VNIVERSITAS BALEARICA

Universitat
de les Illes Balears

Information processing using optoelectronic delayed systems: Influence of an additional delay

Jesús Yelo Sarrión

Master's Thesis

Master's degree in Physics of Complex Systems
at the
UNIVERSITAT DE LES ILLES BALEARS

Academic year 2018/2019

12 July, 2019

UIB Master's Thesis Supervisor: Miguel Cornelles Soriano

Abstract

In this MSc thesis, we study the performance of a delayed optoelectronic reservoir computer. We focus our atention on the system's optimization through the addition of a second temporal delay. We find that the addition of this delay has a positive influence for some benchmark tasks while it remains neutral in other cases. In particular, we evaluate linear and nonlinear memory capacities as well as nonlinear time series prediction of a chaotic Mackey-Glass system and NARMA (*Nonlinear Auto Regressive Moving Average Models*) systems.

First, we carry out a theoretical analysis of the reservoir computer, and precisely of the delay-based RC. The Chapter 2 is then devoted to the numerical implementation of the delayed optoelectronic oscillator. We simulate the addition of a second delay and study its effects through the benchmark tasks. Finally, in Chapter 3 we evaluate experimentally an optoelectronic hardware-based reservoir computer with one and two delays. During this experimental work, besides carrying out the different benchmark tasks, we address the reservoir computer's characterization and look for the best set of parameters.

Acknowledgements

There are no words to express the unconditional support that I have received from my supervisor Miguel. From the very beginning you guided me with the numerical simulations, the experimental setup and until now, the very last day, with the correction of these pages.

I would also like to thank my classmates, especially Lorena for supporting me in all my decisions and being always there for a talk. Also Maria and Pablo for being my family in the island.

Thank you to Pau for your patience and for helping me, so many times, with the experimental setup. Ingo, thank you for listening my progress and for your good advice, also thanks to Apostolos for the calibration of the Mach-Zehnder controller.

Finally, thank you to my parents, if I'm here its because of you, for your support, for helping me in all the decisions I made.

Contents

Chapter 1

Introduction

1.1 History

1.1.1 Computation

In the period between World War I and II, first information processing units were built in order to achieve the increasingly more complex tasks that were no longer possible to be solved by hand. Since then until the beginning of the XXI Century, the scaling law for digital technology proposed by Gordon Moore has revealed as a complete success to keep improving information processing power. Nevertheless, Moore's law cannot be sustained by modern solid-state physics. Multiple CPU cores can't solve the scaling problem either, which is limited by other technical issues as energy dissipation and manufacturing economics [1]. Nowadays, new unconventional information processing techniques are trying to overcome all these challenges. One of the fields in which major efforts are being invested is optical information processing.

1.1.2 Optical Computation

As of 1970, a different vision of information processing emerged based on optical modern theories. In particular, it relied on nonlinear effects or bistability [2, 3]. However, the interest on those devices rapidly decayed due to high power consumption and difficulties when it came to integrating those optical-based systems into already existing machines.

It is the halted acceleration of microprocessor computing power and the parallel evolution of photonic technologies and quantum computing that has reawakened the general interest on opportunities offered by optical information processing [4, 5].

In the same way, when moving to supercomputing needs, the outstanding capability of the light to be generated and propagated with minimum power dissipation, seems to be the answer to heat generation issues, and the key to a lower energy consumption [6].

1.2 Reservoir Computer

1.2.1 Birth of the RC

Taking all the previous considerations into account, huge efforts have been invested in novel computational concepts regarding information processing. These at-

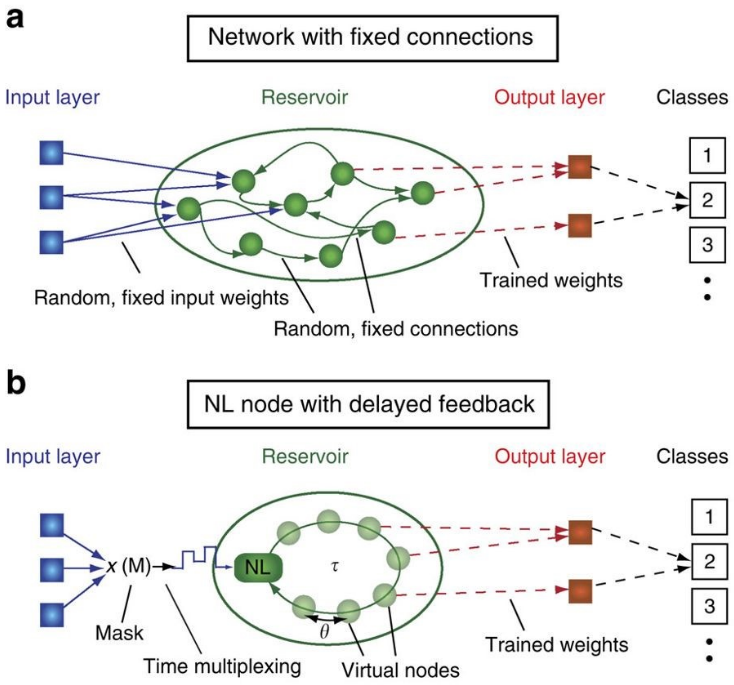


Figure 1.1: **Sketch of RC schemes.** (**a**) Classical RC scheme. (**b**) Scheme of RC utilizing a nonlinear node with delayd feedback. Figure taken from [13].

tempts have given rise to a radically different way of processing information, distant from traditional Turing or von Neumann Machine methods, but as computationally powerful as the transistor-based models [7]. One of these concepts to compute in an unconventional way using random recurrent neural networks was proposed in 2004 which was first introduced as Echo State Network [8] and more generally as reservoir computing (RC) [9]. More recently, L. Larger et al. and Y. Paquot et al., concurrently proposed a way to implement this concept on photonic hardware, the optoelectronic reservoir computer [10, 11]. Reservoir computing is a breakthrough concept, inspired by the brain's ability to process information combined with the advantages of recurrent neural networks and the high performance of machine learning techniques. This bio-inspired approach has demonstrated state-of-the-art performance for computationally demanding tasks such as chaotic time series prediction or speech recognition, among others [10, 11, 12].

