

High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification

Laurent Larger,¹ Antonio Baylón-Fuentes,¹ Romain Martinenghi,¹ Vladimir S. Udaltsov,^{1,2}
Yanne K. Chembo,¹ and Maxime Jacquot¹

¹*FEMTO-ST Institute/Optics Department, CNRS & University Bourgogne Franche-Comté,
15B avenue des Montboucons, 25030 Besançon Cedex, France*

²*Vavilov Optical State Institute, Saint-Petersburg, Russia*

(Received 30 January 2015; revised manuscript received 13 November 2016; published 6 February 2017)

Reservoir computing, originally referred to as an echo state network or a liquid state machine, is a brain-inspired paradigm for processing temporal information. It involves learning a “read-out” interpretation for nonlinear transients developed by high-dimensional dynamics when the latter is excited by the information signal to be processed. This novel computational paradigm is derived from recurrent neural network and machine learning techniques. It has recently been implemented in photonic hardware for a dynamical system, which opens the path to ultrafast brain-inspired computing. We report on a novel implementation involving an electro-optic phase-delay dynamics designed with off-the-shelf optoelectronic telecom devices, thus providing the targeted wide bandwidth. Computational efficiency is demonstrated experimentally with speech-recognition tasks. State-of-the-art speed performances reach one million words per second, with very low word error rate. Additionally, to record speed processing, our investigations have revealed computing-efficiency improvements through yet-unexplored temporal-information-processing techniques, such as simultaneous multisample injection and pitched sampling at the read-out compared to information “write-in”.

DOI: [10.1103/PhysRevX.7.011015](https://doi.org/10.1103/PhysRevX.7.011015)

Subject Areas: Complex Systems, Nonlinear Dynamics,
Photonics

I. INTRODUCTION

Nowadays, digital computers based on the so-called Turing–von Neumann architectures are ubiquitous and deeply integrated in our daily lives. They provide many fast and efficient calculation tasks, from complex scientific computing through networking and communications systems, to smart phone device functionality and service. The demand for greater computational power is naturally always increasing, and as this demand develops, more and more problems are identified as too complex and/or too time consuming, even for the most advanced highly paralleled digital computer farms. Alternative computational paradigms have already been explored for a long time, one obvious direction naturally being suggested by the human brain. However, most of the research dedicated to the brain-inspired computational paradigm has been performed essentially through computer simulations, i.e., through the use of the standard digital Turing–Von Neumann computers that they are actually aimed to replace.

Among the numerous brain-inspired concepts, reservoir computing (RC) [1], also known originally as an echo state

network [2] or a liquid state machine [3] (see Ref. [4] for a review), has recently attracted much attention through an unexpected physical implementation of the hardware [5–8]. Instead of the usual spatially extended networklike dynamical reservoir, purely temporal delay dynamical systems were used as a way to emulate the dynamical complexity traditionally provided by a network of neurons. Delay systems are time-domain, infinite-dimensional dynamics, which have been shown to be able to act as complex dynamical reservoirs, surprisingly and efficiently replacing the traditional spatiotemporal neural-network architectures. They provide technologically tractable hardware solutions, where densely connected spatiotemporal architectures are difficult to design, and they intrinsically benefit from several available highly efficient signal-processing tools and devices developed for modern telecommunications. In addition to the already ground-breaking result of a physical hardware implementation, photonic realizations [9–11] can technically shift the intrinsic millisecond time scale of the brain down to the picosecond or even femtosecond time scale available in standard optical telecommunications, basically gaining 9 to 12 orders of magnitude in the processing speed. The present work finds its place along this line, reporting an electro-optic (EO) phase-delay dynamics built with telecom bandwidth devices and providing ultrafast information processing while implementing reservoir-computing concepts with dedicated hardware.

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Compared to the current state of the art [9], our novel electro-optic phase architecture provides an improvement by a factor of 3, reaching a million words per second when tested on a standard speech-recognition database. As reservoir computing is a universal computational concept, many other complex problems can, in principle, be addressed by our approach, such as real-time information extraction from high bit-rate optical data flow or the prediction of a high-dimensional deterministic time series.

The article is organized as follows: Section II introduces some basic concepts of reservoir computing. Section III bridges these neural-network–based concepts with a time-domain only hardware implementation in the particular case of the proposed electro-optic phase-delay oscillator. Section IV A reports the experimental and numerical results obtained on a first benchmark of spoken digit recognition (TI46). This particular temporal task example gives us the opportunity to discuss the importance of the time-scale control revealed in our work, between the write-in (input layer) of the information to be processed, the intrinsic reservoir (internal layer), and the read-out (output layer). Section IV C is devoted to additional experimental and numerical speech-recognition results obtained from an extended complexity speech-recognition database, namely, the one used for the standard AURORA-2 benchmark. The last section summarizes the results and gives a perspective for future work for practical photonic RC processors.

II. RC CONCEPTS AND DELAY DYNAMICS

Until very recently, RC was essentially developed by the neural-network computing, brain cognitive sciences, and machine learning communities. Since the recent demonstration of its hardware implementation possibility, nonlinear physics and nonlinear dynamics communities have also started to address this topic. In order to provide nonlinear physics insights about the underlying concepts to a broad audience, hereafter we propose a qualitative description of the RC approach in the particular context of its implementation with high-dimensional nonlinear delay dynamics. The latter implementation unconventionally offers the possibility to virtually emulate, in the time domain only, the usual high-dimensional dynamics formed by a spatiotemporal network of coupled dynamical nodes.

A. Standard RC model

An important computational concept used in RC is the nonlinear dynamical expansion of the information to be processed into a higher-dimensional phase space, such that easy linear read-out of this expansion can be efficiently applied (for a review of the concepts briefly recalled here, see Ref. [4]). Figure 1 shows how the information is encoded and injected (write-in step, also called the input layer) into a nonlinear dynamical system (e.g., a network of firing neurons). A transient motion is thus triggered in the

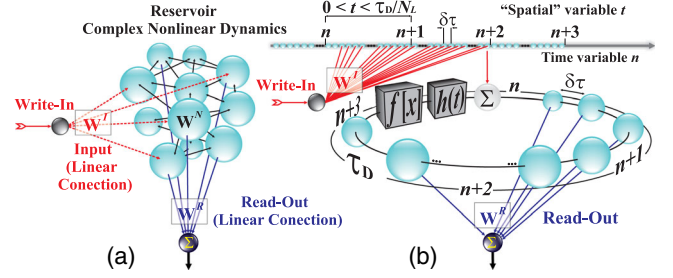


FIG. 1. Principles of RC, with an input mask \mathbf{W}^I spreading the input information onto the RC nodes, and with a read-out \mathbf{W}^R extracting the computed output from the node states. Left diagram: A spatially extended dynamical network of nodes. Right diagram: A nonlinear delayed feedback dynamics emulating virtual nodes which are addressed via time multiplexing. Here, $f(x)$ stands for the nonlinear feedback transformation, and $h(t)$ denotes the loop linear impulse response.

dynamical reservoir, describing a complex trajectory in its high-dimensional phase space. A linear read-out (output layer) operation is then performed on this transient motion so that the expected solution can be extracted and computed. This linear read-out comes conceptually to find, within the reservoir phase space, a hyperplane whose location is characteristic of the problem to be computed. Finding this hyperplane is the goal of the learning procedure, which is to extract the optimized quantities characterizing the hyperplane equation, i.e., the coefficients of a linear combination operated on the phase-space coordinates during the transient dynamics.

Considering a classical network formed by K spatially distributed nodes $\mathbf{x}(n) = [x_1(n) \dots x_k(n) \dots x_K(n)]^T$, the discrete time dynamics of the network (as time n is increased) involves the coupling of each node with the other ones according to the network coupling matrix \mathbf{W}^N . In addition to the internal dynamics, the network also evolves because of the injection of the input information $\mathbf{u}(n)$ to be processed. The information injection is ruled by an input layer coupling matrix \mathbf{W}^I . This results in a network dynamics for each node amplitude x_k at time n , which can be expressed as follows:

$$x_k(n) = f_{\text{NL}} \left[\sum_{j=1}^K w_{kj}^N x_j(n-1) + \sum_{q=1}^Q w_{kq}^I u_q(n) \right], \quad (1)$$

where f_{NL} is the nonlinear node sensitivity on the accumulated stimuli coming from both the network internal connectivity and the input information connectivity. A popular form for f_{NL} is the tanh function, also known as the sigmoid function. With this formalism, the read-out operation extracting the computed M -dimensional solution signal $\mathbf{y}(n)$ consists in a linear combination of the node states:

$$y_m(n) = \sum_{k=1}^K w_{mk}^R x_k(n). \quad (2)$$

Contrary to standard recurrent neural-network (RNN) computing approaches (with which learning can be very difficult, time consuming, and not necessarily converging), learning in RC is typically obtained from a very fast regression operation that is always converging. For RC, this learning procedure concerns the read-out only, whereas in RNN, it generally aims at also optimizing the coupling coefficients of many other layers, i.e., the input and output layers, as well as the internal connectivity of the network. This strong difference makes RC learning fast and reliable. Surprisingly, it was also found to be computationally very efficient in many benchmark tasks—comparable to and, in some cases, even surpassing RNN techniques [4].

