

Boosting reservoir computer performance with multiple delaysS. Kamyar Tavakoli ^{1,*} and André Longtin ^{1,2,†}¹*Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ontario, Canada K1N6N5*²*Centre for Neural Dynamics and AI, University of Ottawa, Ottawa, Ontario, Canada K1N6N5*

(Received 8 June 2023; accepted 1 April 2024; published 8 May 2024)

Time delays play a significant role in dynamical systems, as they affect their transient behavior and the dimensionality of their attractors. The number, values, and spacing of these time delays influences the eigenvalues of a nonlinear delay-differential system at its fixed point. Here we explore a multidelay system as the core computational element of a reservoir computer making predictions on its input in the usual regime close to fixed point instability. Variations in the number and separation of time delays are first examined to determine the effect of such parameters of the delay distribution on the effectiveness of time-delay reservoirs for nonlinear time series prediction. We demonstrate computationally that an optoelectronic device with multiple different delays can improve the mapping of scalar input into higher-dimensional dynamics, and thus its memory and prediction capabilities for input time series generated by low- and high-dimensional dynamical systems. In particular, this enhances the suitability of such reservoir computers for predicting input data with temporal correlations. Additionally, we highlight the pronounced harmful resonance condition for reservoir computing when using an electro-optic oscillator model with multiple delays. We illustrate that the resonance point may shift depending on the task at hand, such as cross prediction or multistep ahead prediction, in both single delay and multiple delay cases.

DOI: [10.1103/PhysRevE.109.054203](https://doi.org/10.1103/PhysRevE.109.054203)**I. INTRODUCTION**

Predicting the time evolution of chaotic systems has been a challenging research topic for many decades due to their complexity and sensitive dependence on initial conditions. Various methods have been developed to learn and predict chaotic time series, including artificial neural networks and phase space reconstruction methods [1–3]. One of the innovative models for chaotic time series prediction is based on the brain-inspired formalism developed by Jaeger and Haas [4]. This reservoir computing (RC) model proposes a network with fixed randomly initialized weight connections. In this framework, an input is injected into the neurons, and only the weights of the connections from the reservoir to the output layer (readout weights) are trained. This system has demonstrated its effectiveness in classifying chaotic signals, including inputs comprising a combination of two chaotic signals from the Lorenz system with different parameters [5].

Appeltant *et al.* [6] later introduced an innovative approach for time series prediction and speech recognition that utilizes a single nonlinear node in the form of a nonlinear delay-differential system, which benefits from the high-dimensional properties of delay systems. This delay-based RC method has been extensively studied in various domains, including quantum dot lasers [7], electronic [8], photonic, and optoelectronic devices [9,10]. This architecture has been shown to achieve the fast voice speech recognition when implemented with an optoelectronic device [11]. Furthermore, this architecture can be used for characterizing chaotic and hyperchaotic time series from data [12].

The rich dynamical properties of delay systems, such as a relatively large Kolmogorov-Sinai entropy in chaotic regimes, have thus opened up many possibilities for practical machine-learning applications. Certain architectures clearly demonstrate the importance of high-dimensional, nonlinear dynamics in RC. In the Lang-Kobayashi (LK) system, the modulation type significantly influences the reservoir's capability to map inputs to higher dimensions. It has been demonstrated that phase modulation is more effective than intensity modulation in utilizing these high-dimensional properties. This was further analyzed through studying the behavior of the steady states of the LK system [13].

Another example is the recently introduced deep time-delay RC, which enhances the time-series prediction capability [14]. This architecture displays characteristics analogous to those of convolutional neural networks, with each layer featuring a nonlinear delay system. Notably, in such multi-layered RC systems, the conditional Lyapunov exponents are influenced by the bifurcation point, which has implications for memory capacity [15]. In a more recent development, a neural network was introduced, utilizing modulated feedback loops [16]. This system incorporates the dynamics of a nonlinear delay differential equation where the feedback nonlinearity is applied to the sum of delayed terms with time-modulated strengths. By utilizing such features, gradient descent can be used to perform more sophisticated tasks.

The choice of time delay is a critical factor that significantly impacts the performance of a time delay RC. Generally, optimal delays cannot be determined *a priori* and may depend on the particular task; unsuitable delays can, in fact, degrade memory capacity (MC) and make time-series predictions difficult [17,18].

*stava089@uottawa.ca

†alongtin@uottawa.ca

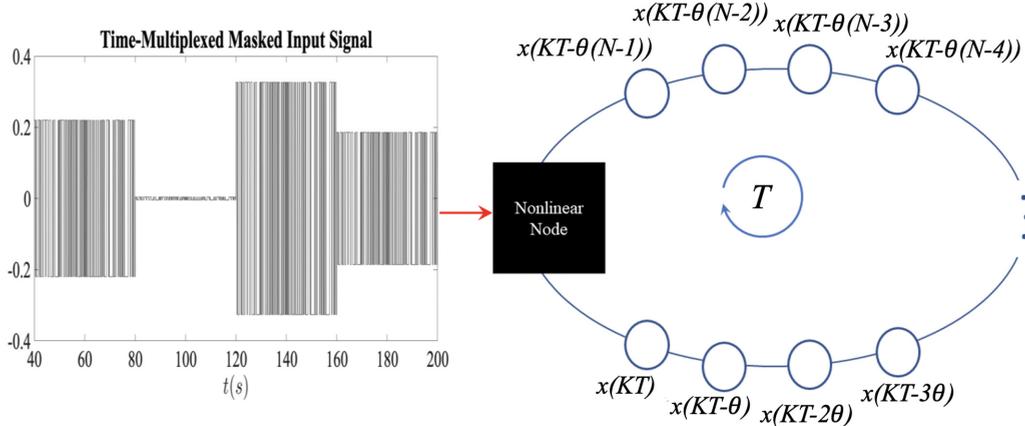


FIG. 1. Schematic of the single delay RC. The time-multiplexed masked input is injected into the nonlinear node, and travels along the loop with a clock cycle of T , here made equal to the delay τ . The virtual nodes provide time series that are combined in the output layer of the reservoir computer. These nodes are separated by a time interval of $\theta = \frac{T}{N}$. The time arguments for the virtual nodes illustrate how they span the state of the nonlinear delay system across the time interval $(K-1)T$ to KT , where K is a positive integer. Only the first and last virtual nodes are shown.

In this paper, we investigate the performance of RC with multiple delays for various tasks. We further explore the optimal choice of parameters for this multidelay system and study resonance phenomena in different scenarios, such as cross-prediction and multistep ahead prediction. We illustrate that employing multiple delays can have the benefit of increasing linear MC.

