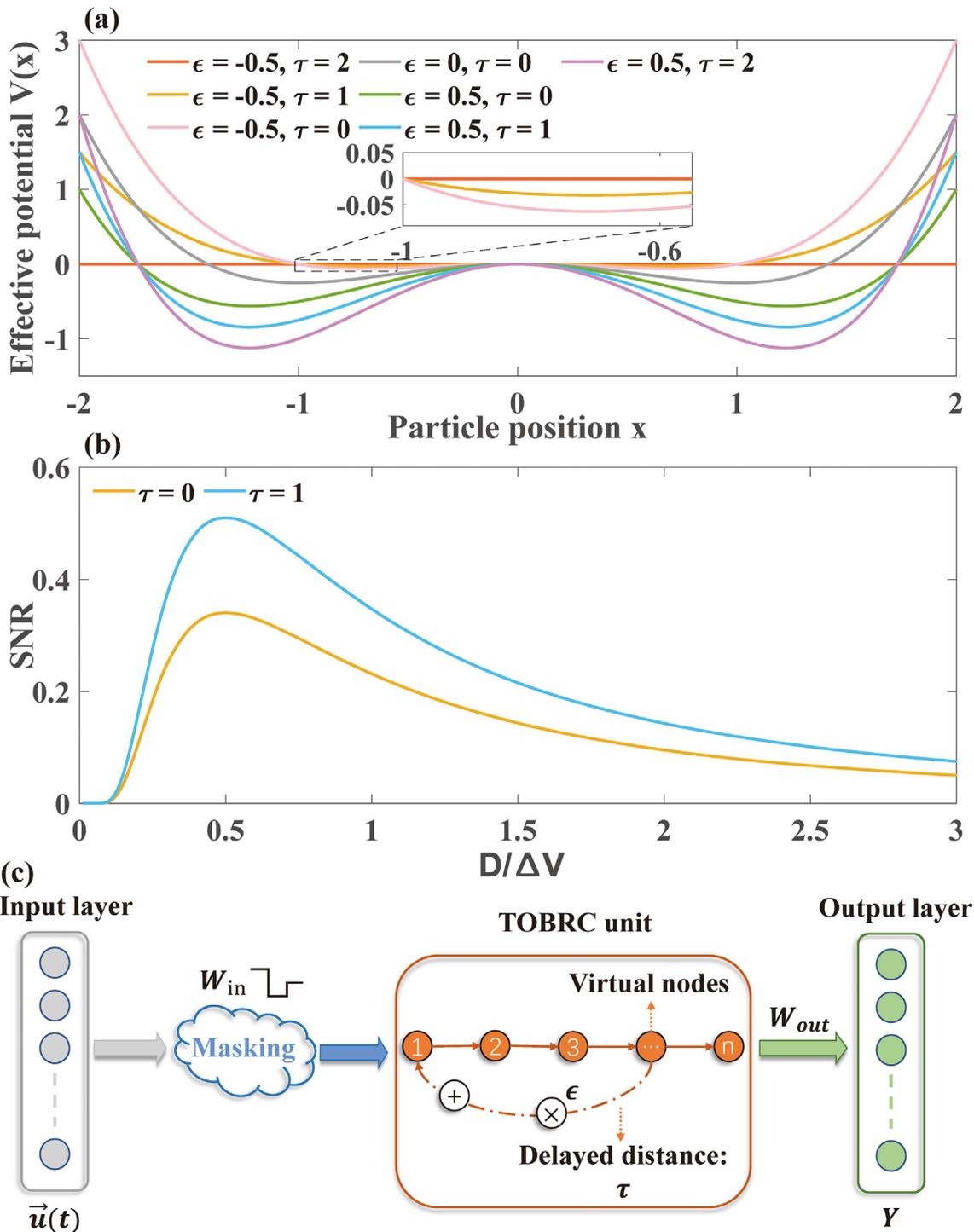


Fig. 2. (a) Effective potential $V(x)$ without external driving force. (b) The output SNR of the TOBRC versus $D/\Delta V$ with $\epsilon = 0.5$. (c) Generalized schematic of the TOBRC unit-based physical RC model.

Optuna has been used as the hyperparameter optimization software owing to its three advantages: First, it is defined-by-run, referring to the design in which the search space is dynamically constructed. Therefore, when a multihyperparameter optimization is conducted, Optuna designs internal variables without user participation. Second, Optuna provides efficient algorithms with a pruning mechanism by searching and performance estimation strategies. Specifically, Optuna features both relational sampling, where the correlations between multiple parameters are exploited, and independent sampling, where the parameters

are evaluated separately. The two sampling strategies are useful for highly correlated hyperparametric optimizations; thus, Optuna is suitable for optimizing overdamped bistable systems. Third, Optuna is lightweight and easy to setup. The setup only includes importing the package, designing the optimization function with target hyperparameters, and the loss function.

In the comparison among Optuna and other hyperparameter optimization algorithms, Optuna has been shown to outperform the other algorithms by obtaining higher or nearly equal objective values in

over 50 out of 56 black-box tests [64]. In addition, Optuna received an error rate <0.1 in the evaluations of pruning and distributed optimization. Hence, Optuna is a powerful optimization algorithm for this study. When benchmarking both the OBRC and TOBRC, the parameters to be optimized include a and b . For the TOBRC, two extra parameters ϵ and τ need to be optimized. As mentioned in Section 3.1, $\epsilon < 0$ is not conducive to maintaining the bistability and noise robustness of the system. Therefore, the optimization range of ϵ is $[0, 1]$. As to a , b , and τ , their optimization ranges are set based on the different tasks described in Section 4. Once the optimization starts, Optuna will not terminate until the loss function converges to a stable value, that is, no numerical change in the loss function occurs in the next 50 iterations.

