



HAL
open science

Minimum complexity integrated photonic architecture for delay-based reservoir computing

Mohab Abdalla, Clément Zrounba, Raphael Cardoso, Paul Jimenez, Guanghui Ren, Andreas Boes, Arnan Mitchell, Alberto Bosio, Ian O'Connor, Fabio Pavanello

► **To cite this version:**

Mohab Abdalla, Clément Zrounba, Raphael Cardoso, Paul Jimenez, Guanghui Ren, et al.. Minimum complexity integrated photonic architecture for delay-based reservoir computing. *Optics Express*, 2023, 31 (7), pp.11610. 10.1364/OE.484052. hal-04047354

HAL Id: hal-04047354






<https://hal.science/hal-04047354v1>

Submitted on 27 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Minimum complexity integrated photonic architecture for delay-based reservoir computing

MOHAB ABDALLA,^{1,2,*}  CLÉMENT ZROUNBA,¹ RAPHAEL CARDOSO,¹ PAUL JIMENEZ,¹ GUANGHUI REN,^{1,2}  ANDREAS BOES,^{2,3,4}  ARNAN MITCHELL,²  ALBERTO BOSIO,¹ IAN O'CONNOR,¹ AND FABIO PAVANELLO¹ 

¹Univ. Lyon, Ecole Centrale de Lyon, INSA Lyon, UCB Lyon, CPE Lyon, CNRS, Lyon Institute of Nanotechnology, UMR5270, Ecully, 69130, France

²Integrated Photonics and Applications Centre (InPAC), School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

³Institute for Photonics and Advanced Sensing (IPAS), University of Adelaide, Adelaide, SA 5005, Australia

⁴School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia

*mohab.abdalla@ec-lyon.fr

Abstract: Reservoir computing is an analog bio-inspired computation scheme for efficiently processing time-dependent signals, the photonic implementations of which promise a combination of massive parallel information processing, low power consumption, and high-speed operation. However, most of these implementations, especially for the case of time-delay reservoir computing, require extensive multi-dimensional parameter optimization to find the optimal combination of parameters for a given task. We propose a novel, largely passive integrated photonic TDR scheme based on an asymmetric Mach-Zehnder interferometer in a self-feedback configuration, where the nonlinearity is provided by the photodetector, and with only one tunable parameter in the form of a phase shifting element that, as a result of our configuration, allows also to tune the feedback strength, consequently tuning the memory capacity in a lossless manner. Through numerical simulations, we show that the proposed scheme achieves good performance -when compared to other integrated photonic architectures- on the temporal bitwise XOR task and various time series prediction tasks, while greatly reducing hardware and operational complexity.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Much interest is currently directed towards neuro-inspired computing paradigms, which are essentially machine learning frameworks for processing information in an intertwined, brain-inspired manner, not limited by the transfer of information from memory to processor, commonly known as the Von-Neumann bottleneck. Reservoir computing (RC) is one such type of analog computing which has garnered widespread interest since it was introduced (independently) by Jaeger as "echo state networks" [1], and Maass et al. [2] as "liquid state machines", offering a simplified model which is easier to train when compared to other recurrent neural network (RNN) approaches.

In general, RC schemes consist of 3 layers: an input layer where data is injected, a reservoir layer where the input signal drives the multiple dynamical nodes, and an output layer where the responses of the nodes are captured, linearly combined, and trained for the desired task. RC is essentially a simplified RNN where only the weights of the output layer are trained, and the input and internal weights are set and fixed to values that depend on the desired dynamical regime. The simplified training is due to the projection of the input data onto a higher dimensional state-space by a nonlinear dynamical system; thereby making it easier to find planes that can linearly separate

the different classes of data and allowing for simple regression techniques to solve complex nonlinear tasks such as chaotic time series prediction. Time-delay reservoir computing (TDRC), first introduced in [3], is a footprint-friendly scheme for hardware implementations, requiring only a single dynamical node connected to itself with a delay line. Through sampling this node N times in the span of one input clock cycle, the time-multiplexed responses can be viewed as the individual responses of N neurons. We refer the reader to [4] for a concise overview of RC principles. Consequently, RC has thus far enjoyed a multitude of hardware demonstrations across many technologies and platforms [5], especially in photonics using bulk optics [6–10], and, more recently, on photonic integrated circuits (PICs) [11–14]. For the latter, this is in large part due to improved performance when compared to electronic approaches in terms of power consumption, speed, footprint, and cost [11]. One particular strength of using RC in photonics is the fact that the input and internal weights need not be tuned. From a hardware point of view, this means that the RC framework is robust to fabrication variations. With this in mind, however, most coherent (single wavelength) nanophotonic systems suffer from sensitivity to environmental factors such as temperature fluctuations, limiting the RC operation time and making it difficult to have one set of weights that are reusable. In general, the parameters are optimized every time the photonic RC is utilized due to ambient fluctuations, especially in the cases of all-optical feedback. To the authors' knowledge, this remains an area to be explored with only a few recent examples in the literature proposing techniques to solve this issue, such as training for a given range of wavelengths corresponding to the range of thermal fluctuations in a controlled setting [15], or using other machine learning techniques such as transfer learning [16]. Thus, while the cost of optimizing the weights themselves is minimal, it is paid for by the need for optimizing the system parameters, often requiring the scanning of a multi-dimensional parameter space. For example, the proposed VCSEL scheme in [17] and the microring-based scheme in [14] both require 4 parameters to be optimized. Furthermore, while it is useful for some parameters such as the bitrate or input power to be tuned and optimized for the purpose of finding the global optimum for each specific task, they may not be readily available degrees of freedom for general-purpose RC within an applications setting. This may be of interest for designing reservoirs for a specific target application, where those parameters would be more or less fixed. On the other hand, for more general-purpose RC applications, the search for local optima within a more constrained parameter space gives a better idea of the usability of the design, while giving a fairly accurate picture of the information processing capabilities of the system.

Considering the above, we propose a novel photonic architecture based on an asymmetric Mach-Zehnder interferometer (MZI) for TDRC with only one tunable parameter: a phase shifting element, and a nonlinearity provided by the photodetector, as its output intensity is proportional to the square of the electric field which describes the node states. We show that such a minimum complexity approach (i.e. using a minimal number of simple hardware components and control parameters) is sufficient for obtaining good performance on the various tasks investigated. Our approach enables GSa/s processing speeds which are only limited by the photodetector electronics, and we consider the Lithium-Niobate-on-Insulator (LNOI) platform [18–20] to leverage the low waveguide losses that enable an on-chip feedback loop, in addition to high-speed on-chip modulation.

