Performance Enhancement of Reservoir Computing Based on Fabry-Pérot Laser as All-Optical Reconfigurable Nonlinear Node

Mladen Ž. Banović, Jasna V. Crnjanski, Marko M. Krstić, and Dejan M. Gvozdić

(Invited Paper)

***Abstract*—The paper explores the performance enhancement of a photonic reservoir computer through the reconfigurability of all-optical nonlinear node based on Fabry-Pérot laser diode. By benchmarking various tasks with differing computational demands, we demonstrate how reconfigurability offers flexible computational capabilities, enabling the reservoir computer to adapt to specific tasks while striking an optimal balance between nonlinear transformation and linear memory. We achieve the state-of-the-art results using a basic delay-line reservoir computer concept with a small number of virtual nodes (∼30 nodes).**

***Index Terms*—All-optical, delay-based reservoir computing, reconfigurable activation function, semiconductor laser bistability, time-series prediction.**

I. Introduction

THE burgeoning era of Big Data analytics, driven by the Internet of Things Cyber-Physical Society, heralds an unprecedented demand for computing capabilities. As we confront the limitations of Moore's Law and Dennard's Scaling Law, the pursuit of enhanced computational prowess intensifies, all while aiming to promote environmental sustainability [1]. In this landscape, analog optical computing emerges as a compelling alternative [2]. Advancements in photonic Ising machines and vector/matrix manipulation pave the way for the evolution of Reservoir Computing (RC) – a recurrent neural network (RNN) architecture inspired by the human brain's information processing mechanisms. By leveraging machine learning within the RC framework, efficient sequential information processing is enabled, making it suitable for real-time analysis and prediction of complex and chaotic dynamic systems, all while minimizing training costs [3], [4].

Manuscript received 1 March 2024; revised 5 June 2024; accepted 30 June 2024. Date of publication 3 July 2024; date of current version 16 November 2024. This work was supported in part by the Science Fund of the Republic of Serbia (All-optical Reservoir Computer Architecture based on Laser Bistability) under Grant 7750121 and in part by the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia under Contract Number 451-03-65/2024-03/200103. *(Corresponding author: Dejan M. Gvozdić.)*

The authors are with the University of Belgrade-School of Electrical Engineering, 11120 Belgrade, Serbia (e-mail: mladen@etf.bg.ac.rs; jasna.crnjanski@etf.bg.ac.rs; marko.krstic@etf.bg.ac.rs; gvozdic@etf.bg.ac.rs).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2024.3422327.

Digital Object Identifier 10.1109/JLT.2024.3422327

The applications of RC span over diverse disciplines, illustrating its versatility and relevance across various domains such as economics [5], ecology [6], [7], power engineering [8], and medicine [9], [10]. In ICT, RC applications range from speech and context recognition [11], [12], which are among the most crucial, to communication encryption [13], human identification [14], advanced modulation formats generation [15], and fiber transmission equalization [16]. While RC's primary focus is on sequential data prediction, it also demonstrates efficacy in addressing classification tasks [17], [18].

One advanced feature of RC, compared to other artificial neural networks, is that only the readout coefficients of the output layer need to be trained and optimized using a simple learning algorithm such as linear regression, while the input and reservoir layers remain fixed. This simplification in the network's architecture is further extended by implementing RC as time-delayed dynamical systems [19], based on a single nonlinear node (NNode) subjected to a delayed feedback loop of round-trip time τ, rather than on the usual structure of multiple connected nodes. It allows for the excitation of several thousand virtual neurons and processing of incoming information in a highly dimensional reservoir.

Given its vast bandwidth, potential for WDM parallelization, integration capabilities, power efficiency, and low latency, photonics has emerged as a promising platform for RC hardware implementation, especially if data processing can be conducted directly in the optical domain, leading to the concept of all-optical RC [3]. However, in photonic implementations, achieving the separation of different inputs and their mapping by nonlinear transformation onto distinct reservoir states poses a significant challenge due to the weakly interacting nature of photons [20].

To date, numerous delayed-feedback RC photonics architectures have been proposed, theoretically investigated and experimentally demonstrated, differing in terms of platform (optoelectronic or all-optical), NNode nonlinearity profile, network complexity (single or multiple NNodes or delay-lines), and the level of algorithm or network parameter optimization. In numerical studies, the network is simulated in the time domain using rate or traveling wave rate equations based on real-time parameters such as node separation θ and feedback round-trip time τ. However, some theoretical studies do not consider analog

signals but rather their sampling values [21], [22], [23], [24], categorizing them as studies based on digital RC model.

The basic delay-feedback RC experimental implementations with a single optoelectronic NNode are commonly based on Mach Zehnder modulators (MZM) [25], [26], [27], [28], [29], [30], which provide sine or sine-squared nonlinear profiles, depending on weather the value of electric field or optical power, respectively, represents the quantity of the virtual nodes. Additionally, a readout photodiode (PD) can serve as a nonlinear element performing a quadratic transformation of the node electromagnetic field into photocurrent of RC output layer [11], [31]. More complex architectures, investigated numerically, also utilize a readout PD, either for establishing single/double optoelectronic feedback to a single [32] or mutually coupled LDs [33]. In these simulations, the input signal is directly injected into the nonlinear element through current modulation, eliminating the need for optical signal preprocessing. Advanced architectures based on 2 MZMs as dual NNodes and double feedback loops are also investigated [34], [35]. Generally, optoelectronic schemes generate minimal noise and typically offer excellent memory capacities [11]. However, the optical to electrical and vice-versa conversion entails a significant trade-off in speed and power efficiency, which additionally occurs due to amplification of photocurrent signals [32], [33].

The majority of basic RC architectures with all-optical NNodes incorporate nonlinear transformation via injection-locking in different types of laser diodes (LDs) [12], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], ring lasers [48], [49] and micro-ring resonators (MRRs) [50]. Borgis et al. [51] utilize Fabry-Pérot laser diodes (FP-LDs) as both master and slave lasers, numerically investigating parallel processing across modes under multimode injection. Harkhoe et al. [52] numerically investigate the single-mode injection with a zero detuning in a FP-LD. Tang et al. [40] employ experimental setup with single-mode injection in a FP-LD to explore the impact of frequency detuning and pump current, finding that the smallest prediction error is achieved for the smallest detuning. Similar conclusions regarding the smallest prediction error occurring at zero detuning have been also drawn in other studies [41], [53].

In addition to lasers, all-optical nonlinearity in RC schemes is achieved by using semiconductor optical amplifiers (SOAs) [54], [55], reflective SOAs as a nonlinear mirror [45], erbium-doped fiber amplifiers [56], and semiconductor saturable absorber mirror (SESAM) [57]. SOAs introduce too much amplified spontaneous emission noise, while SESAM, despite its weak quasi-linear nonlinearity, provides better RC performance owing to a stabilizing negative curvature profile [57]. SOAs and MRRs exhibit similar hyperbolic-like all-optical nonlinear profiles.

In more complex all-optical implementations, architectures with double feedback loops were numerically examined to achieve higher operation speeds and reduce prediction errors [58]. Furthermore, two-tasks parallelization was achieved by investigating mutually delay-coupled LDs [59], [60], [61], double optical injection in a single NNode based on VCSEL [62] or parallel NNodes based on integrated LDs with separate delay lines [63]. Additional degree of freedom leading to improved results was attained by combining VCSEL with 2 or 4 MZMs and by using polarization tracking [64]. Complex setups incorporating two connected reservoirs with a single NNode are also investigated [65], [66]. Additionally, a simple delay-based photonic reservoir, serving as a preprocessor for deep neural networks can also contribute to error reduction in prediction tasks [46].

The effectiveness of RC in predictive tasks hinges on a small number of training coefficients and hyperparameters, with minor variations significantly impacting performance. Hyperparameter optimization algorithms like grid search [28], [53], Bayesian [24] and entropy-based optimizations [67], and gradient boosting [65] are common, alongside training methods such as the Kalman filter [38], online training with gradient descent [29] or recursive least squares [68]. Moreover, bifurcation diagrams can aid in determining optimal parameter values [69].