4. Results and discussion

4.1. STM benchmark

The STM benchmark is a prevalent task widely employed to quantitatively estimate the STM ability of physical RC systems [26–28]. The training data for the STM task can be obtained from the following equation:

$$y_{STM}(t, d) = u(t - d), \quad (18)$$

where u denotes the binary input data, which are composed of random digits of 0 or 1. y_{STM} and d are the target sequence and memory delay, respectively. For each independent testing, input data $u(t)$ containing 2500 binary codes was generated, of which 1500 codes were used for training W_{out} and 1000 codes were used for evaluation. Notably, in both the training and evaluation processes, 500 codes were used as wash out to eliminate the influence of the initial transient dynamics in the RC [31]. Additionally, the masking matrix W_{in} in Eq. (12) was set to 1. In the evaluation process, the correlation between the predictive output of the RC and the actual target sequence can be expressed as follows:

$$cor(d)^2 = \frac{Cov[y(t, d), y_{out}(t)]^2}{Var[y(t, d)]Var[y_{out}(t)]}, \quad (19)$$

where $y(t, d)$ denotes the target sequence, which is a part of $y_{STM}(t, d)$. $y_{out}(t)$ represents the predictive output from RC. $y_{out}(t)$, Cov , and Var represent the predictive output, covariance, and variance, respectively. cor^2 is the square of the correlation coefficient, which ranges from 0 to 1. The higher the cor^2 value, the better the prediction. For general systems with the STM ability, cor^2 will eventually decay to zero with an increase in the memory delay d [65]. This well-known fading memory process is affected by the memory capacity of the system, which can be mathematically expressed as [27]:

$$C = \sum_{d=1}^{d_{max}} cor(d)^2, \quad (20)$$

where C is the memory capacity. Because the objective of the optimization algorithm is to find the minimum value of the loss function, $\frac{1}{C}$ should be set as the loss function in Optuna. Because C reflects the memory capacity of a parameter determined system, the optimization process include all the delay targets from $d = 0$ to $d = d_{max}$. Ideally, d_{max} for calculating C is positive infinity [65], which is not realistic for physical RC. Therefore, d_{max} is generally set as the maximum d that satisfies $cor^2 > 0.01$ [25,26]. For the OBRC, d_{max} was 20 [31], whereas d_{max} was 30 for the TOBRC in our testing. Therefore, to ensure a fair comparison, $d_{max} = 20$ was applied to the loss function of Optuna and $d_{max} = 30$ was applied to calculate C in the evaluation stage. According to our test experience, the change in C is negligible with a small step size in changing n . Thus, following the previous study [31], we adopt the normalized virtual node number $N = \frac{n}{20}$ instead of n . It is

worth noting that N is an integer and the actual virtual node number is still n . Because the maximum N explored is $N_{max} = 20$, the optimization range of τ is $[0, n_{max}]$, that is, $[0, 400]$. The optimization ranges of a and b are $[0, 1]$, in accordance with a previous research on the OBRC [31,32].

First, we investigated the relationship between N and memory capacity using the STM benchmark C_{STM} . As shown in Fig. 3(a), with an increase in N , the C_{STM} value of the TOBRC increases from 10.7 ($N = 1$) to 20.6 ($N = 20$), whereas the C_{STM} value of the OBRC fluctuates at approximately 2.7. The greater C_{STM} value of the TOBRC can be attributed to the existence of a time-delayed feedback loop, which introduces additional previous information back to the virtual nodes; thus, memory fading is postponed. When N increases, it indicates an increase in the virtual nodes with past information, thereby increasing C_{STM} . In contrast, owing to the lack of a time-delayed feedback loop, C_{STM} of the OBRC is significantly lower and does not increase with increasing N . Fig. 3(b) depicts the relationship curves between d and cor^2 of the TOBRC with different N . The attenuation of cor^2 to 0 requires a higher d with an increase in N , which confirms the previous explanation. Therefore, in terms of the performance, a greater N value is preferred for the TOBRC. However, a greater N requires a longer input period or higher sampling frequency, which may increase the requirements for peripherals of the physical RC systems.

Then, contributions of ϵ and τ to C_{STM} of the TOBRC were evaluated. Different from the discussion of N , the parameters a , b , ϵ , and τ are correlated, which can affect the structure of the effective potential $V(x)$. To avoid the influence of parameter correlation on $V(x)$, all the parameters, except the one discussed, were fixed as the optimized parameters under the condition of $N = 15$. Fig. 3(c) shows the relationship curves between cor^2 and d under different ϵ . When ϵ increases, a greater d is required to reduce cor^2 to 0. This is because increasing ϵ is essentially equivalent to increasing the weight of the previous feedback information, which alleviates the decline in cor^2 . It is worth noting that increasing ϵ also means an increase in the feedback signal amplitude, which increases the Joule heat generated in the reservoir part. In addition, a wider optimization range of ϵ burdens the optimization algorithm. Hence, we suggest keeping the optimization range of ϵ at $[0, 1]$ in the practical configuration of the TOBRC. As for the other parameter, the relationship curves with different τ values between cor^2 and d are depicted in Fig. 3(d). With the increase in τ , the area under the cor^2 - d curve first increases and then decreases. Interestingly, the τ value corresponding to the maximum area is the value closest to the actual node number n ($N = 15$, i.e., $n = 300$). This indicates that, except for the first code, each code receives the feedback information of the previous code in the initial stage of its duration, thus enhancing the memory of the past information. If τ is far less than n , excessive information of the first code is retained within the duration of the first code, which cannot contribute to the memory effect. If τ is much greater than n , many nodes in the duration of the second code cannot receive the information of the first code. Therefore, one can expect that the optimized τ is close to n .

Although it can be inferred from the description in Section 3.1 that the TOBRC should have stronger SR effect induced noise robustness than OBRC owing to its deeper effective potential, a quantitative verification is still lacking. Thus, we performed an STM benchmark with input data containing four different noise levels. As shown in Fig. 4(a–c), with an increase in N , the C_{STM} value of the TOBRC increases from approximately 10 ($N = 1$) to approximately 20 ($N = 20$) for $D = 0.001, 0.01$, and 0.1. By comparison, when $D = 0.001, 0.01$, and 0.1, the average values of C_{STM} for OBRC were 2.42, 1.98, and 1.02, respectively, showing a downtrend. Notably, this downtrend is different from that reported previously [31], in which C_{STM} of OBRC could be improved by increasing N . This is because N was adjusted while the other parameters were fixed in the previous study, which led to the offset of the optimal system parameters. In this work, because of the optimization algorithm, the discussed C_{STM} is the upper limit of the tested RC. Therefore, it should be