B. RC based on delay differential dynamics

Recently [5], an unconventional hardware implementation was successfully proposed as the dynamical system of RC: Instead of the standard spatiotemporal structure consisting in a dynamical network of interconnected nodes, the use of nonlinear delay dynamics was proposed. It is an attractive hardware solution, especially when photonic implementation is concerned, for which high-speed optical telecom devices can potentially provide unprecedented processing speed [6–9]. Delay dynamics are indeed known as having an infinite-dimensional phase space. To illustrate this, let us consider the simplest scalar model for delay systems, of the following form:

$$\tau \dot{x}(t) = -x(t) + f_{\text{NL}}[x(t - \tau_D)], \quad (3)$$

where τ_D is the delay and τ the characteristic response time limiting the fastest time scales allowed in the oscillation loop. In the common “large delay” case, one has $\tau \ll \tau_D$, which is well known for enabling the emergence of high-dimensional chaotic attractors [12]. The infinite-dimensional character of such a delay dynamical system can be straightforwardly explained by the kind of initial conditions actually required for the unique definition of a given trajectory: It consists of a functional $x(t)$, with $t \in [-\tau_D; 0]$, belonging to an infinite-dimensional space. In the framework of RC, the expectation is that, compared to the classical case of neural networks, information injection into such delay dynamics can provide similar expansion properties, also allowing for computational capabilities. Conceptually and mathematically, this emulation of delay dynamics into a spatiotemporal dynamics has indeed been known for more than two decades [13]. The same emulation concept was also recently used by our group to demonstrate the existence of chimera states (complex spatiotemporal patterns) in delay dynamics [14]. This emulation makes use of a virtual space variable

corresponding to a short time scale σ of the dynamics [of the order of τ in Eq. (3)], whereas the long time scale of the order of the delay τ_D defines a discrete time evolution of the functional from one time-delay interval to the next, i.e., $\{x_\sigma(n) = x(t)|t = n\tau_D + \sigma\}$ with $n \in \mathbb{N}$ and $\sigma \in [0; \tau_D]$. With this in mind, it appears more convenient to rewrite the usual delay differential equation (DDE) (3) into an integral convolution form involving the impulse response $h(t)$, the latter characterizing the linear differential process ruling the dynamics in the delayed feedback loop:

$$x(t) = \int_{-\infty}^t h(t - \xi) f_{\text{NL}}[x(\xi - \tau_D)] d\xi. \quad (4)$$

The previous linear impulse response $h(t)$ takes the usual form $h(t) = (1/\tau)e^{-t/\tau}$ for $t \geq 0$ (and 0 for $t < 0$ because of causality) in the particular case of Eq. (3). Note that here Eq. (4) is indeed more general than Eq. (3) since it describes any kind of linear filtering in the delay oscillation loop not restricted to the case of a first-order differential process as in Eq. (3).

From Eq. (4), it can then be shown that delay dynamics is also described by a more straightforward discrete time map for the functional, thus highlighting the spatiotemporal analogy:

$$x_k(n) = x_k(n-1) + \int_{\sigma_k - \tau_D}^{\sigma_k} h(\sigma_k - \sigma) f_{\text{NL}}[x_\sigma(n-1)] d\sigma, \quad (5)$$

where $x_k(n) = x[t(n, \sigma_k)]$, with $t(n, \sigma_k) = n\tau_D + \sigma_k$, points to the multiple scale features of such dynamics (a short time σ of the order of τ , and a long time n as an integer multiple of τ_D). Such a multiple time-scale description precisely highlights the space-time analogy with the standard discrete time RC model in Eq. (1). We recover the discrete time dynamics as an iteration of the functional from $(n-1)$ to n , and the usual spatial coupling in the network of neurons is now revealed by the impulse response $h(t)$ in the integral term. The latter fulfills the role of the coupling (matrix \mathbf{W}^N) between the nodes, within the continuous virtual space in σ . The temporal width of the impulse response is representative of the typical coupling distance involved in the usual network of neurons.

However, Eqs. (3)–(5) correspond to a so-called autonomous dynamics. RC is a computing paradigm that deals with the processing of time-dependent information through transient dynamics. Thus, it is implicitly a nonautonomous system. Typically, the information to be processed is mixed within the delayed feedback loop by adding the encoded input information to the feedback signal $x(t)$. The two contributions together then form the argument of the nonlinear transformation $f_{\text{NL}}[\cdot]$.

The information injected into the virtual space σ of delay dynamics is achieved through a time division multiplexing

(TDM) technique, for which the virtual node position $k \in [1; K]$ is defined as a time position $\sigma_k = (k-1)\tau_D/K$ within a time-delay interval. This viewpoint defines a constant node spacing $\delta\tau = \tau_D/K$ along the time-delay interval [5] for all the K nodes of the emulated network. Similarly to the information contribution in Eq. (1), a given virtual node $x_k(n)$ receives an input information signal contribution (write-in) which can be defined as $\sum_q w_{kq}^I u_q(n) \delta(t - n\tau_D - \sigma_k)$, where $\delta(t) = 1$ at $t = 0$ and it equals zero at any other time. In the particular case of photonic delay dynamics as reported in this article, the unit pulse $\delta(t)$ is replaced by a physically more realistic rectangular temporal window $p_{\delta\tau}(t) = 1$ for $t \in [0; \delta\tau]$ and zero elsewhere (also known as an “order zero” sample and hold operation). This results in a dynamical law for the transient according to the following equation [to be identified with Eq. (1)]:

$$x_k(n) = x_k(n-1) + \int_{\sigma_k - \tau_D}^{\sigma_k} h(\sigma - \sigma_k) \times f_{\text{NL}}[x_\sigma(n-1) + \rho \cdot u_\sigma^I(n-1)] d\sigma, \quad (6)$$

where the input signal $u_\sigma^I(n)$ is a stepwise constant function defined according to the TDM principle performing the distribution of each input information sample $\mathbf{u}(n)$ over each of the virtual nodes defined within a time-delay interval:

$$u_\sigma^I(n) = \sum_{k=1}^K \left[\sum_{q=1}^Q w_{kq}^I u_q(n) \right] p_{\delta\tau}(\sigma - \sigma_k). \quad (7)$$

In Eq. (6), we note the important incremental contribution to the theoretical interpretation of delay-based RC systems. Writing a convolution product for the reservoir dynamics indeed provides a more accurate description compared to previous work. In the past, analytical descriptions of the virtual spatial coupling in delay dynamics were presented either through the approximation provided by a Euler discrete integration time step of the continuous dynamics [5] or, roughly, by neglecting the temporal nearest-neighbor coupling provided by the impulse response function (coupling was then induced through the hypothesis of an asynchronous configuration of the cyclic injected TDM samples with respect to the delay [7,8]). Equation (6) provides a rigorous analytical description of the delay dynamics in a mathematical form that exhibits a close analogy to the original ESN model (recalled in Sec. II A).

In the delay-based RC demonstrations proposed so far, it was generally assumed that each input sample $\mathbf{u}(n)$ was spread through the input layer (input connectivity matrix \mathbf{W}^I) over the full time-delay duration τ_D . One goal of this work is to report on the possibility to extend the internal memory of the delay dynamics compared to the temporal features of the input information. Practically, this offers

enhanced correlation-length capabilities (or “memory”) for the reservoir, up to several time delays. Instead of a 1:1 spanning of each input sample over all the K nodes of a time delay, such an extended delay memory (EDM) is emulated through an $N_L:1$ ratio defined as the number N_L of input samples spanned over one time delay. For a fixed node distance $\delta\tau$ and a fixed number of virtual nodes, this implies that the time delay has to be increased by a factor N_L , thus leading to the temporal parameter scaling equation $\tau_D = N_L K \delta\tau$. These quantities rule the node spacing, the number of emulated nodes, the number of internal layers, and the time delay. The consequence is that the discrete time dynamics previously derived in Eq. (6) will develop a temporal mixing due to the delayed feedback loop, now between input samples $\mathbf{u}(n)$ and $\mathbf{u}(n + N_L)$ instead of $\mathbf{u}(n)$ and $\mathbf{u}(n + 1)$ (see Fig. 1, where $N_L = 3$). Later, we will discuss some issues related to this EDM concept for delay dynamics.