The advanced and usually rather complex RC architectures, supported by network parameter optimization, have shown significant potential in enhancing RC performance across diverse benchmark tasks focused on achieving high levels of nonlinear transformation or superior linear memory [70]. The concept of RC reconfigurability has been recognized as beneficial in case of magnetic metamaterials exploiting an array of interconnected magnetic nanorings [71]. It may provide an additional degree of freedom in designing superior RC networks by allowing exploitation of the dynamical system versatility and keeping the network relatively simple. However, the overview of previous works has shown that reconfigurability of the NNode is not systematically studied, although various nonlinear functions have been proposed.

In this paper, we are focused to provide a detailed and systematic study of all-optical reconfigurability of the NNode. We employ a basic RC architecture with a single nonlinear node and a delay line [12] to probe the boundaries of NNode reconfigurability for performance improvement, with respect to standard RC benchmark tasks. The NNode relies on the transient response of a FP-LD, settled in the bistability regime through optical injection of the input signal with a specific frequency detuning relative to a side longitudinal mode. Optimal nonlinear profile within family of sigmoid-like and PReLU profiles, with gradually increasing threshold and saturation points, is selected by adjusting the frequency detuning. Since our simulation is based on assumption that time separation between nodes is sufficiently long to prevent node overlap or symbol interference, we implement digital model, which accounts for the stream of optical signal samples, while directly exploiting nonlinear transfer function derived by transient solving of multimode rate equations [72]. The study of NNode reconfigurability is supported with optimization of RC network parameters, also including analysis of RC performance on virtual nodes number.

In Chapter 2, we present our RC model and the forms of used nonlinearities. In Chapter 3, we outline the benchmark tasks performed. In Chapter 4, we discuss the impact of reconfigurable nonlinearity on the RC results in benchmark tasks and provide clear graphical comparison of our performance benchmarks alongside those reported in the literature.

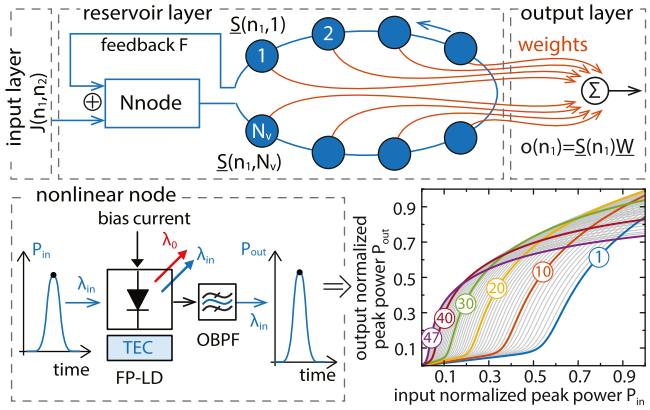

Fig. 1. Reservoir architecture composed of three layers: input layer, reservoir layer, and output layer. Schematic of the proposed hardware realization of nonlinear node based on Fabry-Pérot laser diode under optical injection and resulting family of normalized nonlinear functions for frequency detuning $\Delta\omega$ values starting from $-36\ \Omega$ (#1) to $-13\ \Omega$ (#47) with step $0.5\ \Omega$. Circled numbers represent ordinal number of nonlinear function with respect to $\Delta\omega$.

II. Reservoir Computer Model

A. Reservoir Computer Architecture

The reservoir computer architecture based on a single delay line with a single NNode is presented in Fig. 1. In the input layer, the input dataset $u(n_1)$ is multiplexed using a mask $M(n_2)$ to diversify the data entering to reservoir, $J(n_1, n_2)$:

$$J(n_1, n_2) = Au(n_1)M(n_2) + J_0 \tag{1}$$

where A and J_0 are scaling and offset parameters, respectively, n_1 is a counting iterator of the input dataset, while n_2 counts from 1 to the number of elements in the mask vector determined with the number of virtual nodes N_v in the reservoir.

The core of the reservoir is a nonlinear element $\mathcal{G}$, which activates the input data in the reservoir layer with a simple ring topology and generates the reservoir states $\underline{S}(n_1, n_2)$, forming the total state matrix. The connection between reservoir states is achieved through a desynchronization process between delay time and mask period, resembling a RNN and leading to simplified calculations [3]. In the discrete domain, the node separation time θ becomes equal to 1, while the delay time becomes $N_v + 1$. The size of delay line is determined by the connection established between reservoir state $\underline{S}(n_1, n_2)$ and the state entering the reservoir before $N_v + 1$ steps [3], [23], [55]. The strength of the recurrent connection is determined by the feedback parameter F, defining the percentage of the nonlinear element's response coupled with the data J entering the reservoir. The interactions between nonlinearity and feedback create elements of the reservoir state matrix:

$$\underline{S}(n_1, n_2) = \begin{cases} \mathcal{G}\left(F\underline{S}(n_1 - 1, n_2 - 1) + J(n_1, n_2)\right), & 2 \le n_2 \le N_v \\ \mathcal{G}\left(F\underline{S}(n_1 - 2, N_v) + J(n_1, n_2)\right), & n_2 = 1. \end{cases} \tag{2}$$

Reservoir states are formed by the first element from the input dataset $u(n_1 = 1)$. Coupling of the states starts from the next element $u(n_1 = 2)$, except for $n_2 = 1$, since the coupling distance is $N_v + 1$. An extra column with bias values of 1, $\underline{S}(n_1, N_v + 1) = 1$, is appended as the last column to introduce a constant offset in the predicted values, thereby smoothing the dynamics of the reservoir [12].

Finally, in the output layer, the predicted output data o is determined as a linear combination of reservoir states $o = \underline{S}\underline{W}$. The reservoir output is compared with the known target output $\hat{o}$, and the goal of training is to find the weight matrix $\underline{W}$ that minimizes the error metric. This is most efficiently achieved through the application of ridge regression [22], [73]:

$$\underline{W} = \left(\underline{S}^T\underline{S} + \lambda\underline{I}\right)^{-1}\underline{S}^T\hat{o}, \tag{3}$$

where $\underline{I}$ refers to the identity matrix, and λ is the Tikhonov regularization parameter, which depends on the correlation $\underline{S}^T\underline{S}$ and should be adjusted separately for each reservoir [73]. Since this paper aims to demonstrate the impact of the reconfigurable nonlinearity, including the optimized reservoir parameters, we restrict the network degree of freedom and fix λ to $\lambda = 5 \times 10^{-6}$ for all benchmark tests [23].

B. All-Optical Reconfigurable Nonlinear Node

The computational concept of a digital RC shown in Fig. 1 relays on the basic configuration of all-optical architecture proposed by Brunner et al. [12], with the major novelty related to implementation of a NNode.

Instead of the standard injection-locked DFB [74] or FP-LD [52], we use a optically injected FP-LD operating in the dispersive bistability regime [72]. The fundamental mechanism behind this bistability is the refractive index and injection-locked side-mode gain defect alteration, induced by change of the carrier density in the active region of the LD. This change is caused by variation of injected optical power P_{in} at certain negative frequency detuning $\Delta\omega$ with respect to the side-mode frequency of the free-running LD, and supported by the contributions of unlocked modes, including the free-running dominant mode [75]. Under the injection composed of short optical pulses, variation of P_{in} produces a bijective nonlinear dependence of the LD output signals' peak power P_{out} at the frequency corresponding to the input signal. The output signal is extracted with an optical band pass filter (OBPF) to prevent the contribution of other longitudinal modes and corresponding noise emission. The nonlinear profile of $P_{out}(P_{in})$ optical transfer characteristic strongly depends on the injected light frequency detuning and the LD bias current [72], providing the possibility to in-situ adapt or reconfigure the nonlinear node to a task-at-hand [76], [77]. These activation functions are experimentally verified in [78].

The theory behind bistability and corresponding NNode output is thoroughly discussed in [72], [75]. Here we focus only on numerically simulated and analytically fitted nonlinear transfer characteristics of optically injected FP-LD with the same material and geometrical characteristics as in [72]. A temperature controller (TEC) shown in Fig. 1 is used to prevent the optical spectrum drifting and consequently a change in the applied

frequency detuning. For small laser bias current of $2I_{\text{th}}$, where $I_{\text{th}} = 8.2\,\text{mA}$ represents the threshold current, and under intramodal injection of Gaussian pulses with FWHM = 60 ps in the side mode $m = -3$ (red spectrum side), a change in $\Delta\omega$ from $-36\ \Omega$ to $-13\ \Omega$ with a step of $0.5\ \Omega$ ($\Omega = 10^{10}$ rad/s) produces a family of 47 sigmoid- and PReLU-like profiles with gradually decreasing threshold points, shown in Fig. 1, after normalization to the range from 0 to 1. The activation functions are labeled with ordinal numbers, starting with #1 corresponding to the detuning $\Delta\omega = -36\Omega$.

