Information Processing Capacity of a Single-Node Reservoir Computer: An Experimental Evaluation

Benedikt Vettelschoss, André Röhm, and Miguel C. Soriano, *Senior Member, IEEE*

***Abstract*— Physical dynamical systems are able to process information in a nontrivial manner. The machine learning paradigm of reservoir computing (RC) provides a suitable framework for information processing in (analog) dynamical systems. The potential of dynamical systems for RC can be quantitatively characterized by the information processing capacity (IPC) measure. Here, we evaluate the IPC measure of a reservoir computer based on a single-analog nonlinear node coupled with delay. We link the extracted IPC measures to the dynamical regime of the reservoir, reporting an experimentally measured nonlinear memory of up to seventh order. In addition, we find a nonhomogeneous distribution of the linear and nonlinear contributions to the IPC as a function of the system operating conditions. Finally, we unveil the role of noise in the IPC of the analog implementation by performing *ad hoc* numerical simulations. In this manner, we identify the so-called edge of stability as being the most promising operating condition of the experimental implementation for RC purposes in terms of computational power and noise robustness. Similarly, a strong input drive is shown to have beneficial properties, albeit with a reduced memory depth.**

***Index Terms*— Analog computing, delay systems, information processing, measurement noise, nonlinear dynamical systems, reservoir computing (RC).**

I. Introduction

NEURO-INSPIRED information processing has gained considerable interest in the recent past as research has been accelerated by dramatic breakthroughs in machine learning. At the same time, the high computational costs incurred by these technologies demand more efficient hardware-based implementations. Moreover, neuromorphic computing challenges our understanding of computation as defined by Turing [1]. It is not clear how the well-defined concepts of conventional *von-Neumann* computation carry over to the online processing of input streams accomplished by biological brains. Despite this lack of theoretical understanding, neuro-inspired computing in alternative substrates promises a speedup compared with its digital precursors that is not tied to implementations *in silico* [2]. A number of analog electronic [3]–[6], photonic [7]–[16], and other implementations based on unconventional physical substrates [17]–[19] have been presented. Especially photonic approaches promise low-power consumption, high processing speeds [8], or parallel operation with thousands of network nodes [14]–[16].

These proposals largely employ the principle of reservoir computing (RC) [20]–[23]. In machine learning, RC provides an efficient approach to the training of recurrent neural networks (RNNs), and an appealing alternative to standard training schemes like backpropagation through time. In contrast to these schemes, in RC, one does not attempt to trace gradients through the RNN but instead restricts learning to a set of output weights. In this manner, RC circumvents issues like vanishing or exploding gradient problems and achieves competitive results in nonlinear classification and prediction tasks while employing simple learning techniques like linear regression. The recurrent network is used merely as a temporal kernel that maps an input to a high-dimensional feature space, and it may, thus, be replaced by a broad class of dynamical systems. This makes RC especially appealing to physical implementations [24].

A line of RC research that is particularly suitable for experimental realization concerns the definition of a network in time, rather than in space. This idea has been first proposed under the term *delay-coupled RC* (delay RC) in [3]. Instead of a recurrent network, in delay RC one uses only a single-nonlinear node that is coupled to itself via a delay line. *Virtual* nodes are then created by injecting input into, and reading output from, the nonlinear node at distinct times (see Section II-B1). Consequently, for a hardware implementation of delay RC one only has to build a single-nonlinear element instead of a whole network.

Since the computational capabilities of neuromorphic systems are fundamentally different from the well-quantified computational power of digital computers, their quantification requires a new theory of information processing in dynamical systems. The *information processing capacity* (IPC) introduced by Dambre *et al.* [25] provides such a quantitative

Manuscript received September 29, 2020; revised May 24, 2021 and September 10, 2021; accepted September 19, 2021. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska–Curie Agreement (POST DIGITAL) under Grant 860360; in part by the Spanish State Research Agency under Project PID2019-109094GB-C22/AEI/10.13039/501100011033; and in part by the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in Research and Development under Grant MDM-2017-0711. The work of Miguel C. Soriano was supported by the Spanish Ministerio de Economía, Industria y Competitividad, through a Ramon y Cajal Fellowship, under Grant RYC-2015-18140. *(Corresponding authors: Benedikt Vettelschoss; Miguel C. Soriano.)*

Benedikt Vettelschoss was with the Instituto de Física Interdisciplinar y Sistemas Complejos, Campus Universitat de les Illes Balears, E-07122 Palma, Spain. He is now with the IDLab-AIRO – imec – Ghent University, B-9052 Ghent, Belgium (e-mail: benedikt.vettelschoss@ugent.be).

André Röhm was with the Instituto de Física Interdisciplinar y Sistemas Complejos, Campus Universitat de les Illes Balears, E-07122 Palma, Spain. He is now with the Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan.

Miguel C. Soriano is with the Instituto de Física Interdisciplinar y Sistemas Complejos, Campus Universitat de les Illes Balears, E-07122 Palma, Spain (e-mail: miguel@ifisc.uib-csic.es).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3116709.

Digital Object Identifier 10.1109/TNNLS.2021.3116709

theory. The IPC framework evaluates a driven dynamical system's capacity for computation by measuring the quality of functions of past inputs that can be constructed from the system's state vector. For linear functions, this amounts to the *linear memory capacity* [26], which assesses a system's capability of simple recall and is used widely throughout the RC community. The IPC generalizes the linear memory capacity to nonlinear functions of a system's inputs. By applying their formalism to a range of simulated dynamical systems, Dambre *et al.* [25] demonstrate how the IPC allows assessing the distinct ways in which a system stores and nonlinearly transforms information from an input stream (see Section II-D and [25]).

The IPC has gained renewed interest in the recent past and has since been applied to further simulation studies of various systems, see [27], [28], where in particular its dependence on hyperparameters has been elucidated. Kubota *et al.* [29] dissect the popular NARMA10 benchmark task into the information processing capacities required to solve it. From an applied perspective, this is particularly useful, since knowing the capacity profile of interest enables one to search for a computational substrate, or a dynamical regime within a substrate, that fulfills these requirements. A step in this direction has been made by Dale *et al.* [30], who explore the dynamical regimes of a set of reservoir computers, including a Mackey–Glass numerical example. They quantify the set of *behaviors* and associated computational properties of different substrates. In their work, they rely on multiple measures of which the linear memory capacity is one and do not employ the full IPC framework. Despite this resurgence of activity, experimental evaluations of the IPC for physical systems are scarce. The first attempt at such an evaluation of the IPC in a photonics setup can be found in [31]. The analysis in [31] is restricted to a subset of the IPC, namely, evaluating capacities to construct linear and quadratic polynomials of the input. Capacities up to the quartic degree were analyzed in [32] albeit with a relatively low total memory capacity experimentally measured.

In this contribution, we address the question of how the information processing capabilities of a physical substrate depend on its dynamical operating regime. We go beyond previous works by presenting a complete evaluation of the IPC of a physical reservoir computer, reporting capacities up to degree 7 and with a total memory capacity beyond 80. Our system follows the paradigm of delay RC and is composed of a single-analog Mackey–Glass oscillator and a digitally implemented delay line as described in [3] and [33]. We characterize the computational resources this system provides by evaluating the IPC in various dynamical regimes and show the dependence of the extracted measures on a set of reservoir hyperparameters, namely, the input and feedback scalings. By evaluating the complete spectrum of capacities for the Mackey–Glass system, we are able to link its different dynamical regimes to distinct ways in which it processes information. Moreover, by performing numerical simulations, we investigate the influence of noise (which inevitably affects every physical realization) on the computations carried out by the system. We find that the presence of noise does not simply lead to a homogeneous performance degradation but instead exerts a nontrivial effect on the way the system computes.

II. Methods

A. Reservoir Computing

The principle idea of RC can be traced back to the works of Kirby [20]. It also found some resonance in computational neuroscience [21]. Still, RC started to gain traction within machine learning with the works of Jaeger [22] and Maass *et al.* [23], who independently developed the theoretical foundations of the field.

A reservoir computer consists of a randomly connected RNN—the reservoir—and the input and output weight vectors, $\mathbf{W}^{\text{in}}, \mathbf{W}^{\text{out}}$, that are used to feed a signal to the network and compute a target activation from its state variables. In contrast to conventional RNN training procedures, in RC weights connecting the neurons in the reservoir among each other remain unchanged. Training is restricted to the output weights $\mathbf{W}^{\text{out}}$. In this computing paradigm, the reservoir is used as a high-dimensional random feature space with memory that provides nonlinear transformations of a finite input history. This makes RC a valuable model applicable to computation in dynamical systems in general and not restricted to RNNs. Numerous implementations have been proposed, replacing the RNN with other dynamical substrates or even considering physical substrates for the reservoir.