Compared to the single layer essentially adopted so far in the literature and for which the former theoretical description was detailed, one needs to slightly redefine some of the equations of the RC processing model:

- (i) The number of virtual nodes (per layer) is still defined as K , which is imposed by the number of columns of the input connectivity matrix \mathbf{W}^I performing (through TDM) the spreading of the input information over one single layer.
- (ii) The recurrence time scale defined by the time delay τ_D now covers N_L successive reservoir vectors $\{x_k(n), x_k(n+1), \dots, x_k(n+N_L-1) | k = 1, \dots, K\}$, instead of a single reservoir vector $\{x_k(n) | k = 1, \dots, K\}$. The reservoir vector with K nodes then corresponds to the following definition: $x_k(n) = x(t)$ with $t = n(\tau_D/N_L) + \sigma_k$ and $\sigma_k = (k-1)\tau_D/(N_L K) \in [0; \tau_D/N_L]$.
- (iii) The delay-reservoir dynamics thus has to be rewritten accordingly:

$$x_k(n) = x_k(n - N_L) + \int_{\sigma_k - \tau_D}^{\sigma_k} h(\sigma - \sigma_k) \times f_{\text{NL}}[x_\sigma(n - N_L) + \rho \cdot u_\sigma^I(n - N_L)] d\sigma. \quad (8)$$

In addition to the photonic architecture and to the EDM concept, we finally propose a third original contribution to the field of delay-based photonic RC. This contribution provides a way for the read-out to be temporally designed. We have seen that a relative timing is involved between the intrinsic time scale of the reservoir (and its internal time constants τ, τ_D, \dots) and the clock used to inject the information to be processed according to the TDM principle (node separation $\delta\tau$, node numbers K , EDM scaling N_L). In previous works, the read-out operation was always assumed to be necessarily clocked synchronously with the clock already imposed at the write-in [rhythm at the

information injection, fixing the node separation through the input sampling rate $(\delta\tau)^{-1}$. The read-out is, however, an operation that is fully independent from any previous RC information processing steps, according to the unidirectional information flow from the input layer through the reservoir layer and to the read-out layer. The read-out $\mathbf{y}(n)$ can thus be expressed either in the same way as in Eq. (2) if a synchronous write-in vs read-out sampling is adopted (same node positions σ_k), or one can define asynchronous node positions σ_k^R for the read-out:

$$y_m(n) = \sum_{k=1}^K w_{mk}^R x\left(n \frac{\tau_D}{N_L} + \sigma_k^R\right), \quad \text{with}$$

$$\sigma_k^R \in \left[0; \frac{\tau_D}{N_L}\right]. \quad (9)$$

We thus explored the influence of a slightly different read-out sampling period $\delta\tau^R$ compared to the one at the write-in $\delta\tau$, $\delta\tau^R = (1 + \varepsilon)\delta\tau$. Subsequently, we observed, for a specific relative desynchronization value ε , a significantly improved RC processing efficiency by more than 1 order of magnitude. Such an improvement was confirmed both in numerics and experiments.

III. EXPERIMENTAL SETUP

Experimental nonlinear delay photonic systems intended for signal or information processing have already been studied during the last two decades—however, for totally different applications such as ultrahigh spectral purity microwave generation [15] or physical layer optical encryption using broadband chaotic waveforms [16]. Among the different setups reported in the literature, our group contributed to several versions of the so-called optoelectronic chaos oscillator [17]. The latest version was especially designed in the framework of 10 Gb/s optical chaos communication, and it succeeded in establishing, through the realistic conditions of a field experiment, the state of the art in speed and transmission quality thanks to highly controllable differential phase shift keying (DPSK) optical communication techniques [18]. Beyond the operational chaos communication demonstration, this so-called “phase chaos setup” is characterized by temporally nonlocal delayed feedback, which provided both additional virtual space-time coupling features and novel bifurcation phenomena [19,20]. The phase chaos setup with its bandwidth capability and high controllability thus appeared to be of great potential interest in the framework of RC photonic processing as well.

A. EO phase-delay dynamics

The setup in which RC is performed through the transient motion of an electro-optically modulated phase of a laser beam is depicted in Fig. 2. The delayed feedback

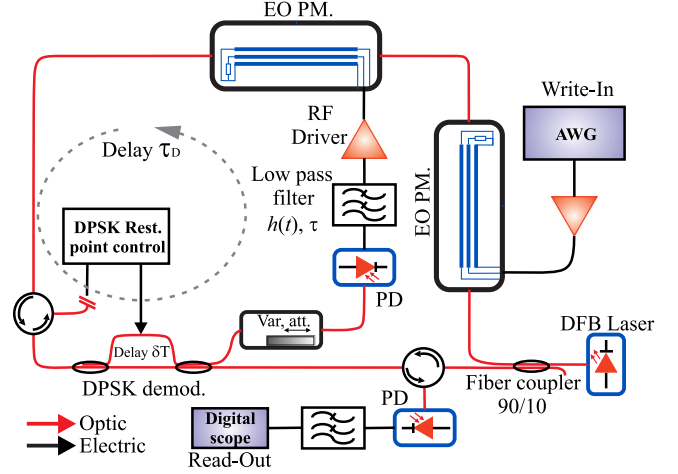


FIG. 2. EO phase setup involving two integrated optic phase modulators followed by an imbalanced Mach-Zehnder DPSK demodulator providing a temporally nonlocal, nonlinear, phase-to-intensity conversion. The information to be processed by this delay photonic reservoir is provided by a high-speed arbitrary waveform generator (AWG). The response signal from the delay dynamics is recorded by an ultrafast real-time digital oscilloscope at the bottom of the setup, after the circulator, followed by an amplified photodiode and a filter.

loop is inspired by the standard optoelectronic chaos architecture [17], which provides the “recurrent” character of the dynamics. This oscillator makes use of two EO phase modulators (PM), where electrical to optical phase conversion efficiency is denoted V_π (also technically referred to as half-wave voltage). The first EO PM allows us to imprint the initial electrical domain information $u_\sigma^l(n)$ onto the optical phase of the laser light beam carrier, whereas the second EO PM combines information for the nonlinear delayed recurrent feedback $x_\sigma(n)$ additively to this input. The resulting total light phase modulation is $\varphi(t) = x_\sigma(n) + \rho u_\sigma^l(n)$, where ρ is an amplification factor scaling the information signal compared to the feedback one. The recurrent dynamics of the RC concept is provided by the delayed optoelectronic feedback loop. This loop consists of the following:

- (i) A significant time delay τ_D issued from the serial combination of optical fiber pigtailed and small electrical links connecting the different electronic, optoelectronic, and EO devices of the loop. This results in a total closed-loop delay measured at $\tau_D \approx 63.33$ ns.
- (ii) A fiber-based passive imbalanced Mach-Zehnder interferometer, which actually consists in a conventional DPSK telecom demodulator, with a time imbalance $\delta T \approx 402.68$ ps. This demodulator converts the input phase modulation $\varphi(t)$ nonlinearly into an output intensity modulation, according to the standard two-wave intensity interference function (with an offset shift ensuring $f_{NL}[0] = 0$, and with a gain factor β):

$$f_{\text{NL}}[\varphi] = \beta \{ \cos^2[\varphi(t) - \varphi(t - \delta T) + \Phi_0] - \cos^2 \Phi_0 \}, \quad (10)$$

in which the imbalance δT is much greater than the fastest time scales present in $\varphi(t)$. In the recurrent delayed dynamics, this provides a strong enough nonlinear transformation through the modulation of the interference condition at the DPSK output. Compared to previous photonic RC, the phase chaos setup not only performs the same neighboring node coupling through τ and thus through $h(t)$ as explained by Eq. (6), it also introduces an additional “nonlocal” coupling between farther nodes separated by the DPSK time imbalance $\delta T \gg \delta\tau$. It is worth noticing the role of the offset phase $\Phi_0 = \omega_0 \delta T$, which is to fix the shape of the nonlinear function profile around which the nonlinear transformation is operated. This actual shape can be set to linear or quadratic for $\Phi_0 = \pi/4$ or 0, modulo $\pi/2$, respectively. For $\Phi_0 = \pi/4$ and a strong enough phase modulation span (parameter ρ), the actual shape can even be of cubic type. The parameter Φ_0 is practically adjusted either through the central wavelength of the laser $\lambda_0 = 2\pi c/\omega_0$ or through a fine-tuning of the time imbalance. Fiber-based DPSK are indeed usually equipped with a heating wire rolled around one interferometer arm, thus allowing us to finely and actively control the absolute differential phase shift, i.e., the DPSK rest point. Such an active control is implemented in our setup, using the backward optical path in the DPSK demodulator and injecting, along this path, the same but unmodulated laser beam extracted from a 10/90 fiber coupler as shown in Fig. 2. Two optical circulators are used to separate and combine the forward modulated light beam and the backward unmodulated one.

The parameter β is a normalized weighting factor for the nonlinear delayed feedback; it scales linearly with the different electro-optic and optoelectronic conversion efficiency (V_π and S), as well as with the electronic gain G and the forward path optical power κP_0 . The latter power is practically tuned through a variable optical attenuator (attenuation factor κ) in order to experimentally adjust β to the optimal value determined from numerical simulations.

- (iii) Telecom-amplified photodiodes with sensitivity S are used to convert the intensity fluctuations at the DPSK demodulator output into electronic signals, either for the delayed optoelectronic feedback applied to the second EO PM or to record the signal $x_\sigma(n)$. The latter signal represents the reservoir response to the transient that is triggered by the input information, and it is recorded by a digital scope for further off-line processing corresponding to learning and testing of the read-out layer.