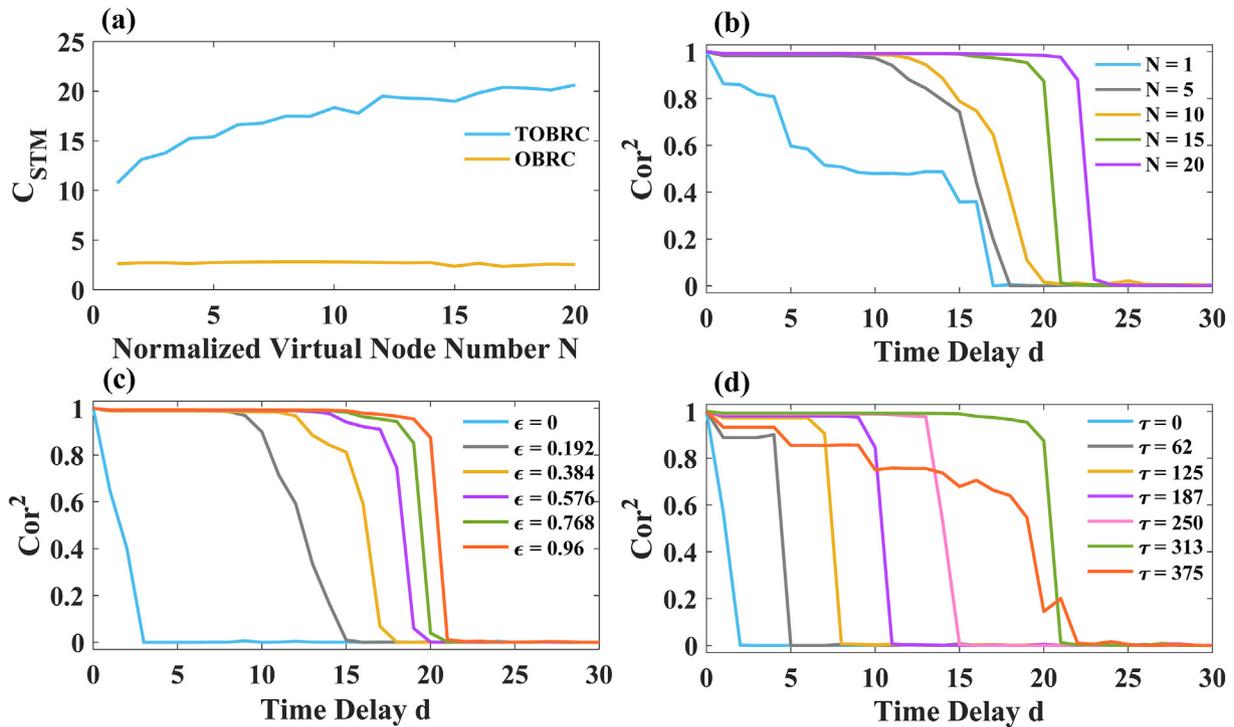


Fig. 3. (a) Memory capacities C_{STM} of TOBRC and OBRC versus N . (b) Relationship curves between d and cor^2 of TOBRC with different N . When the other optimized parameters are fixed under the condition of $N = 15$, (c) relationship curves between d and cor^2 of TOBRC with different (c) ϵ and (d) τ . For all cases shown, the noise intensity $D = 0$.

understood as the decrease in the upper limit of the OBRC performance, which is caused by its limited anti-noise ability. When $D = 1$, as displayed in Fig. 4(d), C_{STM} of the TOBRC decreases markedly from 2.13 ($N = 1$) to 0.37 ($N = 20$) with an increase in N . In addition, the

average value of C_{STM} for the OBRC is 0.18. This lower C_{STM} compared with that of a smaller D is due to the dominance of the large noise in the input data, which significantly affects the memory effect. In particular, for the TOBRC, owing to the time-delayed feedback loop, its

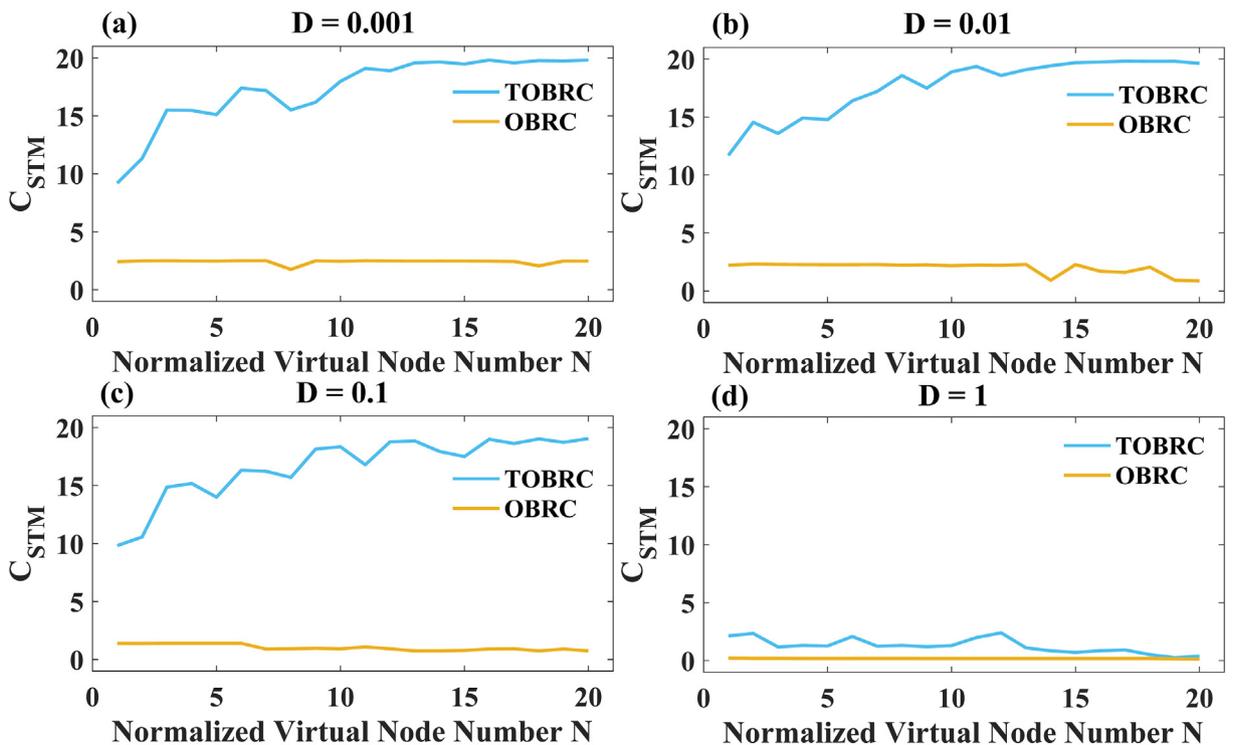


Fig. 4. Memory capacities C_{STM} of TOBRC and OBRC with noise intensity D of (a) 0.001, (b) 0.01, (c) 0.1, and (d) 1 in input data.

nodes are subject to past information dominated by large noise. As a result, an increase in N reduces C_{STM} of the TOBRC. In conclusion, it is quantitatively proven that the TOBRC has a stronger noise robustness than the OBRC on the STM benchmark. Moreover, based on above discussion, the selection strategy of N for tasks mainly requires ensuring the STM ability: in the case of low noise intensity, selecting an N value as high as possible is preferred for realizing better STM capability. In a noise-dominated environment, a smaller N should be selected to avoid any further impact of noise.