For implementation of NNode in reservoir computer model we use a custom-defined analytical function [72]:

$$P_{\text{out}} = b_1 \ln\left\{1 + b_2 P_{\text{in}} + b_3 \ln\left[1 + \left(e^{b_4 P_{\text{in}}} - 1\right)^{b_5}\right]\right\} + b_6 P_{\text{in}} \tag{4}$$

with coefficients $b_k = b_k(\Delta\omega)$ and $k = 1, \ldots 6$ fitted to numerically calculated data points corresponding to each of the normalized transfer functions obtained for particular $\Delta\omega$ in the family of functions shown in Fig. 1. Therefore, reversible switching among different activation functions can be achieved by using frequency detuning $\Delta\omega$ as a control parameter.

III. Benchmark Tasks

The overall RC performance is estimated numerically by determining task-independent indicators such as memory capacity and memory quality, and by general benchmark testing on standard datasets for time-series prediction.

The general procedure is straightforward: for each test the dataset is split into two sets, one for training and the other for testing. In the training set, the initial 100 samples are excluded and used to initialize the reservoir. Similarly, within the testing set, 100 samples are discarded to warm the reservoir and serve as a buffer. The reservoir output $o(n_1)$ is computed using input values from the test set and the known weight coefficients acquired during training. This output is compared with the expected target output $\hat{o}(n_1)$ from the test set to calculate the Normalized Mean Square Error (NMSE):

$$\text{NMSE} = \frac{\left\langle \left(\hat{o}(n_1) - o(n_1)\right)^2 \right\rangle}{\left\langle \left(\hat{o}(n_1) - \langle \hat{o}(n_1) \rangle\right)^2 \right\rangle}, \tag{5}$$

or Symbol Error Rate (SER), representing the ratio of incorrectly predicted symbols and total number of tested symbols.

A. NARMA10

The Non-Linear Auto-Regressive Moving Average (NARMA) represents one of the fundamental tests for evaluating the quality of RC [34]. The NARMA10 dataset is generated based on the following analytical form:

$$\hat{o}(n_1 + 1) = 0.3\hat{o}(n_1) + 0.05\hat{o}(n_1) \sum_{i=0}^{9} \hat{o}(n_1 - i) + 1.5u(n_1 - 9)u(n_1) + 0.1, \tag{6}$$

where $u(n_1)$ are randomly selected values from the range 0 to 0.5 with a uniform distribution. The task is to predict the behavior of this 10^{th}-order nonlinear system in the next step $\hat{o}(n_1 + 1)$. Training and testing utilize sets of 1000 elements each. Averaging is performed for 10 different input datasets, due to the inherent randomness of the input dataset [34].

B. Santa Fe Laser

Santa Fe laser dataset represents a time series of intensities experimentally obtained for a far-infrared laser operating in a chaotic state [61]. The task is to predict one step ahead in a chaotic time series. The dataset consists of 9000 elements, with 3000 used for training and 1000 for testing [61].

C. Nonlinear Channel Equalization

The Nonlinear Channel Equalization (NCE) task considers intersymbol interference and noise [11]. The input signal $d(n_1)$ represents a random sequence with values from the set $\{-3, -1, 1, 3\}$, 10 consecutive elements of which are used to form a new sequence $q(n_1)$:

$$\begin{aligned} q(n_1) = {} & 0.08d(n_1 + 2) - 0.12d(n_1 + 1) + d(n_1) \\ & + 0.18d(n_1 - 1) - 0.1d(n_1 - 2) \\ & + 0.091d(n_1 - 3) - 0.05d(n_1 - 4) \\ & + 0.04d(n_1 - 5) + 0.03d(n_1 - 6) + 0.01d(n_1 - 7). \end{aligned} \tag{7}$$

Finally, the reservoir input sequence $u(n_1)$ takes into account 2^{nd} and 3^{rd} nonlinear distortion and Gaussian noise:

$$u(n_1) = q(n_1) + 0.036q^2(n_1) - 0.011q^3(n_1) + \text{noise}. \tag{8}$$

Noise is typically chosen to achieve a signal-to-noise ratio (SNR) of $12 - 32$ dB. The task of the reservoir is to reconstruct $\hat{o}(n_1) = d(n_1 - j)$ for the input signal $u(n_1)$, i.e., to predict the input sequence j steps backward based on the output sequence. The value of j varies in the literature, with one case being $j = 0$, which involves directly predicting $d(n_1)$ based on $u(n_1)$ [11], [26], [37], [55], [57], [79]. Another case is $j = 2$, where the prediction is made for $d(n_1 - 2)$ based on $u(n_1)$ [22], [48], [49], [80]. Since $d(n_1)$ consists of a random sequence, the results are averaged over 5 different sequences [57]. The training set consisted of 3000 samples and the test set size was 6000 samples [11].

D. Memory Capacity

Linear memory in RC refers to its capability to reconstruct input data that entered the delay line k steps earlier. Thus, for the linear memory, the targeted output of the reservoir is the delayed input $\widehat{o_k}(n_1) = u(n_1 - k)$. While the parameter k theoretically spans from 1 to infinity, in practice, contributions to memory capacity are not significant for k values greater than the number of virtual nodes.

Memory Capacity (MC) is a measure of the normalized correlation between the predicted output from the reservoir and the target value across delayed steps k [11]:

$$\text{MC}_k = \text{corr}\left[o_k(n_1), \widehat{o_k}(n_1)\right] = 1 - \text{NMSE}. \tag{9}$$

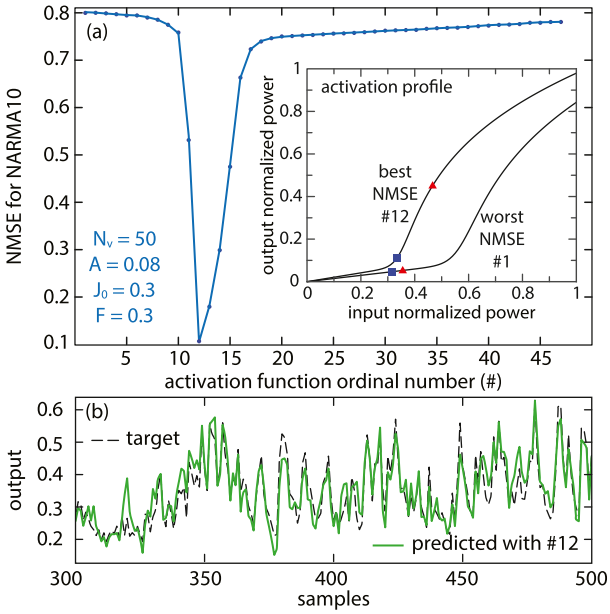

Fig. 2. (a) NMSE with respect to the activation function ordinal number for the NARMA10 dataset. Inset shows activation functions yielding the best NMSE (#12) and the worst NMSE (#1). The smallest and largest values entering the reservoir are marked with square and triangle markers, respectively. (b) Comparison between target and predicted signal for activation #12 and $N_v = 50$ virtual nodes.

The total capacity of linear memory is obtained by summing the memory capacities for each delay k [19]:

$$\mathrm{MC}_{\mathrm{tot}} = \sum_{k=1}^{\infty} \mathrm{MC}_k. \tag{10}$$

The memory capacity plotted against delayed input step k forms the memory curve. The total memory capacity is determined by the area under this curve, as defined by (10).

For testing the RC memory capacity the input data $u(n_1)$ takes random values from a uniform distribution in the interval $[0, 1]$ [57], eliminating the need to separate the data into training and test sets. A total of 2200 data points are utilized.