2. Reservoir architecture and operation principle

The integrated photonic reservoir is based on an asymmetric MZI used in a feedback configuration by means of a delay line. The asymmetric MZI is based on two 3-dB directional couplers and different arm lengths (3.0 mm and 1.5 mm), as shown in Fig. 1. The top ports of the MZI are connected to each other by a spiral waveguide of length 4.55 cm, which introduces delay and thus short-term memory to the system. A phase shifting element on the bottom MZI arm controls both the feedback phase and feedback strength in this configuration, thereby essentially tuning the

memory (without coupling optical power out of the system in the process as with using optical attenuators), thanks to the coupling modulation scheme [21].

Considering an input optical field $E_{in}(t) = A_{in} \exp(i\omega t)$ with amplitude A_{in} and $\omega = 2\pi c/\lambda_0$, where λ_0 is the source wavelength, it enters the first coupler at $t = 0$, and considering the 3-dB couplers as point couplers, we can describe the fields' evolution in time everywhere in the system using the scattering matrix approach (see Fig. 1):

$$\begin{pmatrix} E_1(t) \\ E_2(t) \end{pmatrix} = \sqrt{\alpha_c} \begin{pmatrix} -i\kappa & r \\ r & -i\kappa \end{pmatrix} \begin{pmatrix} E_{in}(t) \\ E_{fb}(t) \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} E_{out1}(t) \\ E_{out2}(t) \end{pmatrix} = \sqrt{\alpha_c} \begin{pmatrix} \sqrt{\alpha_1} r e^{-i\beta L_1} & -\sqrt{\alpha_2} i \kappa e^{-i(\beta L_2 + \Phi)} \\ -\sqrt{\alpha_1} i \kappa e^{-i\beta L_1} & \sqrt{\alpha_2} r e^{-i(\beta L_2 + \Phi)} \end{pmatrix} \begin{pmatrix} E_1(t - \tau_1) \\ E_2(t - \tau_2) \end{pmatrix} \quad (2)$$

$$E_{fb}(t) = \sqrt{\alpha_{fb}} E_{out1}(t - \tau_{fb}) \exp(-i\beta L_{fb}) \quad (3)$$

where α_c is the fraction of optical power exiting from the coupler (considered equal for both ports), $\alpha_{1,2,fb} = 10^{-AL/10}$ are the overall fractions of power after waveguide propagation for a loss factor A [dB/m] and the respective waveguide lengths $L_{1,2,fb}$ [m], which are the lengths of the upper MZI arm, bottom MZI arm, and the feedback loop, respectively, κ and r are the cross and through field coupling coefficients, respectively, $\beta = 2\pi n_{eff}/\lambda_0$ [m⁻¹] is the propagation constant of the guided mode with effective refractive index n_{eff} , $\tau_{1,2,fb}$ [s] are the delay times of the upper MZI arm, bottom MZI arm, and delay line, respectively, and Φ [rad] is the applied phase shift on the bottom arm. The $-i$ in front of κ results from the $\pi/2$ phase shift encountered when the field is crossing in the coupler.

The choice of the spiral waveguide length L_{fb} is important for enabling the desired maximum memory of the system. Normally, the feedback length would be constrained by the desired operation speed through one of two techniques: (i) Matching the sample hold duration with the delay time of the feedback, (ii) using a slightly longer sample hold duration than the delay time of the feedback. The first case is useful when the temporal distance between the 'virtual' nodes is smaller than the timescale of the nonlinearity such as the electronic implementation in [3], which creates a forward coupling of these nodes in addition to remembering their previous states by equating the delay time to the bit period. On the other hand, when the timescale of the nonlinearity is too fast such that it can be considered instantaneous in the system [6,7], using (i) will result in the nodes remembering only their own previous states (provided that the system does not reach steady state during one bit period) and so they become completely disconnected from each other. This can be alleviated by mismatching the sample hold duration with respect to the delay line (ii), which allows the nodes to remember the previous state of their neighboring node instead of their own (for a de-synchronization time of one node distance). However, as discussed in [22], it is not necessary for the delay time to be constrained by these two regimes for a variety of applications, especially those that do not require a large short-term memory. These constraints are due to considering the network equivalents of the TDRC scheme.

In fact, the memory capacity, discussed later in section 3.1, is significantly affected by the ratio of delay time to the sample hold duration. In [23] it has been shown that a resonance between the delay time and the sample hold duration can even be detrimental to the memory of the system for some tasks, specifically when they are integer multiples of each other. However, the specific components of the total metric are affected differently and thus the detriment in performance is task-dependent. Considering the above, we leverage these insights to design the delay line of our system for good performance and to reduce footprint. At an input sample rate $B = 1$ Gbit/s (i.e. sample hold time $\tau_B = 1$ ns), the equivalent length is $L_T = c\tau_B/n_g \approx 13.24$ cm. While the study in [22] uses the opto-electronic model as their basis, their findings show the impact of the ratio of

the delay time to the input clock cycle on the memory capacity, which are also applicable in our case, the only difference being that the feedback phase is also considered in our case of optical feedback. Therefore we have done a few sweeps around $L_T/L_{fb} = 3$ while avoiding the resonant condition of having the exact integer value. Our design choice of $L_T/L_{fb} = 2.91$ thus reflects a region where indeed the memory capacity has been reduced below its maximum, but is still enough for performing the nonlinear tasks presented here, while saving around $3\times$ on footprint. The second point to consider in this architecture is the choice of an asymmetric MZI as opposed to a symmetric one, where $\tau_1 = \tau_2$. As shown in Fig. 2 (b), the dynamics of the system are more interesting than that of the symmetric case in Fig. 2 (a) due to the different number of delays introduced in the system. Using an asymmetric MZI, there is one additional delay which enriches the dynamics further and this temporal mismatch enables the system to provide a more complex spectro-temporal response. These rich dynamics correspond to a more interesting mapping of input to output, thereby allowing the reservoir to solve highly nonlinear tasks more effectively.