Despite the disparate ways to implement the reservoir, Jaeger [22] and Maass *et al.* [23] identify a set of constraints a substrate must satisfy in order to provide input–output relationships, that are useful for computations on an input stream. The first of these constraints is referred to as *fading memory* [23], [34] or the *echo state property* in [22]. The fading memory property ensures that the reservoir state depends only on the inputs from some finite time window into the past. This requirement essentially means that the system possesses a single globally attracting fixed point in its state space. The second of these constraints, called the *pairwise separation property* [23], states that for a different history of driving inputs, the reservoir states should be sufficiently distinct. These two conditions are antagonistic and often tuning the reservoir parameters to the desired operating regime implies a tradeoff between them.

B. Experimental Setup

1) Single-Node Reservoir Computing: Our physical implementation of a reservoir computer follows the paradigm of delay RC [3], [35]. Instead of a network of neuron-like elements, delay RC employs only a single nonlinear node that is coupled to itself via a feedback loop of time τ. A set of N *virtual* nodes is then created by considering the response of the nonlinear node at equidistant time slots defined by the node temporal distance θ. Measuring the single nonlinear node at times separated by θ gives us access to the reservoir state as defined by the values of the virtual nodes.

Since now the nodes are defined in time, rather than in space, we need to adapt the procedure used to drive the reservoir with an input signal $u(t)$. First, in order to distribute

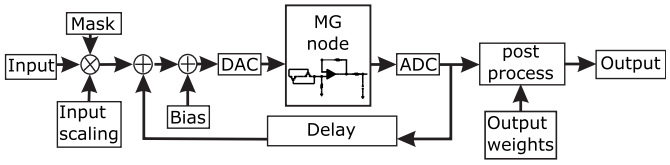


Fig. 1. Schematic of the experimental setup for a single-node reservoir computer with delay. The Mackey–Glass nonlinear node is implemented via an analog electronic circuit. DAC stands for digital-to-analog converter, ADC for analog-to-digital converter, and MG for Mackey–Glass.

the input to every virtual node, we apply a sample and hold procedure. This procedure yields an input stream $I(t)$ that is constant during a time T_{in}, called the data injection time. To implement the temporal analog of the input weights $\mathbf{W}^{\text{in}}$, we define a step function that is constant for the node separation θ and periodic with period T_{in}

$$M(t) = \mathbf{W}_i^{\text{in}} \quad \text{for } (i-1)\theta \leq (t \bmod T_{\text{in}}) < i\theta \tag{1}$$

where $i = 1, \ldots, N$ enumerates the virtual nodes and $T_{\text{in}} = N\theta$ sets the data injection time. This $M(t)$ is called the *masking function*. The stream $J(t) = I(t)M(t)$, resulting from the multiplication of the input stream with the masking function is then injected into the nonlinear node sequentially.

The interplay of delay τ, node separation θ, data injection time T_{in}, and the time constant of the nonlinear node T_{nl} gives rise to a multitude of virtual network topologies that have been discussed in detail in [36] and [37]. In this study, we will be operating in the so-called map limit regime, where $T_{\text{nl}} \ll \theta$. This choice of θ allows the virtual nodes to reach their steady state for each of the consecutive steps of $J(t)$. Choosing injection time $T_{\text{in}} = \tau$ would lead to a virtual network in which every node is only coupled to itself and the input. Such a connection structure is disadvantageous from the viewpoint of RC as it leads to a reservoir response that is not sufficiently diverse. We may, however, induce a coupling among nodes by choosing $T_{\text{in}} \neq \tau$ [38]–[41]. The topology is then determined by the mismatch between data injection time T_{in} and delay τ. In the following, we will use $\tau = T_{\text{in}} + \theta$, i.e., $\tau = (N+1)\theta$. As elaborated in [35] and [37], this choice yields a topology in which the virtual nodes are arranged on a ring.

2) Analog Mackey–Glass Nonlinearity: The nonlinear node we use in the delay RC setup consists of an analog Mackey–Glass nonlinearity [42], [43] coupled to a Raspberry Pi. The Mackey–Glass nonlinearity can be implemented electronically according to [33]. Our setup is schematically shown in Fig. 1. In the analog part of the RC implementation, the time constant of the nonlinear node is $T_{\text{nl}} \sim 5\ \mu$s and the virtual node separation is $\theta \sim 83\ \mu$s, i.e., $T_{\text{nl}} \ll \theta$. Preprocessing and postprocessing of the input and output are done digitally on the Raspberry Pi. The digital part of the setup is also used to implement the delay line and feed the sequential input to the nonlinearity, to which it is coupled through a digital-to-analog and analog-to-digital converter (DAC and ADC, respectively). Thus, we are considering a hybrid digital–analog implementation, in which the analog Mackey–Glass nonlinear node is the physical substrate. The experimentally measured Mackey–Glass nonlinearity and the corresponding numerical fit are shown in Appendix A.

To set the operating region along the nonlinearity for information processing, we introduce a bias Φ and an input scaling γ. These two parameters, respectively, determine the center and extent of the operating region. In addition, we denote by β the feedback scaling that determines the influence of the activations from one delay time before on the current reservoir state. When $T_{\text{nl}} \ll \theta$, the analog physical system can be modeled as a continuous-time difference equation [44]. Considering the input injection, the RC parameters and the Mackey–Glass nonlinearity, the system state $x(t)$ can be written as [33]

$$x(t) = \frac{a(\gamma\, J(t) + \beta x(t-\tau) + \Phi)}{1 + b^p(\gamma\, J(t) + \beta x(t-\tau) + \Phi)^p} \tag{2}$$

where the values of the parameters a, b, and p are fit from the experimentally measured nonlinearity, as detailed in Appendix A. In the experiments, the virtual nodes are measured at the end of each node separation time θ such that we get a single value per node in each delay cycle from the continuous system output.

C. Numerical Simulations

In order to provide additional insights on the experimental assessment of the information processing capabilities of the physical substrate under study, we perform detailed numerical simulations of a matching model of the analog Mackey–Glass nonlinearity. For our choice of the parameters τ, θ, and T_{in}, taking measurements at fixed time intervals from a system evolving according to (2) may be formulated as iterating the map of virtual nodes x_i on discrete time $k \in \mathbb{N}_+$

$$x_i(k) = \frac{a\left(\gamma\, W_i^{\text{in}} u(k) + \beta x_{i-1}(k-1) + \Phi\right)}{1 + b^p\left(\gamma\, W_i^{\text{in}} u(k) + \beta x_{i-1}(k-1) + \Phi\right)^p} \tag{3}$$

for $i = 2, \ldots, N$ and

$$x_1(k) = \frac{a\left(\gamma\, W_1^{\text{in}} u(k) + \beta x_N(k-2) + \Phi\right)}{1 + b^p\left(\gamma\, W_1^{\text{in}} u(k) + \beta x_N(k-2) + \Phi\right)^p}. \tag{4}$$

The ring topology of the virtual network is reflected by the coupling of node $x_1(k)$ to node $x_N(k-2)$ through (4), while all other nodes receive feedback from the previous node measured one time step before [(3)] [37], [38]. In Section III, we will study the influence of the feedback scaling β and input scaling γ on the information processing capabilities of the system.

D. Information Processing Capacity

A general framework to assess the IPC of input driven dynamical systems has been introduced by Dambre *et al.* [25]. Its foundations lie in the assessment of a system's ability to compute functions $z(k)$ of the input history. Using linear estimators

$$\hat{z}(k) = \sum_{i=1}^{N} W_i^{\text{out}} x_i(k) \tag{5}$$

a system's capacity to reconstruct the function z from its state variables x_i is quantified in terms of the mean-square error (MSE)

$$\mathrm{MSE}_K[\hat{z}] = \frac{1}{K}\sum_{k=1}^{K}(\hat{z}(k) - z(k))^2. \tag{6}$$

A system's capacity C to reconstruct the function z is

$$C[z] = 1 - \frac{\min_{W^{\mathrm{out}}} \mathrm{MSE}_K[\hat{z}]}{\langle z^2\rangle_K}. \tag{7}$$

Using this definition of capacity, we can assess memory and nonlinear transformations carried out by the system depending on our choice of the function z, for which the capacity is evaluated. Evaluating (7) for $z(k) = u(k-3)$, for example, quantifies the system's capacity to recall an input that occurred three-time steps before. Similarly, evaluating the capacity for $z(k) = u(k-3)u(k-2)$ characterizes the system's memory and ability to carry out nonlinear transformations on past inputs.