E. Memory Quality

Since the total memory capacity doesn't indicate the constancy of the memory curve, i.e., the number of steps k for which it maintains a rectangular shape [81], Memory Quality (MQ), as an additional task-independent property, is used to address the accuracy and integrity of information stored in the reservoir memory [82]:

$$\mathrm{MQ} = \frac{1}{\mathrm{MC}_{\mathrm{tot}}} \sum_{k=1}^{[\mathrm{MC}_{\mathrm{tot}}]} \mathrm{MC}_k. \tag{11}$$

For the frequently used NARMA10 dataset, it's crucial to retain the values of the ten preceding steps, necessitating a strong memory of these ten steps, while anything beyond that is irrelevant. However, in other tasks, remembering only one or two previous steps may suffice.

IV. Results

For all benchmark tasks, the goal is to determine the reservoir parameters A, J_0 and F, as well as to select the activation function that optimizes the chosen performance metric, e.g. prediction accuracy, memory capacity or quality. To maintain consistency in parameter values across all datasets, normalized activation functions are utilized, and the datasets themselves are normalized to a range from 0 to 1. Additionally, a uniform distribution of random values between 0 and 1 is applied to each element of the mask M. The results are averaged over 100 different random masks to mitigate the impact of specific individual mask values, similarly as in [23].

The employed algorithm for determining the optimal parameters A, J_0 and F for task-dependent tests is straightforward to implement but computationally intensive. The algorithm performs a grid search in a 3D space, with the constraint that the data entering reservoir $J(n_1, n_2)$ does not exceed 1 due to activation normalization. Each parameter set is tested across all 47 activation functions, and the set yielding the minimal prediction error, regardless the activation function, is defined as optimal. Finally, the impact of increasing the number of virtual nodes is assessed for the obtained parameters.

A. NARMA Test

Fig. 2(a) illustrates the impact of activation function profile on NMSE for the NARMA dataset, using optimal reservoir parameters $A = 0.08$, $F = 0.3$ and $J_0 = 0.3$ for $N_{\mathrm{v}} = 50$ virtual nodes. The graph highlights that the lowest NMSE < 0.15 is obtained for activation function #12. The inset displays the profiles of this best (#12) and the worst (#1) performing activation functions. The range of input values for these functions, delineated by square and triangle markers, is dictated by reservoir parameters. In Fig. 2(b), the predicted test output is shown alongside a comparison to the target output. The smaller value of $A = 0.08$ indicates utilization of only a fraction of the activation function's linear portion. The optimal activation function exhibits the highest slope around the offset parameter J_0, promoting input values clustering in this region and leading to distinct output separation. Additionally, due to the steep slope and feedback influence, the input value range for activation functions is expanded, resulting in an expansion of reservoir state value regions. Conversely, activation function #1 with a lower slope in the region of input value clustering, yields poorer NMSE performance, indicating less distinct reservoir state separation.

In Fig. 3, we illustrate the dependency of NMSE on the number of virtual nodes and compare our results with those from the literature. Our approach can provide competitive accuracy for small number of virtual nodes (≤ 50). For $N_{\mathrm{v}} = 30$, better accuracy is achieved only in a digital reservoir with dual input, and consequently dual masking of both input signals, particularly when considering both the current input and the input delayed by 9 steps [23]. However, in contrast to previous findings [22], [81], our model exhibits a different trend: increasing the number of nodes does not significantly reduce NMSE. For instance, with $N_{\mathrm{v}} = 30$ nodes, the best-case NMSE is 0.1278, while with 200 nodes, it is 0.105. This is due to a small gain

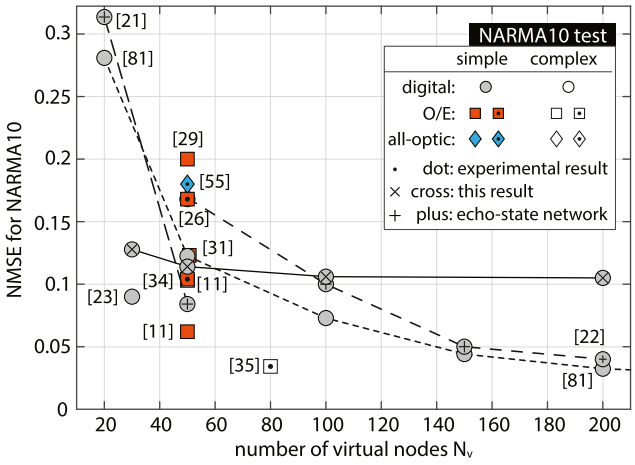

Fig. 3. NMSE for NARMA10 dataset with respect to the number of virtual nodes for digital reservoir computer proposed in this paper and for available results from the literature. Results from the literature are categorized based on the RC implementation platform: digital (circle markers), optoelectronic (square markers), and all-optical (diamond markers), as well as in terms of complexity (shaded markers correspond to simple RC architectures mainly based on a single NNode and a delay loop). Results with experimental verification are denoted with black dots inside markers.

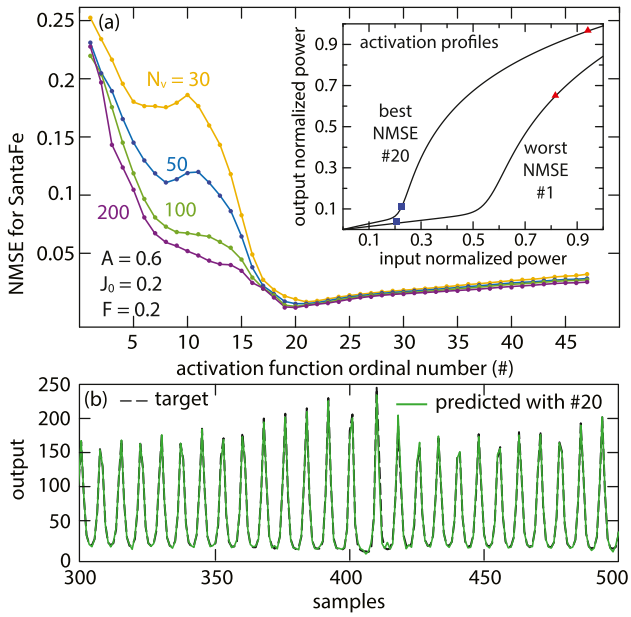

Fig. 4. (a) NMSE with respect to activation function ordinal number for the Santa Fe dataset and different number of virtual nodes (30, 50, 100, 200). Inset shows the range of input values for the activation function yielding the smallest (#20) and the largest (#1) NMSE value. (b) Comparison between target and predicted signal for activation #20 and $N_v = 50$ virtual nodes.

coefficient A, which limits the range of utilized node nonlinearity and together with memory of network, significantly affects NMSE of NARMA test. Nevertheless, the network parameters in our approach are already optimized to provide relatively small values of NMSE, even for small number of nodes (30) using the basic delay-line architecture. A modified architecture with an additional delay-line [23], would provide further reduction of NMSE and more significant dependence on node number.

B. Santa Fe Test

For the Santa Fe laser dataset, optimal values of reservoir parameters are $A = 0.6$, $J_0 = 0.2$, and $F = 0.2$. Fig. 4 illustrates the NMSE dependency on the activation function ordinal number for various numbers of virtual nodes and comparison between the predicted and target outputs. Notably, activation functions near #20 consistently yield the best NMSE values regardless of the number of nodes. Increasing N_v from 30 to 400 leads to a negligible NMSE reduction from 0.0082 to 0.0029, while prolonging the simulation time. The reduction in NMSE by increasing N_v is more significant for activation functions from #1 to #20, but these functions do not yield the best NMSE values. The inset in Fig. 4 shows the profiles of the best (#20) and the worst (#1) performing activation functions. Unlike for the NARMA dataset (Fig. 2), sigmoid-like activation profiles yield superior results for the Santa Fe dataset compared to PReLU type profiles.

A comparison with results from the literature is presented in Fig. 5, alongside the impact of increasing the number of nodes on NMSE. In our case, achieving NMSE values below 0.01, considered as the state-of-the-art [37], [46], requires only 30 NNodes. Lower NMSE values are typically attained in studies employing either a more complex architecture or a more intricate optimization process, which includes parameter tuning and training [24], [38], [59], [60].