II. MULTIDELAY RESERVOIR COMPUTER (MDRC)

We focus on delay RC built from the electro-optic oscillator (EOO) model. In its simplest form, it is modeled as a first-order nonlinear delay differential equation with one delay. A schematic of the operation of the single delay reservoir with virtual nodes is shown in Fig. 1. The output of the nonlinear node (i.e., the laser) is sent through a delay loop (e.g., an optical fiber), and samples of the activity at specific virtual nodes along this loop are versions of the output of the nonlinear node delayed by a fraction of the whole loop delay τ . The high dimension of the phase space dynamics is such that the different virtual nodes behave approximately as independent variables. This is certainly the case in the chaotic regime, but also in the fixed point regime during transients caused by driving the system unidirectionally with inputs generated, e.g., by external chaotic systems, and more so when such inputs are time-multiplexed with random masks, as we will now describe.

The single delay RC model operates by injecting multiplexed input into a nonlinear node, which then propagates through a series of virtual nodes equally spaced along the delay loop with a clock-cycle of T . The RC receives a time-varying input, such as a chaotic time series, in the form of a sequence of constant samples of the time series (sample-and-hold values). Each sample, lasting one clock-cycle, advances the input-driven dynamics through the delay loop. This sampling interval can be different from the time delay that couples the virtual nodes together. Figure 1 shows via the time arguments the states of the virtual nodes along the K th cycle.

The constant value of the input during a sample is further modulated by a piecewise-constant random binary signal known as the mask; the same mask is used for each cycle. There are N piecewise-constant sections within each clock cycle T , corresponding to the N virtual nodes. The jumps between the mask values serve to excite complex transients in the nonlinear node, whose state would otherwise converge to a fixed point determined by the current value of the input sample. The product of the sampled input time series (e.g., from a chaotic system), $u(t)$, with this random binary mask constitutes the input forcing signal $J(t)$ to the RC. The amplitude of $J(t)$ is controlled by the scaling factor γ , which needs to be big enough to perturb the RC, but not too large to overwhelm it.

A. Stability analysis of multidelay EOO model with filter

In this paper, the single delayed feedback is replaced by multiple feedback paths, each with its own delay. For simplicity, we consider the case of equally spaced delays. As we will see, the spacing between these delays is an important parameter. Recent work has shown that increasing the spacing between delays can effectively suppress chaotic dynamics, resulting in the emergence of periodic or even fixed point behavior in first-order and higher-order nonlinear delay differential equations [19,20]. In addition to the modified version of the model that incorporates multiple delayed feedback pathways, we consider a filter term. The system's dynamics is governed by the following equations:

$$\begin{aligned} \dot{x} &= -x(t) - \delta y(t) + \frac{\beta}{M} \sum_{i=1}^M \sin^2(x(t - \tau_i) + \phi) \\ \frac{dy}{dt} &= x(t), \end{aligned} \quad (1)$$

where β is the gain coefficient, ϕ is the phase offset, and M represents the number of delays. Due to the presence of multiple timescales, this system can exhibit slow-fast dynamics. The filter term acts to eliminate any nonzero steady-state

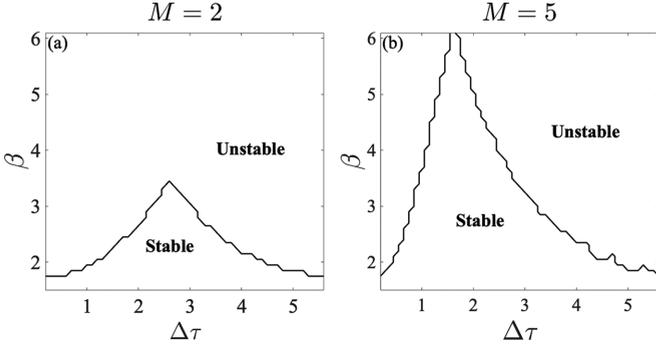


FIG. 2. Linear stability analysis of Eq. (1) for a multidelay RC for a system with (a) two delays and (b) five delays. As the number of delays increases, the peak of the stability region shifts towards smaller values of delay spacing $\Delta\tau$ and larger feedback strengths.

solution for the variable x . Throughout this study, we considered $\delta = 1$ and $\phi = \frac{2\pi}{5}$. For the MDRC case, we picked delays as

$$\tau_i = \tau_{\min} + i\Delta\tau, \quad (2)$$

where $\Delta\tau$ is the spacing between the periodically arranged delays and $i = 1 \dots M$. We note here that this spacing is part of the dynamics of the core of the RC, and determines its eigenvalues and transients at the edge of chaos. It should not be confused with the other use of $\Delta\tau$ in the context of nonlinear time series analysis, namely, the embedding delay applied to a low-dimensional time series to embed it into a higher dimension and analyze its dynamical invariants such as correlation dimension and predictability.

We first consider the intrinsic dynamics of the nonlinear node, independently of the time-multiplexed input and of the output weight matrix applied to its virtual nodes. The nonlinear core of a RC typically operates in a stable fixed point regime, but close to the destabilization point where a bifurcation to a limit cycle occurs. This avoids the divergence of the information about the input that would be caused by the sensitivity to initial conditions in a chaotic regime or the imposition of regularity in a limit cycle regime.

The stability diagram of Eq. (1) around its unique fixed point is plotted in Fig. 2 for different β and $\Delta\tau$, for two as well as five delays. Increasing the number of delays, the peak of the stability region moves towards the lower spacing between delays and higher feedback coefficient β . For example, for $M = 2$, the peak is located at $\Delta\tau = 2.66$ and $\beta = 3.6$, while for $M = 5$ the peak is located at $\Delta\tau = 1.6$ and $\beta = 6.1$. Note, however, that for both numbers of delays, the smallest spacings lead to loss of stability at smaller rather than larger β . For larger $\Delta\tau$, the boundary of the stabilized regime tends towards that for the single delay case. Here we refer to this latter regime as the large spacing regime. The regime where fixed point destabilization requires a stronger gain will be referred to as the short spacing regime.

B. Role of time delays in architecture

The multidelay configurations that we tested are robust to adding small random values to the delays, as this had little effect on the destabilization boundary. The following

sections will show that the regime needed for time series prediction can depend on the specifics of that task, such as the nature of the time series to be predicted.

The dynamical equations for the EOO with filter can be expressed as follows when masked input is injected into the nonlinear node:

$$\begin{aligned} \dot{x} &= -x(t) - \delta y(t) + \frac{\beta}{M} \sum_{i=1}^M \sin^2(x(t - \tau_i) + \phi) + \gamma J(t) \\ \frac{dy}{dt} &= x(t), \end{aligned} \quad (3)$$

where γ is the input scaling and $J(t)$ denotes the multiplexed and masked version of the discrete input data $u(k)$. Here we considered a binary mask, consisting of a series of randomly chosen -1 or 1 values. More sophisticated mask functions, such as a chaotic mask or a multilevel mask, could be used to improve the performance of the RC [21]. We considered 200 virtual nodes that are evenly distributed with the temporal distance summing up to $T = 40$ (in other words, $\theta = 0.2$). The states of the nodes are stored in a matrix X . Ideally, we aim to find a weight matrix that linearly combines the node states in a way that minimizes the difference between the predicted output and the target output. In our paper, we employ ridge regression to calculate this weight matrix, thereby mitigating the risks of overfitting:

$$\min (||XW - o||_2^2 + \lambda ||W||_2^2). \quad (4)$$

Here o represents the target output, λ is the Tikhonov regularization parameter, and $||\cdot||_2$ is the Euclidean norm. We set the Tikhonov regularization parameter λ to 10^{-8} to mitigate overfitting. The minimization in the preceding equation leads to the desired output weight matrix for the RC:

$$W = (X^T X + \lambda I)^{-1} (X^T o). \quad (5)$$

While linear regression is a well-established method for this minimization in RC, it has been shown that using an adaptive training weight matrix can offer significant improvements in reducing prediction errors [22].