1.2.2 Conventional RC

A conventional Reservoir Computer, as shown in Fig. 1.1(a) is made of three specific sections, each one of them in charge of a specific role. The so-called "Input layer" receives the input data. The "reservoir" encloses a recurrent neural network made up of a large number of fixed and randomly interconnected nonlinear nodes, that in consequence, also exhibit internal feedback loops. The signal is sent from the input layer to a given number N of reservoir nodes through random weights w_N. When excited by the input signal, the network not only processes the information but also exhibits complex transient dynamics, *i.e.* states in which no stable equilibrium is reached. It is this response of the network that is evaluated by the "output layer" via a linear weighted sum of N individual node states.

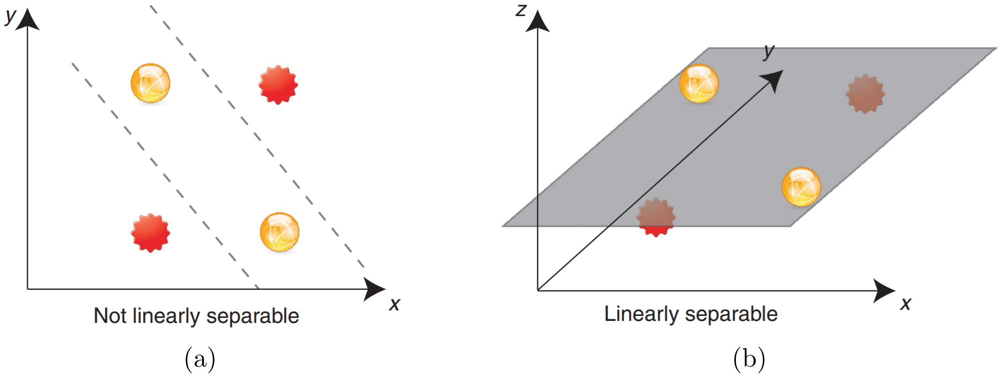


Figure 1.2: Illustration of linear separability. A non linear mapping from a small dimensional space into a high dimensional space facilitates classification. (a) Two dimensional space where red stars and blue spheres cannot be separated with a single straight line. (b) In a three dimensional space, spheres and stars can be separated by a single linear hyperplane. Figure taken from [13].

The transient dynamical response of the reservoir, plays a major role in processing the information from the input layer. Without input, the reservoir is typically set to operate in an asymptotically stable (fixed point) state, *i.e.* not too far from a bifurcation point. Those behaviors have also been found in physiological systems from where the RC is inspired [14].

The particular characteristics of the reservoir and the large number of dynamical elements in it pave the way to perform any kind of complex task. Despite the already mentioned degrees of freedom, a training procedure of the weights is still needed in order to perform any computational task.

Being the training of the recurrent neural network the main difficulty of this procedure, the weights between nodes in RC are kept fixed, thus untrained, as the ones linking the input layer to the reservoir. Consequently, there is only need to train the output layer with the help of already known signals in a supervised learning procedure. As a result, the Reservoir Computer is then able to process unseen signals.

1.2.3 Properties of a RC

A key issue of the Reservoir Computer is the specific nonlinear transformation of the input signal in order to efficiently solve a given task. To this end, a reservoir should fulfill several properties.

Firstly, the reservoir must be able to nonlinearly transform the (usually) low-dimensional input signal into a high-dimensional state space. This new configuration is obtained by means of the large number of interconnected reservoir nodes in the recurrent neural network. Thanks to this high-dimensional space, the different tasks can be performed in a more efficient way, *e.g.* the classification of states [13]. This high-dimensional mapping is illustrated in Fig. 1.2.

In order to obtain not only robust results but also reproducible, if the inputs are similar, the transient response of the system must be similar for similar outputs. This property is known as approximation property. In the same way, for different inputs, the transient response should be different enough (separation property) [13].

Finally, the RC should also exhibit a fading-memory (also known as short-term memory). The idea is that recent inputs of the signal need to have an influence on the present reservoir state. Depending on the parameters of the system, the number of recent inputs that the system can remember varies. The fading memory property can turn out to be crucial depending on the nature of the performed task. For instance, in the case of a time-series prediction, the upcoming step is usually related to a given number of previous inputs. Thus, and in order to reach an optimal prediction, the system should have the ability to retain those inputs [9].

1.2.4 One single NL hardware node RC

Even if highly demanding tasks can be performed, the previously described concept of RC remains technologically challenging and often unrealistic. Here we will focus on a specific architecture, and more efficient, concept of RC. The recurrent Neural Network, typically composed of a large number of nodes, can be substituted by a dynamical system comprising a single nonlinear node plus a delayed self-feedback [13]. In this equivalent system, as shown in Fig. 1.1(b), the reservoir is built within the delay line. The resulting loop is divided into N "virtual nodes" receiving data from the input layer via time multiplexing. Consequently, the addition of this delay τ results in a large dimensionality of the dynamics of the system. Even after this strong simplification of the hardware, the system fulfills the requirements for proper operation as high dimensionality and fading memory [13].

1.3 Delay-based Reservoir Computer

1.3.1 Description

In this section, we will perform an in-depth description of the RC's delayed scheme. Having changed the Recurrent Neural Network for a delayed line of time τ, now N equidistant points separated in time by $\theta = \tau/N$ act as virtual nodes as shown in Fig. 1.3. When the information is sent to the nodes, each of them acquire a value that characterize the transient response of the reservoir to a certain input at a given time. Being T the characteristic time of the nonlinear node, the condition $\theta < \tau$ is preferred if one wants to exploit the bandwidth of the system. In this case the system does not have the time to reach an asymptotic value. Therefore the states of the virtual nodes depend on the states of neighbouring nodes.