To evaluate a dynamical system's IPC in general, that is not in relation to a particular function of past inputs, Dambre *et al.* define the Hilbert space $\mathcal{H}_{U^h}$ of functions $z : U^h \to \mathbb{R}$, where U^h denotes the space of inputs h steps into the past. If we can find a denumerable basis for this space, evaluating the capacity for each basis function gives us a measure of the system's IPC. Moreover, by noting which basis functions yield a high capacity we obtain a profile that characterizes the kinds of computation carried out by a dynamical system. To construct $\mathcal{H}_{U^h}$, we need an inner product from which we can successively derive a distance and norm. This inner product is given to us by the expectation with respect to the input probability distribution $p(u)$

$$\langle z, z'\rangle = E_{U^h}[zz'] = \int_0^h p(u_{k-k'})z(u_{k-k'})z'(u_{k-k'})dk' \tag{8}$$

where z and z' are functions in the space $\mathcal{H}_{U^h}$ [25]. Hence, feeding our system with i.i.d. inputs drawn from a chosen probability distribution $p(u)$, we may construct an appropriate basis and evaluate the IPC for all of its members. Here, as in [25], we choose a uniform distribution over the interval $[-1, 1]$ for the input and products of normalized Legendre polynomials as a corresponding basis. We gain insights into how past inputs are transformed in our system, by associating to each product of Legendre polynomials the sum of degrees, denoted as d, of the individual basis functions. We obtain the IPC by computing the sum of capacities evaluated for basis functions belonging to each of these linear ($d = 1$) and nonlinear ($d > 1$) memory degrees. The total capacity evaluated in this way is bounded from above by the number of readout variables, here N. Further details on the procedure to estimate the IPC are given in Appendix B.

III. Results

Along this section, we will present experimental and numerical results for the assessment of the IPC of the Mackey–Glass based reservoir computer. In both, experiment and simulation, we create a virtual network of $N = 100$ nodes arranged in a ring topology. Hence, throughout this study, the maximally achievable measurement of the total IPC will be 100. To feed the input stream to the reservoir, the input weights $\mathbf{W}^{\mathrm{in}}$ for the mask are drawn from the standard normal distribution $\mathcal{N}(0, 1)$. In order to identify the relationship between the dynamical regimes and the IPC, we conduct experiments and simulations for five combinations of input scaling γ and feedback scaling β (see Eqs. 2–4). For each $\gamma - \beta$ parameter combination, we scan a set of 20 values of the bias Φ in the range $[12, 725]$, spanning the domain of the Mackey–Glass nonlinearity (see the range of the x-axis in Fig. 11 of Appendix A).

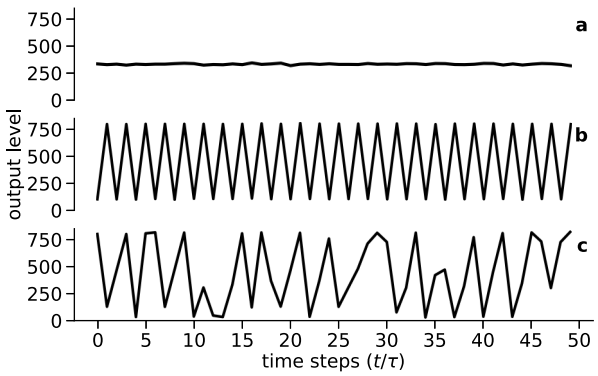


Fig. 2. Time series obtained experimentally from the system operating without external input $\gamma = 0$ in different dynamical regimes. (a) Fixed point dynamics for $\beta = 0.2$ and $\Phi = 540$. (b) Periodic dynamics for $\beta = 0.4$ and $\Phi = 360$. (c) Chaotic dynamics for $\beta = 0.8$ and $\Phi = 90$.

A. Assessment of the IPC: Experimental Results and Matching Numerical Simulations

In the absence of external input, our system's *repertoire* of dynamical regimes is controlled by the feedback scaling β and the bias Φ. By increasing β and tuning the bias Φ accordingly, a part of phase space undergoes a transition from a zero fixed point to a nonzero fixed point. An example of an experimentally recorded time series of a nonzero fixed point can be seen in Fig. 2(a). As shown in Fig. 2(b), the state of the system can then transition into oscillatory dynamics that finally can give rise to chaos, see Fig. 2(c), in a period-doubling cascade, similar to the continuous flow case shown in [42] and [45]. During this cascade, we observe a plethora of fixed points, periodic and chaotic dynamics among which we may select by a proper choice of the experimentally accessible parameters β and Φ. An important aspect in the comparison between experimental and numerical results can already be identified in Fig. 2(a). The experimentally recorded time series for stable fixed point dynamics shows small fluctuations, which are due to noise. We can already anticipate that the addition of noise in the numerical simulations will be required to obtain a quantitative agreement between experimental and numerical results in the assessment of the IPC.

In order to characterize the dependence of the dynamical regimes on the operating point, we run the system without external input (i.e., with an input scaling of $\gamma = 0$) and obtain a bifurcation diagram of the autonomous dynamics as a function of the bias Φ. We will first focus on the case $\beta = 0.4$, a regime in which oscillatory and fixed-point dynamics exist

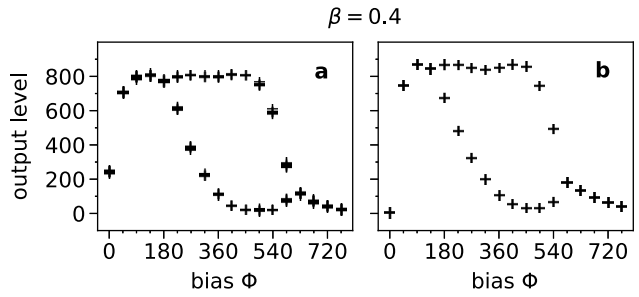


Fig. 3. Bifurcation diagrams obtained from (a) experiment and (b) noiseless numerical simulation for a feedback scaling of $\beta = 0.4$. Both plots show the dynamics of a single node operating autonomously (input scaling $\gamma = 0$) for distinct values of the bias Φ. The last 50 values of a 5000-step time series are plotted.

for different values of the bias Φ. Fig. 3 shows the bifurcation diagrams obtained in experiment [Fig. 3(a)] and simulation [Fig. 3(b)] by scanning Φ in the range [0, 765]. The bifurcation diagram in Fig. 3 is obtained by measuring the response of a single virtual node for a total of 5000 timesteps and displaying its last 50 activations for each value of the bias Φ in a scatter plot. From the bifurcation diagrams in Fig. 3, we can identify the periodic regime for $\Phi \in (180, 630)$. In this region, the system displays oscillatory dynamics of period 2.

We now proceed to evaluate the capacities of the Mackey–Glass reservoir. Fig. 4(a) shows the IPC measures obtained from the experimentally recorded time series for a feedback scaling of $\beta = 0.4$ and an input scaling of $\gamma = 20$. To present our results, we sum capacities evaluated for basis functions of the same polynomial degree d and show the total capacity in a bar plot, indicating the degree by the different colors. Hence, the first degree [yellow in Fig. 4(a) and throughout], corresponds to the classical "linear memory." Changing the bias Φ changes both the distribution of memory among the degrees and the total capacity. We observe a region of vanishing capacities for intermediate values of the bias Φ. The drop in the IPC for intermediate values of Φ coincides with the region of periodic dynamics in the bifurcation diagram of Fig. 3. We find that the maximal information processing capacities are reached for values of the bias Φ on both the left-hand side ($\Phi \simeq 198$) and the right-hand side ($\Phi \simeq 612$) of the periodic region. Changes in the capacity distribution are sufficiently smooth and therefore let us identify trends in capacity development. In agreement with previous works, we observe that periodic dynamics are disadvantageous for RC. If the reservoir has noncontractive dynamics, its state no longer strictly depends on the input history alone [46], violating the *fading memory* property.

Along with the experimental results presented in Fig. 4(a), we show in Fig. 4(b) capacities evaluated numerically for matching operating conditions. The window of vanishing capacities for intermediate values of the bias Φ persists. In contrast to the experimental assessment of the IPC, we observe that outside this region the total IPC approaches the maximum of $N = 100$ (theoretical bound) for several values of the bias Φ. The apparent disagreement between the evaluation of the IPC for the experiment [Fig. 4(a)] and for the

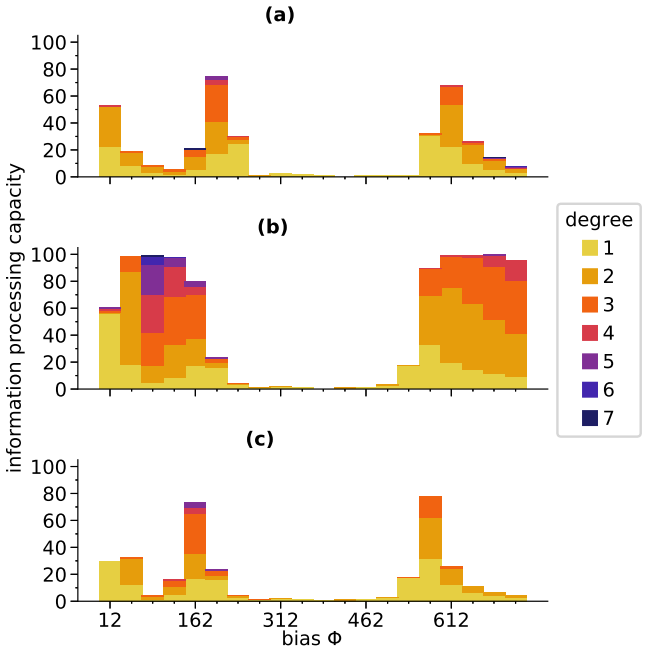


Fig. 4. Information processing capacities (y-axis) evaluated in (a) experiment, (b) noiseless simulation, and (c) simulation with added Gaussian noise with a standard deviation of $\sigma = 5$ for feedback scaling $\beta = 0.4$ and input scaling $\gamma = 20$ over 20 different values of the bias ϕ (x-axis). Colors indicate different constituting degrees of the IPC. Due to the system size, the maximum theoretically obtainable IPC was 100.