- (iv) A 566-MHz cutoff low-pass filter is used to “slow down” the internal delay dynamics time scale, resulting in a characteristic response time $\tau \approx 284$ ps; in other words, it results in the definition of the impulse response $h(t)$ involved in Eqs. (4)–(6). The motivation for this filter, which slows down the actually available 10-GHz bandwidth of the unfiltered optoelectronic feedback loop, is governed on one side by empirically determined [5] optimal distance between the virtual nodes of a delay reservoir ($\delta\tau = \tau/5 \approx 56.8$ ps), and on the other side by the maximum sampling frequency of our AWG (Tektronix 7122C) up to 24 GS/s. The input sampling rate $(\delta\tau)^{-1} = 17.6$ GHz deduced from one-fifth of the low-pass filter characteristic time scale indeed fulfilled these theoretical and technical scaling requirements.

Motivated by a qualitative comparison of performances with respect to the existing literature, we tried to adjust the number of EDM within the time delay so that our experiment also involves a number of nodes per layer close to the 400 most often used in the literature [5,6,9]. The time delay $\tau_D = 63.33$ ns, resulting in a total number of nodes $\tau_D/\delta\tau = 5\tau_D/\tau \approx 1113$; we chose $L = 3$ as the EDM factor, thus leading to $K = 371$ nodes per layer.

- (v) A broadband telecom RF driver with gain G is used for amplification before applying the nonlinear delayed filtered feedback to the second EO PM. A similar driver is also used between the AWG and the first EO PM to ensure a strong enough modulation span [weight ρ in Eq. (6), practically set through the adjustable output level of the AWG].

The previous nonlinear delay optoelectronic phase oscillator can typically be modeled by an integro-differential delay equation [19,21], with a nonlocal nonlinearity, and with a driving term corresponding to the information to be processed. This results in a nonautonomous DDE system modeled by two coupled equations:

$$\begin{aligned} \tau \frac{dx}{dt}(t) &= -x(t) + \frac{1}{\theta} y(t) + f_{\text{NL}}[\varphi(t - \tau_D)], \\ \frac{dy}{dt}(t) &= x(t), \end{aligned} \quad (11)$$

where $\varphi(t) = x_\sigma(n) + \rho \cdot u_\sigma^l(n)$ is a phase modulation issued from both first and second EO PM (for write-in and delayed feedback, respectively). The supplementary variable $y(t)$ originates from the necessary bandpass filtering feature performed in any broadband optoelectronic amplified process (low cutoff frequency of ca. 50 kHz, resulting in $\theta \approx 3$ μ s). In addition to the non-local character of δT , this bandpass feature is also unusual in the common delay dynamical systems such

as the ones modeled by Eq. (3). Such a bandpass feature is a source of supplementary dynamical complexity, as illustrated by the many new interesting fundamental dynamical properties they allow for, such as chaotic breathers [22], chimera states [14], or period-one stable asymmetric oscillations [23]. In the context of RC, more complexity is also expected to provide enhanced processing capabilities.

B. Role and setting of the delay dynamics parameters

The previously described physical parameters are directly accessible in experiment, so they can be adjusted to optimize the performance of the RC processor. This optimization concerns the sizing of the virtual network such as the node separation $\delta\tau$, the number K of such nodes per virtual network layer, and the number of these layers, N_L , allowing for the EDM feature.

The node distance $\delta\tau$ within a layer, according to the discrete time dynamics described in Eq. (5), can be viewed as directly related to the local span of the impulse response $h(t)$ defined by the linear filter limiting the feedback bandwidth, and thus it is ruled by the characteristic time scale τ of this filter. This makes a connection between node separation and filter bandwidth, as already mentioned through an empirically defined optimal node distance of ca. $\tau/5$. However, similarly to what was introduced in Ref. [24] with a multiple delayed feedback, here the nonlocal character of the nonlinear transformation adds another contribution to the virtual network connectivity with an additional coupling distance related to δT .

Beyond the structural definition of the virtual network (size and spatial coupling), one also has to set the (discrete time) dynamical properties of the emulated network, which are indirectly set by the amplitude parameters of the delayed feedback dynamic. In conventional network-based RC, these properties are essentially defined by two features: the so-called spectral radius of the network, which can also be viewed as a bifurcation parameter for the corresponding autonomous network, and the nonlinear function ruling the contribution of the other nodes, when one node has its states updated from n to $(n+1)$.

The spectral radius typically has to be set below unity in order to ensure a stable steady state in an autonomous regime (i.e., when no external drive is applied or, in other words, when the reservoir has no information to process). A value above unity would indeed correspond to the emergence of a self-oscillation. Such auto-oscillations are considered to prevent an important required property for reservoir dynamics, referred to as the fading memory property [3]. In the case of a differential delay dynamics [5,6,9], the limit for the stability of a steady state typically occurs at a Hopf bifurcation usually controlled through the linear gain of the delayed feedback. Applying this criteria to our EO phase-delay dynamics, one obtains

$$\left| \left[\frac{df_{\text{NL}}}{d\varphi} \right]_{\varphi=0} \right| = |2\beta \sin(2\Phi_0)| < 1. \quad (12)$$

The strength of the fading memory property then directly depends on the strength of the recurrence in the network, which recurrence strength is tuned by the spectral radius in the standard network reservoir, or the linear feedback gain for a delay reservoir. For a more quantitative measure of the recurrence-induced memory, one has to calculate the real part of the eigenvalues for the linearized dynamics around the stable steady state of concern: The (small-amplitude) transient duration scales as its inverse. For very low feedback levels [or for low recurrence strength, $f'_{\text{NL}}(0) \rightarrow 0$], the memory is very small. Contrarily, if the feedback is strong enough to get close to the bifurcation threshold [$|f'_{\text{NL}}(0)| \rightarrow 1^-$], the memory can grow to infinity, and the transient cannot fade fast enough to properly process the information according to the RC concepts. However, such feedback level tuning criteria have to be considered as qualitative ones only since they refer to a small amplitude linear approximation. Under realistic operations, this approximation is poorly valid since large amplitude nonlinear transients are actually triggered. A finer setting of the feedback strength was practically obtained through exhaustive numerical simulations under many different parameter conditions ($\beta, \Phi_0, \rho, \dots$). It is worth noticing that optimal values are generally not universal for a delay reservoir, but they have different optimal settings depending on the particular task to be performed (classification, prediction, ...). However, for more complex delay reservoir topologies, it was recently found [25] that “universal” computational power, i.e., without the requirement for task-dependent parameter optimization, can be obtained. Work is currently in progress to experimentally confirm this theoretical result.

Beyond its influence on the linear stability condition as expressed in Eq. (12), the parameter Φ_0 also determines the actual large amplitude shape of the nonlinear transformation f_{NL} for the discrete time dynamics as described by Eq. (6). The role of this function shape can be compared to the traditional sigmoid (or tanh) nonlinear function often used in standard discrete-time RNN computing. However, depending on the value of the parameter Φ_0 , our setup provides a few technical differences. As already stated, the choice of Φ_0 and ρ allows us to choose different polynomial approximations for this nonlinear function: $\Phi_0 \simeq 0$ modulo $\pi/2$ corresponds to the scan of the DPSK interferometer around constructive or destructive interferences, thus leading to a nearly quadratic nonlinearity; $\Phi_0 = \pm\pi/4$ modulo $\pi/2$ sets the average operating point in the middle, between a constructive and a destructive interference, resulting in either a quasilinear operation (small ρ values) or a cubic operation (for large enough ρ). If relevant, any intermediate situation for Φ_0 can of course be chosen, in principle. For the classification task of concern in this paper, we found

(consistently with Refs. [6,9]) that a nearly quadratic operation corresponds to the most efficient computational performance. Contrarily, note that, for prediction tasks, a quasilinear operating point was found to be optimal in the literature [6,9].

Because the different parameter settings lead to inter-dependent RC properties, and in the absence of a global self-consistent theory for the optimization of RC, intensive numerical simulations based on Eq. (11) are required to identify the best operating conditions for a given task. In our work, we have focused on two benchmark classification tasks for spoken digit recognition, which allowed us to identify and confirm the qualitative optimization criteria for ultrafast photonic delay setup, as they were described above. However, a successful mathematical optimization theory was recently proposed in single delay nonlinear dynamics, in the case of a prediction task [26].

IV. TESTING ELECTRO-OPTIC PHASE RC WITH SPEECH RECOGNITION

Up to now, most RC processing efficiency tests have been performed through algorithmic implementations of the concept. Such tests thus involve hardware whose capabilities are necessarily limited by the ones of conventional Turing–von Neumann machines (standard digital computers), the same machines that RC approaches are actually aimed to replace. Dedicated hardware can actually open the path to ground-breaking efficient computing machines, providing the expected gain in speed and energy efficiency as an attractive perspective compared to the current limitations. In this work, we aim to provide a step toward this objective by proposing ultrafast photonic hardware.