As in the traditional case, the delay-based approach is also characterized by three main different sections. In the input layer, a continuous $u(t)$ or discrete $u(k)$ flow of data is transformed into a continuous function $\mathbf{I}(t)$, where each step is held for a time τ. These states are, one by one, multiplied by a random mask $\mathbf{J}(t) = \mathbf{M}(t)\mathbf{I}(t)$ as illustrated in Fig. 1.4. Now, in the reservoir, the new stream of data $\mathbf{J}(t)$ is added to the previous one $\mathbf{x}(t - \frac{\tau}{N}(N-i))$ and then computed by the nonlinear function F. This final set of data can be delivered to the virtual nodes and the operation is repeated again. The system would be governed by the following equation:

$$\mathbf{x}(t) = F\left(\gamma \mathbf{J}(t) + \beta \mathbf{x}(t-\tau) + \phi\right) \tag{1.1}$$

where τ is the delay time, β and γ are the feedback and input scaling parameters and ϕ is the phase of the system.

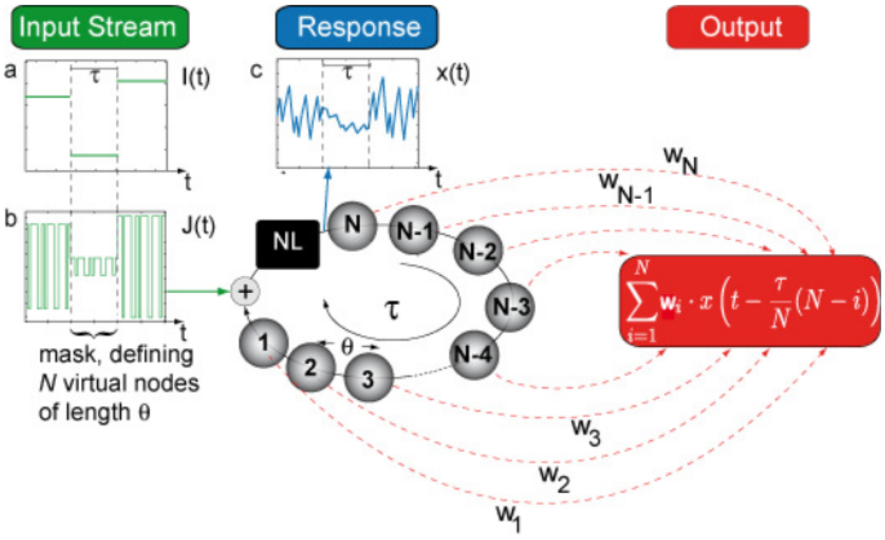


Figure 1.3: Schematic representation of reservoir computing with a delayed feedback system and a single nonlinear node.

The output layer remains unchanged when compared with the classical RC. After being trained, each of the virtual nodes is related with a given weight $\mathbf{w}_i$. The output data can be described as a linear combination of the states of the nodes:

$$\mathbf{o}(t) = \sum_{i=1}^{N} \mathbf{w}_i \cdot \mathbf{x}(t - \frac{\tau}{N}(N-i)) \tag{1.2}$$

The whole process, highlighting the output layer weights, is described in Fig. 1.3. In the training procedure, $\mathbf{o}(t)$ must minimize the distance to a given target sequence $\mathbf{y}(t)$.

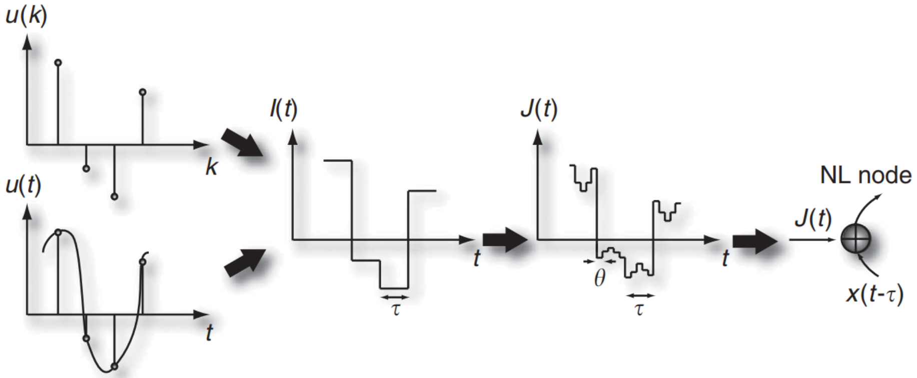


Figure 1.4: Scheme of the input data preparation and masking procedure. Starting either from a time-continuous or time discrete input stream, after going under a sample and hold operation, we obtain a constant stream $\mathbf{I}(t)$. After being multiplied by a random mask $\mathbf{M}(t)$, it is feed to the virtual nodes. Figure taken from [13].

1.4 Benchmark tasks

1.4.1 Memory capacity

The fading memory is the property to retain a certain number of inputs within the reservoir activity after a certain time. In order to compare the fading memory of different systems, a quantitative value of the memory capacity is obtained by a simple methodology. In this case, the input to the reservoir is set as an identically distributed random sequence, $u(n)$, uniformly distributed between $[-1, 1]$. Setting the output target of the RC as the "i" previous input, $y_i(n) = u(n - i)$, proceeding in this way, the system output after training will give as a result the prediction, $o_i(n)$ of an input that was presented a time "i" before at the reservoir input. The memory function $m(i)$ is described as follows:

$$m(i) = corr[o_i(n), y_i(n)] = \frac{< y_i(n)o_i(n) >_n^2}{\sigma^2(y_i(n))\sigma^2(o_i(n))} \tag{1.3}$$

where $<>_m$ is the mean over all n values and σ denotes the standard deviation. The memory capacity MC is then defined as the sum of the memory function $m(i)$, with i going to infinity:

$$MC = \sum_i^\infty m(i) \tag{1.4}$$

The MC gives a good understanding of the system's ability to retain previously seen values [15].