corresponding numerical simulations [Fig. 4(b)] justifies the extension of the numerical model. A minimal change to the model is the consideration of readout noise, which is known to degrade the performance in analog reservoir computers [5], [47]. Therefore, the capacities will be estimated from the noisy state variables as follows:

$$\hat{z}'(k) = \sum_{i=1}^{N} W_i^{\text{out}}(x_i(k) + \sigma\xi_i(k)) \tag{9}$$

where $\xi_i(k)$ is a Gaussian random variable drawn from the standard normal distribution $\mathcal{N}(0, 1)$ and σ is the noise strength. We find that the best matching noise has a standard deviation of $\sigma = 5$, as detailed in Appendix C. With added Gaussian readout noise, the numerical results, an example of which is given in Fig. 4(c), are in good agreement with the capacities measured in the experiment, see Fig. 4(a). This is an indication that the observed experimental capacities are influenced by noise as they can be recovered in the numerical simulations by including the readout noise in the modeling. We note, however, that there seems to be a slight shift along the Φ axis throughout the bifurcation diagrams and capacity profiles between the numerical and experimental results in this study. We attribute this effect to the deviations between the constructed analog nonlinearity and the ideal Mackey–Glass nonlinearity, as shown in Appendix C, Fig. 11.

B. Influence of the Input Scaling on the IPC

Here, we evaluate capacities for the whole range of Φ, $\beta = 0.4$, and two additional values of the input scaling γ,

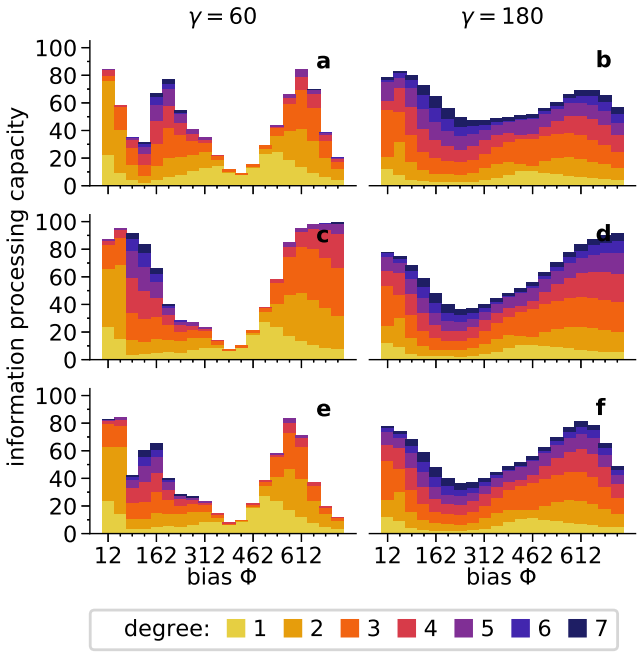


Fig. 5. Information processing capacities (y-axis) evaluated in (a) and (b) experiment, (c) and (d) noiseless simulation, and (e) and (f) simulation with added Gaussian noise with a standard deviation of $\sigma = 5$ for feedback scaling $\beta = 0.4$ over 20 different values of the bias ϕ (x-axis), comparing input scaling $\gamma = 60$ (left) and $\gamma = 180$ (right). Colors indicate different constituting degrees of the IPC. Due to the system size, the maximum theoretically obtainable IPC was 100.

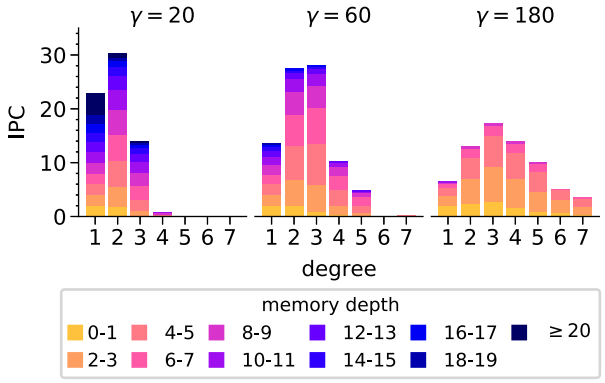


Fig. 6. Distribution of the experimentally measured IPC (y-axis) over basis function degrees for feedback scaling $\beta = 0.4$, bias $\Phi = 612$, and increasing values of the input scaling γ. Colors indicate the memory depth, i.e., a splitting of the total IPC per degree by the maximum delay entering the constituting basis functions. Darker colors correspond to longer memory, higher degrees to more nonlinear processing.

namely, $\gamma = 60$ and $\gamma = 180$. Capacities have been evaluated from experimental data in Fig. 5(a) and (b); from noiseless simulations in Fig. 5(c) and 5(d); and from numerical simulations with readout noise in Fig. 5(e) and (f). We again find evidence of the quantitative agreement between the experimental assessment of the IPC, see Fig. 5(a) and (b), and the IPC for the numerical simulations with readout noise, see Fig. 5(e) and (f). In contrast, the IPC for the noiseless numerical simulations shows larger total capacities for wider ranges of Φ. The noise susceptibility of the IPC for different operating conditions depends on the underlying dynamical regime and on the experimental signal-to-noise ratio, as will be discussed in more detail in Section III-D. Here, we turn our attention to the influence of the input scaling γ.

The IPC of the system depends on both the dynamical properties of the autonomous reservoir and on the properties of the external input driving. When the input scaling γ is small, the underlying intrinsic dynamics of the reservoir dominate the distribution of the capacities over the Φ range. Therefore, the existence of a periodic window in the dynamics (see Fig. 3) is directly related to a drop in the IPC for intermediate values of Φ, as observed in Fig. 5(a). When the input scaling γ is large, the nonlinear transformation of the external input driving dominates the IPC and the influence of Φ is less marked than for small input scaling. This is evidenced in Fig. 5(b) by the more homogeneous distribution of the IPC as a function of Φ.

Regarding the IPC degree contributions, changing the input scaling γ has a multitude of effects. A closer inspection of (2) shows that choosing Φ sets the point on the nonlinearity around which the input is injected, while the input scaling γ controls the range around the operating point to which inputs will be mapped. With the inputs being in the range $[-1, 1]$, the operating range is therefore roughly $[\Phi - \gamma, \Phi + \gamma]$. In Fig. 5(a), for $\gamma = 60$, we see significant contributions of degree larger than or equal to four emerging for some portions of the Φ-range, which were not present in the IPC measured for $\gamma = 20$ in Fig. 4(a). In Fig. 5(b), for $\gamma = 180$, significant capacities of degree 5 and beyond can be found across the whole spectrum, while linear memory decreases steadily. We note that, in the case of large input scalings γ, inputs get mapped to an increasing section of the nonlinearity. This has two important consequences. First, the influence of the bias Φ is reduced for larger γ. While $\gamma = 20$ shown in Fig. 4(a) had a pronounced drop of the IPC for the periodic window, this gradually disappears for higher γ in Fig. 5(a) and (b). Second, the IPC contains more linear and lower degree components for smaller γ. This is intuitively clear when considering that, in the limit case for very small input scaling γ, inputs get mapped to a linear section of the Mackey–Glass function. Conversely, a high input scaling γ produces many high-degree contributions in the IPC, see Fig. 5(b).

The highly nonlinear information processing achieved by increasing the input scaling γ comes, however, at the cost of a reduced memory depth. Fig. 6 demonstrates this effect, showing the experimental capacity measures for the point $\beta = 0.4$ and $\Phi = 612$, and an increasing input scaling γ, where we have grouped the basis functions of the IPC differently. Here, we plot capacities evaluated for degrees 1–7 separately. In order to visualize the maximum delay of which the system retains memories, we split the capacity contributions according to their different memory depths for each degree (color in Fig. 6). In accordance with the results presented in Figs. 4 and 5, we observe in Fig. 6 that for $\gamma = 20$ significant capacities are found mainly for degrees 1–3. Furthermore, here, we can see that the system can recall inputs that occurred more than 20 input steps into the past (dark blue, Fig. 6 for

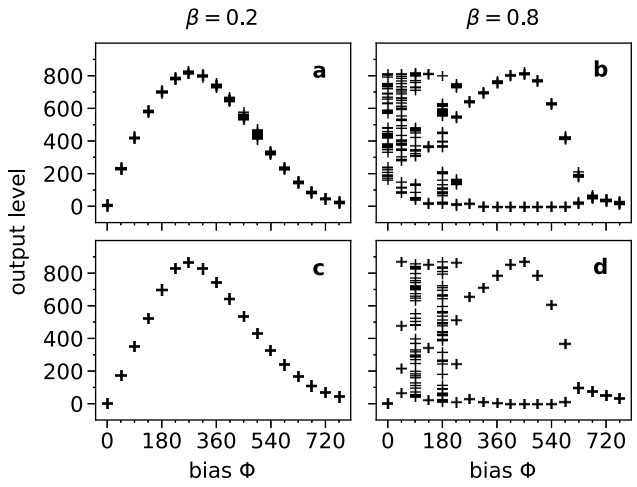


Fig. 7. Bifurcation diagrams obtained from (a) and (b) experiment and (c) and (d) numerical simulation for a feedback scaling of $\beta = 0.2$ (left) and $\beta = 0.8$ (right). All plots show the dynamics of a single node operating autonomously (input scaling $\gamma = 0$) for distinct values of the bias Φ.