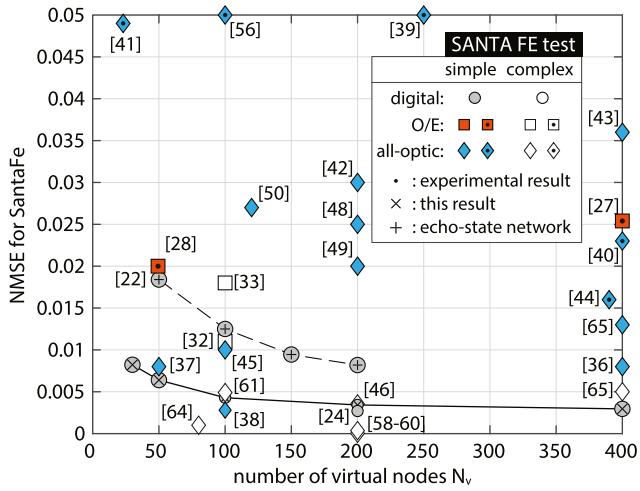

Fig. 5. NMSE for the SantaFe dataset with respect to the number of virtual nodes for digital reservoir computer proposed in this article and for available results from the literature.

C. NCE Test

Fig. 6 presents averaged values of NMSE and SER for the NCE dataset across different SNR values (ranging from 12 to 32 dB), plotted against the ordinal number of the activation function. Since simulations do not show significant error reduction with an increase in the number of nodes, we use relatively small $N_v = 50$ for which optimal reservoir parameters are determined as $A = 0.2$, $J_0 = 0.4$ and $F = 0.2$. The impact

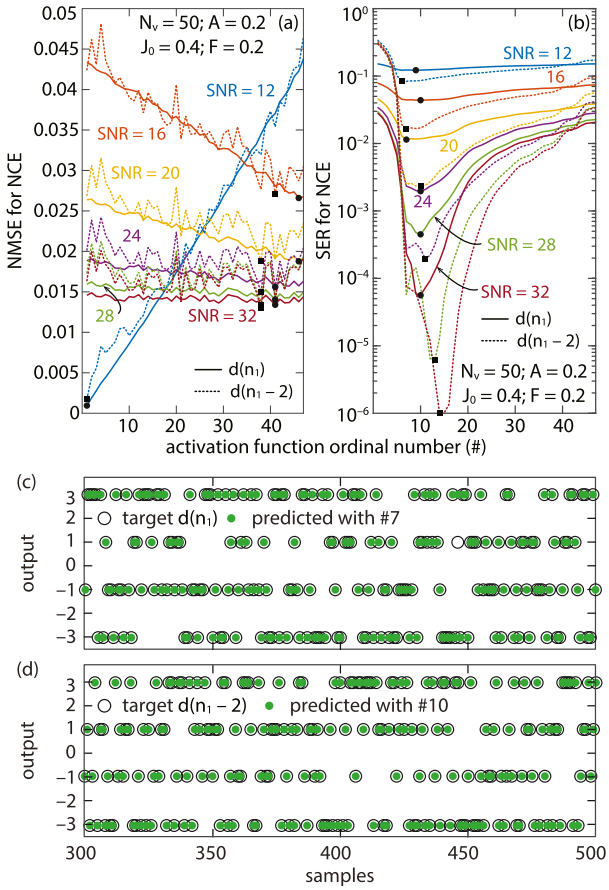

Fig. 6. (a) NMSE in linear and (b) SER in logarithmic scale with respect to activation function ordinal number for NCE dataset with different SNR. Dots and squares indicate optimal activation functions for $d(n_1)$ (solid lines) and $d(n_1 - 2)$ (dashed lines), respectively. Comparison between target and predicted signal for $N_v = 50$ virtual nodes and SNR = 20 dB, for (c) $d(n_1)$ and activation #7, and for (d) $d(n_1 - 1)$ and activation #10.

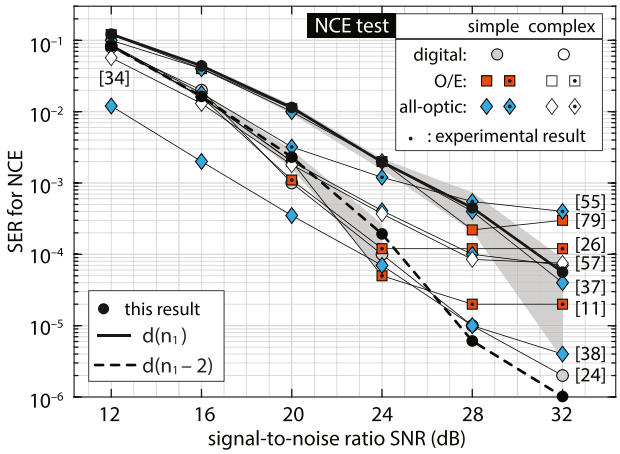

Fig. 7. Minimum SER values for the NCE dataset with respect to SNR for optimal activation function for predicting $d(n_1)$ and $d(n_1 - 2)$, compared with the literature. All literature results refer to the prediction $d(n_1)$. Average SER values are marked with markers and shaded areas represent the range of values obtained in each simulation run.

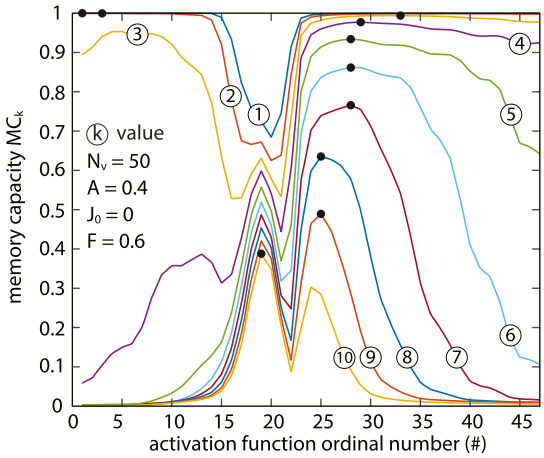

Fig. 8. Memory capacity MC_k with respect to activation function ordinal number and for predicting input for $k = 1$–10 steps in past. Dots indicate the optimal activation function.

of changing the activation function is more pronounced for SER than for NMSE. This influence is even more pronounced when predicting $d(n_1 - 2)$. Finally, a comparison between the optimal activation functions in Figs. 6(a) and (b) reveals that the activation function minimizing NMSE does not necessarily minimize SER. In Fig. 6(c) and (d), the predicted and target outputs are shown for SNR = 20 dB.

In Fig. 7, we depict the minimum SER as a function of SNR, comparing it with published results. The minimum SER was determined from the previous figure for each SNR value, considering the optimal activation function. As the input sequence is a random, the results are significantly influenced by the specific random number generator. Therefore, in addition to markers denoting averaged SER values for 5 independent simulation runs, we use shaded area to represent the range of values obtained in each simulation run. For specific input sequences and high SNR values (> 30 dB), the number of misinterpreted symbols can even be zero.

While our SER results for predicting $d(n_1)$ align with existing literature, predictions for $d(n_1 - 2)$ with SNR > 25 dB outperform previous studies in both NMSE and SER [48], [49], [80]. As shown in Fig. 7, we achieve satisfactory SER values (< 0.01 [60]) for $d(n_1 - 2)$@ SNR > 16 dB and for $d(n_1)$@ SNR ≥ 20 dB. Previous studies with similar simple reservoir architectures reported NMSE of 0.058 at SNR = 30 dB [80], whereas more complex architectures based on delay-decoupled or -coupled reservoirs achieved a minimum NMSE of 0.026. Our results show this low NMSE even for SNR = 28 dB.

D. Memory Capacity

The purpose of memory capacity evaluation is to assess the reservoir effectiveness in retaining past inputs. To investigate potentially high MC, we fix large feedback parameter, $F = 0.6$, and adjust the remaining reservoir parameters to $A = 0.4$ and $J_0 = 0$, ensuring that the data J entering the reservoir falls within the range [0,1], aligning with normalized activation functions. In Fig. 8, we present MC_k with respect to activation function for predicting the last 10 input values, corresponding to k values from 1 to 10.