1.4.2 Mackey-Glass time series prediction

The performance of the RC will be further tested through a nonlinear time series prediction of the chaotic Mackey-Glass system (MG) [16]. The dynamics followed by such a system are governed by a delay differential equation:

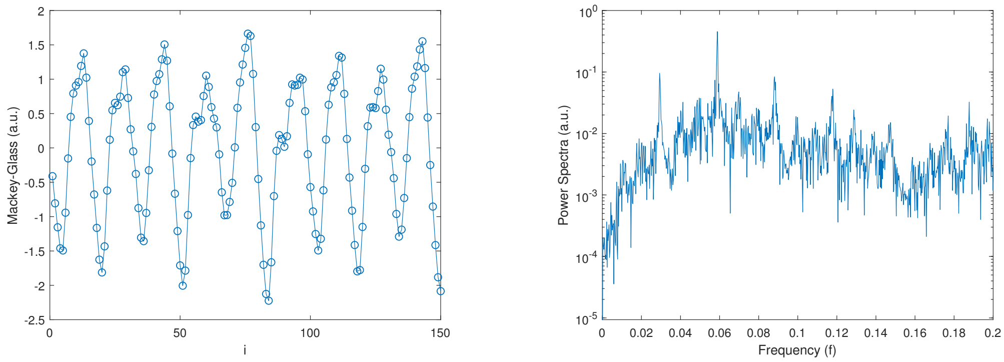
$$\dot{z}(t) = \frac{az(t-\tau)}{1 + z^{10}(t-\tau)} - bz(t),\ t \geq 0 \tag{1.5}$$

The typical constant values introduced to the system are $a = 0.2$, $b = 0.1$ and $\tau = 17$ [12]. A time series $u(n) = z(nT)$, $n = 1$, ..., 4000 of discrete points with sampling time $T = 3$ is obtained. In general, the time series is divided in two different sets: 3000 points for training and 1000 for the test of the system's performance. A segment of the MG time series can be seen in Fig. 1.5(a) and the corresponding power spectra in Fig. 1.5(b).

The main objective is to optimize the one step-ahead prediction of the RC using the discrete values $y(n) = u(n + 1)$ of the Mackey-Glass chaotic system. Being $\tau = 17$ the parameter that provides information about the delayed component and the sampling time $T = 3$, the minimum memory capacity required for this problem is approximately $MC \approx 6$, $(MC\ > \tau/T)$.

1.4.3 NARMA tasks

NARMA (*Nonlinear Auto Regressive Moving Average Models*) tasks are commonly used benchmarks in the field of reservoir computing for evaluating the computational capability of the learning system and specially the memory capacity [17,



(a) MG system

(b) FFT

Figure 1.5: (a) Time series of the Mackey-Glass chaotic system. The circles denote the samples that are kept in $u(t)$. In this time series, three different pseudo-periodicities are visible with periods $\Delta_1 = 3\tau/T$, $\Delta_2 = 6\tau/T$ and $\Delta_3 = 12\tau/T$. (b) Fast Fourier Transform (FFT) of the whole set of data. The power spectra shows frequency resonances at $1/\Delta_i$.

18, 19, 20]. To emulate the desired outputs, the RC must have enough memory of historical inputs and transform them nonlinearly. For the NARMA task, the input $u(k)$ of the system consists of scalar random numbers drawn from a uniform distribution in the interval [0, 0.5] and the target $y(k+1)$ is given by a nth-order recursive formula. In this case, we compared the performance of the system for $n = 2$ and 5. The first one is called NARMA2 and is described by:

$$y_{k+1} = 0.4y_k + 0.4y_k y_{k-1} + 0.6u_k^3 + 0.1 \tag{1.6}$$

In turn, the NARMA5 task evaluates a 5^{th}-order nonlinear dynamical system. It's written as follows:

$$y_{k+1} = 0.3y_k + 0.05y_k \left(\sum_{j=0}^{4} y_{k-j} \right) + 1.5u_{k-4}u_k + 0.1 \tag{1.7}$$

The purpose of both tasks is to approximate the $y(k+1)$ value given $u(k)$ as the input to the reservoir.

1.5 Objectives

This project aims at exploring the influence of an enhanced node connectivity in the performance of delay-based RC. For this purpose, different strategies have been suggested as the addition of multiple delays or tailoring the input mask amongst other [21, 22]. Here we will focus on the addition of a second delay and we will evaluate it in an optoelectronic RC implementation. In delay-based RC, the first delayed loop is usually added in such a way that each of the virtual nodes is connected to the direct neighbour.

In this context, the enhanced connectivity provided by the addition of a second delay is illustrated in Fig. 1.6. Each one of the virtual nodes is then connected to two different nodes in previous temporal states. Proceeding in this way, we obtain a

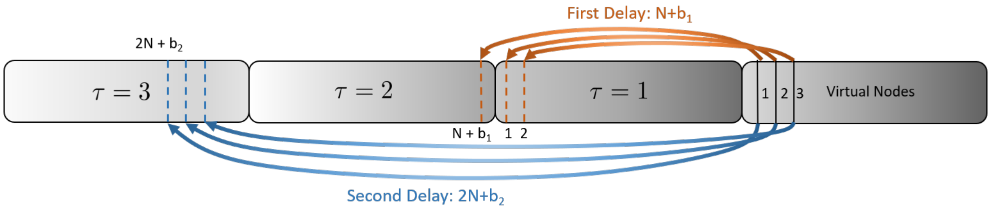


Figure 1.6: Schematic representation of the virtual node connectivity in a delay-based reservoir with two delays. The parameter b_1 allows to shift the first delay, only positive b_1 values are allowed if one wants to create recurrent connections. The condition for the second delay is less restrictive and b_2 can take both positive or negative values.

reservoir computer with two different delays and a richer virtual node connectivity. The equation governing the RC is therefore given by:

$$\mathbf{x}(t) = F\left(\gamma\mathbf{J}(t) + \frac{\beta}{2}(\mathbf{x}(t-\tau_1) + \mathbf{x}(t-\tau_2)) + \phi\right) \tag{1.8}$$

This new approach allows to manipulate the connectivity of the reservoir and to explore systematically the influence of having a second delay for each of the performed tasks.