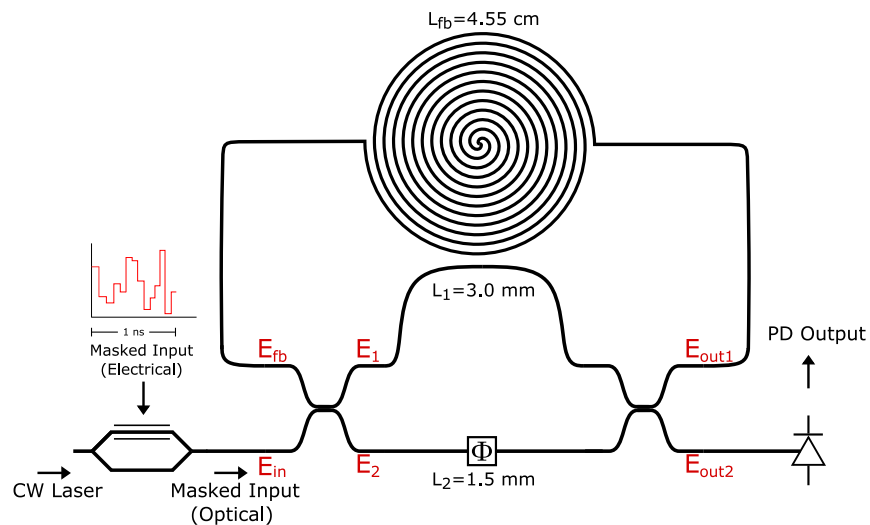


Fig. 1. Proposed architecture: a CW laser is modulated by the electrical input using a Mach-Zender modulator, the reservoir layer consists of the asymmetric MZI connected onto itself with a spiral waveguide and a photodetector, which also performs the readout.

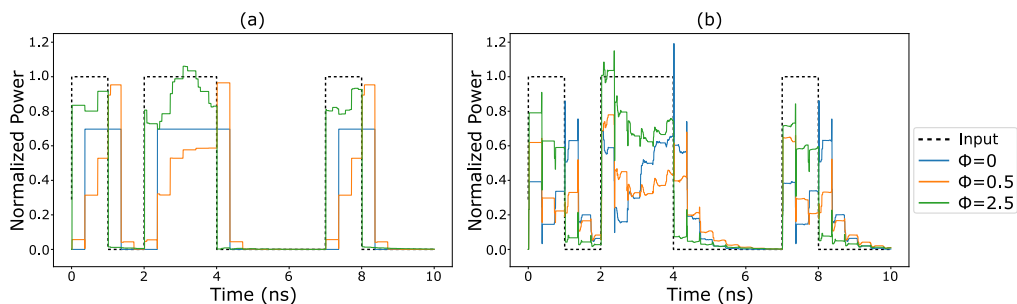


Fig. 2. Simulated dynamical response of the system subjected to different applied phase shifts Φ [rad] to an input bitstream with (a) symmetric MZI, (b) asymmetric MZI

3. Methodology

The inputs of the various tasks are fed one at a time to the simulated system and its response is recorded and then trained on the various tasks using linear regression. All the simulations were carried out with the model presented in 2 and with an open source S-matrix based photonic circuit solver [24] to validate the reliability of our model. In this study, we consider the operation of the phase shifting element up to V_π and divide the interval into 101 points, constituting our applied phase values, to get an accurate view of the trend between the reservoir's predicted results and the applied phase. Masks were applied on the inputs for all the benchmark tasks presented here, with values drawn pseudo-randomly from a uniform distribution on the interval (0,1] corresponding to the number of 'virtual' nodes in the reservoir N_v (determined here by the photodetector's bandwidth). In this way, each mask value can be viewed as the input weight connecting the input layer to its corresponding N^{th} node. For all the tasks, we passed the photonic circuit's response (E_{out2} in Eq. (2)) through photodetectors of 5 GHz, 10 GHz, 20 GHz, and 25 GHz bandwidth (corresponding to 5, 10, 20, and 25 nodes, respectively) to determine the required size of the reservoir for the various tasks, and choose the photodetector that gives the best performance considering all the tasks presented here. Furthermore, the trained output layer was tested on different inputs and their results are presented here. To ensure accurate circuit-level simulations, the simulation timestep Δt was chosen to be 100 times smaller than the span of one input clock cycle, which is also small enough to take into account the short delays of the MZI arms. The simulation parameters are shown in Table 1 (a). All the simulated photodetectors were bandwidth-dependent (incorporating a 4th order butterworth filter) and exhibited Gaussian noise with variance corresponding to the different contributions to noise, as listed in Table 1 (b). The use of standard ordinary least squares regression proved sufficient for the model to generalize and predict accurately the unseen test data, because we considered almost ideal inputs and also injected noise to the training data through the photodetector's response (a form of regularization in itself). For experimental verification, however, Tikhonov regularization or Bayesian regression may need to be employed, where nonidealities and outliers may result in an ill-posed problem when attempting to calculate the matrix pseudoinverse [25].

3.1. Linear memory capacity

The linear memory capacity is one of the fundamental tasks for RC, which aims to test the echo state property by training the reservoir to reconstruct a given input stream of values $\in [0, 1]$ drawn from an independent and identical distribution (i.i.d.) k inputs later. It was first introduced in [26] and is given by:

$$MC_k = \frac{\text{cov}(u(n-k), y_k(n))^2}{\text{var}(u(n))\text{var}(y_k(n))} = 1 - NMSE \quad (4)$$

$$NMSE = \frac{\langle \|y(n) - y_{exp}(n)\|^2 \rangle}{\langle \|y_{exp}(n) - \langle y_{exp}(n) \rangle\|^2 \rangle} \quad (5)$$

where $u(n)$ is the input at discrete timestep n , $y_{exp}(n)$ is the expected value, $NMSE$ is the normalized mean square error, and $MC_k \in [0, 1]$ is the memory capacity for a k bits shift. $MC_k = 1$ corresponds to a perfect recall of the input sequence after k input samples/bits, while $MC_k = 0$ corresponds to the complete absence of any information regarding the input sample/bit k steps into the past.