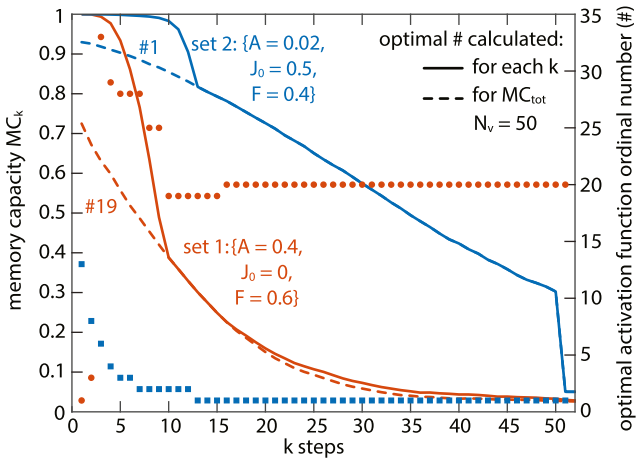

Fig. 9. MC_k for two sets of reservoir parameters. Solid lines represent memory curves for MC_k calculated for optimal activation in each step k. The ordinal number of the selected activation function is displayed on the right axis with circle markers (set 1) and square markers (set 2). Dashed lines show MC_k for only one activation function for which the total memory is the largest (#19) and (#1), for parameter sets 1 and 2, respectively.

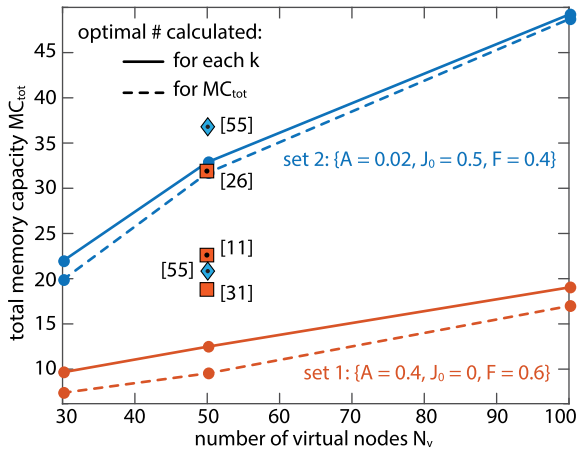

Fig. 10. Total memory capacity MC_{tot} relative to the number of virtual nodes for two combinations of scaling coefficients and a comparison with results from the literature.

TABLE I
MEMORY QUALITY FOR TWO SETS OF RESERVOIR PARAMETERS WITH RESPECT TO THE NUMBER OF VIRTUAL NODES

Memory quality (MQ)	30 nodes	50 nodes	100 nodes
$A = 0.4, J_0 = 0, F = 0.6$	0.7688 (0.8356)	0.5721 (0.7)	0.5613 (0.64)
$A = 0.02, J_0 = 0.5, F = 0.4$	0.7491 (0.8252)	0.7689 (0.7933)	0.7535 (0.7561)

For k values 1 and 2, activation functions with a steep slope at either large input values (ranging from #1 to #10) or small values (ranging from #25 to #47) yield similarly good outcomes. However, as k increases, activation functions positioned between these extremes emerge as optimal. The reservoir demonstrates efficient memory retention, with $MC_k \geq 0.9$ for k ranging from 1 to 4.

To enhance memory capacity we further adjust reservoir parameters. While keeping the relatively high feedback parameter of $F = 0.4$, we decrease the reservoir scaling parameter to $A = 0.02$ to reduce the range of input values, and adjust the offset parameter to $J_0 = 0.5$ thus positioning data in the region of the activation function's steepest slope. Fig. 9 shows memory curves MC_k for predicting delayed input at step k, for two parameter sets, $\{A, J_0, F\}$, set 1: $\{0.4, 0, 0.6\}$, and set 2: $\{0.02, 0.5, 0.4\}$. Dashed lines represent memory curves calculated for a single activation function identical in every k step. To select the optimal activation function, we calculate the total memory for each activation function and choose the one that maximizes the total memory MC_{tot}. For set 1, this is activation function #19, while for set 2, it is #1. Solid lines in Fig. 9 represent memory curves where the optimal activation function, yielding maximal MC_k, is chosen for each k, as read on the left axis in Fig. 9. The ordinal number of the optimal function for each k is denoted with markers and can be read on the right axis in Fig. 9. It can be noted that for sufficiently large number of steps, the optimal activation function saturates and remains fixed (#20 for set 1, #1 for set 2). The critical number of steps before the optimal activation function saturates depends on the reservoir parameters. For set 1 ($A = 0.4$, $J_0 = 0$, and $F = 0.6$) the critical number of steps is $k = 15$. In this case, the activation function #20 that becomes and remains optimal does not correspond to the activation function #19 for which the total memory is maximal, since the contribution to the total memory of these steps is small ($MC_k \leq 0.2$). For set 2 ($A = 0.02$, $J_0 = 0.5$, and $F = 0.4$) and prediction of 13 and more steps, the optimal activation settles at #1, aligning with the activation function providing maximal total memory. This is evidenced by a complete overlap between solid and dashed lines for $k \geq 13$. MC_k versus ordinal number of the activation function for $k \geq 13$, resembles the sharp dependence seen in Fig. 2 for small value of scaling parameter A, where we have a distinctive optimal activation function.

Fig. 10 presents the total memory capacity (MC_{tot}) in comparison with relevant literature results concerning the number of nodes. Total memory capacity increases in line with increasing N_v. We compute MC_{tot} for $k \leq N_v$, as for $k > N_v$, MC_k is more a result of the limited dataset and statistical approach rather than actual memory. Fig. 10 illustrates the contribution of both sets of reservoir parameters, and covers the cases when the activation function changes with respect to step k, and when only one activation function is used (the one yielding the highest total memory). Fig. 9 shows that for the second set of reservoir parameters, there is less difference in total memory between these two cases, diminishing as the number of nodes increases. This underscores how a reconfigurable nonlinear activation function with sufficiently high slope and offset J_0, along with reservoir feedback parameter optimized to provide strong feedback, can significantly enhance memory and total memory capacity in an all-optical RC architecture.

E. Memory Quality

The memory quality was assessed for reservoir parameters sets 1 and 2 with respect to the number of virtual nodes, as shown in Table I. Two MQ values are presented: one corresponding to the case of single activation function, yielding the highest total memory capacity, and the other (in parentheses) corresponding

to optimal activation function determined for each delay step k. As expected, the second value surpasses the first, commending the adaptability of the activation function.

Generally, a smaller parameter A value results in a larger MQ since the shape of the memory curve tends to resemble a rectangle. However, even with a low MC_{tot} value, a high MQ can be achieved, particularly if the number of nodes is small enough. Increasing the MC_{tot} due to the augmentation of the number of nodes results in a decrease in MQ. This is attributed to the contribution of predicting larger backward steps to MC_{tot}, causing the memory curve to exhibit a gradually declining trend. However, for a steep transition between different steps on the memory curve, Moran [83] achieved MQ of approximately 0.86 with $MC_{tot} = 17.6$ for 400 nodes, a value comparable to our result obtained with 30 nodes.

V. Conclusion

In this paper, we investigate how nonlinear node reconfigurability, coupled with network parameter optimization, enhances the performance of photonic reservoir computers. To isolate the impact of reconfigurability and minimize the influence of RC architecture, we focus on the basic delay-line architecture, comprising a single delay-line and all-optical reconfigurable nonlinear node based on Fabry-Pérot laser. Our findings reveal that node reconfigurability can provide both high quality nonlinear transformation and memory, traits typically considered mutually exclusive. Optimal nonlinear profiles, combined with optimized network parameters, yield excellent predictions on chaotic time series like the Santa Fe dataset, demonstrating efficiency of RC in nonlinear transformations, with small NMSE (3.4×10^{-3} for 200 nodes) rivaling state-of-the-art performance. By simply reconfiguring the nonlinear profile, high memory capacity becomes attainable, resulting in reduced NMSE for datasets like NARMA10 and NCE, especially for $d(n-2)$, with notably low number of nodes (30 nodes). Additionally, nonlinear profile optimization allows for competitive total memory capacity and quality, aligning closely with the best results in the literature.

References

[1] K. I. Kitayama, M. Notomi, M. Naruse, K. Inoue, S. Kawakami, and A. Uchida, "Novel frontier of photonics for data processing-photonic accelerator," *APL Photon.*, vol. 4, 2019, Art. no. 090901.