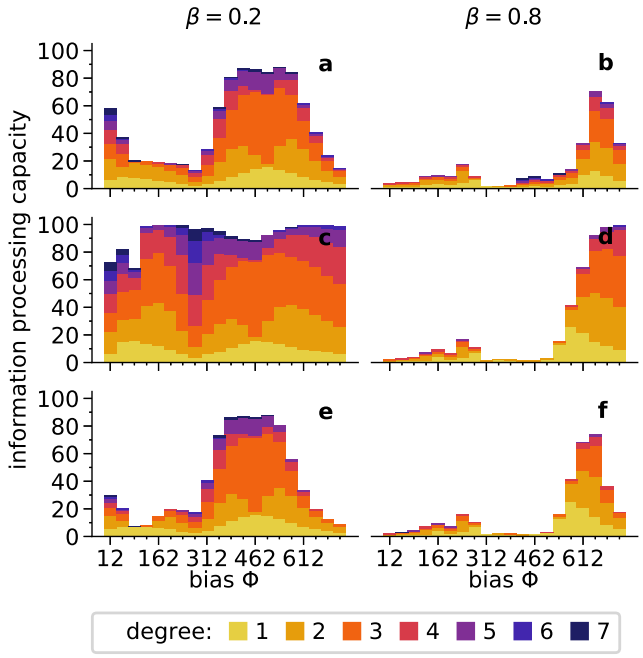


Fig. 8. IPC (y-axis) evaluated in (a) and (b) experiment, (c) and (d) noiseless simulation, and (e) and (f) simulation with added Gaussian noise with a standard deviation of $\sigma = 5$ for input scaling $\gamma = 60$ over 20 different values of the bias ϕ (x-axis), comparing feedback scaling $\beta = 0.2$ (left) and $\beta = 0.8$ (right). Colors indicate different constituting degrees of the IPC.

$\gamma = 20$). Similarly, we see that the experimental system is able to compute basis functions of degrees 2 and 3 of inputs that occurred up to 20 time-steps before for $\gamma = 20$. This memory depth decreases steadily with the emerging higher memory degrees that appear for larger values of the input scaling γ . For an input scaling of $\gamma = 60$, the IPC concentrates in degrees 2 and 3 and significant capacities can be found up to 13 steps into the past. Finally, for $\gamma = 180$, we find capacities up to degree 7, which, however, do not exceed a memory depth of 10. For such high values of the input scaling, memories of past inputs tend to be overwritten due to the large influence of present inputs. Fig. 6, thus, shows the tradeoff between memory and nonlinearity [25]: a dynamical system either stores information for a long time if it tends to be linear or nonlinearly transforms recent inputs but then does not retain a long memory of the past.

C. Further Insights on the Relationship Between the Dynamical Regimes and the IPC

Fig. 7 shows the influence of the feedback strength β on the system's dynamics. For a small feedback strength of $\beta = 0.2$, the system has a single fixed point whose value is determined by Φ, as shown in Fig. 7(a) and (c), for the experiments and the numerical simulations, respectively. For a higher feedback scaling of $\beta = 0.8$, we show in Fig. 7(b) and (d) that the dynamics of the autonomous system may become chaotic in the lower end of the Φ range, see the broad spread across output voltage around $\Phi = 180$. The experimental and numerical bifurcation diagrams for $\beta = 0.8$ also show a window of periodic dynamics for intermediate values of Φ.

Fig. 8 shows the corresponding IPC measures for an intermediate input scaling of $\gamma = 60$ and two values of the feedback strength β, namely, $\beta = 0.2$ and $\beta = 0.8$. We start by discussing the results for $\beta = 0.2$. In this case, the fading memory property is fulfilled for all biases Φ as the dynamics of the autonomous system are always contractive. Accordingly, as shown in Fig. 8(c), the total IPC nearly saturates at the maximum available capacity of 100 for almost the whole range of Φ in the noiseless numerical simulations. However, the experimentally measured IPC in Fig. 8(a) and the IPC for the numerical simulations with readout noise in Fig. 8(e) have a clear range of larger capacity values around $\Phi = 462$. The observed decrease in capacities for other values of Φ is due to the different susceptibility to noise of the IPC for different operating conditions, which can be attributed in this case to the resulting different signal-to-noise ratios, as it will be discussed in more detail in Section III-D.

Next, we discuss the results for a higher feedback scaling of $\beta = 0.8$. In Fig. 8(b), (d), and (f), the corresponding capacity profiles show a drastic decay in memory across all degrees in the lower end of the Φ range due to the chaotic dynamics of the autonomous system. Significant capacities are concentrated around the right-hand side of the bias range at $\Phi = 612$, at the transition between periodic and fixed-point dynamics of the autonomous system.

With the IPC results for $\beta = 0.8$ and $\gamma = 60$, our results convincingly debunk a common misconception regarding the region of optimal performance for RC. The misconception consists of confusing the "edge of stability" with the "edge of chaos." The "edge of stability" denotes the region in the vicinity of a bifurcation, where a globally stable fixed point loses stability. When approaching the bifurcation point from the stable direction, the time it takes a system to relax back to equilibrium increases, thereby endowing the system with long transient memory. For tasks requiring such long memory spans and for which a fast reaction to input is not essential, this regime has been demonstrated to yield

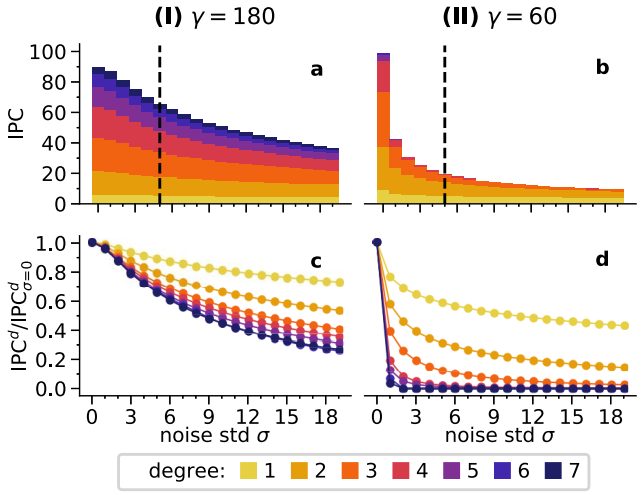


Fig. 9. Decay of the IPC as a function of increasing standard deviation σ of a Gaussian readout noise for $\beta = 0.4$, $\Phi = 687$ in two different operating regimes, set by the input scaling γ. (a) and (b) Decay of the total IPC for the two operating regimes split into the constituting degrees according to the color code. The dashed vertical line marks the noise strength found to optimally reproduce our experimental IPC measurements, as detailed in Appendix C. (c) and (d) Decay of capacities for each degree d (IPC^d) normalized by the respective capacity obtained in the noiseless case ($\text{IPC}^d_{\sigma=0}$).

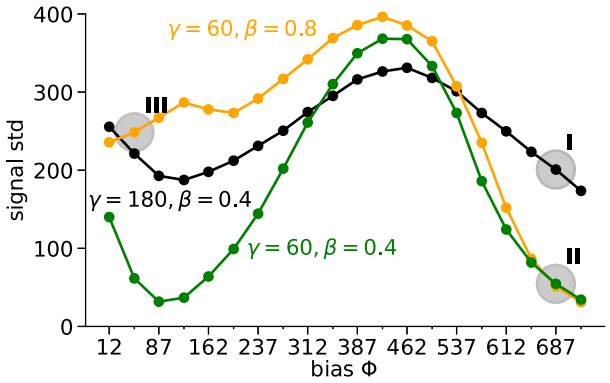


Fig. 10. Standard deviation of the experimentally measured analog signals as a function of the bias Φ for the indicated (γ, β) parameter sets. Marked points I and II correspond to the decay of capacities shown in Fig. 9.

optimal results [48]. In contrast, the "edge of chaos" was introduced by Langton [49] and Packard [50] in the context of Turing computability in cellular automata. It refers specifically to the situation where a system undergoes a transition from an ordered state, a fixed point or limit cycle, to a disordered—i.e., chaotic—state. Inspecting the capacity profiles in Fig. 8(b), (d), and (f) leads us to the conclusion that it is the edge of *stability* that equips a reservoir with computational resources in terms of IPC. Although we observe a small IPC peak at the border between periodic and chaotic dynamics, most of the system's computational power is concentrated at the transition between the fixed point and the periodic regime. Entering the chaotic regime entails a total breakdown of capacities.