Finally, C_{STM} of the proposed TOBRC was compared with those of other advanced RC systems [26–28,66]. Among them, the STM ability and nonlinearity of the spin-torque oscillator RC originate from the relaxation dynamics of the physical unit [26], which represents the physical RC without any time-delayed feedback. The magnonic RC and enhanced magnonic RC are constructed based on the spin-wave delay line [27,28], which represent physical RC with a time-delayed feedback loop. Different from the former classical RC, the quantum RC is a stronger framework based on disordered quantum dynamics [66]. Because the upper bound of memory capacity C is equal to the computational node number of RC system [67,68], which includes the number of true nodes and virtual nodes, the computational node number of each RC is given in Table 1 for fair comparison. Besides, considering the noise and noise suppression methods in the real-world RCs, the noise intensity is set to 0.01 for the OBRC and TOBRC. As shown in Table 1, whether the computational node number is small or large, the C_{STM} value of the TOBRC is significantly higher than the maximum C_{STM} of other classical physical RC systems. When the computational node number is large, the C_{STM} of TOBRC is comparable to that of quantum RC.

4.2. PC benchmark

Because the nonlinearity of the RC cannot be assessed by the STM benchmark, the PC benchmark was employed, which requires both STM ability and strong nonlinearity [27]. The training data of the PC task can be obtained from the following:

$$y_{pc}(t, d) = (u(t - d) + u(t - d + 1) + \dots + u(t)) \bmod 2. \quad (21)$$

The implementation of the input data and optimization algorithm was consistent with the STM benchmark. As evaluation indices, the square of the correlation coefficient cor^2 and memory capacity C_{PC} were calculated using Eqs. (19) and (20), respectively.

First, we performed tests to explore the influence of N on C_{PC} without noise. As illustrated in Fig. 5(a), C_{PC} of the OBRC decreases from 2.8 to 2.3 with an increase in N . This downtrend is similar to that observed in a previous study [31]. This can be explained by the fact that the relaxation of the physical processes is time consuming; thus, there is a certain correlation between the adjacent virtual nodes in a time-multiplexing structure. When N increases, the increased linear correlation reduces system nonlinearity, leading to a decrease in C_{PC} . Inevitably, the nonlinearity of the TOBRC is also negatively affected by a large N .

Table 1

C_{STM} comparison among different physical RC systems.

Physical RC	Computational node number	Maximum C_{STM}
Spin torque oscillator RC [26]	250	3.6
Magnonic RC [27]	20	3.16
Enhanced magnonic RC [28]	40	4.82
Quantum RC [66]	50	20.0
OBRC ($N = 1, D = 0.01$)	20	2.62
OBRC ($N = 20, D = 0.01$)	400	2.52
TOBRC ($N = 1, D = 0.01$)	20	11.69
TOBRC ($N = 20, D = 0.01$)	400	19.62

Best performance of proposed system are in bold.

However, as shown in Fig. 5(a), the C_{PC} value of the TOBRC increases from 3.26 to 6.02 with an increase in N . This reveals that the influence of the STM ability improvement is greater than that of the nonlinearity attenuation caused by increasing N . As shown in Fig. 5(b), increasing N results in area enlargement under the cor^2-d curve, which further confirms the former conclusion.

As depicted in Fig. 5(a), when $\epsilon = 0$, the TOBRC degenerates into the OBRC with a high attenuation rate of cor^2 . Once $\epsilon > 0$, cor^2 decays to 0 at a greater d compared with that under the $\epsilon = 0$ condition. However, the increase in ϵ cannot lead to the continuous expansion of the area under the cor^2-d curve. This may be because increasing the weight of the previous feedback information also increases the correlation between the nodes, which restricts the performance improvement of the TOBRC on the PC benchmark. For the hyperparameter τ , its increase first induces an increase and then a decrease in the area under the cor^2-d curve, which is similar to the trend observed in Fig. 3(d). This indicates that the change in τ only affects the STM ability but cannot influence the nonlinearity of the TOBRC.

Thereafter, the performances of the OBRC and TOBRC are assessed on the PC benchmark with four different noise intensities. In the previous discussion, it is known that the effect of noise with $D \leq 0.1$ on the STM ability of the TOBRC is negligible. Hence, the C_{PC} value of the TOBRC can still be enhanced by increasing N , as displayed in Fig. 6(a–c). However, with the increase in D in TOBRC, C_{PC} corresponding to each N starts to decrease gradually, as shown in Fig. 6. This implies that the noise has a negative impact on the nonlinearity of the TOBRC. For the OBRC, the nonlinearity reduction caused by the increase in N leads to a downtrend in C_{PC} under all the tested noise intensities. In addition, due to the noise, each $C_{PC}-N$ curve moves downward. When $D = 1$, the average C_{PC} of the TOBRC decreases to 0.0847, which is only slightly higher than the average C_{PC} of the OBRC. Combined with the above phenomena, we can obtain the selection strategy of N for the TOBRC to deal with tasks that require strong nonlinearity, and in the case of small D , N should be selected as high as possible. In the case of a large D , it is better to increase the input amplitude to suppress the influence of noise [31], and then a greater N can be selected.

Similar to that described in Section 4.1, C_{PC} of the proposed TOBRC is compared with those of other advanced RC systems [26–28,66]. As shown in Table 2, although the C_{PC} value of the TOBRC is smaller than that of quantum RC, it is markedly greater than the maximum C_{PC} values of the other classical physical RC systems when the computational node number is 400.

4.3. Recognition of MNIST handwritten digits

The MNIST database contains handwritten digital images collected from 500 different people. These handwritten digital images deviate from the standard printing font owing to the difference in penmanship. This irregular deviation can also be considered a time-varying random noise, which makes the MNIST handwritten digit recognition a popular classification task in evaluating the performance of a learning system in real-world applications [69–71]. In this section, we compare the performance of the OBRC and TOBRC on the MNIST handwritten digit recognition task. There are 70,000 handwritten digital images in the MNIST database, of which 60,000 digits belong to the training set and 10,000 digits belong to the test set. Because all the images are grayscale images with 28-by-28 pixels, it is necessary to convert them into one-dimensional data to input RC with a single physical node. In the process of treating handwritten digital images shown in Fig. 7(a), an n -by-28 masking matrix W_{in} is randomly generated with a distribution of $[-0.5, 0.5]$. For the Optuna, if $b > 0.01$, we found that the drifting value in the computer exceeded the valid floating-point value, resulting in an optimization failure. Therefore, we set the optimization ranges of a and b as $[0, 1]$ and $[0, 0.01]$, respectively. In addition, the optimization range of τ was set to $[0, n_{max}]$. As for the evaluation index, the