In addition to the hardware feasibility of a new concept, it is also important to evaluate how such a brain-inspired analog computing approach can offer hardware capabilities in terms of computational power. To evaluate such computational performances, we have implemented classification tasks consisting of speech recognition. Two different speech databases have been used for testing. The first one corresponds to a test already widely used in several recent photonic RC demonstrations—the TI46 test [5–9, 24,27,28]. This first test is important to compare with already available hardware RC results in the literature. The second test was also addressed in this work, based on a very similar task—however, with an extended database for which the classification problem is more realistic and thus also more complex.

A. Database from the TI46 test

1. Preprocessing: 2D input from the Lyon ear model

Spoken digit recognition tasks are standard classification tests, which have been widely used already to evaluate the processing efficiency of RC. Such a test was introduced in

Ref. [28] and reused in several recent RC hardware evaluation under the same condition; it consists in the recognition of pronounced digits between 0 and 9, with the database corresponding to 500 digits extracted from the TI46 speech corpus [29]. The digits are pronounced by five different female speakers uttering the 10 digits 10 times, with the acoustic waveform being digitally recorded at a sampling rate of 12.5 kHz.

Following the usually adopted preprocessing for such kinds of speech recognition tasks, we applied the so-called Lyon ear model before running the RC processing as described previously. The Lyon ear model preprocessing is motivated by a physiologically relevant transformation as identified in the human ear. It converts the 1D sound waveform (pressure variations over time) into a 2D frequency-time matrix called a cochleagram: This graph represents the time evolution, during the pronunciation, of the Fourier spectral power distribution over $Q = 86$ frequency channels. The 2D input information consists in a cochleagram whose columns form the previously introduced input information signal $\mathbf{u}(n) = [u_1(n) \dots u_q(n) \dots u_Q(n)]^T$: The successive columns of the cochleagram matrix \mathbf{M}_u are indexed by the discrete time n . Depending on the duration of the pronounced digit, the number N varies from 32 to 130, with an average of ca. 60.

2. Write-in signal injected into the delay RC

For RC processing as described in Sec. II, a connectivity matrix \mathbf{W}^I is used to spread each of these input samples $\mathbf{u}(n)$ onto each of the virtual nodes (position σ_k) defined within one layer of the delay dynamics. The scalar input waveform $u_{\sigma}^I(n)$ appears as the queued columns of the matrix product $\mathbf{M}_{\text{in}} = \mathbf{W}^I \times \mathbf{M}_u$ (see Fig. 3). According to the matrix multiplication and consistently with Eq. (7), the single-layer signal $u_{\sigma_k}^I(n) | k \in [1; K]$ can be interpreted as a set of K linear combinations of the Q different Fourier frequency components occurring at time n . Each node k of the network is thus excited by a different linear

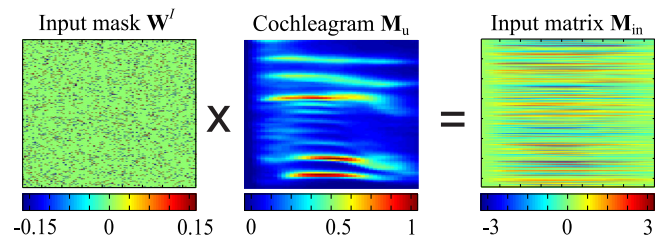


FIG. 3. Illustration of the input information injection into the dynamics. The (sparse and random) $K \times Q$ write-in matrix \mathbf{W}^I performs a spreading of the input cochleagram information represented as a $Q \times N$ cochleagram matrix \mathbf{M}_u . The resulting $K \times N$ matrix \mathbf{M}_{in} defines a scalar temporal waveform $u_{\sigma}^I(n)$ obtained after horizontally queuing the N columns, each of them being formed by the K amplitudes addressing the virtual nodes in one layer.

combination whose coefficients are defined by the nonzero elements in the rows of the connectivity matrix $\mathbf{W}^I = [w_{kg}^I]$. From numerical simulations, it turns out that, for this particular classification task, the optimal connectivity matrix is obtained for a sparsity of ca. 0.1, assuming a random setting of its nonzero elements within the binary set $[-1; 1]$. For scaling issues and consistency over various numerical simulations, the input signal amplitude $u_\sigma^I(n)$ is normalized in such a way that its maximum peak-to-peak deviation for all 500 digits equals unity. The scaling factor ρ was then considered as a normalized amplitude span, independently from the random choice of the connectivity matrix.

The signal $\rho u_\sigma^I(n)$ is programmed in the AWG, which technically seeds the RF driver modulating the first EO PM, at the calculated optimal sampling rate of 17.6 GS/s. The parameter ρ is set through the adjustable amplitude level of the AWG output.

3. Transient response of the delay dynamics

As described previously, $N_L = 3$ EDM layers are introduced in this work. This implies that three successive input samples $u_\sigma^I(n)$ indexed from n to $n + 2$ are needed to fill the delay line, instead of a single one as practiced in previous experiments. Each of these samples for a given n is also spanned over $\tau_D/3$ according to the input connectivity, leading to a time division multiplexing of the input information over $K = 371$ virtual nodes covering the duration $\tau_D/3$, according to Eq. (7). Here, the node separation is imposed by the AWG sampling rate set to 17.6 GHz, theoretically defining the node spacing $\delta\tau = 57.8$ ps. Note that there actually exists a difference between the stepwise constant input as described in Eq. (7) and the physical light phase modulation. Indeed, the AWG has a 24 GS/s maximum sampling rate, but it also has a limited analog bandwidth of 9 GHz, thus preventing it from generating a stepwise constant amplitude since the harmonics, and even the fundamental frequency of the AWG, are necessary filtered out by the AWG analog bandwidth. The actual phase modulation ρu_σ^I is thus a smooth waveform of time. Nevertheless, there is no theoretical requirement for the stepwise constant profile of the input signal, the motivation here being essentially to keep the reservoir dynamics in the permanent transient regime all along the duration of the input information. The latter condition is indeed also fulfilled even with a smooth input waveform, as long as its frequency content is fast enough compared to the characteristic time scale of the delay dynamics.

Each response of the delay dynamics to each of the 500 encoded input spoken digit signals $\{u_\sigma^I(n)|n \in [1; N]\}$ is recorded by a real-time ultrafast digital scope (Lecroy WaveMaster 845Zi-A, 30-GHz analog bandwidth, 80 GS/s) as depicted in the bottom of Fig. 2, with an oversampling allowing for further postprocessing [e.g.,

averaging, or selection of interpolated samples according to the already mentioned asynchronous read-out sampling in Eq. (9), i.e., $\delta\tau^R = (1 + \varepsilon)\delta\tau$. Each sampled response signal $\{x_{\sigma_k^R}(n)|n \in [1; N]\}$ is conditioned numerically in the form of a $K \times N$ matrix \mathbf{M}_x , thus in a similar form compared to the previously introduced input signal matrix \mathbf{M}_{in} for $u_{\sigma_k}(n)$. Processing all the digits of the database thus leads to 500 of such response matrices \mathbf{M}_x .

4. Learning the optimal read-out matrix \mathbf{W}^R

A given response matrix \mathbf{M}_x is representative of the dynamical processing performed by the EO phase-delay dynamics. It consists in the sampling of the transient trajectory triggered by one input signal, with this trajectory taking place in the infinite-dimensional phase space of the delay dynamics. According to the concept of RC, one then needs to learn the coefficients $[w_{mk}^R]$ of a hyperplane in this phase space, which is expected to have a characteristic role for the problem presented to the RC processor. For the particular case of a classification problem as for the spoken digit recognition task, we expect to learn $M = 10$ such sets of coefficients, i.e., one for each of the 10 possible modalities (the 10 digits from zero to nine). The $M \times K$ read-out matrix \mathbf{W}^R can be constructed such that the matrix product $\mathbf{W}^R \cdot \mathbf{M}_x$ results in a so-called target matrix \mathbf{M}_y with dimensions $M \times N$ (see Fig. 4). The correct digit is then expected to be easily identified from the content of \mathbf{M}_y . A simple form for the construction of such a matrix \mathbf{M}_y is to set zeros on each $(M - 1)$ lines related to the wrong modality and to set ones on the line corresponding to the correct modality.