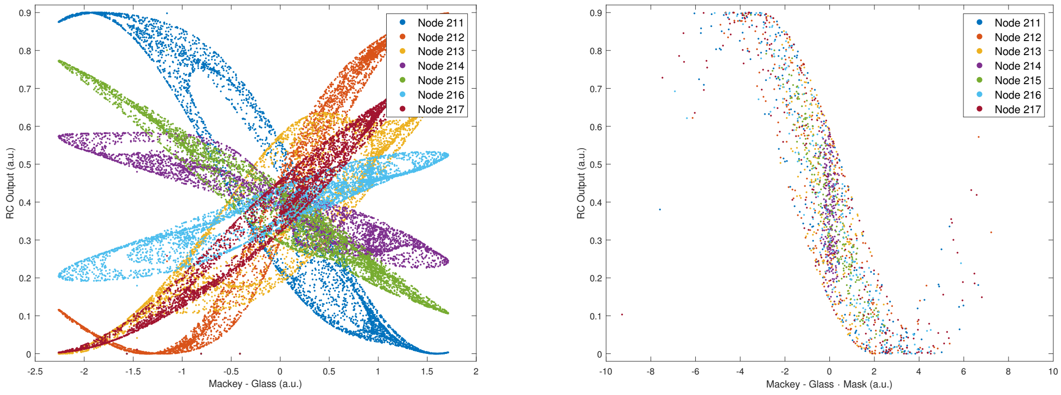
In order to study the influence of the addition of a second delay, in Chapter 2 we will perform numerical simulations. In this part, we will evaluate numerically the response of an optoelectronic reservoir with one and two delays to the benchmark tasks explained in Sec. 1.4. In Chapter 3, we will characterize the dynamics of a novel hardware setup of an optoelectronic reservoir computer that allows to add multiple delays. Finally, we will test this experimental setup with the different benchmark tasks.

Chapter 2

Numerical Simulations

2.1 Optoelectronic Reservoir Computer

We focus our work on the study of a delayed optoelectronic oscillator which has a sine squared nonlinearity [10, 11]. The delay-based approach to RC usually has a ring network connectivity [11, 12].



(a)

(b)

Figure 2.1: Plot of selected virtual nodes of the delayed optoelectronic oscillator for $\gamma = 0.2$, $\beta = 0.6$, $\kappa = 0.9$ and $\phi = 0.7\pi$. In panel (a), the RC output is plotted against the value of $\mathbf{I}(k)$ of the Mackey-Glass time series. In panel (b), the Mackey - Glass values are multiplied by the random mask $\mathbf{M}(k)$ with gaussian distribution.

The network connectivity formed by the virtual nodes of the RC can play an important role in the system performance as it will be shown in the following. Not only we can play with the value of the first delay but we can also enhance the connectivity with the addition of the second one. The value of the first delay has usually been set to $N + 1$ [12] but here we vary the time-shift in order to evaluate its impact. In the case $\tau = N + 1$, the virtual neurons, $\mathbf{x}_i(k) = z(kN + i)$, form a ring topology and can be described by:

$$\mathbf{x}_i(k) = \kappa \sin^2 \left(\gamma \mathbf{J}_i(k) + \beta \mathbf{x}_{i-1}(k-1) + \phi\right), \ i = 2, \ldots, N \tag{2.1}$$

and $\mathbf{x}_1(k) = \kappa \sin^2 \left(\gamma \mathbf{J}_1(k) + \beta \mathbf{x}_N(k-2) + \phi\right)$ where $\mathbf{J}(k)$ is the input stream of data $\mathbf{I}(k)$ multiplied by a gaussian random mask. Thus the nonlinear projected space is composed of the N consecutive outputs of the system at each time step, $\mathbf{x}_i(k) = (\mathbf{x}_1(kN), \ldots, \mathbf{x}_N(kN))$.

Each one of the consecutive outputs of the system can be plotted in the space defined by the input-output mapping. In Fig. 2.1(a), the RC's outputs are plotted against the Mackey - Glass input values and in Fig. 2.1(b) against the values of $\mathbf{J}(k) = \mathbf{I}(k) \cdot \mathbf{M}(k)$. In the latter representation, the sinusoidal nonlinearity becomes apparent. In this same figure, the width of the curve is related with the value of the parameter β while the number of explored periods of the sine squared is linked to the parameter γ. It is by changing the value of the phase ϕ that the curve can be shifted along the sine squared nonlinearity.

After the addition of a second delay, we will evaluate the memory capacity, the nonlinear time-series prediction and the NARMA task for different positions of both delays.