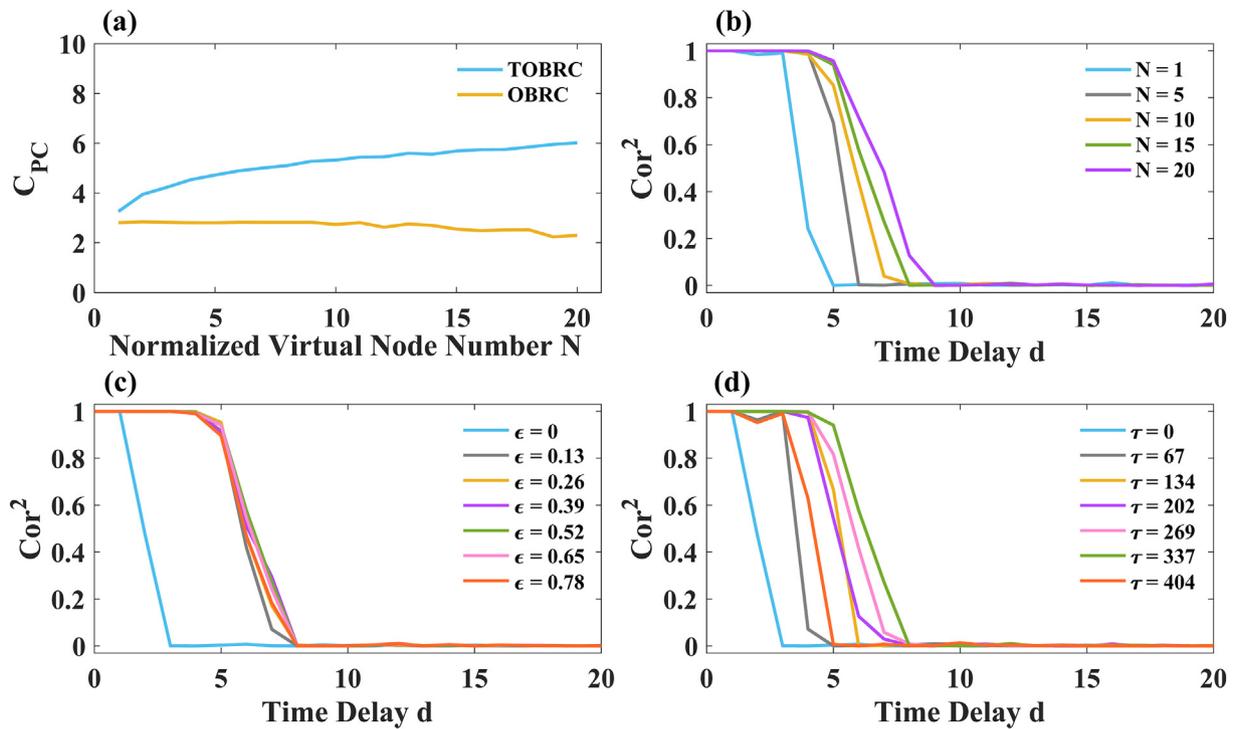


Fig. 5. (a) Memory capacities C_{PC} of TOBRC and OBRC versus N . (b) Relationship curves between d and Cor^2 of the TOBRC with different N . When the other optimized parameters are fixed under the condition of $N = 15$. (c) Relationship curves between d and Cor^2 of TOBRC with different (c) ϵ and (d) τ . For all the cases shown in this figure, the noise intensity $D = 0$.

recognition error rate was utilized, which was also the loss function set in Optuna.

Fig. 7(b) shows a comparison of the recognition error rates between TOBRC and OBRC under three different n values; with the increase in

n , the recognition error rate decreases, which is consistent with the trend reported previously [32]. For further comparison, we selected two advanced RC devices that were benchmarked by the MNIST handwritten digit recognition task. The first is the diffusive memristor

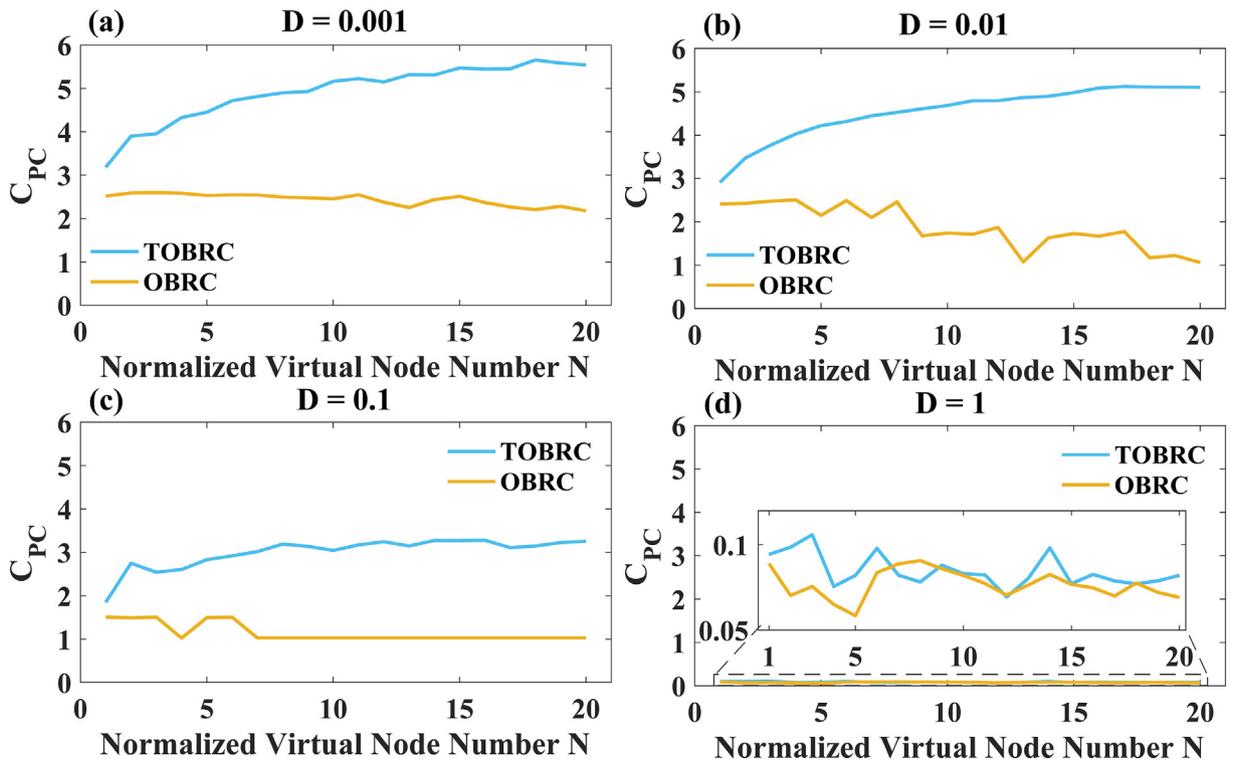


Fig. 6. Memory capacities C_{PC} of TOBRC and OBRC with noise intensity D of (a) 0.001, (b) 0.01, (c) 0.1, and (d) 1 in input data.