A sequence of 4000 samples was constructed from an i.i.d. stream. The target sequence is a k -bits delayed copy of the input, testing the reservoir's ability to faithfully reconstruct the input sequence after k input samples. The model was trained on the first 1000 samples and then tested on the remaining 3000 samples. The performance of specific components of the memory

Table 1. Simulated photonic circuit and photodetector parameters^a

Simulation Parameter	Value
B	1 GSa/s
Δt	10 ps
λ_0	1550 nm
n_{eff}	2.2111
n_g	2.2637
α_c	0.966
A	20 dB/m
P_{in}	10 mW
r	0.7
R_l	100 Ω
I_d	5 nA
NEP_{TIA}	24 pW/ \sqrt{Hz}

^aSimulated photonic circuit parameters: B is the input bit/sample rate, Δt is the simulation timestep, λ_0 is the operation wavelength, n_{eff} and n_g are the effective and group indices for the waveguide, α_c is the directional coupler gain, A is the waveguide propagation loss, and P_{in} is the input optical power. Simulated integrated photodetector parameters: responsivity r , load resistance R_l , dark current I_d , and NEP_{TIA} is the noise equivalent power of the transimpedance amplifier.

capacity (Eq. (4)) are investigated as it gives a better indication of the usability of the stored information, in contrast to just the amount of information stored given by the summation of all the components, given by $MC_{total} = \sum_{k=1}^N MC_k$, where N is the number of nodes. The relevance of this evaluation for tasks requiring specific memory has also been mentioned and taken into account in other works [14,22].

3.2. Temporal bitwise XOR

The temporal bitwise XOR task is a nontrivial, nonlinear memory-specific task commonly used for evaluating RC performance which was first introduced in [27]. For this task, a quasi-ideal bit stream of 4000 bits was generated, where the first 1000 bits were fed to the circuit for training and the rest were used for testing. The target bit streams were constructed by applying the XOR operation on the bit stream and a k time steps shifted version of it, yielding $x[n] \oplus x[n - k]$, where $x[n]$ is the current input bit (similar to the treatment of this task in [13]). The performance (up to $k = 4$) is evaluated with the bit error rate (BER) metric which is the number of wrongly predicted bits over the length of the total bit sequence.

3.3. Mackey-Glass

The Mackey-Glass sequence was first used as a RC benchmark in [28], and is generated from solving the following differential equation numerically using the 4th order Runge-Kutta method:

$$\frac{dy(t)}{dt} = \frac{ay(t - \tau)}{1 + y(t - \tau)^{10}} - by(t) \quad (6)$$

with the commonly used parameters $a = 0.2$, $b = 0.1$, $\tau = 17$, and an integration step of 0.1. The behavior resulting from these chosen parameters is fairly periodic and only slightly irregular in the sense of causing minor fluctuations for each repeated cycle. The training set was 5000

samples long and the test set consisted of another 3000 samples. The task is a one-step ahead prediction. The performance is then evaluated according to the NMSE between the target values and the predicted output.

3.4. Santa Fe

The Santa Fe dataset [29] comprises of data points collected experimentally from a far infrared laser operating in a chaotic regime. This dataset is fairly chaotic in the vicinity of a few data points and fairly cyclic in terms of long-term dynamical behavior. The stream of 4000 data points was divided into 2000 points used for training and the other 2000 for testing. This task is also a one-step ahead prediction. The performance on the test set is then evaluated by NMSE.

3.5. NARMA3

The nonlinear autoregressive moving average (NARMA) is a commonly used benchmark task for RC which mimics a randomly varying signal around a certain average value, similar to noise. It is often used in its 10th order form to test a reservoir with very large memory. Due to the smaller memory in our system, we test the performance on a 3rd order variant of this task, which would show how the system is solving a sufficiently nonlinear task, without imposing further memory requirements than the system is capable of. The discrete difference equation that produces the NARMA3 sequence is given by:

$$y[n] = 0.3y[n-1] + 0.05y[n-1] \sum_{i=1}^3 y[n-i] + 1.5u[n]u[n-3] + 0.1 \quad (7)$$

where the input sequence u is drawn from a uniform distribution [0,0.5]. The task is to predict $y[n]$ given $u[n]$. The performance on the test set is evaluated with NMSE.

3.6. Baseline: asymmetric MZI

To better understand the role of the delay line and its impact on the various tasks presented here, and as such their memory requirements, we proceed to compare the architecture presented in Fig. 1 with just the asymmetric MZI without the feedback spiral waveguide. To that end, we consider the same tasks mentioned above to evaluate the performance of the asymmetric MZI alone on solving them.

4. Results and discussion

4.1. Linear memory capacity

The results for different k time steps shifts in Fig. 3 show the variation of MC_k with respect to applied phase shift. For lower number of nodes, the effect of the phase shift is more pronounced on the memory as can be seen in Fig. 3(a),(b), with $k = 3$ and $k = 4$ improving significantly as the reservoir size increases (Fig. 3(c),(d)). We also show MC_{total} for each reservoir size in Fig. 4(a) and the optimal MC_k obtained for each k up to $k = 10$, as shown in Fig. 4. Beyond $N_v = 5$, the peak MC_k values for many k 's appear to be close to each other. This suggests that further exceeding the studied number of nodes (i.e $N_v = 25$) would not enhance the memory further as it is fundamentally limited by the length of the spiral waveguide with the given input bitrate. Our results for MC_{total} , shown in Fig. 4, are consistent with those presented in [22], as we achieve $MC_{total} \approx 5.5$ which is close to the result obtained in the same work for our chosen ratio of delay time and input clock cycle. Thus, our total memory capacity exceeds the PIC implementation in [30] which requires copies of the delayed input at the modulation stage to exceed its intrinsic $MC_{total} = 1$, and is lower than the one presented in [13] where additional post-processing that merges the responses to previous inputs and the current one increases the

number of virtual nodes and thus the linear memory capacity from an intrinsic $MC_{total} \approx 6$ to $MC_{total} \approx 8$. Furthermore, our MC_k results for 20 nodes are almost equivalent to the simulated result in [12] using 20 on-chip lasers from $k = 1$ to $k = 5$. However, it is lower for further k as our system was not designed for large memory for the purpose of the currently investigated tasks. This can be easily alleviated - thanks to the low losses of LNOI platform - by utilizing a longer spiral length and using the desynchronized regime explained in section 2, and/or possibly using similar pre/post-processing techniques as the ones described above in [13,30].