Previous works have demonstrated that whenever the time delay τ and the clock cycle T are very close or equal, there is a resonance that is detrimental to the performance of RC as it leads to memory degradation [15,17,18]. This resonance effect was investigated analytically using the characteristic equation in Ref. [23]. The long delay approximation of the characteristic roots was used, which implies that the real part of the characteristic equation scales with the inverse of the time delay. Solving for the imaginary parts of the characteristic equation then yields [23]

$$\omega_k \approx \frac{\pi}{\tau} (2k - \nu), \quad (6)$$

with k the index of the eigenvalue and ν equal to either 0 or 1. The relative distance between two successive sample inputs in the drive $J(t)$ is the traversed angle ωT . Whenever T is a multiple of τ , the traversed angle is a multiple of π [17,23]. This results in overlapping responses and thus in the reduction of the available dimension for the RC to map inputs into higher dimensions [23].

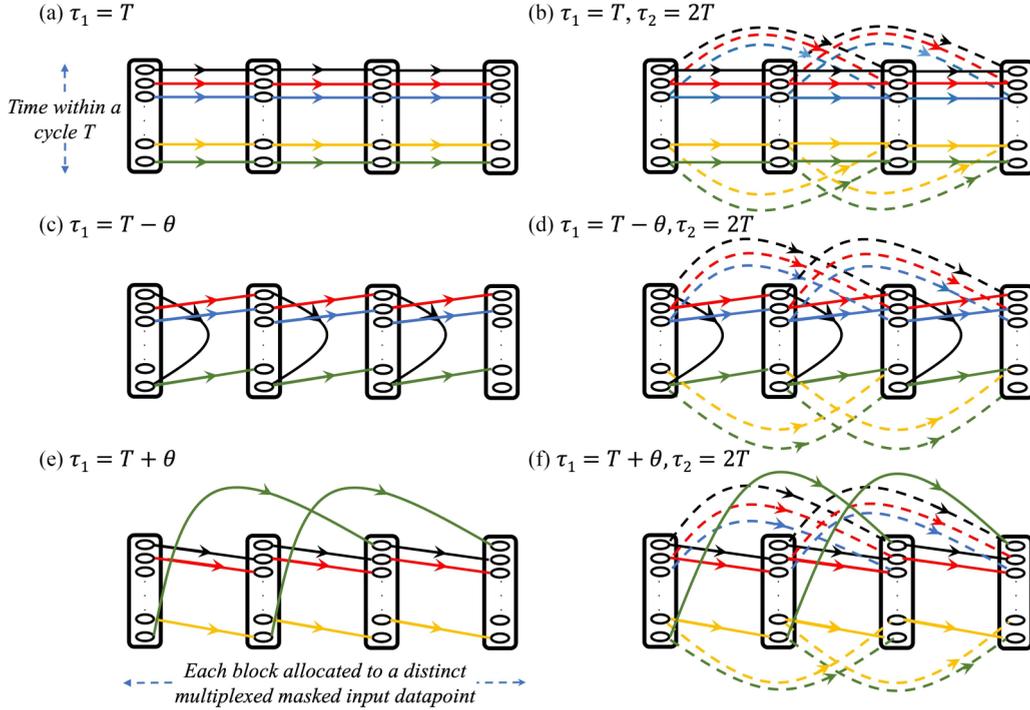


FIG. 3. Visualizing time-delay couplings between virtual nodes in the loop depicted in Fig. 1, showcasing various combinations of single and two delays with different time delay values and spacing between delays. Each vertical block (i.e., cycle of duration T) corresponds to a distinct input data point, which is a sample of the driving (input) time series. Within each block, there are N virtual nodes with a spacing of $\theta = \frac{T}{N}$. A single time delay is assumed in the first column. In (a), the time delay is set to the clock cycle ($\tau = T$). The dynamics can then be seen as mapping (i.e., affecting the dynamics of) each virtual node in one block to itself in the next block. In (c), the time delay is $\tau = T - \theta$, resulting in inner couplings where each virtual node maps backwards by θ . In this case, the dynamics of a given node in one block affects the dynamics of the previous node in the next block. Furthermore, the first node of each block also affects the dynamics of the last node in the same block. The second column depicts cases where each node's dynamics are affected by two delayed states. Here solid lines are associated with the smaller delay, and dashed lines with the larger delay. In (b), $\tau_1 = T$ and $\tau_2 = 2T$ ($\Delta\tau = T$), corresponding to the resonance case for single-step ahead prediction in multidelay reservoir computing; this case is more susceptible to poor performance. In (d), $\tau_1 = T - \theta$ and $\tau_2 = 2T$, so $\Delta\tau = \tau_2 - \tau_1 = T + \theta$. In (f), $\tau_1 = T + \theta$ and $\tau_2 = 2T$, so $\Delta\tau = T - \theta$.

The temporal dependence of virtual nodes on the states of previous nodes, considering different values of time delays and a fixed clock cycle, can be visualized in Fig. 3. The scheme is more subtle compared to the single delay RC. The first column depicts the single delay cases, while the second column shows the more intricate two delay case. The resonance condition is depicted in Fig. 3(a) for a single delay, while Figs. 3(c) and 3(e) illustrate the single delay case but in the desynchronized regime where τ and T are mismatched; this could serve as an alternative to avoid resonance. Figures 3(b), 3(d), and 3(f) illustrate the simplest scheme of MDRC, a system with two delays. In Fig. 3(b), the time delays are multiples of the clock cycle, which may result in poor prediction performance, as we will discuss in the next sections. In Figs. 3(d) and 3(f), where $\Delta\tau = T + \theta$ and $\Delta\tau = T - \theta$, respectively, represent alternatives for the multidelay case. Adjusting the spacing between delays could enhance forecasting accuracy, even when one of the delays is a multiple of the clock cycle. However, as we will see, these results highly depend on the complexity of the data, the autocorrelation of the input data, and the objectives of the reservoir, such as cross-prediction or multistep ahead prediction.

III. EFFECT OF INPUT SIGNAL ON PREDICTION

In this section, we examine this multidelay RC for signals originating from various sources with distinct properties. We show that in some cases, choosing short spacing between delays outperforms large spacing and vice versa. Given that many parameters affect the performance of the RC, we limit our study to parameters near the resonance regimes. The parameters are presented in Table I. We used 10 000 data inputs for the training process, followed by testing on 5000 data inputs for each input. The metric that we used to analyze the

TABLE I. Parameter values used for the simulation of Figs. 7–9.