D. Noise Susceptibility of the IPC for Different Operating Conditions

The addition of Gaussian readout noise in the numerical simulations allows us to recover the approximate distributions of experimentally measured capacities, as shown at the top and bottom rows of Figs. 4, 5, and 8. Comparing the noiseless capacity numerical estimates with the numerical results obtained with noise, however, we also note that the noise-induced decay of capacities does not affect all operating conditions equally.

Fig. 9 shows the numerical results of the IPC as a function of the noise strength σ [see (9)] for two different parameter sets, namely, (I) $\gamma = 180$, $\beta = 0.4$, and $\Phi = 687$ and (II) $\gamma = 60$, $\beta = 0.4$, and $\Phi = 687$. As shown in panels a and b, for both parameter sets I and II, the total IPC decreases for increasing noise strengths. Panels c and d show the decay of capacities for different degrees (IPC^d) separately. In addition, the capacities evaluated for each degree d and varying σ are normalized by the corresponding measurement obtained for $\sigma = 0$ ($\text{IPC}^d_{\sigma=0}$). On the one hand, we find that the high memory degrees decay faster than the low ones. On the other hand, we find that the parameter set I, with $\gamma = 180$, maintains a large IPC for a significantly large range of noise strengths. In the following, we discuss these two findings separately.

Noise can be thought of as an additional but neglected external input to the system. A complete basis for the space of functions of the input history would then need to consist of *products* of Legendre polynomials of both, input and noise. When evaluating linear memory in the presence of noise, we implicitly neglect linear basis functions that are functions only of the noise. Similarly, for quadratic memory, we cannot take into account quadratic basis functions that are functions of the noise. But we also neglect all quadratic basis functions that consist of products of linear Legendre polynomials of input and noise. This effect carries over to the higher memory degrees, where it becomes increasingly severe [25]. As a consequence, we observe the decay shown in Fig. 9(c) and (d), where capacities decay progressively faster with increasing memory degree.

The large noise susceptibility of high IPC degrees only partially explains the varying noise susceptibility of the different operating conditions. An aspect that we have disregarded so far is the fact that different operating conditions have a different signal-to-noise ratio. In Fig. 10, we show the signal standard deviations measured experimentally for three different $\beta - \gamma$ combinations as a function of Φ. The standard deviation of the measured signals varies significantly across different operating conditions. In our experimental setup, the standard deviation of the readout noise is independent of the operating conditions. As a result, the signal-to-noise ratio of the different operating conditions varies following the standard deviation of the measured signals. This effect is also reproduced in the numerical simulations, where the added noise in (9) is independent of the values of x. The points for which we display the decay of capacities due to noise in Fig. 9 are marked as I and II, respectively, in Fig. 10. The operating point I has a larger signal standard deviation than point II. We find that the signal-to-noise ratio plays a key role in the observed difference in the decay rate of the IPC as a function

of the readout noise strength. Thus, to obtain a capacity profile with significant nonlinear memory, that is nevertheless robust to noise, a large input scaling appears as a promising strategy, albeit at the cost of a reduced memory depth. However, not just any signal with a high standard deviation provides noise-robust information processing capabilities. For instance, the point $\gamma = 60, \beta = 0.8$, and $\Phi = 50$, marked as III in Fig. 10 displays a high standard deviation in the reservoir responses due to the presence of chaotic dynamics. Since the fading memory property is violated in this chaotic regime, capacities vanish completely even in the absence of noise, as can be seen in Fig. 8(d).

In summary, the readout noise does not homogeneously deteriorate the IPC of the system. A parameter combination that yields the desired capacity profile in (noiseless) simulation might be of not much use when it comes to experimental implementation due to the presence of noise. Since different tasks demand different nonlinear properties of the reservoir [51], different parameters may be tweaked to obtain a computationally advantageous regime that is robust to noise. Knowing the noise sources of an experimental system can thus be of great help in the process of tuning the system to the right regime. In the case of an experimentally observed low IPC value, the comparison between the numerical simulations with and without readout noise allows us to elucidate whether the low IPC of a given operating point originates from an unsuitable underlying dynamical regime or from a strong susceptibility to noise. Here, we model experimental noise in the simplest possible way—by adding Gaussian observation noise—and find that this is sufficient to achieve a good overall quantitative agreement for the IPC between the numerical simulations and the experiments.

IV. Conclusion

In this work, we have evaluated the IPC proposed in [25] on a physical reservoir computer. While there is a range of simulation studies of the IPC [27], [28] and an initial application to a physical implementation [31], in contrast to these works our study constitutes a complete evaluation of the IPC across multiple dynamical regimes. The evaluated single-node reservoir computer implementation shows information processing capacities beyond 80, nonlinear capacities up to degree 7, and memory depths beyond 20 when appropriately driven.

We connect the different dynamical regimes our system displays without external input to distinct ways in which it processes information. Moreover, we identify how the system parameters can be used to set the system up in a configuration that yields the capacity profile of interest. Our results demonstrate the influence of fading memory, criticality, periodic, and chaotic dynamics on a system's computational capabilities. We find that the transition point between stable and periodic dynamics provides both noise robustness and computational performance. While the edge of stability has been shown to be an optimum operating point for software reservoirs, here we show that this is even more so for hardware reservoirs. While chaotic dynamics can be employed in applications, such as encryption [52], we do not find any computational advantage for RC.

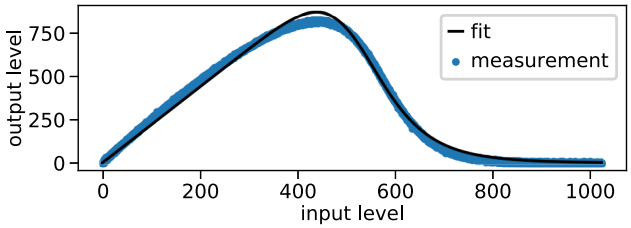


Fig. 11. Experimental Mackey–Glass nonlinearity (blue line) and corresponding fit (black line) used for the simulations according to (10). The fit parameters used in simulations are $a = 2.189$, $b = 0.001839$, and $p = 10.37$.

Furthermore, evaluating the complete capacity profiles allows us to investigate the degradation of capacities due to noise, which is unavoidable in any experimental setup. We find that the susceptibility to noise depends crucially on the dynamical regime the system operates in and demonstrate how higher memory degrees are significantly more affected. To this end, we model the noise observed in experiments as Gaussian readout noise in the corresponding simulation. By comparing with noiseless numerical simulations, we find that adding noise does not simply flatten capacity profiles but may change the distribution of total capacity among the degrees and the shape of the profile when expressed as a function of system parameters. It becomes apparent that the system parameters for an experimental reservoir computer should not be inferred beforehand from noiseless numerical simulations. We show how the information processing capabilities of the system depend on the interplay between the underlying dynamical regime, the input scaling, the operating point, and the noise. The implications of this analysis are relevant to both hardware and software implementations of RC systems.

Our findings demonstrate how to use the framework laid out in [25] to assess the computational utility of the different dynamical regimes of a physical substrate. This framework provides an important tool to connect a task at hand with a computationally useful substrate and can, thus, be of great value in the design of analog neuromorphic information processing technology.

Appendix A
Fitting of the Analog Mackey–Glass Nonlinearity

The electronic Mackey–Glass nonlinearity schematically shown in Fig. 1 is described by the transfer function

$$f(x) = \frac{ax}{1 + b^p x^p} \tag{10}$$

where a, b, and p are parameters that can be determined from the experimentally observed transfer function. Specifically, we use $a = 2.189, b = 0.001839$, and $p = 10.37$ along the numerical simulations. The resulting fit of the experimentally measured Mackey–Glass nonlinearity is shown in Fig. 11. Through analog-to-digital conversion, voltages are quantized to take values between 0 and 1023 (10-bit digitization), corresponding to a range of 0–5 V.

TABLE I
INDEXING OF BASIS FUNCTIONS

linear	quadratic	cubic
0	0,0	0,0,0
1	0,1	0,0,1
2	1,1	0,1,1
3	0,2	1,1,1
4	1,2	0,0,2
5	2,2	0,1,2

APPENDIX B
EVALUATION OF CAPACITIES AND THRESHOLD ESTIMATION

To evaluate capacities, we drive the Mackey–Glass dynamical system with i.i.d. input from a uniform distribution over $[-1, 1]$, and record the reservoir responses. We drive the reservoir for 3×10^5 inputs with a length T_{in} each, of which we discard the first 10^4 to wash out initial transients. Of the remaining data points, we use 5×10^4 as a test set for the evaluation of capacities, such that the linear estimators are trained to construct the basis functions using the remaining 2.4×10^5 data points. To find the output weights $\mathbf{W}^{\text{out}}$ that minimize the MSE in (7), we employ a linear regression with a Tikhonov regularization factor of $\alpha = 10^{-3}$ to avoid overfitting on the training set. Note that we include an additional output weight to account for a constant bias [53]. We normalize the reservoir state matrix before the regression. The same input and training procedures are considered for the experimental and numerical implementations for the sake of comparability.