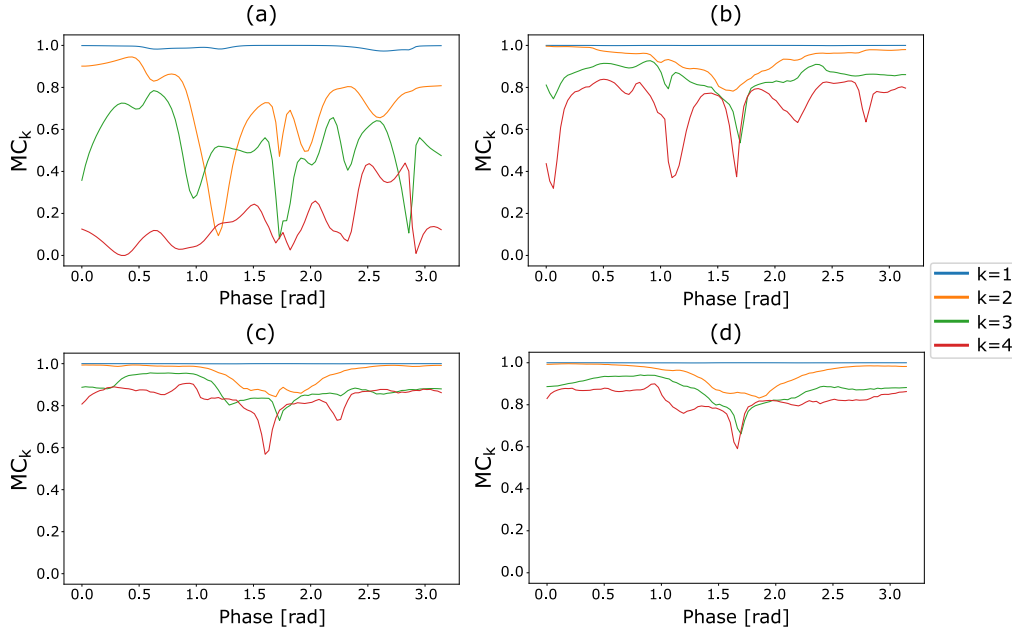


Fig. 3. Performance of the reservoir on solving the linear memory capacity task for different values of applied phase shift Φ for different reservoir sizes: (a) $N_V=5$ (b) $N_V=10$, (c) $N_V=20$, (d) $N_V=25$.

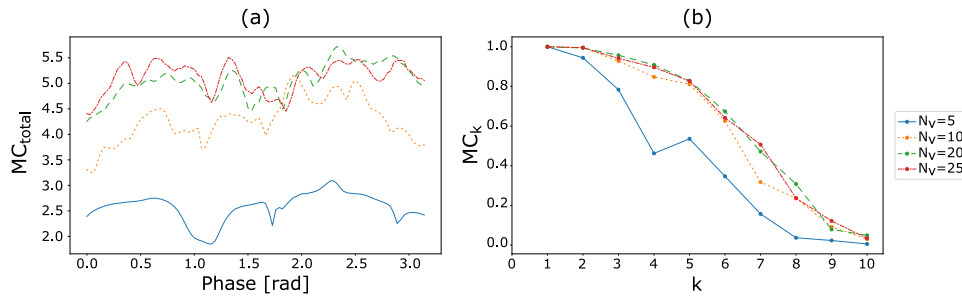


Fig. 4. MC_{total} (a) and peak obtained values of MC_k (b) for different reservoir sizes: $N_V=5$, $N_V=10$, $N_V=20$, and $N_V=25$.

4.2. Temporal bitwise XOR

For the XOR task, the test sequence used consisted of 3000 bits, limiting the resolution of the BER is limited to 0.33×10^{-3} . Thus, a BER below 10^{-3} is considered as acceptable, as shown in purple in Fig. 5, where results are shown for $k = 1$ to $k = 4$ for various reservoir sizes. Similar to

the memory capacity results, the performance on the XOR task mostly improves as the reservoir size scales. It is shown in Fig. 5(a) that it is possible to do the one bit XOR with 5 nodes, and possibly even less. It can be seen from Fig. 5(c) that the architecture can be used successfully for XOR-ing the current input bit with 3 bits into the past, for $N_v = 20$.

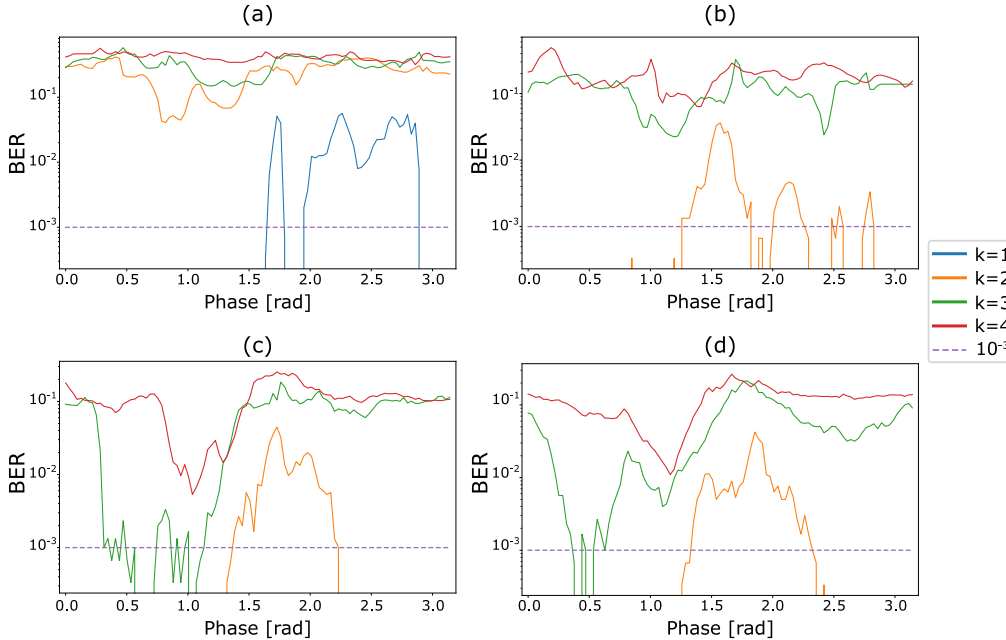


Fig. 5. Performance of the reservoir on solving the temporal bitwise XOR task for different values of applied phase shift Φ [rad] on the MZI arm for different reservoir sizes: (a) 5 Nodes, (b) 10 Nodes, (c) 20 Nodes, (d) 25 Nodes. Where the blue line ($k=1$) is not visible, it is due to BER = 0 everywhere on the plot.