[2] C. Li, X. Zhang, J. Li, T. Fang, and X. Dong, "The challenges of modern computing and new opportunities for optics," *PhotoniX*, vol. 2, 2021, Art. no. 20.

[3] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger, "Tutorial: Photonic neural networks in delay systems," *J. Appl. Phys.*, vol. 124, 2018, Art. no. 152004.

[4] Y. K. Chembo, "Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems," *Chaos*, vol. 30, 2020, Art. no. 013111.

[5] W. J. Wang, Y. Tang, J. Xiong, and Y. C. Zhang, "Stock market index prediction based on reservoir computing models," *Expert Syst. Appl.*, vol. 178, 2021, Art. no. 115022.

[6] J. D. Ser et al., "Randomization-based machine learning in renewable energy prediction problems: Critical literature review, new results and perspectives," *Appl. Soft Comput.*, vol. 118, 2022, Art. no. 108526.

[7] J. Hu, Y. Lin, J. Tang, and J. Zhao, "A new wind power interval prediction approach based on reservoir computing and a quality-driven loss function," *Appl. Soft Comput. J.*, vol. 92, 2020, Art. no. 106327.

[8] B. T. Nadiga, "Reservoir computing as a tool for climate predictability studies," *J. Adv. Model. Earth Syst.*, vol. 13, 2021, Art. no. e2020MS002290.

[9] B. Liu et al., "Nanophotonic reservoir computing for COVID-19 pandemic forecasting," *Nonlinear Dyn.*, vol. 111, pp. 6895–6914, 2023.

[10] L. Chiasson-Poirier, H. Younesian, K. Turcot, and J. Sylvestre, "Detecting gait events from accelerations using reservoir computing," *Sensors*, vol. 22, 2022, Art. no. 7180.

[11] Q. Vinckier et al., "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, pp. 438–446, 2015.

[12] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature Commun.*, vol. 4, 2013, Art. no. 1364.

[13] L. Jin, Z. Liu, and L. Li, "Chain-structure time-delay reservoir computing for synchronizing chaotic signal and an application to secure communication," *Eurasip J. Adv. Signal Process.*, vol. 2022, 2022, Art. no. 68.

[14] E. Picco, P. Antonik, and S. Massar, "High speed human action recognition using a photonic reservoir computer," *Neural Netw.*, vol. 165, pp. 662–675, 2023.

[15] Q. Cai et al., "Modulation format identification in fiber communications using single dynamical node-based photonic reservoir computing," *Photon. Res.*, vol. 9, pp. B1–B8, 2021.

[16] A. Argyris et al., "Comparison of photonic reservoir computing systems for fiber transmission equalization," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, 2020, Art. no. 5100309.

[17] T. Yamaguchi, K. Arai, T. Niiyama, A. Uchida, and S. Sunada, "Time-domain photonic image processor based on speckle projection and reservoir computing," *Commun. Phys.*, vol. 6, 2023, Art. no. 250.

[18] S. Krishnagopal, Y. Aloimonos, and M. Girvan, "Similarity learning and generalization with limited data: A reservoir computing approach," *Complexity*, vol. 2018, 2018, Art. no. 6953836.

[19] L. Appeltant et al., "Information processing using a single dynamical node as complex system," *Nature Commun.*, vol. 2, 2011, Art. no. 468.

[20] D. Brunner, M. C. Soriano, and G. V. d. Sande Eds., *Photonic Reservoir Computing, Optical Recurrent Neural Networks*. Berlin, Boston: De Gruyter, 2019.

[21] H. Jaeger, "Adaptive nonlinear system identification with echo state networks," in *Proc. Adv. Neural Inf. Process. Syst.*, S. Becker, S. Thrun, and K. Obermayer, Eds., 2002, pp. 609–616.

[22] A. Rodan and P. Tiňo, "Minimum complexity echo state network," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 131–144, Jan. 2011.

[23] L. Jaurigue, E. Robertson, J. Wolters, and K. Lüdge, "Reservoir computing with delayed input for fast and easy optimisation," *Entropy*, vol. 23, no. 12, 2021, Art. no. 1560.

[24] J. Yperman and T. Becker, "Bayesian optimization of hyper-parameters in reservoir computing," 2016, *arXiv:1611.05193*.

[25] L. Larger et al., "Photonic information processing beyond turing: An optoelectronic implementation of reservoir computing," *Opt. Exp.*, vol. 20, no. 3, pp. 3241–3249, 2012.

[26] Y. Paquot et al., "Optoelectronic reservoir computing," *Sci. Rep.*, vol. 2, 2012, Art. no. 287.

[27] P. Kumar, M. Jin, T. Bu, S. Kumar, and Y.-P. Huang, "Efficient reservoir computing using field programmable gate array and electro-optic modulation," *OSA Continuum*, vol. 4, pp. 1086–1098, 2021.

[28] M. C. Soriano et al., "Optoelectronic reservoir computing: Tackling noise-induced performance degradation," *Opt. Exp.*, vol. 21, no. 1, pp. 12–20, 2013.

[29] P. Antonik, M. Haelterman, and S. Massar, "Online training for high-performance analogue readout layers in photonic reservoir computers," *Cogn. Computation*, vol. 9, pp. 297–306, 2017.

[30] L. Huang and J. Yao, "Multi-task photonic time-delay reservoir computing based on polarization modulation," *Opt. Lett.*, vol. 47, pp. 6464–6467, 2022.

[31] A. Akrout, P. Antonik, M. Haelterman, and S. Massar, "Towards autonomous photonic reservoir computer based on frequency parallelism of neurons," *Proc. SPIE*, vol. 10089, pp. 120–126, 2017.

[32] W. Liang et al., "Enhanced optoelectronic reservoir computation using semiconductor laser with double delay feedbacks," *Appl. Opt.*, vol. 62, pp. 620–626, 2023.

[33] W. Y. Liang et al., "Design of parallel reservoir computing by mutually-coupled semiconductor lasers with optoelectronic feedback," *Opt. Commun.*, vol. 495, 2021, Art. no. 127120.

[34] Y. Chen et al., "Reservoir computing system with double optoelectronic feedback loops," *Opt. Exp.*, vol. 27, pp. 27431–27440, 2019.

[35] M. Hermans, P. Antonik, M. Haelterman, and S. Massar, "Embodiment of learning in electro-optical signal processors," *Phys. Rev. Lett.*, vol. 117, no. 12, Sep. 2016, Art. no. 128301.

[36] J. Nakayama, K. Kanno, and A. Uchida, "Laser dynamical reservoir computing with consistency: An approach of a chaos mask signal," *Opt. Exp.*, vol. 24, pp. 8679–8692, 2016.

[37] D. Yue et al., "Performance optimization research of reservoir computing system based on an optical feedback semiconductor laser under electrical information injection," *Opt. Exp.*, vol. 27, no. 14, pp. 19931–19939, 2019.

[38] J. Jin et al., "Adaptive time-delayed photonic reservoir computing based on Kalman-filter training," *Opt. Exp.*, vol. 30, pp. 13647–13658, 2022.

[39] K. Takano et al., "Compact reservoir computing with a photonic integrated circuit," *Opt. Exp.*, vol. 26, pp. 29424–29439, 2018.

[40] J.-Y. Tang et al., "Asynchronous photonic time-delay reservoir computing," *Opt. Exp.*, vol. 31, pp. 2456–2466, 2023.

[41] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. d. Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Opt. Exp.*, vol. 28, 2020, pp. 3086–309.

[42] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. d. Sande, "Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics," *Opt. Exp.*, vol. 22, pp. 8672–8686, 2014.

[43] K. Hicke, M. A. Escalona-Morán, D. Brunner, M. C. Soriano, I. Fischer, and C. R. Mirasso, "Information processing using transient dynamics of semiconductor lasers subject to delayed feedback," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 4, Jul.-Aug. 2013, Art. no. 1501610.

[44] J. Vatin, D. Rontani, and M. Sciamanna, "Experimental reservoir computing using VCSEL polarization dynamics," *Opt. Exp.*, vol. 27, pp. 18579–18584, 2019.

[45] X. Li et al., "Performance-enhanced time-delayed photonic reservoir computing system using a reflective semiconductor optical amplifier," *Opt. Exp.*, vol. 31, pp. 28764–28777, 2023.