	Lorenz x			Lorenz z			MG		
	β	γ	$\Delta\tau$	β	γ	$\Delta\tau$	β	γ	$\Delta\tau$
$M = 1$	1.2	0.2		1.35	0.16		1.6	0.28	
$M = 2$	2.3	0.24	2.66	2.2	0.16	2.66	3	0.2	2.66
$M = 2$	1.3	0.2	41	1.5	0.22	39	1.6	0.24	40
$M = 5$	4.1	0.24	1.6	3.5	0.12	1.6	5.5	0.16	1.6
$M = 5$	1.4	0.2	38	1.5	0.2	39	1.6	0.3	40

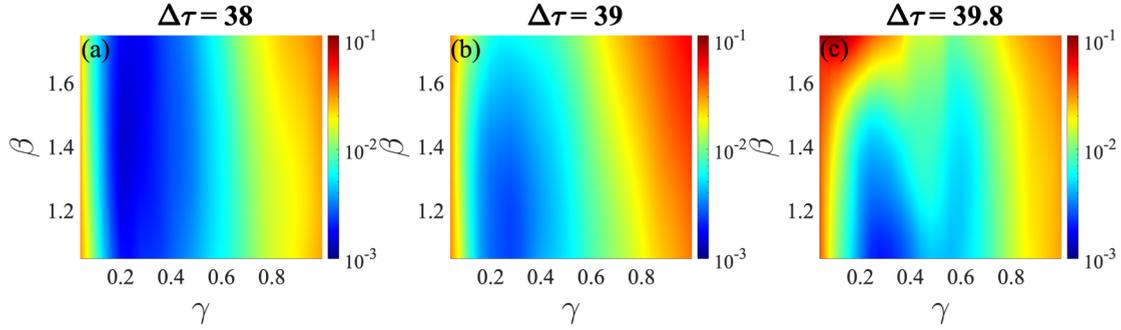


FIG. 4. Results of Lorenz x task prediction errors for three distinct cases of (a) $\Delta\tau=38$, (b) $\Delta\tau=39$, and (c) $\Delta\tau=39.8$, all with $M = 5$ when $\tau_{\min} = 39.8$ and $T = 40$. Each panel displays the input scaling (γ) on the x axis against the feedback strength (β) on the y axis.

error is the normalized root mean square error:

$$\text{NRMSE} = \sqrt{\frac{1}{K} \frac{\sum_{i=1}^K (o(i) - \hat{o}(i))^2}{\text{VAR}(o)}}. \quad (7)$$

Here the sum is over the number of target data points. The variable $\hat{o}(i)$ represents the predicted values, while $o(i)$ corresponds to the target output values. Additionally, we compared the linear MC of the RC, i.e., the ability to recall the past inputs, for the single and multiple delay cases.

Our strategy to choose delay parameters and focus our exploration of the large parameter space is as follows. Our initial stability analysis guided our selection for short delay spacings. We aimed to investigate the differences between short delay spacings—where feedback destabilization boundary is notably higher than in the single delay case—and large delay spacings, where the feedback strength’s destabilization threshold is close to that of a single-delay system.

A. Input from Lorenz model

We begin our study with the Lorenz model, a chaotic dynamical system, characterized by the following equations:

$$\dot{x} = \sigma(y(t) - x(t)), \quad (8)$$

$$\dot{y} = -x(t)z(t) + \rho x(t) - y(t), \quad (9)$$

$$\dot{z} = x(t)y(t) - bz(t). \quad (10)$$

We used the standard parameters $\sigma = 10$, $\rho = 28$, $b = 8/3$, and sampled the data every 0.05 seconds. Initially, we focus on predicting the x component of the Lorenz system with a one-step ahead prediction. The x component of the Lorenz system has the highest complexity and its autocorrelation function reveals little correlation between data points.

Finding the optimal parameters for the implementation of a multidelay system in the RC framework poses a significant challenge, as it requires precise determination of values for γ , β , $\Delta\tau$, and τ_{\min} . To address this, we categorize the problem based on the spacing between delays, which is depicted in Fig. 2, illustrating the impact of different spacings on system behavior. For short delay spacings, we set $\Delta\tau$ to a value that ensures the largest feedback strength maintains the system in a stable fixed point at steady state. As a case in point, we chose $\Delta\tau = 2.66$ for systems with two delays and $\Delta\tau = 1.6$ for systems with five delays, based on observed peaks of

destabilization with the highest feedback strength for each case.

For large delay spacings, we followed a more empirical approach. We first analyzed the prediction errors by varying the feedback strengths (β) and input scalings (γ) across different delay spacings ($\Delta\tau$). In these cases, The minimum time delay, τ_{\min} , is set to $T - \theta$, i.e., close to the clock cycle, which is set to 40 in all of our studies. This particular setting of τ_{\min} is based on previous findings that indicate improved performance in single delay cases under similar conditions. Figure 4 displays the prediction errors for three different cases of $\Delta\tau = 38$, $\Delta\tau = 39$, and $\Delta\tau = 39.8$ for $M = 5$.

It is evident from Fig. 4 that the optimum parameters for feedback strength and input scaling change with different spacing between delays. For $\Delta\tau = 38$ [Fig. 4(a)], the optimal parameters are identified as $\beta = 1.4$ and $\gamma = 0.2$. As $\Delta\tau$ increases to $\Delta\tau = 39$ [Fig. 4(b)], the optimum parameters shift to $\beta = 1.15$ and $\gamma = 0.28$, and for $\Delta\tau = 39.8$ [Fig. 4(c)], the ideal settings found to be $\beta = 1.05$ and $\gamma = 0.28$. Based on these findings, Fig. 5 plots the prediction error using these parameters across varying spacings between delays and different minimum time delays to identify the optimum spacing between delays near clock cycle $T = 40$.

It can be inferred from Fig. 5 that harmful resonance occurs when $\Delta\tau = T$, evidenced by the increased error across all panels. The minimum error is observed with the parameters used in Fig. 5(a). We applied these parameters to four different samples, each characterized by unique initial conditions for Lorenz x cases. These cases are illustrated in Fig. 6, demonstrating their behavior with increasing minimum time delays. This figure reveals a consistent trend in prediction errors across these parameters, leading to the conclusion that the fluctuations are intrinsic to the variations of τ_{\min} and not due to lack of averaging. After finding the approximated optimum feedback strength (β), input scaling (γ), and spacing between delays ($\Delta\tau$) through this process, we investigated the impact of different minimum time delays on the performance of RC, and the results are presented in Fig. 7.