4.3. Mackey-Glass

The results in Fig. 6(a) are for differently sized reservoirs under applied phase shift. One of the interesting features in the curve is that the performance is only minimally affected by the number of nodes N_v , presented here, which suggests that only a small memory is required for this task. We obtain NMSE = 0.0056 which is close to the optimum value obtained in [14] and even when compared to a bulk setup [31].

Another interesting point this result shows is minimal dependence on the varying phase shift, which prompts further investigation into which part of the architecture is responsible for the obtained NMSE performance. It was found that equivalent performance is obtainable by training the input data with linear regression, without going through the photonic circuit. This is discussed later in section 4.6.

4.4. Santa Fe

For the Santa Fe timeseries prediction, the results in Fig. 6(b) show a minimum NMSE of 0.038 using 25 nodes, which is close to the simulated result in the nonlinear microring approach in [14] (NMSE=0.038) and better than the experimental result in the multiple cavities approach based on a feed-forward photonic neural network in feedback reported in [32] (NMSE=0.06). It is also slightly better than the experimental result mentioned in [13] (NMSE=0.049), which was achieved by increasing the laser pump current and the semiconductor optical amplifier (SOA)

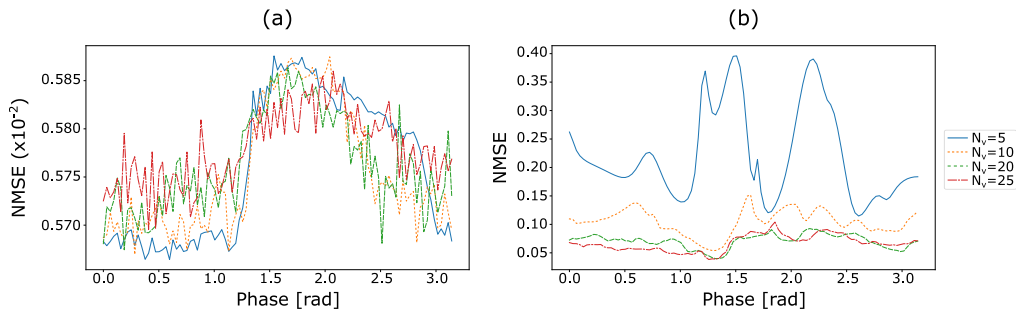


Fig. 6. Performance of the reservoir (NMSE) under applied phase shift on one-step ahead prediction time series tasks: (a) Mackey-Glass, and (b) Santa Fe

current, using 23 virtual nodes, albeit with additional postprocessing techniques. Furthermore, we also obtain better prediction results than the approach in [12] where they reported a minimum $NMSE \approx 0.01$ using 40 on-chip lasers with small external cavities of 10 mm.

4.5. NARMA3

For the NARMA3 task, beyond $N_v = 5$ the results show only a slight dependence on the number of nodes for all values of phase shift, and is especially the case around $\Phi = 0.5$ rad. However, it is much more strongly influenced by the phase shifter's effects of altering the memory from $\Phi = 1$ rad to $\Phi = 2.5$ rad. Considering the memory of our system, a low $NMSE = 0.096$ is obtained using as few as 10 Nodes.

4.6. Baseline: asymmetric MZI

Performing the same numerical investigations on the MZI alone without the feedback loop helps in understanding the delay's role further. For our operation speed of 1 GSa/s, it can be seen that tasks requiring a memory of one sample/bit into the past are achievable, which is not surprising since at some point the current sample interacts with the previous input sample due to the differences between the arm lengths, and consequently the asymmetric MZI alone is sufficient. Such tasks are the memory capacity and XOR tasks for $k = 1$, where $MC_1 \approx 1.0$ everywhere for all phase shifts and for all reservoir sizes N_v . The XOR operation for $k = 1$ is successful beyond a certain value of phase shift, due to destructive interference at this value of applied phase (Fig. 8(a), a similar behavior is seen in (b) as well). For tasks requiring deeper memories the MZI fails completely: $MC_{k>1} \approx 0$ for all N and phase shift, $BER \approx 0.5$ XOR for $k > 1$. For the Santa Fe task, the performance degrades considerably as shown in Fig. 8(b) with $NMSE \approx 0.34$ being the best value achieved. For the NARMA3 task (Fig. 8(c)), it fails completely with $NMSE \approx 0.7$ everywhere on the plot.

For the Mackey-Glass one-step ahead prediction task, we get equivalent performance ($NMSE = 0.00587$) with the MZI alone as shown in Fig. 8(d), and in fact it is also similar to the performance obtained when training the input data itself (masked and unmasked) using linear regression, where we also found no degradation in $NMSE$ for all N_v considered ($NMSE = 0.00583$). We find this result particularly interesting, since it shows that solving the one-step ahead task can be done with 5 trainable features and using a regression on the input data itself. Considering two and three steps ahead predictions on the same task, the full reservoir architecture only slightly outperforms ($NMSE_{k=2} = 0.0114$, $NMSE_{k=3} = 0.0170$) the almost equivalent result of training on both the input dataset directly, and passing it through just the MZI ($NMSE_{k=2} = 0.0119$, $NMSE_{k=3} = 0.0182$). According to these results, the one-step ahead Mackey-Glass task is

linearly separable and is not a significant challenge for the RC framework, unless much lower NMSE values (ex. $<10^{-3}$) are obtained.