In practice, as we are using a finite dataset, our capacity estimates are affected by a systematic positive error that leads to an overestimation of small capacities. Overestimation of small capacities has been treated in [25] by regarding capacities as significant only if they exceed a predefined threshold.

We now proceed to describe the procedure used to estimate degree-specific thresholds for the individual memory degrees. As in [25], we assume that capacities decrease monotonically with the increasing degree of the individual polynomials and with an increasing delay of the polynomial argument. Using this assumption, we enumerate the Legendre polynomials for a chosen degree as follows. We write the delays of the individual participating Legendre polynomials as tuples, for example, the quadratic basis function $P_1(u_{t-2})P_1(u_{t-3})$ that consists of the product of linear Legendre polynomials of the input delayed by 2 and 3 timesteps would be written as $(2, 3)$. Basis functions that consist of a higher degree polynomial evaluated for a single value of the delay are denoted by repeating indices in the tuple. For instance, the quadratic Legendre polynomial $P_2(u_{t-1})$ of the input delayed by 1 timestep is written $(1, 1)$. Similarly, the degree 3 basis function $P_1(u_{t-1})P_2(u_{t-2})$ is written as the tuple $(1, 2, 2)$. Table I displays the order we assign to index tuples constructed in this way: for a given index tuple, the successor is obtained by increasing the leftmost entry that is smaller than the entry to its right by one. If all tuple entries are equal, the rightmost entry is incremented and all other entries are set to 0.

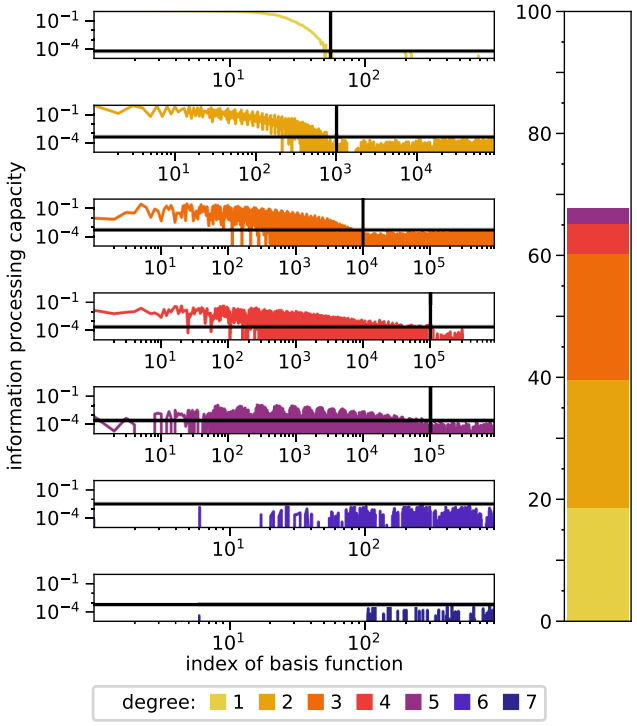


Fig. 12. Threshold estimation procedure for input scaling $\gamma = 60$, feedback scaling $\beta = 0.4$, and bias $\Phi = 575$. Vertical lines mark the last index that is considered significant. Horizontal lines mark the estimated threshold, i.e., the maximum capacity evaluated for an index after the last significant index. The right bar accounts for the total memory capacity, separated in the different degrees.

Fig. 12 shows the decay of capacities when the basis function index is increased. We observe a general decrease in evaluated capacity down to a bottom level around which capacities for later indices fluctuate. This bottom level is the noise floor of insignificant capacities originating from, e.g., the use of finite data or experimental noise. We ensure that the noise floor is excluded from the computation of capacities by identifying its starting index and determining the maximum evaluated capacity for later indices. In the heuristic capacity searches, we use twice the value of the bottom level as a threshold. The thresholds strongly decrease with the length of the training and testing sets, so getting as many data points as possible is advised. The choice of reservoir parameters (γ, β, Φ), however, does not have a significant influence on the estimates. Therefore, we use the parameters $\gamma = 60, \beta = 0.4$, and $\Phi = 575$ to obtain an estimate for experimental thresholds and simulation thresholds. The thresholds for the estimation of the information processing capacity (IPC) for each degree correspond to the horizontal lines in Fig. 12. The corresponding estimated thresholds up to degree 7 are shown in Table II. For each degree, as we go through the corresponding index set, increasing the delay of an individual basis function, we abort the search for capacities once the evaluated capacities are found to be insignificant for

TABLE II
ESTIMATED EXPERIMENTAL THRESHOLDS

degree	threshold
1	6×10^{-5}
2	4×10^{-4}
3	3.1×10^{-4}
4	4.1×10^{-4}
5	5×10^{-4}
6	3.3×10^{-3}
7	6×10^{-4}

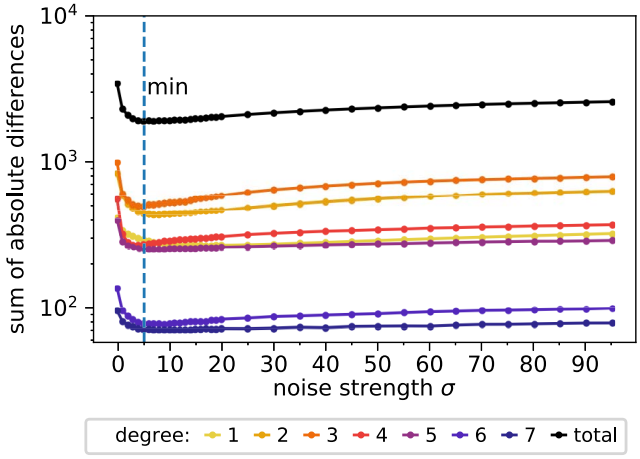


Fig. 13. Absolute differences between experiment and simulation information processing capacities as a function of the standard deviation σ of Gaussian noise added to reservoir states in the numerical simulations, see (9).

ten consecutive delay values. The total IPC is then given by the sum of all the measured capacities, as shown in the right bar of Fig. 12.

APPENDIX C
DETERMINING THE BEST MATCHING NOISE STRENGTH

To determine the best matching noise strength to reproduce the experimental results in simulation, we run simulations using parameters γ, β, and Φ as we used them in the experiments. After obtaining the reservoir states by iterating (3) and (4), we add Gaussian noise with a standard deviation σ in the range $[0, 95)$ to the reservoir states [see (9)]. We then evaluate capacities in a heuristic search using the thresholds estimated in an experiment to guarantee a fair comparison. As a measure of agreement, we take the sum of the absolute differences of all evaluated basis functions for all parameter sets.

Inspecting the bifurcation diagrams in Figs. 3 and 7 and the capacity profiles in Figs. 4, 5, and 8, we note that the dynamics in experiment seem to be consistently shifted by one value of the bias Φ to the right with respect to the numerical simulations. To acknowledge this shift, we do not compare experiment and simulation using the same parameter sets, but instead, compare basis functions for Φ appropriately shifted. Fig. 13 shows the result of this comparison by plotting the sum of absolute differences as a function of noise standard deviation σ. We observe that the differences quickly decrease when adding noise until we hit a minimum at $\sigma = 5$ and then slowly increase again.