It can be seen in Fig. 7(a) that when $\tau_{\min} < 15$ or so, the performance of the RC is superior for both large and small spacing between delays compared to the single delay case. However, as the minimum time delay value increases and approaches the clock cycle $T = 40$, the performance of the RC with single delay surpasses that of the RC with multiple delays with large spacing. Notably, the

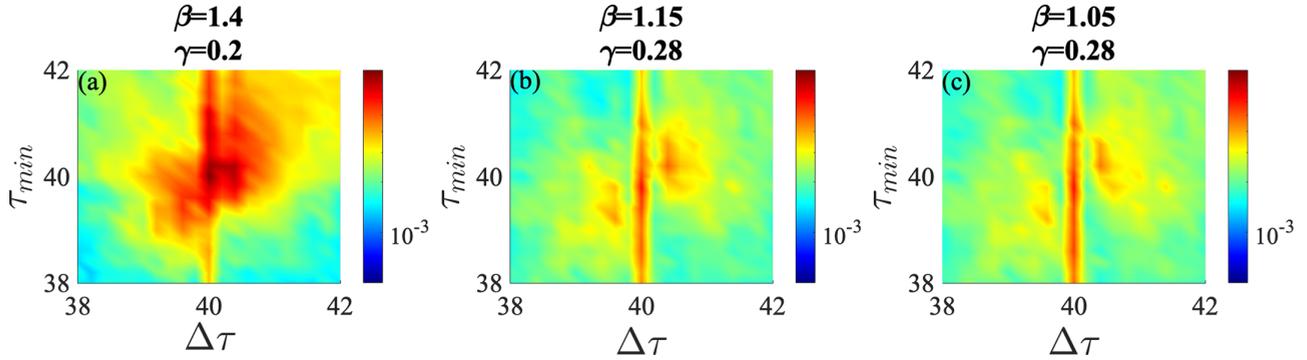


FIG. 5. Prediction error plots over a range of delay intervals ($\Delta\tau$) and minimum delays (τ_{\min}) centered on the values used in Fig. 4 for Lorenz x task after fixing β and γ to their respective optimal values found for each of the equivalent panels in Fig. 4 (that resulted in the lowest error for each case).

RC with five delays and small spacing exhibits the best performance.

In Fig. 7(b), we present the cross-prediction error for the z component of the Lorenz signal, with the x component as the input. Cross prediction is feasible for the Lorenz system when using the x component as the input [24], but the errors are significantly larger, as seen here. The sharp drop in error is observed at $\tau_{\min} = T - \theta$ and $\tau_{\min} = 2T - \theta$.

When considering the z component of the Lorenz system as the input, the trends observed are similar to those observed when the input was the x component, as depicted in Fig. 8. As the minimum time delay approaches the clock cycle, the error increases across all cases. When the spacing between the delays is set to be equal to the clock cycle, i.e., at the resonance condition, the error increases.

B. Input from Mackey-Glass model

The Mackey-Glass (MG) equation is a nonlinear delay differential equation that exhibits high-dimensional chaotic behavior whenever the time delay is large compared to its

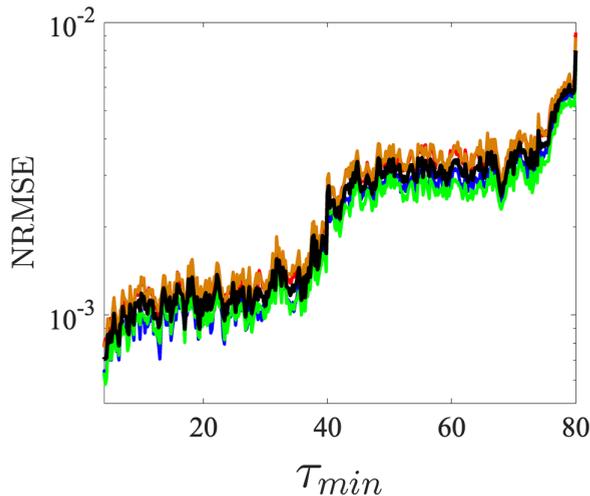


FIG. 6. Impact of minimum time delay (τ_{\min}) on the Lorenz x time series prediction, illustrated for four different samples, each obtained from unique initial conditions. The black line depicts the average of the samples' errors.

response time. Here, we consider the standard parameters for the MG model of $a = 0.1$ and $b = 0.2$ and the time delay of 17 in the following equation:

$$\frac{dx}{dt} = -ax(t) + \frac{bx(t-\tau)}{1+x^{10}(t-\tau)}. \quad (11)$$

For these parameters, the MG solution exhibits a much longer autocorrelation time than the Lorenz system used above. The differences in the prediction properties likely relate to the differing correlation times between these two systems. We illustrate in Fig. 9 the performance of the RC for the MG model input, where we try to predict 34 steps ahead for this equation [$o(i) = u(i+34)$], where u represents the input and $o(i)$ represents the target output data point, which is equivalent to the input data point 34 steps ahead].

In contrast to the Lorenz time series, the use of large spacing between delays yields superior results for the MG input. As shown in Fig. 9, the prediction error for the MG time-series with five delays and a large delay spacing can be up to two orders of magnitude better than the error for the single delay case or for short delay spacing. However, it is worth noting that there is a specific case where the error drops drastically for both short delay spacing and the single delay case. This occurs when $\tau_{\min} = 79.8$ (equivalent to $2T - \theta$), similar to the Lorenz z when the input was Lorenz x . The results for predicting the MG time series exhibit significant fluctuations when altering the minimum value of the time delay. This behavior was also evident in Fig. 6, and does not reflect poor averaging of the error due to limited data. These fluctuations, however, diminish in amplitude as more delays are incorporated into the equations. We have compared the prediction errors for scenarios with five and ten delays, with large spacing between delays, as illustrated in Fig. 10. It can be observed that increasing the number of delays from five to ten results in a higher minimum error for the ten-delay scenario compared to that observed for five delays even though the error fluctuates less with τ_{\min} .

C. Narma10 task

The Narma10 task is a standard task used to evaluate RC performance. It involves generating a discrete target output based on an input sequence u_k , which is randomly drawn from

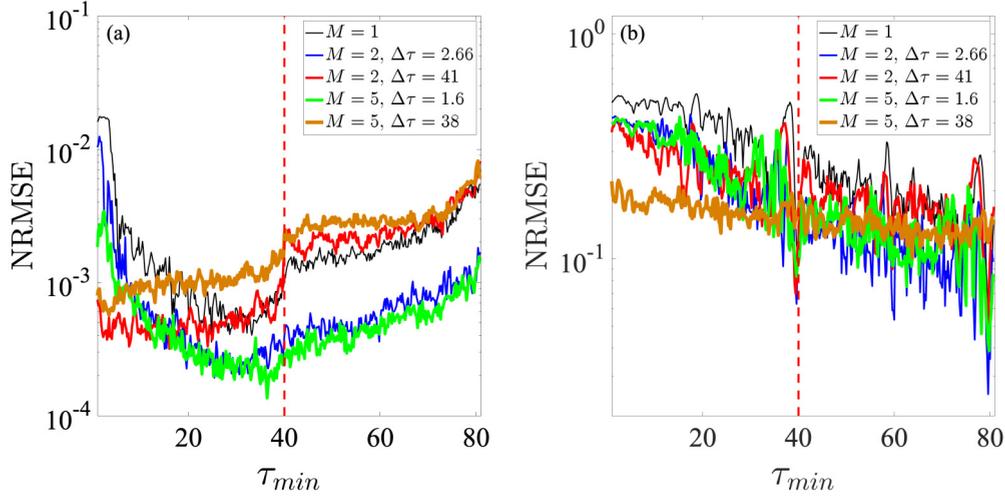


FIG. 7. MDRC can improve prediction and cross-prediction compared to single delay RC. (a) One-step ahead time series prediction error of the Lorenz x component as a function of τ_{min} for 1, 2, and 5 time delay RC. (b) Same as in (a) but for one-step ahead cross-prediction of the Lorenz z component with the input being the Lorenz x component. The vertical red dashed lines indicate the value of the cycle time $T = 40$. Parameters are in Table I and $\theta = 0.2$.