2.2 Memory Capacity

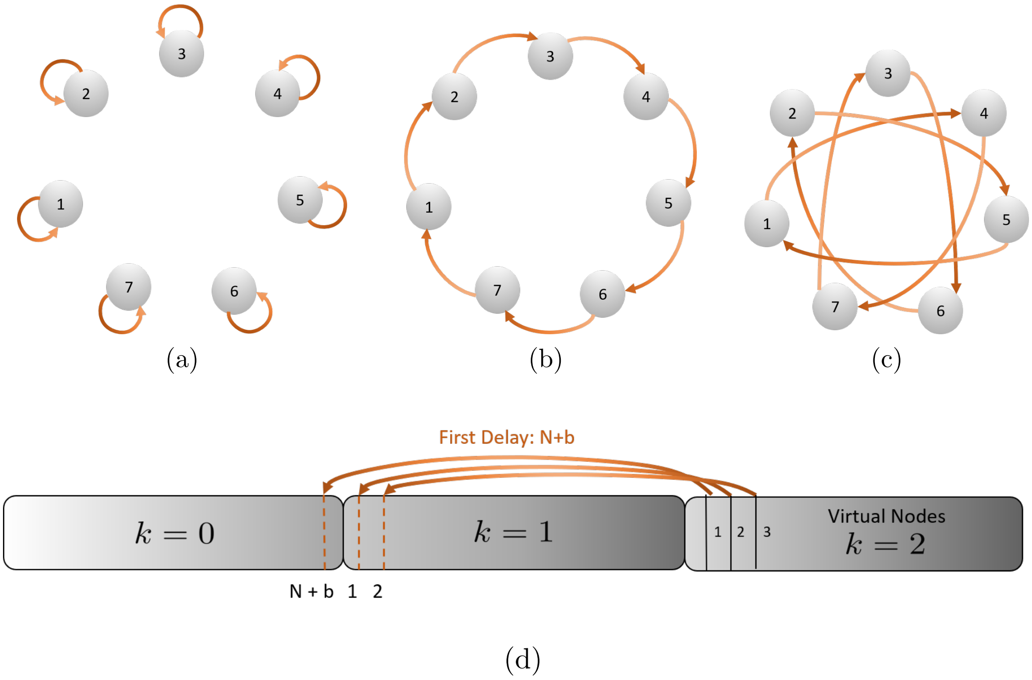
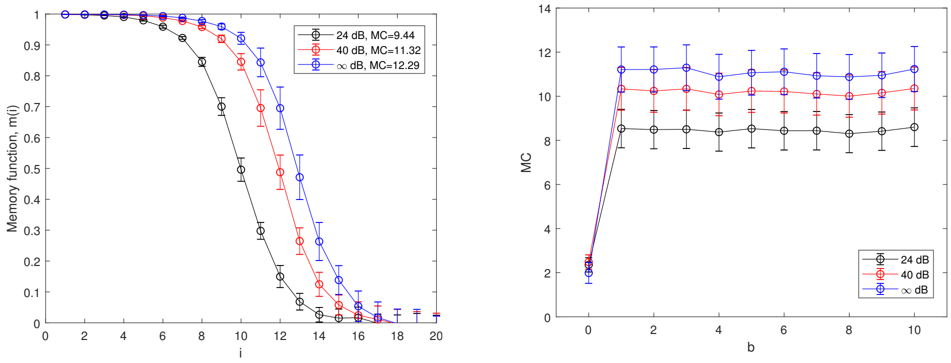


Figure 2.2: Nodes distribution in a 7 nodes network with first delay set to: (a) $\tau = N$ *i.e*, each node i receives information from itself, (b) $\tau = N + 1$ where all the nodes are connected to their left neighbour and (c) $\tau = N + 3$. All these figures are equivalent to the temporal representation depicted in (d) setting $b = 0$ for (a), $b = 1$ for (b) and $b = 3$ for (c).

Firstly, we numerically evaluate the behavior of the memory function for the optoelectronic system with one single delay. In Fig. 2.2, we show two examples of different network configurations that can be built using one delay and different time shifts between the length of the reservoir (N) and the value of the delay (τ). In Fig. 2.2(b), we show the resulting configuration of having $\tau = N + 1$, *i.e.* a connectivity from node i in $k = 2$ to the $i - 1$ node in a precedent state $k = 1$

($N + 1$ previous nodes). However, other configurations can be obtained with the delay-based RC and other time shift. For instance, in Fig. 2.2(c) we show how node i is connected with the $i - 3$ node in the previous state ($N + 3$ previous nodes). Those figures are equivalent to the temporal representation shown in figure (d).

In the following, we show the numerical results for the memory function and the memory capacity (see Section 1.4.1) when the delay is shifted from N to $N + 10$. Figure 2.3(a), shows the memory function for the ideal noise-free system (∞ dB) as well as the influence of noise in the memory function of the system: the higher the signal-to-noise ratio (SNR), the longer will be the fading memory. The analysis of the SNR is performed in order to compare with the experimental set up in the next chapter. Here we show how a different value of the SNR gives rise to different results. More specifically, the noise degrades the memory capacity of the system.



(a) Memory function

(b) Memory capacity

Figure 2.3: (a) Shape of the memory function for $b = 1$, *i.e.* with a delay set to $N + 1$. Figure done with $N = 246$, $\beta = 1.0$, $\gamma = 0.3$, $\kappa = 0.9$ and $\phi = 0.4\pi$. The numerical evaluation has been done for different values of the SNR. (b) Memory capacity for different values of b and for three values of the SNR. The other parameters are equivalent to those from (a).

In Fig. 2.3(b), the memory capacity in function of the parameter b, *i.e.* when the first delay is set to $N + b$, is shown. For $b = 0$, the memory capacity is close to zero since the virtual nodes are not connected as it can be seen in Fig. 2.3(a). Nevertheless, for $1 < b < 10$ the numerical simulations show that the MC is approximately constant. Thus, relating those results to Fig. 2.2, we see that no matter how the connectivity of the network is, for a given SNR and with one single delay, the memory capacity will remain constant, provided $b \geq 1$.

Addition of a second delay

The influence of the recent inputs varies with the addition of a second delay. In the following, the first delay will be set at $N + 1$, and the second one will be shifted. The second delay at $2N + b$ can be set in a given reference state varying the parameter b. Even with the addition of the second delay, the temporal information about the input that the system can remember after a certain time still depends on the SNR.