Table 2
 C_{PC} comparison among different physical RC systems.

Physical RC	Computational node number	Maximum C_{PC}
Spin torque oscillator RC [26]	250	3.1
Magnonic RC [27]	20	1.27
Enhanced magnonic RC [28]	40	2.23
Quantum RC [66]	225	6.0
OBRC ($N = 1, D = 0.01$)	20	2.80
OBRC ($N = 20, D = 0.01$)	400	2.30
TOBRC ($N = 1, D = 0.01$)	20	2.91
TOBRC ($N = 20, D = 0.01$)	400	5.10

Best performance of proposed system are in bold.

RC without time-delayed feedback, which has over 1000 real physical nodes [29]. The second is the in-material RC with a time-delayed feedback function, which forms a large number of physical nodes through self-organizing nanowire networks [30]. Table 3 presents a comparison of the TOBRC with 1000 virtual nodes and other systems. The recognition error rate of the TOBRC is comparable to that of In-material RC and significantly lower than that of diffusive memristors RC. Notably, in addition to the reservoir part, the other components of the physical RC contribute to its learning ability [26,27]. Hence, if the physical implementation of the TOBRC includes electronic components other than the reservoir, it can be expected that its performance may be further improved.

5. Conclusions

In this work, an RC model based on a time-delay feedback overdamped bistable stochastic resonance system (TOBRC) was proposed. In addition, the Optuna algorithm was adopted to adaptively optimize the highly correlated hyperparameters of the TOBRC and OBRC. For the evaluation, we compared the STM ability and nonlinearity of the OBRC and TOBRC using STM and PC tasks. The effects of the virtual node number and time-delayed feedback hyperparameters on the proposed physical RC model were discussed in detail. The results verify the positive contribution of the time-delayed feedback structure to the TOBRC, which helped make the STM ability and nonlinearity of the TOBRC stronger than those of OBRC in all discussed cases. Moreover, compared with some recently reported physical RC systems, TOBRC exhibited significant advantages in terms of the STM ability and nonlinearity. Finally, we evaluated the learning ability of the TOBRC on a practical

Table 3
Recognition error rate comparison among different physical RC.

Physical RC	Recognition error rate
Diffusive memristors RC [29]	0.17
In-material RC [30]	0.096
OBRC	0.166
TOBRC	0.098

Best performance of proposed system are in bold.

challenging dataset using the MNIST handwritten digit recognition benchmark. Compared with the OBRC under three different virtual node numbers, the TOBRC could produce a lower recognition error rate, which is competitive with advanced RC systems with large amounts of physical nodes.

However, the noise type discussed in this study is limited. In the future, the influence of real-world noise, such as colored noise [72] and levy noise [73], on the TOBRC will be worth studying. In addition, with the deep network becoming a mainstream technology in the field of machine learning [71,74,75], research on the integration of TOBRC and deep learning frameworks is another promising direction.

CRedit authorship contribution statement

Zhuozheng Shi: Software, Investigation, Visualization, Writing – original draft. **Zhiqiang Liao:** Conceptualization, Methodology, Validation, Visualization, Writing – original draft. **Hitoshi Tabata:** Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Zhiqiang Liao would like to acknowledge the support provided by the Quantum Science and Technology Fellowship Program of the University of Tokyo. This research was supported by the Basic Research Grant (Hybrid AI) of the Institute for AI and Beyond of the University of Tokyo and JSPS KAKENHI [Grant Number JP20H05651].

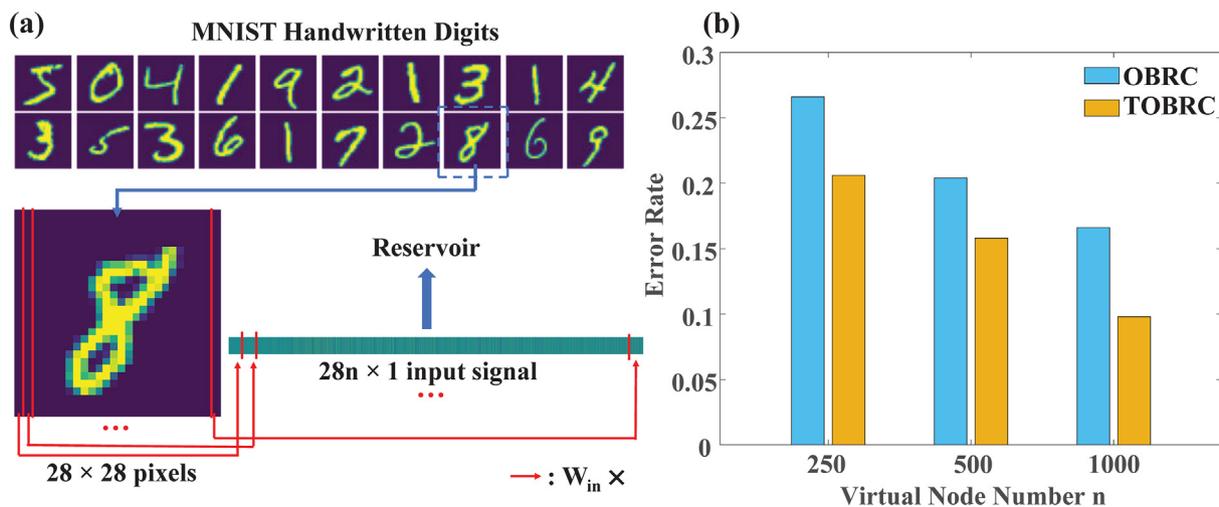


Fig. 7. (a) Workflow of treating handwritten digital images before inputting them into the reservoir. (b) Error rate of recognition when virtual node numbers are 250, 500, and 1000.