a uniform distribution in the interval $[0,0.5]$. The equation for the Narma10 task is

$$o_{n+1} = 0.3o_n + 0.05o_n \left(\sum_{i=0}^9 o_{n-i} \right) + 1.5u_{n-9}u_n + 0.1. \quad (12)$$

The above equation illustrates the dependency of target output values on past target output data. As noted earlier, a larger spacing between delays is observed to enhance prediction accuracy for correlated inputs, such as those in the MG model. To identify optimal time delays and spacing, we plotted the prediction error for the Narma10 task in Fig. 11. We considered minimum time delays ranging from 36 to 44 and $\Delta\tau$ values ranging from 76 to 84. The most accurate forecasting was achieved when $\Delta\tau = 2T \pm \theta$ and τ_{min} was equal to or

close to T . A significant increase in forecasting error was observed at $\Delta\tau = 2T$ regardless of the value of τ_{min} . This error increase occurs periodically when $\Delta\tau$ is an integer multiple of the clock cycle, which corresponds to the resonance case for the multidelay system [Fig. 3(b)].

To evaluate the performance of the multidelay system, we compare its MC with that of the single delay case in Fig. 12. To analyze the MC, the target is set equal to the i th input in the past [$o(k) = u(k - i)$, where u is the input] and the correlation between the target output and the recalled input is numerically computed:

$$m(i) = \text{corr}[o_i(k), \hat{o}_i(k)]. \quad (13)$$

Results indicate that the MDRC can recall up to 20 steps with a correlation between the recalled value and the target close

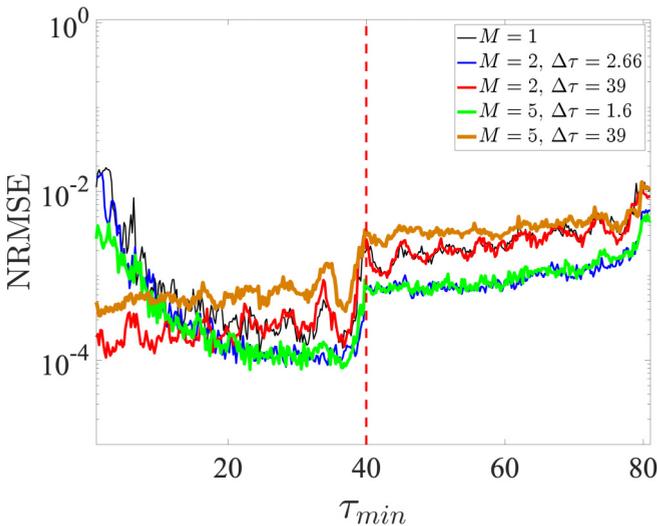


FIG. 8. One-step ahead time series prediction of the Lorenz z component when the input is the Lorenz z component for different τ_{min} . The parameters used are in Table I.

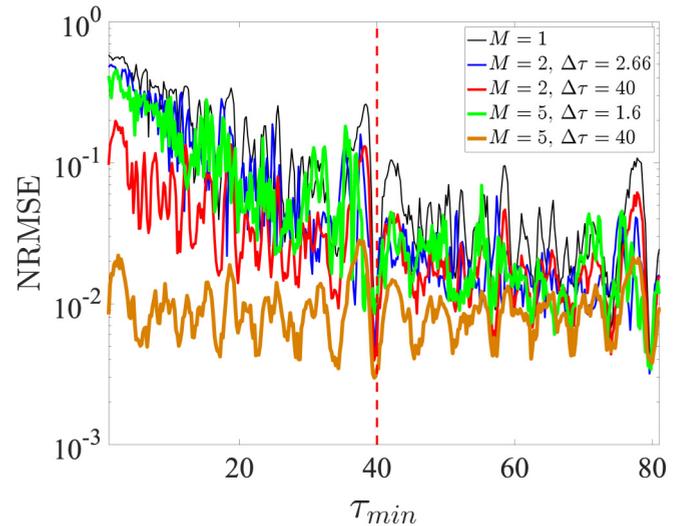


FIG. 9. The normalized root mean square (NRMSE) for a 34-step ahead prediction of the Mackey-Glass time series. The spacing between delays, denoted as $\Delta\tau$, is given in the inset.

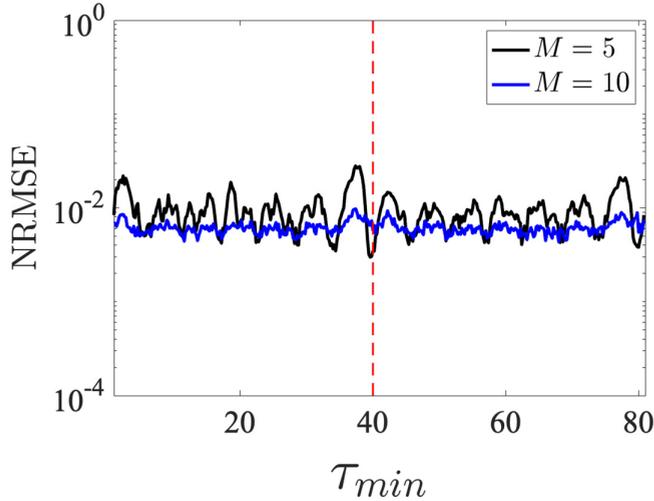


FIG. 10. A comparison of prediction errors in systems with five delays (black) and ten delays (blue) is depicted for the Mackey-Glass task.

to or equal to 1, while the single delay RC can only recall up to 11 steps before. The total MC can be determined by calculating the area under the graph, which is clearly greater for the MDRC. Overall, the results demonstrate that the MDRC outperforms the single delay RC in terms of MC.

We further compared the performance of the RCs in the $\gamma - \beta$ parameter space in Fig. 13. We considered three cases: single delay ($\tau_{\min} = 68.4$), five delays with short spacing between them ($\Delta\tau = 1.6$ and $\tau_{\min} = 68.4$), and five delays with large spacing between them ($\Delta\tau = 79.8$ and $\tau_{\min} = 40$). The values of τ for the single delay case, and τ_{\min} for the five delays case with short spacing, are chosen to minimize the error. However, it remains unclear why these specific delay values yield the lowest errors for both the single delay and multiple delay cases with short spacing. Like in the case of the MG input, we observe that the prediction for multiple delays with large spacing outperformed the other two cases.