Practically, the learning requires the use of a subset of digits from the whole database of 500 spoken digits, the complementary part of this subset being used later for the testing of untaught digits. One of the most attractive features of RC is its easy and always converging learning technique to compute the optimal read-out matrix \mathbf{W}_{opt}^R . This is basically performed through a simple ridge regression minimizing the error for all possible \mathbf{W}^R matrices in

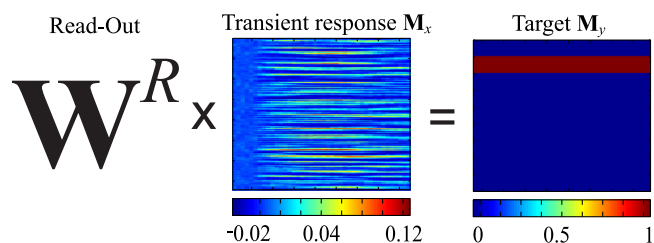


FIG. 4. Illustration of the expected optimized read-out processing through a $(M \times K)$ matrix \mathbf{W}^R , left multiplying the transient response $(K \times N)$ matrix \mathbf{M}_x , thus resulting in an easy-to-interpret target $(M \times N)$ matrix \mathbf{M}_y . The latter matrix is aimed at designating the right answer for the digit to be identified (the second line in this example, indicating digit “1”).

order to obtain the proper target for each digit of the training subset:

$$\mathbf{W}_{\text{opt}}^R = \underset{\mathbf{W}^R}{\text{argmin}} \|\mathbf{W}^R \cdot \mathbf{M}_x - \mathbf{M}_y\|^2 + \lambda \|\mathbf{W}^R\|^2, \quad (13)$$

where λ is the regression parameter to compensate for a necessarily ill-posed problem. The matrices \mathbf{M}_x and \mathbf{M}_y are the horizontally concatenated matrices gathering all the experimental response matrices of the training subset and all their corresponding targets, respectively. Practically, we randomly chose 475 digits for the training subset, with the 25 remaining ones being left for testing. This partitioning is essentially guided by the recent implementations of the same test, thus allowing for a more confident relative performance evaluation for the RC processing efficiency. Cross-validation was performed through the repetition of 20 learning procedures, according to the choice of 20 nonintersecting subsets of 25 testing digits, and their corresponding training subsets. This allowed us to obtain an average computational performance measure in which each of the 500 digits has been used once for testing.

The optimal read-out matrix was calculated off-line using the experimentally recorded transient response matrices \mathbf{M}_x . This calculation was done by a Matlab routine executing the ridge regression through a standard More-Penrose matrix inversion, according to the following expression:

$$\mathbf{W}_{\text{opt}}^R = \mathbf{M}_y \cdot \mathbf{M}_x^T (\mathbf{M}_x \mathbf{M}_x^T - \lambda \mathbf{I}_K)^{-1}, \quad (14)$$

where \mathbf{I}_K is the K -dimensional unity matrix, and $(\cdot)^T$ refers to the matrix transpose.

5. Testing and word error rate (WER) calculation

Testing is also done off-line (its physical implementation could, however, be realized on-line through the design of a standard analog signal correlator, thus preserving, for the whole testing step, the full analog processing speed of the photonic reservoir). Right after computing the optimal read-out matrix $\mathbf{W}_{\text{opt}}^R$, we evaluated the result of the RC processing on each of the testing digits by calculating $\tilde{\mathbf{M}}_y = \mathbf{W}_{\text{opt}}^R \mathbf{M}_x$. This consists of a set of standard circular correlation operations between each column of the read-out, as well as the reservoir response. The corresponding result of the matrix product is expected to resemble the target matrix \mathbf{M}_y . One naturally does not obtain a perfectly correct line with ones and wrong lines with zeros; however, integrating (summing) the elements of each line of $\tilde{\mathbf{M}}_y$ over n leads to a score for each modality (see Fig. 5). The RC processing is then simply determined by the modality showing the best score. This procedure is depicted in Fig. 6, which can be compared with the known answer in order to count the number of errors. A word error rate is

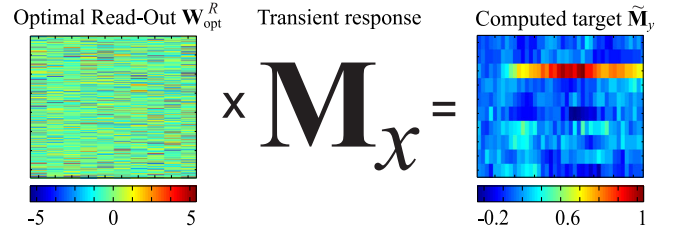


FIG. 5. Example of an imperfect “reservoir-computed” target answer while testing the optimal read-out $\mathbf{W}_{\text{opt}}^R$ on an untrained digit of response \mathbf{M}_x . However, the digit “2” clearly appears as the most obvious answer for this untrained tested digit.

finally obtained after 20-times repetition of learning and testing (cross-validation), thus giving a WER over the 500 digits of the original database.

B. Numerical and experimental results

As indicated previously, intensive numerical simulations of Eq. (11) have been performed in order to identify the optimal parameter conditions for best RC performances. These simulations allowed us to find qualitative criteria that are fully consistent with previously obtained results [5,6,9]. Quantitatively, excellent RC classification performances have been obtained both numerically and experimentally with WER close to 0%, thus confirming previous results from other photonic setups. Moreover, the EO phase-delay dynamics setup allowed for the use of relatively fast processing (though not yet optimized), reaching state-of-the-art classification speed: For an average duration of 60 samples for one spoken digit, the analog processing can be performed within $1.3 \mu\text{s}$, thus resulting in a processing speed of the order of 1 million words per second.

The numerical simulations used to set the experimental parameter have given the following optimized operating conditions:

- (i) Nearly quadratic nonlinearity is found to be optimal for this classification task; i.e., an offset phase Φ_0 close to 0 modulo $\pi/2$ is obtained numerically [see Fig. 7(a)]. It is worth noticing that an operation close

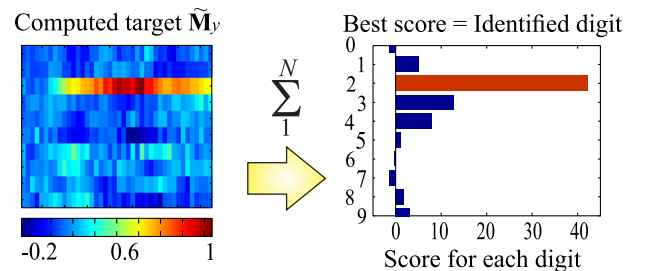


FIG. 6. Illustration of the decision procedure for the computed answer. The temporal amplitudes of the actual target are summed over time for each line (or modality), i.e., for each of the 10 possible digits. The right modality is then declared as the one with the highest sum.

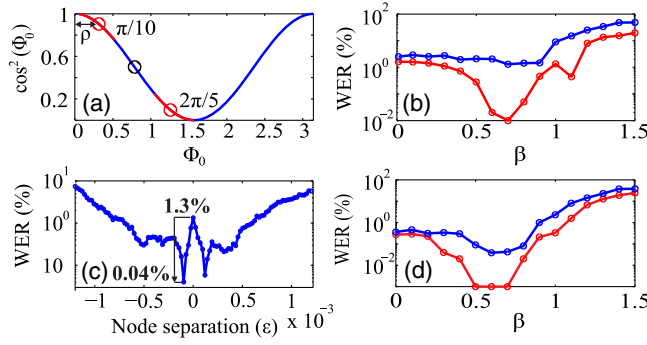


FIG. 7. Numerical and experimental results for the parameter optimization with the TI46 database. (a) The \cos^2 static nonlinear transformation function and its scanned portion in red, under the best operating points close to a minimum or a maximum. (b) WER vs β parameter, under synchronous write-in and read-out, i.e., $\delta\tau = \delta\tau^R$. The red line is the numerics; the blue line is experimental (best: 1.3%). (c) WER as a function of the relative read-out vs write-in asynchrony quantified as $\varepsilon = \delta\tau^R/\delta\tau - 1$. (d) WER vs the β parameter, under asynchronous write-in and read-out. The red line is the numerics; the blue line is experimental (best: 0.04%).

to a minimum or a maximum led to comparable performance optimizations compared to the linear operating point ($\pi/4$ modulo $\pi/2$), with a WER improvement factor of the order of 10. We also notice that the minimum WER is not obtained for a purely parabolic operating condition but always for a slightly shifted position from the minimum or the maximum, e.g., $\pi/10$ or $2\pi/5$ instead of 0 and $\pi/2$. This indicates that a linear contribution might be needed for an optimal processing of this classification task.

- (ii) For each offset phase value $\Phi_0 \in [0; \pi]$, we numerically scanned the delayed feedback gain parameter $\beta \in [0.01; 1.4]$. Again, we consistently found an optimal value slightly below the instability threshold for the fixed point, as described by Eq. (12) in our setup. An optimal value of $\beta \approx 0.7$ was obtained for the optimal, nearly quadratic, offset phase $\Phi_0 \approx 2\pi/5$ [Fig. 7(b)].
- (iii) The weight ρ for the input signal injected into the first EO PM allows for a more or less wide scan of the nonlinear function profile, thus controlling the actual nonlinear expansion of the information into the delay dynamics phase space. A value close to unity confirmed that large-amplitude injected information is in favor of optimal performances. This unit peak-to-peak amplitude is compared to the delayed feedback signal peak-to-peak amplitude, $x_\sigma(n)$, which is of the order of $|2\beta \sin(2\Phi_0)| \ll 1$ (for $\Phi_0 \approx \pi/2$). Excessive values of ρ are detrimental to RC processing because they can trigger transient self-oscillations if β is above the minimum Hopf threshold (at $\beta = 0.5$ in this setup).