In this section, it will be shown that the addition of a second delay improves the memory capacity of the Reservoir Computer. In order to show this, a numerical

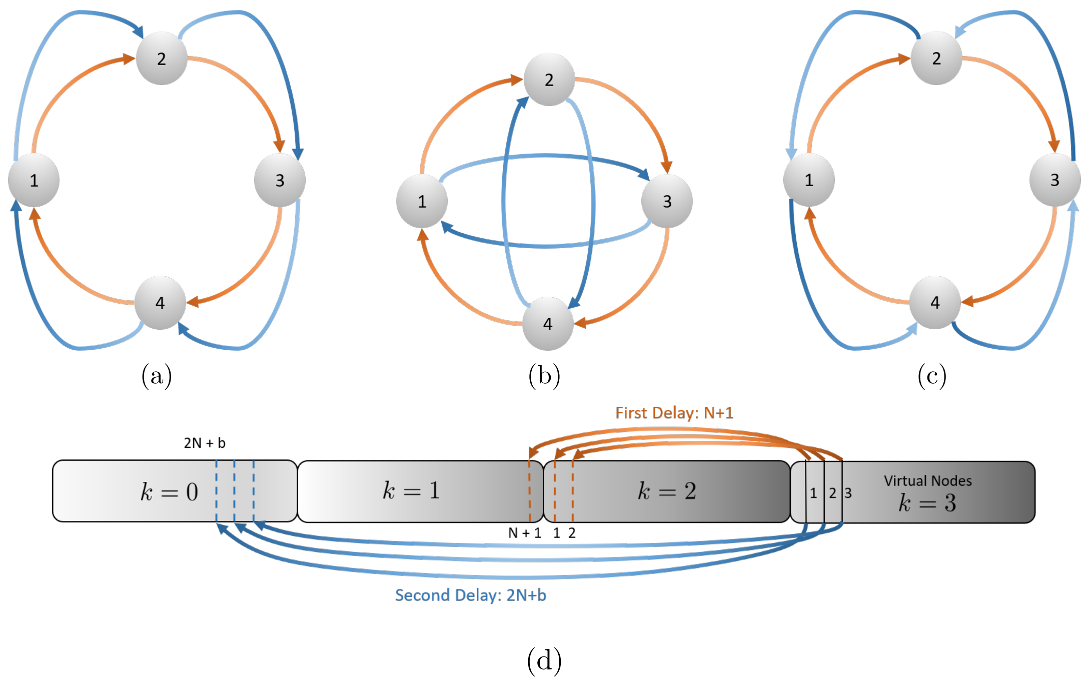
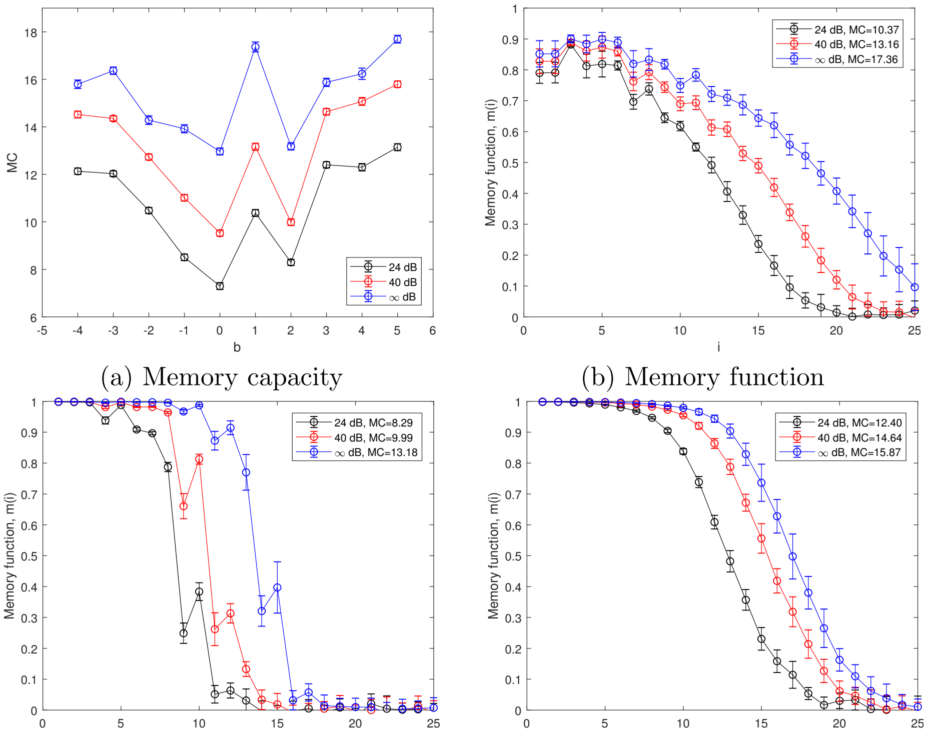


Figure 2.4: Nodes distribution in a 4 nodes network with first delay (red) set at $N + 1$ and the second delay (blue) in (a) $2N + 1$, (b) $2N + 2$, (c) $2N + 3$. For a) and (c), the connectivity is the same but with the direction of the second delay inverted. Interestingly, in (b), the second delay is divided in two different clusters that are disconnected between them.

simulation has been performed. The second delay was set as shown in Fig. 1.6. Varying the parameter b, the MC also varies. The evaluation has been done for different values of the SNR.

In this case, as seen in Fig. 2.5(a), when varying the parameter b for the second delay, the memory capacity for a given SNR suffers important changes. It decreases when approximating to $b = 0$ from negative values, for $b = 1$ there is a large peak and then it decreases for $2N + 2$. After this value of the time shift ($b \geq 3$), the memory capacity starts to increase again and eventually saturates.