References

- [1] Lv Z, Wang Y, Chen J, Wang J, Zhou Y, Han S-T. Semiconductor quantum dots for memories and neuromorphic computing systems. *Chem Rev.* 2020;120:3941–4006.
- [2] Marković D, Mizrahi A, Querlioz D, Grollier J. Physics for neuromorphic computing. *Nat Rev Phys.* 2020;2:499–510.
- [3] Windhorst U. On the role of recurrent inhibitory feedback in motor control. *Prog Neurobiol.* 1996;49:517–87.
- [4] Wigström H. A neuron model with learning capability and its relation to mechanisms of association. *Kybernetik.* 1973;12:204–15.
- [5] Windhorst U. Auxiliary spinal networks for signal focussing in the segmental stretch reflex system. *Biol Cybern.* 1979;34:125–35.
- [6] Perruquía A, Yu W. Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: an overview. *Neurocomputing.* 2021;438:145–54.
- [7] Mücke NT, Bohté SM, Oosterlee CW. Reduced order modeling for parameterized time-dependent PDEs using spatially and memory aware deep learning. *J Comput Sci.* 2021;53:101408.
- [8] Begy V, Barisits M, Lassnig M, Schikuta E. Forecasting network throughput of remote data access in computing grids. *J Comput Sci.* 2020;44:101158.
- [9] Khalil K, Eldash O, Kumar A, Bayoumi M. Economic LSTM approach for recurrent neural networks. *IEEE Trans Circuits Syst Express Briefs.* 2019;66:1885–9.
- [10] Zainab M, Usmani AR, Mehrban S, Hussain M. FPGA based implementations of RNN and CNN: a brief analysis. 2019 international conference on innovative computing (ICIC). IEEE; 2019. p. 1–8.
- [11] Nepomnyashchiy O, Khantimirov A, Galayko D, Sirotnina N. Method of recurrent neural network hardware implementation. *Computer science on-line conference.* Springer; 2020. p. 429–37.
- [12] Siegelmann HT, Sontag ED. Turing computability with neural nets. *Appl Math Lett.* 1991;4:77–80.
- [13] Liu J, Sun T, Luo Y, Yang S, Cao Y, Zhai J. Echo state network optimization using binary grey wolf algorithm. *Neurocomputing.* 2020;385:310–8.
- [14] Jaeger H. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report; 2001. p. 13.
- [15] Maass W, Natschläger T, Markram H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 2002;14:2531–60.
- [16] Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science.* 2004;304:78–80.
- [17] Doan NAK, Polifke W, Magri L. Physics-informed echo state networks. *J Comput Sci.* 2020;47:101237.
- [18] Song Z, Wu K, Shao J. Destination prediction using deep echo state network. *Neurocomputing.* 2020;406:343–53.
- [19] Guo Y, Yang X, Wang Y, Wang F, Chen B. Online robust echo state broad learning system. *Neurocomputing.* 2021;464:438–49.
- [20] Nakajima K. Physical reservoir computing—an introductory perspective. *Jpn J Appl Phys.* 2020;59:060501.
- [21] Nakajima K, Fischer I. Reservoir computing: theory, physical implementations, and applications. Springer; 2021.
- [22] Dejonckheere A, Dupont F, Smerieri A, Fang L, Oudar J-L, Haelterman M, et al. All-optical reservoir computer based on saturation of absorption. *Opt Express.* 2014;22:10868–81.
- [23] Larger L, Soriano MC, Brunner D, Appeltant L, Gutiérrez JM, Pesquera L, et al. Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Opt Express.* 2012;20:3241–9.
- [24] Dupont F, Schneider B, Smerieri A, Haelterman M, Massar S. All-optical reservoir computing. *Opt Express.* 2012;20:22783–95.
- [25] Furuta T, Fujii K, Nakajima K, Tsunegi S, Kubota H, Suzuki Y, et al. Macromagnetic simulation for reservoir computing utilizing spin dynamics in magnetic tunnel junctions. *Phys Rev Appl.* 2018;10:034063.
- [26] Tsunegi S, Taniguchi T, Nakajima K, Miwa S, Yakushiji K, Fukushima A, et al. Physical reservoir computing based on spin torque oscillator with forced synchronization. *Appl Phys Lett.* 2019;114:164101.
- [27] Watt S, Kostylev M. Reservoir computing using a spin-wave delay-line active-ring resonator based on yttrium-iron-garnet film. *Phys Rev Appl.* 2020;13:034057.
- [28] Watt S, Kostylev M, Ustinov AB, Kalinikos BA. Implementing a magnonic reservoir computer model based on time-delay multiplexing. *Phys Rev Appl.* 2021;15:064060.
- [29] Midya R, Wang Z, Asapu S, Zhang X, Rao M, Song W, et al. Reservoir computing using diffusive memristors. *Adv Intell Syst.* 2019;1:1900084.
- [30] Milano G, Pedretti G, Montano K, Ricci S, Hashemkhani S, Boarino L, et al. In materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks. *Nat Mater.* 2021;21:195–202.
- [31] Liao Z, Wang Z, Yamahara H, Tabata H. Low-power-consumption physical reservoir computing model based on overdamped bistable stochastic resonance system. *Neurocomputing.* 2021;468:137–47.
- [32] Liao Z, Wang Z, Yamahara H, Tabata H. Echo state network activation function based on bistable stochastic resonance. *Chaos Solitons Fractals.* 2021;153:111503.
- [33] Vemuri G, Roy R. Stochastic resonance in a bistable ring laser. *Phys Rev A.* 1989;39:4668.
- [34] Spagnolo B, Valenti D, Guarcello C, Carollo A, Adorno DP, Spezia S, et al. Noise-induced effects in nonlinear relaxation of condensed matter systems. *Chaos Solitons Fractals.* 2015;81:412–24.
- [35] Suo J, Wang H, Shen X, Dong H. System design and circuit implementation of stochastic resonance for underwater weak signals detection. 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). IEEE; 2018. p. 1–6.
- [36] Kang Y, Liu R, Mao X. Aperiodic stochastic resonance in neural information processing with Gaussian colored noise. *Cogn Neurodyn.* 2021;15:517–32.
- [37] McNamara B, Wiesenfeld K. Theory of stochastic resonance. *Phys Rev A.* 1989;39:4854.
- [38] Hänggi P, Talkner P, Borkovec M. Reaction-rate theory: fifty years after kramers. *Rev Mod Phys.* 1990;62:251.
- [39] Qiao Z, Shu X. Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery. *Chaos Solitons Fractals.* 2021;145:110813.
- [40] Wang Z, Yang J, Guo Y. Unknown fault feature extraction of rolling bearings under variable speed conditions based on statistical complexity measures. *Mech Syst Signal Process.* 2022;172:108964.