- (iv) The effect of an asynchronous read-out (sampling of the transient response) with respect to the write-in (rhythm imposed by the AWG) was explored, both numerically and also from the experimentally recorded responses. We discovered a significant RC processing enhancement by a factor of more than 10, for a very small relative ε deviation between $\delta\tau$ and $\delta\tau^R$, of the order of 10^{-4} . Such an improvement was moreover obtained for either positive or negative relative asynchrony, the exact synchrony actually leading to a nonoptimal RC classification when considering moderate asynchrony ($|\varepsilon| \approx 5 \times 10^{-4}$) [see Fig. 7(c)]. The origin of this phenomenon has not been explained yet and was not discovered earlier, to the best of our knowledge. Because of the extremely small time-scale shifts of concern, this phenomenon is necessarily related to fast response times and thus to the continuous-time information mixing through the impulse response $h(t)$. We anticipate that efficient information extraction from the input signal actually requires asynchrony, possibly related to globally nonuniform sampling issues (we anticipate that a theoretical explanation of this fact might come from irregular sampling theory and/or compressive sensing theory).
- (v) Finally, the concept of EDM for a delay-based reservoir was also found to be of practical interest for improving the RC processing efficiency. Its motivation was initially inspired by the long time delays compared to the fast response times of photonic telecom devices, essentially due to the too-long fiber pigtailed in these devices. This resulted in an initial scaling of the delay dynamics with more than 1000 virtual nodes. Introducing three EDM layers enabled us to process information with a comparable number of nodes (a few hundred) with respect to existing literature. Simulating the processing with an arbitrary integer number ruling the EDM concept has indicated that this concept indeed improved the performances compared to the single-delay memory topology reported up to now. Additional numerical simulations showed us that an optimal scaling can even be obtained for six time EDM. This might be related to some internal correlation of human speech, which could be more efficiently revealed by the use of the EDM RC topology.

C. Database from the AURORA-2 test

The AURORA-2 test is a standardized benchmark developed in the framework of speech recognition for mobile-phone communications. The original acoustic waveforms are extracted from the TIDigit corpus database (from the Linguistic Data Consortium) as for the TI46 test. It is, however, much richer in terms of speaker variability

and thus more challenging in terms of recognition difficulty. The database is constructed of 326 speakers (111 men, 114 women, 50 boys, and 51 girls), each speaking approximately 77 digit sequences from one (isolated) to seven digits. The sequences were made up of the following digits: zero, oh, one, two, three, four, five, six, seven, eight, and nine. On this basis of a set of clean speeches, the AURORA-2 test consists in the construction of test and training subsets, in which different kinds of noise can be added with different noise levels. The whole AURORA-2 thus forms a highly complex test of significant size (more than 28000 digits) corresponding to realistic conditions [30].

For the sake of simplicity and also for ease of comparison with the previous TI46 test popular in photonic RC literature, we have restricted this second test to clean and isolated digits from the AURORA-2 test. Nevertheless, this second test processed through our photonic RC involved a training set of 2412 clean digits (folder *train/clean* of the AURORA-2 test set), and a testing set of 289 untaught digits (folder *test/clean1* in the AURORA-2 test set). This allowed us to essentially focus on the influence of the enhanced speech variability compared to the TI46 test (digits spoken by females only) and thus to focus on the performance impact with our photonic RC. It is worth noticing that performance on clean sets is available in the literature [30] and, moreover, also in the recent case of an advanced RC-based recognition algorithm [31]. In the latter reference, a complex multistate architecture and very high number of nodes were used (up 20 thousand), resulting in performances very close to the state of the art, with the best WER of about 1% on clean digits conditions.

We have performed, both numerically and experimentally, a speech recognition test on the clean and isolated digits subset of the AURORA-2 database. The parameters of our photonic reservoir are the following: number of nodes = 1000, the feedback gain is $\beta = 0.7$, the offset phase is $\phi = 2\pi/5$, and the scalar factor is $\rho = 1$. We have also used 3 times EDM so that we have indeed $1000 \times 3 = 3000$ nodes per time delay. One should notice that these

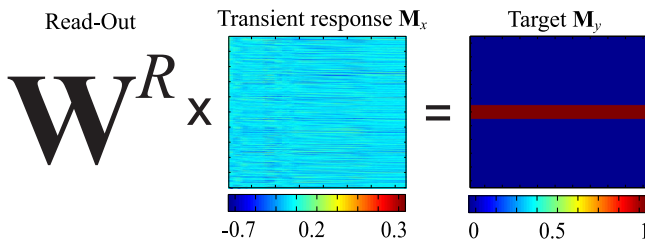


FIG. 8. Illustration of the expected optimized read-out processing for the AURORA-2 test. This illustration is to be compared with Fig. 4 (TI46 test). The latter matrix is aimed at designating the right answer for the digit to be identified (here, the digit “5”).

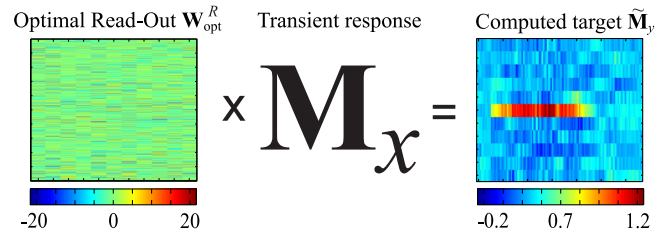


FIG. 9. Example of an imperfect “reservoir-computed” target answer while testing the optimal read-out W^R_{opt} for the AURORA-2 test (to be compared to Fig. 5 for the TI46 test). The answer appears to be the digit “5” here.

parameters have not been optimized and were essentially chosen from the exhaustive parameter scan already done for the TI46 test. Very similar qualitative signal processing is obtained compared to the TI46 test, as can be observed in Figs. 8–10. Despite the simplicity of our approach (single delayed feedback reservoir, dedicated analog hardware necessarily including processing noise and drifts, and moderate number of nodes compared to the advanced algorithmic RC approach in Ref. [31]) and despite the absence of any fine optimization of the operating parameters, we obtained WER below the state of the art but of the same order of magnitude (numerically 1.2% for the learning set WER and 4.5% for the testing set WER; experimentally 6.6% for the learning WER, and 8.9% for the testing WER). The more difficult character of this second test can be illustrated through the speech recognition performance obtained from a direct read-out on the input data. This relevant test highlights the contribution of the reservoir (effect of nonlinear dynamical expansion in a high-dimensional phase space) since it consists in applying a linear read-out only to the input information after the input layer. While 6% WER can already be obtained with the TI46, one only gets 40% when the clean and isolated subset of the AURORA-2 test is concerned. In both cases, the photonic reservoir clearly shows the efficiency of the nonlinear transient expansion performed by the delay dynamics.

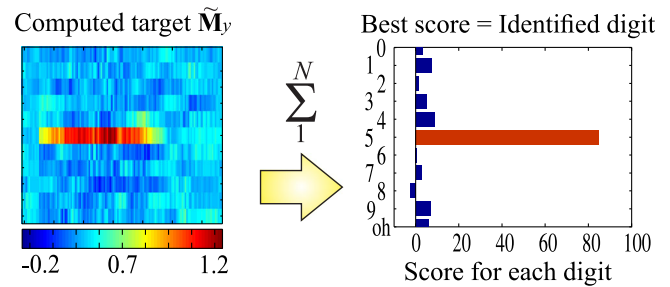


FIG. 10. Decision procedure for the computed answer for the AURORA-2 test (to be compared to Fig. 6 for the TI46 test). The outcome of the test is the digit “5.”

V. CONCLUSION

This work has demonstrated the potential of photonic delay dynamics for ultrafast and efficient RC processing techniques, with a novel electro-optic phase-delay dynamics architecture. A rigorous modeling correspondence of the nonlinear delay dynamics with the original RC concept was proposed, thus theoretically supporting the efficiency of the experimental demonstration. The reported photonic implementation of the RC concept proposes several attractive features: speed related to a structure based on optical telecommunication devices; processing efficiency demonstrated in the particular case of speech recognition; processing universality, according to the universal computing machine concept of RC; design simplification and accurate modeling, through its pure temporal signal-processing architecture compared to complex spatio-temporal attempts for photonic RC processors [10,32,33]. The reported photonic RC architecture also has an important chip integration potential, through the use of conventional photonic devices. The latter are indeed now available in many libraries for photonic integrated chips making use of hybrid III-V and silicon photonics technologies [34]: Lasers, phase modulators, delay lines, and detectors together could provide a “reservoir on a chip,” as could additional in-line signal-processing functions such as correlators for a direct on-chip write-in and read-out. The temporal character of the proposed delay-dynamics approach indeed enables the use of many efficient signal-processing techniques available from modern telecommunication technologies (advanced filtering and modulation techniques), thus offering numerous practical solutions for a powerful implementation of a time-domain RC concept with ultrafast analog photonic systems.