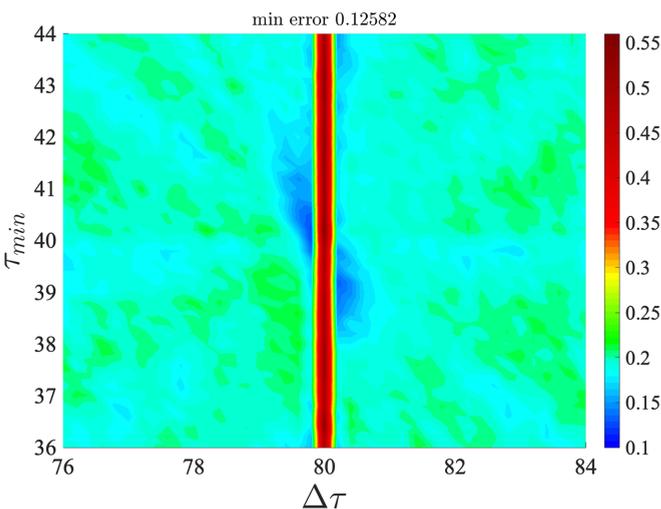


FIG. 11. Prediction error of a RC system with five delays, as a function of the parameters τ_{\min} and $\Delta\tau$, where the EOO parameters are $\beta = 1.35$ and $\gamma = 0.5$.

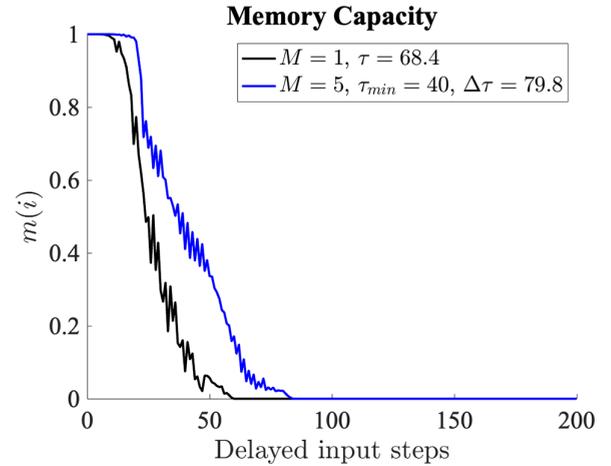


FIG. 12. Memory capacity of a RC system for the Narma10 task, comparing single delay (black) and five delays (blue) schemes. For the single delay case, parameters were set to $\beta = 1.5$, $\gamma = 0.25$, and $\tau_{\min} = 68.4$. For the five delays case, parameters were set to $\beta = 1.35$, $\gamma = 0.5$, $\tau_{\min} = 40$, and $\Delta\tau = 79.8$. These parameters were chosen because they provide the lowest error on the respective Narma10 prediction tasks.

The system maintains a higher linear MC in this scenario, resulting in higher prediction accuracy. The use of multiple delays with large spacing necessitates higher input scaling compared to the other cases, indicating the need for incorporating higher nonlinearity. Furthermore, employing multiple delays with large spacing offers the advantage of a broader range of parameter choices in the $\gamma - \beta$ space.

IV. DISCUSSION

The practical application of multidelay architectures in real-world systems has been demonstrated through various studies. A particularly notable example is observed in semiconductor lasers with optical feedback, where fiber random gratings are utilized as an alternative to traditional single reflective devices, such as mirrors [25]. Additionally, the EOO model with double feedback loops has been examined in Ref. [26]. In that study, it was found that incorporating double feedback loops significantly increases the complexity of system dynamics, a change primarily attributed to the introduction of an additional delay. However, it is essential to note that the nonlinear feedback function used in this referenced study is different from the one we employ in our research. Moreover, our model, in contrast to the one in Ref. [26], includes a filter in its equations. While we acknowledge the unique challenges and limitations associated with the physical realization of multidelay architectures, our research is mainly focused on a numerical analysis. The primary aim of our paper is to elucidate the impact of multiple delays on the performance of RC systems, thereby laying a foundation for future practical applications and advancements in this field.

While time delay RCs have demonstrated success in predicting chaotic time series through numerical and experimental approaches, identifying optimal time delays is critical for tackling more complex objectives, such as multistep

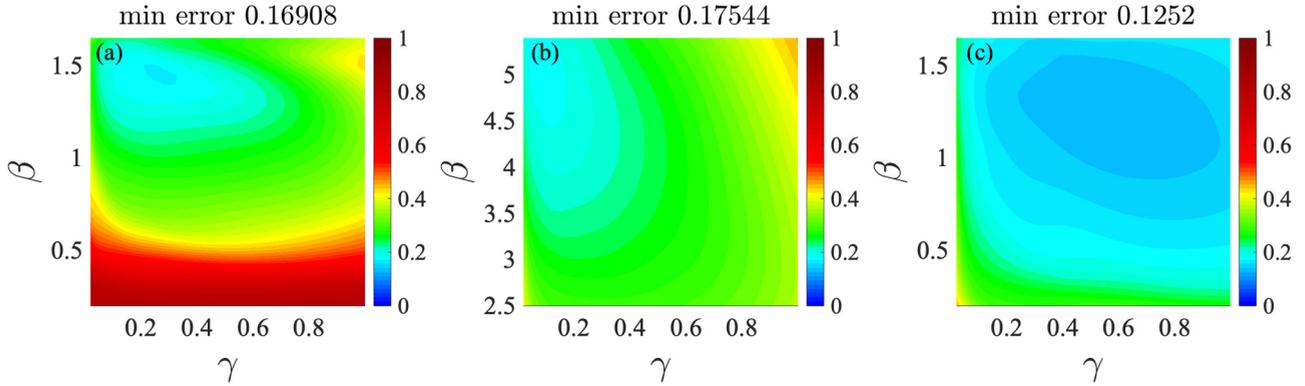


FIG. 13. The Narma10 prediction error of a RC system for different reservoir configurations. (a) Prediction error for a single delay case ($M = 1$) and $\tau = \tau_{\min} = 68.4$. (b) Prediction error for a five delay case ($M = 5$) with $\Delta\tau = 1.6$, and $\tau_{\min} = 68.4$. (c) Prediction error for five delays ($M = 5$), $\Delta\tau = 79.8$, and $\tau_{\min} = 40$. In each panel, the performance is shown for different input scaling and feedback strength values.

ahead prediction or prediction of individual signals within a mixed input signal. Further research is needed to explore ways of improving the performance of RC for more complicated prediction challenges. In particular, our results show that a thorough understanding of the input data properties is essential to successfully implement a multidelay RC for time series prediction.

Predicting multiple steps ahead for high-order complex signals like the Lorenz signal can be challenging. When predicting one step ahead, selecting a time delay equal to the clock cycle of the RC often leads to poor outcomes. However, the optimal time delay may vary depending on the correlation between the input and output. For example, in the case of cross prediction where the input is Lorenz x and the target is Lorenz z , we found that the resonance shifts to a smaller time delay value, and significant improvement is observed at $\tau = T$, as shown in Fig. 7. Previous studies have shown that using a delayed input method can improve RC performance for cross-prediction, and the lowest error can be achieved by delaying the input by multiple steps due to the correlation between the x component and the z component [24].