- [41] Yang J, Yang C, Zhuang X, Liu H, Wang Z. Unknown bearing fault diagnosis under time-varying speed conditions and strong noise background. *Nonlinear Dyn.* 2022;1–17.
- [42] Liu J, Hu B, Wang Y. Optimum adaptive array stochastic resonance in noisy grayscale image restoration. *Phys Lett A.* 2019;383:1457–65.
- [43] Liao Z, Jin S, Kuwahata A, Sekino M, Tabata H. Coherent detection stochastic resonance assisted biomagnetometer for measuring magnetocardiography at room temperature. *Appl Phys Express.* 2021;14:097001.
- [44] Ren Y, Pan Y, Duan F. SNR gain enhancement in a generalized matched filter using artificial optimal noise. *Chaos Solitons Fractals.* 2022;155:111741.
- [45] Duan F, Duan L, Chapeau-Blondeau F, Ren Y, Abbott D. Binary signal transmission in nonlinear sensors: stochastic resonance and human hand balance. *IEEE Instrum Meas Mag.* 2020;23:44–9.
- [46] Yashima J, Kusuno M, Sugimoto E, Sasaki H. Auditory noise improves balance control by cross-modal stochastic resonance. *Heliyon.* 2021;7:e08299.
- [47] Levin JE, Miller JP. Broadband neural encoding in the cricket cerebellar sensory system enhanced by stochastic resonance. *Nature.* 1996;380:165–8.
- [48] Hänze M, McMurtrie G, Baumann S, Malavolti L, Coppersmith SN, Loth S. Quantum stochastic resonance of individual Fe atoms. *Sci Adv.* 2021;7:eabg2616.
- [49] Liao Z, Ma K, Tang S, Sarker MS, Yamahara H, Tabata H. Phase locking of ultra-low power consumption stochastic magnetic bits induced by colored noise. *Chaos Solitons Fractals.* 2021;151:111262.
- [50] Liao Z, Ma K, Tang S, Yamahara H, Seki M, Tabata H. Nonbistable rectified linear unit-based gain-dissipative Ising spin network with stochastic resonance effect. *J Comput Sci.* 2022;62:101722.
- [51] Duan L, Duan F, Chapeau-Blondeau F, Abbott D. Stochastic resonance in hopfield neural networks for transmitting binary signals. *Phys Lett A.* 2020;384:126143.
- [52] Liu X, Duan L, Duan F, Chapeau-Blondeau F, Abbott D. Enhancing threshold neural network via suprathreshold stochastic resonance for pattern classification. *Phys Lett A.* 2021;403:127387.
- [53] Duan L, Duan F, Chapeau-Blondeau F, Abbott D. Noise-boosted backpropagation learning of feedforward threshold neural networks for function approximation. *IEEE Trans Instrum Meas.* 2021;70:1–12.
- [54] Ikemoto S, DallaLibera F, Hosoda K. Noise-modulated neural networks as an application of stochastic resonance. *Neurocomputing.* 2018;277:29–37.
- [55] Ikemoto S. Noise-modulated neural networks for selectively functionalizing sub-networks by exploiting stochastic resonance. *Neurocomputing.* 2021;448:1–9.
- [56] Lu S, He Q, Zhang H, Kong F. Enhanced rotating machine fault diagnosis based on time-delayed feedback stochastic resonance. *J Vib Acoust.* 2015;137.
- [57] Manohar SG, Muhammed K, Fallon SJ, Husain M. Motivation dynamically increases noise resistance by internal feedback during movement. *Neuropsychologia.* 2019;123:19–29.
- [58] Longo DM, Selimkhanov J, Kearns JD, Hasty J, Hoffmann A, Tsimring LS. Dual delayed feedback provides sensitivity and robustness to the NF- κ B signaling module. *PLoS Comput Biol.* 2013;9:e1003112.
- [59] Kuriki Y, Nakayama J, Takano K, Uchida A. Impact of input mask signals on delay-based photonic reservoir computing with semiconductor lasers. *Opt Express.* 2018;26:5777–88.
- [60] Guillozic S, L’Heureux I, Longtin A. Small delay approximation of stochastic delay differential equations. *Phys Rev E.* 1999;59:3970.
- [61] Shao R-H, Chen Y. Stochastic resonance in time-delayed bistable systems driven by weak periodic signal. *Physica A.* 2009;388:977–83.
- [62] Gammaitoni L, Hänggi P, Jung P, Marchesoni F. Stochastic resonance. *Rev Mod Phys.* 1998;70:223.
- [63] Kumar P, Erturk VS, Almusawa H. Mathematical structure of mosaic disease using microbial biostimulants via caputo and atangana-baleanu derivatives. *Results Phys.* 2021;24:104186.
- [64] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining.* 2019. p. 2623–31.
- [65] Jaeger H. Short term memory in echo state networks. *Technical Report GMD Report 152.* German National Research Center for Information Technology; 2002.
- [66] Fujii K, Nakajima K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys Rev Appl.* 2017;8:024030.
- [67] Dambre J, Verstraeten D, Schrauwen B, Massar S. Information processing capacity of dynamical systems. *Sci Rep.* 2012;2:1–7.
- [68] Kubota T, Takahashi H, Nakajima K. Unifying framework for information processing in stochastically driven dynamical systems. *Phys Rev Res.* 2021;3:043135.
- [69] Shibata T, Shinohara T, Ashida T, Ohta M, Ito K, Yamada S, et al. Linear and symmetric conductance response of magnetic domain wall type spin-memristor for analog neuromorphic computing. *Appl Phys Express.* 2020;13:043004.

- [70] Chen M-R, Chen B-P, Zeng G-Q, Lu K-D, Chu P. An adaptive fractional-order BP neural network based on extremal optimization for handwritten digits recognition. *Neurocomputing*. 2020;391:260–72.
- [71] Furuhashi G, Niiyama T, Sunada S. Physical deep learning based on optimal control of dynamical systems. *Phys Rev Appl*. 2021;15:034092.
- [72] Liao Z, Ma K, Sarker MS, Tang S, Yamahara H, Seki M, et al. Quantum analog annealing of gain-dissipative ising machine driven by colored Gaussian noise. *Adv Theory Simul*. 2022;5:2100497.
- [73] Liao Z, Ma K, Tang S, Sarker MS, Yamahara H, Tabata H. Influence of levy noise on subthreshold synchronization of spintronic stochastic neurons. *Results Phys*. 2021;27:104475.
- [74] Zhang Z, Gu GX. Finite-element-based deep-learning model for deformation behavior of digital materials. *Adv Theory Simul*. 2020;3:2000031.
- [75] Li Z, Tanaka G. Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction. *Neurocomputing*. 2022;467:115–29.