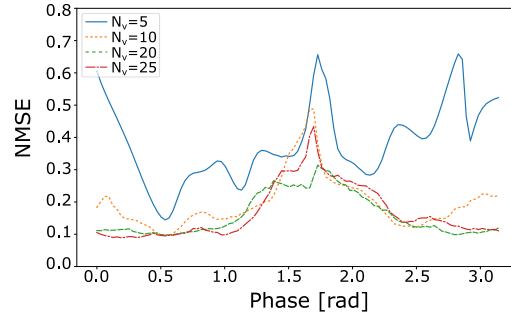


Fig. 7. Performance on the NARMA3 task for various N_v

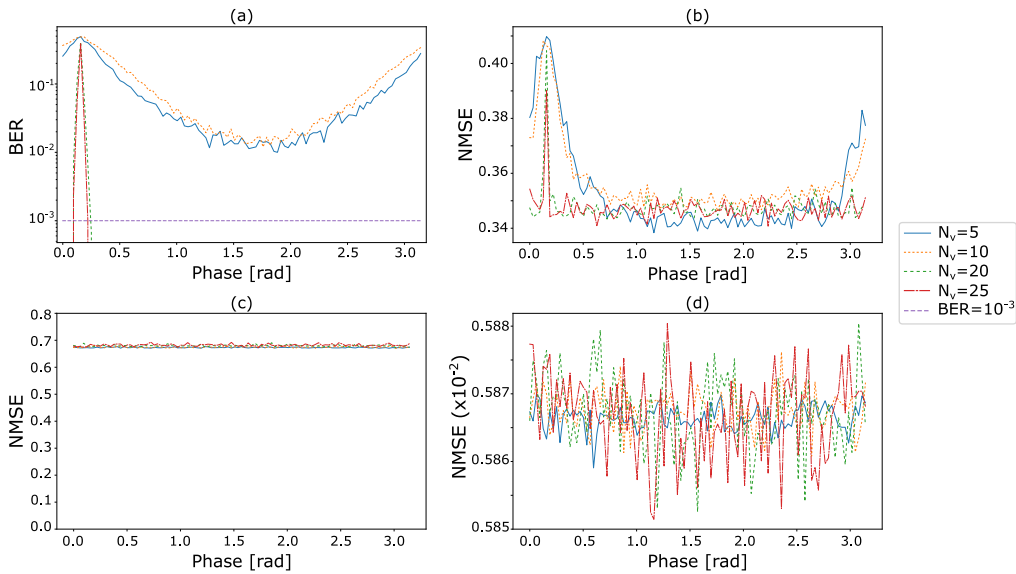


Fig. 8. Performance of the different tasks using only the asymmetric MZI under varying phase shift: (a) XOR for $k=1$, (b) Santa Fe, (c) NARMA3, (d) Mackey-Glass.

4.7. Further discussion

For all the tasks presented here, it can be seen that a photodetector with 20 GHz bandwidth -yielding 20 virtual nodes- is sufficient for obtaining the best performance on this architecture. Furthermore, the variation in prediction accuracy (under applied phase shift) for the several tasks presented are strongly related to the phases of the signals travelling into the spiral from the two MZI arms. To further explain this notion, we refer back to the memory capacity results in section 4.1. According to the value of Φ , interference occurs at the output and input couplers, where the incoming signal also participates. Due to the low losses, multiple round trips can occur within both paths, which can yield either constructive or destructive interference over one or multiple round trips. This directly influences the virtual nodes' connectivity matrix, and has a much stronger effect for lower number of nodes such as $N_v = 5$, as can be seen from the larger

variability in Fig. 3 (a). Increasing the number of nodes allows more information to survive after each roundtrip, which especially enhances tasks requiring deeper memories. The memory eventually saturates when there is no longer any representation of the information inside the system for further past inputs (Fig. 3(c) and (d)). Naturally, this behavior is also pronounced in other tasks (Fig. 6(b) and Fig. 7).

In addition, there are multiple advantages for using the proposed RC scheme, and we illustrate this by briefly discussing the other PIC implementations from the literature. First is the fully integrated low-loss delay line, which is made possible by considering low-loss platforms such as LNOI, which entails less power loss coupling into and out of the chip, similar to [13], and in contrast to [14] when using an external feedback. Second is high-speed operation, limited only by the photodetector bandwidth, whereas other architectures employing relatively slow nonlinearities (especially thermal nonlinearities in case of silicon-on-insulator) can significantly lower computation speeds [30]. Third is the multiple timescales approach we used, which has also been leveraged in [32], however our architecture reduces complexity in terms of number of phase shifters needed to be controlled while also obtaining better results on the Santa Fe task. Compared to [12] which uses up to 40 on-chip lasers, the memory MC_k of our system is close to the one they obtain using 20 on-chip lasers in the range of $k = 1$ to $k = 5$ (Fig. 4 (b)). The similarly passive architecture in [11], which is also relying only on the photodetector nonlinearity, is however limited in scalability by usage of physical nodes instead of virtual ones, and the need to change the ratio of interconnection delay and bit period for solving different tasks, and since the former is fixed this entails changing the bitrate of the input stream for tasks requiring different memories, such as the bitwise XOR with multiple bits in the past. However, in our case as has been shown in section 4.2, only a phase shifting element is required. In fact it is even possible to do the XOR for $k = 1$ to $k = 3$ at the same value of phase shift, thus requiring only a change in the applied output weights to perform the three different tasks. The architecture in [13], consisting of a distributed Bragg reflector laser and amplifiers, as well as integrated delay lines, achieved similar performance on the Santa Fe task after additional post-processing.

On the other hand, it is also important to consider that simulation setups and learning algorithms can differ between different investigations, for example sometimes ridge regression is employed instead of linear regression. It is therefore not straightforward to compare these different examples from the literature, which is why such a comparison is beyond the scope of this work. Instead, we aim to shed light on the fact that RC with matching performance to the above examples can be done on-chip with passive components, without the need for nonlinearities beyond that of the photodetector, with minimum active components (no amplifiers or multiple laser sources), and with only one tunable phase shifter as a tunable parameter. Using only one phase shifter that is relatively easy to control, as opposed to multiple parameters, can enable on-chip stabilization using optical feedback techniques [33], which could potentially allow photonic RC that is robust to ambient fluctuations, without the need to retrain constantly. Further adding to the system complexity may indeed boost the system performance beyond simpler architectures such as the one presented in this work. We believe this work could therefore serve as a baseline in terms of performance for the given system and hardware requirements, and that future works could enable performance improvements that warrant the use of higher complexity PIC RC schemes.