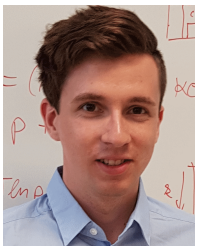
REFERENCES

[1] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *J. Math.*, vol. 58, nos. 345–363, p. 5, 1936.
[2] G. Tanaka *et al.*, "Recent advances in physical reservoir computing: A review," *Neural Netw.*, vol. 115, pp. 100–123, Jul. 2019.
[3] L. Appeltant *et al.*, "Information processing using a single dynamical node as complex system," *Nature Commun.*, vol. 2, no. 13, p. 468, Sep. 2011.
[4] M. S. Kulkarni and C. Teuscher, "Memristor-based reservoir computing," in *Proc. IEEE/ACM Int. Symp. Nanosc. Archit. (NANOARCH)*, Jul. 2012, pp. 226–232.
[5] M. C. Soriano *et al.*, "Delay-based reservoir computing: Noise effects in a combined analog and digital implementation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 2, pp. 388–393, Feb. 2015.
[6] D. Canaday, A. Griffith, and D. J. Gauthier, "Rapid time series prediction with a hardware-based reservoir computer," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 28, no. 12, Dec. 2018, Art. no. 123119.
[7] 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.
[8] 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, p. 1364, Jan. 2013.
[9] M. A. A. Fiers *et al.*, "Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 344–355, Feb. 2014.
[10] K. Vandoorne *et al.*, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Commun.*, vol. 5, p. 3541, Mar. 2014.
[11] J. Vatin, D. Rontani, and M. Sciamanna, "Experimental reservoir computing using VCSEL polarization dynamics," *Opt. Exp.*, vol. 27, no. 13, pp. 18579–18584, Jun. 2019.
[12] A. Argyris, J. Bueno, and I. Fischer, "Photonic machine learning implementation for signal recovery in optical communications," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 8487.
[13] K. Takano *et al.*, "Compact reservoir computing with a photonic integrated circuit," *Opt. Exp.*, vol. 26, no. 22, pp. 29424–29439, 2018.
[14] M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, "Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction," *Phys. Rev. X*, vol. 10, no. 4, Nov. 2020, Art. no. 041037.
[15] P. Antonik, N. Marsal, and D. Rontani, "Large-scale spatiotemporal photonic reservoir computer for image classification," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jun. 2020, Art. no. 2924138.
[16] R. M. Nguimdo, P. Antonik, N. Marsal, and D. Rontani, "Impact of optical coherence on the performance of large-scale spatiotemporal photonic reservoir computing systems," *Opt. Exp.*, vol. 28, no. 19, pp. 27989–28005, 2020.
[17] M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer, "Evolving carbon nanotube reservoir computers," in *Proc. Int. Conf. Unconventional Comput. Natural Comput.* Cham, Switzerland: Springer, 2016, pp. 49–61.
[18] K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, "Information processing via physical soft body," *Sci. Rep.*, vol. 5, no. 1, Sep. 2015, Art. no. 10487.
[19] J. Torrejon *et al.*, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, pp. 428–431, Jul. 2017.
[20] K. Kirby, "Context dynamics in neural sequential learning," in *Proc. Florida AI Res. Symp.*, 1991, pp. 66–70.
[21] P. F. Dominey, "Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning," *Biol. Cybern.*, vol. 73, no. 3, pp. 265–274, Aug. 1995.
[22] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks-with an erratum note," German Nat. Res. Center Inf. Technol., Bonn, Germany, GMD Tech. Rep. 148, 2001, p. 13, vol. 148, no. 34.
[23] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comput.*, vol. 14, no. 11, pp. 2531–2560, 2002.
[24] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, 2009.

[25] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Sci. Rep.*, vol. 2, Apr. 2012, Art. no. 514.
[26] H. Jaeger, "Short term memory in echo state networks," GMD-German Nat. Res. Inst. Comput. Sci., Bonn, Germany, GMD Rep. 152, 2002.
[27] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, "Optimal nonlinear information processing capacity in delay-based reservoir computers," *Sci. Rep.*, vol. 5, no. 1, Oct. 2015, Art. no. 12858.
[28] K. Harkhoe and G. Van der Sande, "Task-independent computational abilities of semiconductor lasers with delayed optical feedback for reservoir computing," *Photonics*, vol. 6, no. 4, p. 124, Dec. 2019.
[29] T. Kubota, H. Takahashi, and K. Nakajima, "A unifying framework for information processing in stochastically driven dynamical systems," 2019, *arXiv:1906.04608*. [Online]. Available: http://arxiv.org/abs/1906.04608
[30] M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer, "A substrate-independent framework to characterize reservoir computers," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 475, no. 2226, Jun. 2019, Art. no. 20180723.
[31] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Exp.*, vol. 20, no. 20, pp. 22783–22795, 2012.
[32] J. Pauwels, G. Verschaffelt, S. Massar, and G. Van der Sande, "Distributed Kerr non-linearity in a coherent all-optical fiber-ring reservoir computer," *Frontiers Phys.*, vol. 7, p. 138, Oct. 2019.
[33] I. Estébanez, I. Fischer, and M. C. Soriano, "Constructive role of noise for high-quality replication of chaotic attractor dynamics using a hardware-based reservoir computer," *Phys. Rev. A, Gen. Phys.*, vol. 12, no. 3, Sep. 2019, Art. no. 034058.
[34] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Trans. Circuits Syst.*, vol. CAS-32, no. 11, pp. 1150–1161, Nov. 1985.
[35] 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, no. 15, Oct. 2018, Art. no. 152004.
[36] L. Appeltant *et al.*, "Reservoir computing based on delay-dynamical systems," Ph.D. dissertation, Appl. Phys. Res. Group, Instituto de Física Interdisciplinar y Sistemas Complejos, Vrije Univ. Brussel, Ixelles, Belgium, 2012.
[37] S. Ortín, L. Pesquera, G. Van der Sande, and M. C. Soriano, "Time delay systems for reservoir computing," in *Photonic Reservoir Computing: Optical Recurrent Neural Networks*, D. Brunner, M. C. Soriano, and G. Van der Sande, Eds. Berlin, Germany: Walter de Gruyter GmbH & Co KG, 2019, ch. 5, pp. 117–152.
[38] Y. Paquot *et al.*, "Optoelectronic reservoir computing," *Sci. Rep.*, vol. 2, no. 287, Feb. 2012, Art. no. 287.
[39] P. Antonik, M. Haelterman, and S. Massar, "Brain-inspired photonic signal processor for generating periodic patterns and emulating chaotic systems," *Phys. Rev. A, Gen. Phys.*, vol. 7, no. 5, May 2017, Art. no. 054014.
[40] F. Stelzer, A. Röhm, K. Lüdge, and S. Yanchuk, "Performance boost of time-delay reservoir computing by non-resonant clock cycle," *Neural Netw.*, vol. 124, pp. 158–169, Apr. 2020.
[41] S. Ortín and L. Pesquera, "Delay-based reservoir computing: Tackling performance degradation due to system response time," *Opt. Lett.*, vol. 45, no. 4, pp. 905–908, 2020.
[42] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
[43] A. Namajūnas, K. Pyragas, and A. Tamaševičius, "An electronic analog of the Mackey–Glass system," *Phys. Lett. A*, vol. 201, no. 1, pp. 42–46, 1995.
[44] B. Mensour and A. Longtin, "Chaos control in multistable delay-differential equations and their singular limit maps," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 58, no. 1, p. 410, 1998.
[45] L. Junges and J. A. C. Gallas, "Intricate routes to chaos in the Mackey–Glass delayed feedback system," *Phys. Lett. A*, vol. 376, nos. 30–31, pp. 2109–2116, Jun. 2012.
[46] C. Gallicchio and A. Micheli, "Architectural and Markovian factors of echo state networks," *Neural Netw.*, vol. 24, no. 5, pp. 440–456, 2011.
[47] M. C. Soriano *et al.*, "Optoelectronic reservoir computing: Tackling noise-induced performance degradation," *Opt. Exp.*, vol. 21, no. 1, pp. 12–20, Jan. 2013.
[48] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, "Re-visiting the echo state property," *Neural Netw.*, vol. 35, pp. 1–9, Nov. 2012.
[49] C. Langton, "Computation at the edge of chaos: Phase transition and emergent computation," Los Alamos Nat. Lab., Los Alamos, NM, USA, Tech. Rep. LA-UR–90-379, 1990.
[50] N. H. Packard, "Adaptation toward the edge of chaos," *Dyn. Patterns Complex Syst.*, vol. 212, p. 293, Apr. 1988.
[51] M. C. Soriano, D. Brunner, M. Escalona-Morán, C. R. Mirasso, and I. Fischer, "Minimal approach to neuro-inspired information processing," *Frontiers Comput. Neurosci.*, vol. 9, p. 68, Jun. 2015.
[52] L. Keuninckx, M. C. Soriano, I. Fischer, C. R. Mirasso, R. M. Nguimdo, and G. Van der Sande, "Encryption key distribution via chaos synchronization," *Sci. Rep.*, vol. 7, no. 1, Apr. 2017, Art. no. 43428.
[53] M. Lukoševičius, "A practical guide to applying echo state networks," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 659–686.

Benedikt Vettelschoss was born in Göttingen, Germany, in 1992. He received the B.Sc. degree in cognitive science from Osnabrück University, Osnabrück, Germany, in 2019, and the M.Sc. degree in physics of complex systems from the University of the Balearic Islands, Palma, Spain, in 2020. He is currently pursuing the Ph.D. degree with Ghent University, Ghent, Belgium.

His research interests include information processing in dynamical systems, (physical) reservoir computing, and biological self-organization.



André Röhm was born in Wolgast, Germany, in 1989. He received the Ph.D. degree in physics from Berlin Institute of Technology, Berlin, Germany, in 2018.

He is currently a Project Researcher with the Department of Information Physics and Computing, The University of Tokyo, Tokyo, Japan. He has worked on the modeling of quantum-dot-based laser devices, laser networks, reservoir computing, and symmetry breaking in coupled oscillators.

Miguel C. Soriano (Senior Member, IEEE) was born in Benicarlo, Spain, in 1979. He received the Ph.D. degree in applied sciences from Vrije Universiteit Brussel, Brussels, Belgium, in 2006.

He currently holds a tenure-track position at the University of the Balearic Islands, Palma, Spain. He has authored or coauthored over 65 research papers in international refereed journals. His main research interests include nonlinear laser dynamics and information processing based on reservoir computing.

Dr. Soriano is a member of the IEEE Photonics Society.