[46] I. Bauwens, G. V. d. Sande, P. Bienstman, and G. Verschaffelt, "Using photonic reservoirs as preprocessors for deep neural networks," *Front. Phys.*, vol. 10, 2022, Art. no. 1051941.

[47] Y. Liu, Z. M. Wu, S. L. Tan, and G. Q. Xia, "Improving the processing rate of VCSEL-based reservoir computing through adopting variable polarization information injection," *Opt. Laser Technol.*, vol. 161, 2023, Art. no. 109200.

[48] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. D. Sande, "Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3301–3307, Dec. 2015.

[49] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. d. Sande, "Reducing the phase sensitivity of laser-based optical reservoir computing systems," *Opt. Exp.*, vol. 24, pp. 1238–1252, 2016.

[50] H. Zhang et al., "Integrated photonic reservoir computing based on hierarchical time-multiplexing structure," *Opt. Exp.*, vol. 22, no. 25, pp. 31356–31370, 2014.

[51] A. Bogris, C. Mesaritakis, S. Deligiannidis, and P. Li, "Fabry-Perot lasers as enablers for parallel reservoir computing," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 2, Mar./Apr. 2021, Art. no. 7500307.

[52] K. Harkhoe and G. V. d. Sande, "Task-independent computational abilities of semiconductor lasers with delayed optical feedback for reservoir computing," *Photonics*, vol. 6, 2019, Art. no. 124.

[53] J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer, "Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback," *Opt. Exp.*, vol. 25, pp. 2401–2412, 2017.

[54] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," *IEEE Trans. Neural Netw.*, vol. 22, no. 9, pp. 1469–1481, Sep. 2011.

[55] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Exp.*, vol. 20, no. 20, pp. 22783–22795, Sep. 2012.

[56] R. M. Nguimdo, E. Lacot, O. Jacquin, O. Hugon, G. V. d. Sande, and H. G. d. Chatellus, "Prediction performance of reservoir computing systems based on a diode-pumped erbium-doped microchip laser subject to optical feedback," *Opt. Lett.*, vol. 42, pp. 375–378, 2017.

[57] A. Dejonckheere et al., "All-optical reservoir computer based on saturation of absorption," *Opt. Exp.*, vol. 22, pp. 10868–10881, 2014.

[58] Y. Hou et al., "Prediction performance of reservoir computing system based on a semiconductor laser subject to double optical feedback and optical injection," *Opt. Exp.*, vol. 26, pp. 10211–10219, 2018.

[59] Y. S. Hou, G. Q. Xia, E. Jayaprasath, D. Z. Yue, W. Y. Yang, and Z. M. Wu, "Prediction and classification performance of reservoir computing system using mutually delay-coupled semiconductor lasers," *Opt. Commun.*, vol. 433, pp. 215–220, 2019.

[60] Y. S. Hou, G. Q. Xia, E. Jayaprasath, D. Z. Yue, and Z. M. Wu, "Parallel information processing using a reservoir computing system based on mutually coupled semiconductor lasers," *Appl. Phys. B: Lasers Opt.*, vol. 126, 2020, Art. no. 40.

[61] D. Cai, Y. Yang, P. Zhou, and N. Li, "Enhanced prediction performance of reservoir computing based on mutually delay-coupled semiconductor lasers via parameter mismatch," *Electronics*, vol. 11, 2022, Art. no. 2577.

[62] X. Tan, Y. Hou, Z. Wu, and G. Xia, "Parallel information processing by a reservoir computing system based on a VCSEL subject to double optical feedback and optical injection," *Opt. Exp.*, vol. 27, pp. 26070–26079, 2019.

[63] C. Sugano, K. Kanno, and A. Uchida, "Reservoir computing using multiple lasers with feedback on a photonic integrated circuit," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan./Feb. 2020, Art. no. 1500409.

[64] 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. Exp.*, vol. 27, pp. 23293–23306, 2019.

[65] J. Y. Tao, Z. M. Wu, D. Z. Yue, X. S. Tan, Q. Q. Zeng, and G. Q. Xia, "Performance enhancement of a delay-based reservoir computing system by using gradient boosting technology," *IEEE Access*, vol. 8, pp. 151990–151996, 2020.

[66] H. Hasegawa, K. Kanno, and A. Uchida, "Parallel and deep reservoir computing using semiconductor lasers with optical feedback," *Nanophotonics*, vol. 12, pp. 869–881, 2023.

[67] T. L. Carroll, "Optimizing reservoir computers for signal classification," *Front. Physiol.*, vol. 12, 2021, Art. no. 685121.

[68] P. Antonik, M. Hermans, F. Duport, M. Haelterman, and S. Massar, "Towards pattern generation and chaotic series prediction with photonic reservoir computers," *Proc. SPIE*, vol. 9732, pp. 21–32, 2016.

[69] M. Zhang, Z. Liang, and Z. R. Huang, "Hardware optimization for photonic time-delay reservoir computer dynamics," *Neuromorphic Comput. Eng.*, vol. 3, 2023, Art. no. 014008.

[70] M. Inubushi and K. Yoshimura, "Reservoir computing beyond memory-nonlinearity trade-off," *Sci. Rep.*, vol. 7, 2017, Art. no. 10199.

[71] I. T. Vidamour et al., "Reconfigurable reservoir computing in a magnetic metamaterial," *Commun. Phys.*, vol. 6, 2023, Art. no. 230.

[72] J. Crnjanski, M. Krstić, A. Totović, N. Pleros, and D. Gvozdić, "Adaptive sigmoid-like and PReLU activation functions for all-optical perceptron," *Opt. Lett.*, vol. 46, no. 9, pp. 2003–2006, 2021.

[73] F. Wyffels, B. Schrauwen, and D. Stroobandt, "Stable output feedback in reservoir computing using ridge regression," in *Proc. Int. Conf. Artif. Neural Netw.* 2008, pp. 808–817.

[74] D. Z. Yue, Z. M. Wu, Y. S. Hou, and G. Q. Xia, "Effects of some operation parameters on the performance of a reservoir computing system based on a delay feedback semiconductor laser with information injection by current modulation," *IEEE Access*, vol. 7, pp. 128767–128773, 2019.

[75] M. M. Krstić, J. V. Crnjanski, and D. M. Gvozdić, "Injection power and detuning-dependent bistability in Fabry–Perot laser diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 18, no. 2, pp. 826–833, Mar./Apr. 2012.

[76] J. V. Crnjanski, I. Teofilović, M. M. Krstić, and D. M. Gvozdić, "Application of a reconfigurable all-optical activation unit based on optical injection into a bistable Fabry– Perot laser in multilayer perceptron neural networks," *Opt. Lett.*, vol. 49, no. 5, pp. 1153–1156, Mar. 2024.

[77] J. Crnjanski, M. Banović, I. Teofilović, M. Krstić, and D. Gvozdić, "Application of adaptive activation unit based on injection-locked lasers in machine learning tasks," in *Proc. IEEE Photon. Soc. Summer Topicals Meeting Ser.*, 2023, pp. 1–2.

[78] P. Atanasijević et al., "Adaptive all-optical sigmoid activation functions for photonic neural networks using Fabry-Perot laser diodes under optical injection," in *Proc. IEEE Opt. Fiber Commun. Conf. Exhib.*, 2024, pp. 1–3.

[79] S. Boshgazi, A. Jabbari, K. Mehrany, and M. Memarian, "Virtual reservoir computer using an optical resonator," *Opt. Mater. Exp.*, vol. 12, no. 3, pp. 1140–1153, 2022.

[80] S. Ortín and L. Pesquera, "Reservoir computing with an ensemble of time-delay reservoirs," *Cogn. Computation*, vol. 9, pp. 327–336, 2017.

[81] L. Appeltant, "Reservoir computing based on delay-dynamical systems," Ph.D. dissertation, Vrije Universiteit Brussel, Ixelles, Belgium, Universitat de les Illes Balears, Palma, Spain, 2012.

[82] M. Hermans and B. Schrauwen, "Memory in linear recurrent neural networks in continuous time," *Neural Netw.*, vol. 23, no. 3, pp. 341–355, 2010.

[83] M. A. E. Morán, "Computational properties of delay coupled systems," Ph.D. dissertation, Universitat de les Illes Balears, Palma, Spain, 2012.