In the context of the MDRC analyzed in our paper, it is more crucial to choose delays with spacing unequal to the clock cycle (see Fig. 11). For the Narma10 task, we observed that the error decreased significantly around the resonance case. However, we should note that this might not be true for multiple step-ahead prediction.

In general, our results reveal that the RC prediction performance is expected to improve as the number of delays increases. This is often true, especially when the minimum time delay approaches the resonance condition, which, in the single delay case, occurs when the time delay equals the clock cycle. However, there may be values of the single time delay for which the RC exhibits performance comparable to multidelay systems. In the context of multidelay cases, the resonance condition is most likely observed when the spacing between delays is equal to the clock cycle. Overall, it is likely that a delay configuration can be found for enhanced MDRC performance which presents a broader parameter space compared to the single delay case.

We should note that the relationship between parameters and time delays does not always adhere to a straightforward, logical pattern and can be challenging to pinpoint precisely. This complexity might explain why the observed phenomenon is absent in some cases. This increased parameter space enhances the chances of achieving optimal reservoir performance, depending on the specific properties of the task.

While choosing the proper time delay is important, it is not the only factor that can enhance system performance. One potential approach to enhancing MC involves adding layers to the RC architecture. By considering the same number of virtual nodes for each layer, the total number of nodes increases, which in turn results in a reduction in prediction error [15]. Furthermore, it has been shown that feeding the input into multiple parallel lasers and then collecting the states of the virtual nodes from all lasers can further enhance the RC's performance [9]. Initial investigations along these lines reveals that the same benefits of these architecture changes can occur in the context of the MDRCs investigated here, a direction to be further explored in the future.

Finally, our results only apply to the case where delays are picked with a uniform spacing from one another. Small deviations from this scheme lead to similar results (not shown), but it is clear that it would be of interest to investigate whether a MDRC with random delays leads to better results. There are many parameters underlying the operation of these RCs. While we have presented dominant features of their performance, a full description of the dependence of the error on the type and parameters of the node dynamics as well as the type and parameters of the external input is beyond the scope of the present paper, and may itself benefit from machine learning techniques for parameter tuning.

ACKNOWLEDGMENT

The authors acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support in funding this research.

- [1] L. Maguire, B. Roche, T. McGinnity, and L. McDaid, Predicting a chaotic time series using a fuzzy neural network, *Inf. Sci.* **112**, 125 (1998).
- [2] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke, Prediction in chaotic nonlinear systems: Methods for time series with broadband fourier spectra, *Phys. Rev. A* **41**, 1782 (1990).
- [3] Gong Xiaofeng and C. H. Lai, Improvement of the local prediction of chaotic time series, *Phys. Rev. E* **60**, 5463 (1999).
- [4] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [5] S. Krishnagopal, M. Girvan, E. Ott, and B. R. Hunt, Separation of chaotic signals by reservoir computing, *Chaos* **30**, 023123 (2020).
- [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] J.-Y. Tang, B.-D. Lin, J. Yu, X. He, and C. Wang, Parallel time-delay reservoir computing with quantum dot lasers, *IEEE J. Quantum Electron.* **58**, 8100109 (2022).
- [8] 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 Trans. Neural Networks Learn. Syst.* **26**, 388 (2015).
- [9] C. Sugano, K. Kanno, and A. Uchida, Reservoir computing using multiple lasers with feedback on a photonic integrated circuit, *IEEE J. Sel. Top. Quantum Electron.* **26**, 1500409 (2020).
- [10] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, *Sci. Rep.* **2**, 287 (2012).
- [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* **7**, 011015 (2017).
- [12] D. Wenkack Liedji, J. H. Talla Mbé, and G. Kenne, Classification of hyperchaotic, chaotic, and regular signals using single nonlinear node delay-based reservoir computers, *Chaos* **32**, 123126 (2022).
- [13] K. Kanno, A. A. Haya, and A. Uchida, Reservoir computing based on an external-cavity semiconductor laser with optical feedback modulation, *Opt. Express* **30**, 34218 (2022).
- [14] B. Penkovsky, X. Porte, M. Jacquot, L. Larger, and D. Brunner, Coupled nonlinear delay systems as deep convolutional neural networks, *Phys. Rev. Lett.* **123**, 054101 (2019).
- [15] M. Goldmann, F. Köster, K. Lüdge, and S. Yanchuk, Deep time-delay reservoir computing: Dynamics and memory capacity, *Chaos* **30**, 093124 (2020).
- [16] F. Stelzer, A. Röhm, R. Vicente, I. Fischer, and S. Yanchuk, Deep neural networks using a single neuron: Folded-in-time architecture using feedback-modulated delay loops, *Nat. Commun.* **12**, 5164 (2021).
- [17] T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, Role of delay-times in delay-based photonic reservoir computing, *Opt. Mater. Express* **12**, 1214 (2022).
- [18] F. Stelzer, A. Röhm, K. Lüdge, and S. Yanchuk, Performance boost of time-delay reservoir computing by non-resonant clock cycle, *Neural Networks* **124**, 158 (2020).
- [19] S. K. Tavakoli and A. Longtin, Multidelay complexity collapse, *Phys. Rev. Res.* **2**, 033485 (2020).
- [20] S. Kamyar Tavakoli and A. Longtin, Dynamical invariants and inverse period-doubling cascades in multidelay systems, *Chaos* **31**, 103129 (2021).
- [21] J. Nakayama, K. Kanno, and A. Uchida, Laser dynamical reservoir computing with consistency: An approach of a chaos mask signal, *Opt. Express* **24**, 8679 (2016).
- [22] J. Jin, N. Jiang, Y. Zhang, W. Feng, A. Zhao, S. Liu, J. Peng, K. Qiu, and Q. Zhang, Adaptive time-delayed photonic reservoir computing based on Kalman-filter training, *Opt. Express* **30**, 13647 (2022).
- [23] F. Köster, S. Yanchuk, and K. Lüdge, Insight into delay based reservoir computing via eigenvalue analysis, *J. Phys. Photonics* **3**, 024011 (2021).
- [24] L. Jaurigue, E. Robertson, J. Wolters, and K. Lüdge, Reservoir computing with delayed input for fast and easy optimisation, *Entropy* **23**, 1560 (2021).
- [25] Y. Xu, M. Zhang, L. Zhang, P. Lu, S. Mihailov, and X. Bao, Time-delay signature suppression in a chaotic semiconductor laser by fiber random grating induced random distributed feedback, *Opt. Lett.* **42**, 4107 (2017).
- [26] M. W. Lee, L. Larger, V. Udaltsov, É. Genin, and J.-P. Goedgebuer, Demonstration of a chaos generator with two time delays, *Opt. Lett.* **29**, 325 (2004).