5. Conclusion

We have proposed an integrated photonic architecture for RC which leverages the low losses of the LNOI platform to enable a fully integrated delay line and with only one phase shifting parameter to tune the feedback phase and the feedback strength simultaneously. The delay line was designed to be compact enough while still delivering performance that is equivalent or slightly better than other PIC implementations for a comprehensive body of tasks. Further enhancement of the memory is possible by increasing the length of the spiral waveguide, at the cost of footprint. This

approach also provides more efficient utilization of power and the information stored inside the reservoir layer, when compared to other photonic implementations requiring an optical attenuator block in the feedback loop to tune the feedback strength, where light is simply coupled out of the system. We conclude that minimum complexity RC designs can also open the doors towards robust RC in ambient conditions by only requiring the stabilization of one parameter, thereby increasing the longevity of each training cycle and possibly allowing the deployment of photonic RC in real-world settings and applications. Lastly, we believe this work can also serve as a baseline to be compared against for more complex photonic RC systems, since there could indeed be room for performance improvement through using more interesting configurations which exhibit more system or hardware complexity. The exploration of nonlinearities in the LNOI platform would be a good choice, as its nonlinearities are on the timescale of the optical cycle.

Funding. H2020 Marie Skłodowska-Curie Actions (801512); International Associated Laboratory in Photonics between France and Australia (LIA ALPhFA); Agence Nationale de la Recherche (ANR-20-CE39-0004).

Acknowledgments. This work was supported by the ECLAUSion project which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 801512. We acknowledge the support of the International Associated Laboratory in Photonics between France and Australia (LIA ALPhFA). This work was also carried out within the framework of the PHASEPUF project supported by the French "Agence Nationale de la Recherche" under project number ANR-20-CE39-0004.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. H. Jaeger, "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* **148**, (2001).
2. W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comput.* **14**(11), 2531–2560 (2002).
3. 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**(1), 468 (2011).
4. Z. Konkoli, *Reservoir Computing* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2017), pp. 1–12.
5. G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks* **115**, 100–123 (2019).
6. Y. Paquot, F. Dupont, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Sci. Rep.* **2**(1), 287 (2012).
7. Q. Vinckier, F. Dupont, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica* **2**(5), 438–446 (2015).
8. 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**(1), 011015 (2017).
9. J. Vatin, D. Rontani, and M. Sciamanna, "Experimental reservoir computing using vcsel polarization dynamics," *Opt. Express* **27**(13), 18579–18584 (2019).
10. Y. Chen, L. Yi, J. Ke, Z. Yang, Y. Yang, L. Huang, Q. Zhuge, and W. Hu, "Reservoir computing system with double optoelectronic feedback loops," *Opt. Express* **27**(20), 27431 (2019).
11. K. Vandoorne, P. Mechet, T. V. Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nat. Commun.* **5**(1), 3541 (2014).
12. 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**(1), 1 (2019).
13. K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Opt. Express* **28**(3), 3086 (2020).
14. G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Opt. Express* **30**(1), 522 (2022).
15. E. Gooskens, F. Laporte, C. Ma, S. Sackesyn, J. Dambre, and P. Bienstman, "Wavelength dimension in waveguide-based photonic reservoir computing," *Opt. Express* **30**(9), 15634–15647 (2022).
16. I. Bauwens, K. Harkhoe, P. Bienstman, G. Verschaffelt, and G. V. der Sande, "Transfer learning for photonic delay-based reservoir computing to compensate parameter drift," *Nanophotonics* (2022).
17. X. X. Guo, S. Y. Xiang, Y. H. Zhang, L. Lin, A. J. Wen, and Y. Hao, "Four-channels reservoir computing based on polarization dynamics in mutually coupled vcsels system," *Opt. Express* **27**(16), 23293–23306 (2019).

18. A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (Inoi) for photonic integrated circuits," *Laser Photonics Rev.* **12**(4), 1700256 (2018).
19. X. Han, L. Chen, Y. Jiang, A. Frigg, H. Xiao, T. G. Nguyen, A. Boes, J. Yang, G. Ren, Y. Su, A. Mitchell, and Y. Tian, "Integrated subwavelength gratings on a lithium niobate on insulator platform for mode and polarization manipulation," *Laser Photonics Rev.* **16**(7), 2200130 (2022).
20. X. Han, Y. Jiang, A. Frigg, H. Xiao, P. Zhang, T. G. Nguyen, A. Boes, J. Yang, G. Ren, Y. Su, A. Mitchell, and Y. Tian, "Mode and polarization-division multiplexing based on silicon nitride loaded lithium niobate on insulator platform," *Laser Photonics Rev.* **16**(1), 2270001 (2022).
21. W. D. Sacher, W. M. J. Green, S. Assefa, T. Barwicz, H. Pan, S. M. Shank, Y. A. Vlasov, and J. K. S. Poon, "Coupling modulation of microrings at rates beyond the linewidth limit," *Opt. Express* **21**(8), 9722–9733 (2013).
22. T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, "Role of delay-times in delay-based photonic reservoir computing," *Opt. Mater. Express* **12**(3), 1214–1231 (2022).
23. F. Köster, D. Ehlert, and K. Lüdge, "Limitations of the recall capabilities in delay-based reservoir computing systems," *Cognitive Computation* (2020).
24. F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework pytorch," *Sci. Rep.* **9**(1), 5918 (2019).
25. D. Li, M. Han, and J. Wang, "Chaotic time series prediction based on a novel robust echo state network," *IEEE Trans. Neural Netw. Learning Syst.* **23**(5), 787–799 (2012).
26. H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," (2002).
27. N. Bertschinger and T. Natschläger, "Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks," *Neural Comput.* **16**(7), 1413–1436 (2004).
28. H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science* **304**(5667), 78–80 (2004).
29. A. Weigend and N. Gershenfeld, "Results of the time series prediction competition at the santa fe institute," in *IEEE International Conference on Neural Networks*, (1993), pp. 1786–1793 vol.3.
30. M. Borghi, S. Biasi, and L. Pavesi, "Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations," *Sci. Rep.* **11**(1), 15642 (2021).
31. J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer, "Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback," *Opt. Express* **25**(3), 2401 (2017).
32. M. Nakajima, K. Tanaka, and T. Hashimoto, "Scalable reservoir computing on coherent linear photonic processor," *Commun. Phys.* **4**(1), 20 (2021).
33. C. Sun, M. Wade, M. Georgas, S. Lin, L. Alloatti, B. Moss, R. Kumar, A. H. Atabaki, F. Pavanello, J. M. Shainline, J. S. Orcutt, R. J. Ram, M. Popovic, and V. Stojanovic, "A 45 nm cmos-soi monolithic photonics platform with bit-statistics-based resonant microring thermal tuning," *IEEE J. Solid-State Circuits* **51**(4), 893–907 (2016).