Aside from these many technological and computing potentials, photonic RC still requires important efforts to improve its fundamental understanding, which is obviously needed for a full and self-consistent implementation capable of providing a future breakthrough in brain-inspired computing machines. Perspectives are indeed numerous and challenging in this respect, a few of which have been proposed in this article: through the accurate modeling of the continuous time processing with delay dynamics, through the suggestion for EDM structures, and through the discovery of the tiny but efficient asynchronous sampling between information write-in and dynamical response read-out. Work is also in progress to implement a recently proposed parallel architecture of a smaller delay reservoir [25], allowing for a universal processing of such architecture without the need for the currently time-consuming and task-dependent parameter optimization. Additional research directions concern the investigation of a fully consistent RC system combining both digital and analog subsystems, in which all the different RC processing steps can be integrated (e.g., a digital part taking care of the tasks sequencing, learning and testing coordination, and an

analog photonic part providing the ultrafast RC processing core of the dynamical reservoir).

Last but not least, work is also in progress to address the possibility to design unsupervised learning architectures for these delay-based photonic RC processor. This would make such brain-inspired processors universal and independent from the current algorithmic tool required for the ridge regression in the learning step. We anticipate that particular nonlinear dynamics properties such as pattern formation and emergence of chimera states [35] might be of great importance for this highly challenging perspective in nonlinear dynamics for brain-inspired computing machines.

ACKNOWLEDGMENTS

This work was supported by the European project PHOCUS (FP7 Grant No. 240763), by the Labex ACTION program (Contract No. ANR-11-LABX-01-01), by the BiPhoProc ANR project (ANR-14-OHRI-0002-02), and by CONACyT/MÉXICO under Project No. 290788 and Grant No. 314603.

-
- [1] D. Verstraeten, B. Schrauwen, M. D’Haene, and D. Stroobandt, *An Experimental Unification of Reservoir Computing Methods*, *Neural Netw.* **20**, 391 (2007).
 - [2] H. Jaeger and H. Haas, *Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication*, *Science* **304**, 78 (2004).
 - [3] W. Maass, T. Natschläger, and H. Markram, *Real-Time Computing without Stable States: A New Framework for Neural Computation Based on Perturbations*, *Neural Comput.* **14**, 2531 (2002).
 - [4] M. Lukoševičius and H. Jaeger, *Reservoir Computing Approaches to Recurrent Neural Network Training*, *Computer Sci. Rev.* **3**, 127 (2009).
 - [5] 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. (London)* **2**, 1 (2011).
 - [6] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutiérrez, L. Pesquera, C. R. Mirasso, and I. Fischer, *Photonic Information Processing beyond Turing: An Optoelectronic Implementation of Reservoir Computing*, *Opt. Express* **20**, 3241 (2012).
 - [7] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, *Optoelectronic Reservoir Computing*, *Sci. Rep.* **2**, 287 (2012).
 - [8] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, *All-Optical Reservoir Computing*, *Opt. Express* **20**, 22783 (2012).
 - [9] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, *Parallel Photonic Information Processing at Gigabyte per Second Data Rates Using Transient States*, *Nat. Commun.* **4**, 1364 (2013).

- [10] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, *Parallel Reservoir Computing Using Optical Amplifiers*, *IEEE Trans. Neural Networks* **22**, 1469 (2011).
- [11] M. A. A. Fiers, T. Van Vaerenbergh, F. Wyffels, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, *Nanophotonic Reservoir Computing with Photonic Crystal Cavities to Generate Periodic Patterns*, *IEEE Trans. Neural Networks Learning Syst.* **25**, 344 (2014).
- [12] M. Le Berre, É. Ressayre, A. Tallet, and H. M. Gibbs, *High-Dimension Chaotic Attractors of a Nonlinear Ring Cavity*, *Phys. Rev. Lett.* **56**, 274 (1986).
- [13] F. T. Arecchi, G. Giacomelli, A. Lapucci, and R. Meucci, *Two-Dimensional Representation of a Delayed Dynamical System*, *Phys. Rev. A* **45**, R4225 (1992).
- [14] L. Larger, B. Penkovskiy, and Y. L. Maistrenko, *Virtual Chimera States for Delayed-Feedback Systems*, *Phys. Rev. Lett.* **111**, 054103 (2013).
- [15] X. S. Yao and L. Maleki, *High Frequency Optical Sub-carrier Generator*, *Europhys. Lett.* **30**, 1525 (1994).
- [16] A. Argyris, D. Syvridis, L. Larger, V. Annovazzi-Lodi, P. Colet, I. Fischer, J. Garcia-Ojalvo, C. R. Mirasso, L. Pesquera, and A. K. Shore, *Chaos-Based Communications at High Bit Rates Using Commercial Fiber-Optic Links*, *Nature (London)* **438**, 343 (2005).
- [17] L. Larger and J. M. Dudley, *Optoelectronics Chaos*, *Nature (London)* **465**, 41 (2010).
- [18] R. Lavrov, M. Jacquot, and L. Larger, *Nonlocal Nonlinear Electro-Optic Phase Dynamics Demonstrating 10 gb/s Chaos Communications*, *IEEE J. Quantum Electron.* **46**, 1430 (2010).
- [19] R. Lavrov, M. Peil, M. Jacquot, L. Larger, V. S. Udaltsov, and J. M. Dudley, *Electro-Optic Delay Oscillator with Nonlocal Nonlinearity: Optical Phase Dynamics, Chaos, and Synchronization*, *Phys. Rev. E* **80**, 026207 (2009).
- [20] L. Weicker, T. Erneux, M. Jacquot, Y. Chembo, and L. Larger, *Crenelated Fast Oscillatory Outputs of a Two-Delay Electro-Optic Oscillator*, *Phys. Rev. E* **85**, 026206 (2012).
- [21] L. Larger, *Complexity in Electro-Optic Delay Dynamics: Modeling, Design, and Applications*, *Phil. Trans. R. Soc. A* **371**, 20120464 (2013).
- [22] M. Peil, M. Jacquot, Y. C. Kouomou, L. Larger, and T. Erneux, *Routes to Chaos and Multiple Time Scale Dynamics in Broadband Bandpass Nonlinear Delay Electro-Optic Oscillators*, *Phys. Rev. E* **79**, 026208 (2009).
- [23] L. Weicker, T. Erneux, O. D’Huys, J. Danckaert, M. Jacquot, Y. Chembo, and L. Larger, *Slow-Fast Dynamics for Time-Delay Problems: Theory and Experiments*, *Phil. Trans. R. Soc. A* **371**, 20120459 (2013).
- [24] R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, *Photonic Nonlinear Transient Computing with Multiple-Delay Wavelength Dynamics*, *Phys. Rev. Lett.* **108**, 244101 (2012).
- [25] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, *Stochastic Nonlinear Time Series Forecasting Using Time-Delay Reservoir Computers: Performance and Universality*, *Neural Netw.* **55**, 59 (2014).
- [26] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, *Optimal Nonlinear Information Processing Capacity in Delay-Based Reservoir Computers*, *Sci. Rep.* **5**, 12858 (2015).
- [27] R. Martinenghi, A. B. Fuentes, M. Jacquot, Y. K. Chembo, and L. Larger, *Towards Optoelectronic Architectures for Integrated Neuromorphic Processors*, *Proc. SPIE Int. Soc. Opt. Eng.* **8989**, 89890K (2014).
- [28] D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. Van Campenhout, *Isolated Word Recognition with the Liquid State Machine: A Case Study*, *Inf. Proc. Lett.* **95**, 521 (2005).
- [29] T. Instrument and N. I. S. T., *TI46 Word Speech Database Speaker-Dependent Isolated Word Corpus*, <https://catalog.ldc.upenn.edu/LDC93S9>.
- [30] H. Hirsch and D. Pearce, *Applying the Advanced ETSI frontend to the Aurora-2 Task*, in version 1 (2006).
- [31] A. Jalalvand, F. Triefenbach, D. Verstraeten, and J.-P. Martens, *Connected Digit Recognition by Means of Reservoir Computing*, in Proc. Interspeech, 2011, pp. 1725–1728.
- [32] C. Mesaritakis, V. Papataxiarhis, and D. Syvridis, *Micro Ring Resonators as Building Blocks for an All-Optical High-Speed Reservoir-Computing Bit-Pattern-Recognition System*, *J. Opt. Soc. Am. B* **30**, 3048 (2013).
- [33] K. Vandoorne, P. Mechet, T. Van 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**, 3541 (2014).
- [34] H. Zhang, X. Feng, B. Li, Y. Wang, K. Cui, F. Liu, W. Dou, and Y. Huang, *Integrated Photonic Reservoir Computing Based on Hierarchical Time-Multiplexing Structure*, *Opt. Express* **22**, 31356 (2014).
- [35] L. Larger, B. Penkovskiy, and Y. Maistrenko, *Laser Chimeras as a Paradigm for Multistable Patterns in Complex Systems*, *Nat. Commun.* **6**, 7752 (2015).