For $b = 0$ each node of the second delay is connected with itself in a previous state, and therefore there is no transfer of information between the neighbouring nodes and the memory is reduced. The memory function has been plotted for $b = 1$, 2 and 3 in Figs. 2.5(b)-(d). For $b = 1$, the first and second delays are in $N + 1$ and $2N + 1$ respectively. As seen in Fig. 2.5(b), the memory function does not reach the highest value but it is kept for several values i around 0.9 and then it decreases slowly. In the case $b = 2$ (Fig. 2.5(c)), the memory function starts constant at the highest value but, *e.g.* for the highest value of the SNR, it suddenly collapses for $i \approx 10$. This memory capacity loss is due to the division of the reservoir connectivity provided by the second delay in two different clusters. This phenomenon occurs for all even N and $b = 2$, as in this case with $N = 246$. A schematic representation of this type of connectivity is shown in Fig. 2.4(b). In the case $b = 3$, the memory capacity is partially recovered and the shape of the memory function is similar to the one obtained for one single delay. For $b \geq 3$, the same shape for the memory function is repeated again and again but slightly extended to the right, what also results in an



(a) Memory capacity

(b) Memory function

(c) Memory function

(d) Memory function

Figure 2.5: a) Memory capacity for $N = 246$, $\beta = 1.0$, $\gamma = 0.3$, $\kappa = 0.9$ and $\phi = 0.4\pi$. The first delay is set to $N + 1$ while the second one varies with $2N + b$. The simulation has been done for different values of the SNR. b) Shape of the memory function for the case $b = 1$, c) for $b = 2$ and d) for $b = 3$.

increase of the memory capacity.

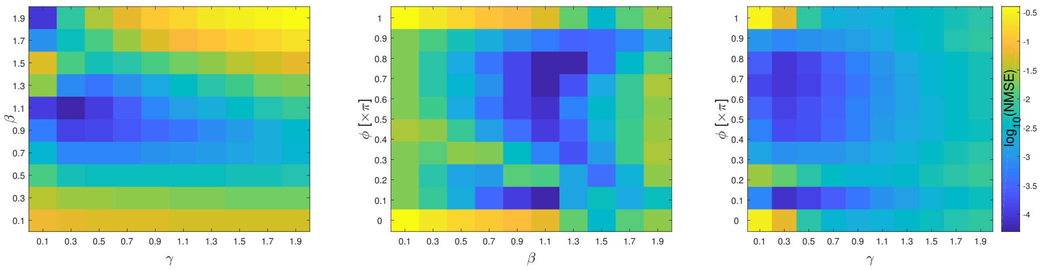
2.3 Chaotic time series prediction

In this section, the RC performance has been tested through the one step ahead prediction task of the well known chaotic Mackey-Glass oscillator described in Sec. 1.4.2. The training has been performed with the first 3000 values from a set of 4000 data while the prediction test has been done with the remaining 1000 values. In the following the NMSE will be described as:

$$\text{NMSE} = \frac{\frac{1}{N}\sum_{i=1}^{N}(o(i) - y(i))^2}{\sigma(y)} \tag{2.2}$$

where y are the targets and o the output of the RC.

Firstly, the system has been tested for one single delay with a number of nodes $N = 246$, the results are shown in Fig. 2.6 for different parameters combination. In order to obtain the parameters that give rise to the minimum NMSE in the prediction, three parameters of Eq. (2.1) have been modified, in this case β, γ and the phase ϕ. Firstly, Fig. 2.6(a) was obtained with $\phi = 0.1\pi$ and varying γ and β. The second plot (b) has been done fixing γ at the optimum value extracted from (a) while in Fig. 2.6(c), the value of β is the best one obtained in (b). The parameter κ



(a) (b) (c)

Figure 2.6: Heat maps for the Mackey - Glass system one step ahead prediction test with one single delay. Each map gives the better combination of parameters (a) $\gamma - \beta$ ($\phi = 0.1\pi$), (b) $\phi - \beta$ ($\gamma = 0.3$) and (c) $\phi - \gamma$ ($\beta = 1.1$). The best value of the NMSE is obtained in (b) for the combination $\beta = 1.1$, $\gamma = 0.3$ and $\phi = 0.1\pi$ with $\log_{10}(NMSE) = -4.09$. All the figures have the same color range for the NMSE values. These results are the average over 10 random masks.

is in all cases fixed at 0.9. The SNR of this system is set at 40dB and each NMSE is obtained averaging for 10 trials.

As seen in the heat maps, the variation of a single parameter has a strong influence in the prediction accuracy. The result varies from $\log_{10}(NMSE) \approx -0.5$ to a minimum of -4.09. We find that the best result for the MG one step ahead prediction for one delay is $\log_{10}(NMSE) = -4.09$ (*i.e.*, NMSE=$8.13 \cdot 10^{-5}$) for the combination $\beta = 1.1$, $\gamma = 0.3$ and $\phi = 0.1\pi$.

Addition of a second delay

In order to test the influence of the network connectivity, now we add a second delay to the RC. In these simulations, the first and second delays are set in $N+1$ and $2N+7$ respectively. Following the same methodology that we considered for the case of one delay, we build three heat maps for the different parameters combinations.

The set of equations describing this new scheme for the RC can be written as follows:

$$\mathbf{x}_i(k) = \begin{cases} \kappa \sin^2\left(\gamma \mathbf{J}(k) + \frac{\beta}{2}\left(\mathbf{x}_N(k-2) + \mathbf{x}_{N-6}(k-3)\right) + \phi\right), & i = 1; \\ \kappa \sin^2\left(\gamma \mathbf{J}(k) + \frac{\beta}{2}\left(\mathbf{x}_{i-1}(k-1) + \mathbf{x}_{N-7+i}(k-3)\right) + \phi\right), & i = 2, \ldots, 7; \\ \kappa \sin^2\left(\gamma \mathbf{J}(k) + \frac{\beta}{2}\left(\mathbf{x}_{i-1}(k-1) + \mathbf{x}_{i-7}(k-2)\right) + \phi\right), & i = 8, \ldots, N; \end{cases} \tag{2.3}$$

where both delays share the same value of the parameter β. Firstly, in order to evaluate the influence of the SNR on the chaotic time series prediction task, we simulate the behavior of the $\log_{10}(NMSE)$ against the parameter γ for different fixed values of the parameters β and ϕ. The result of this simulation is shown in Fig. 2.7.

For low γ and a SNR of 40 dB, there exists a small region where the performance of the system with two delays is better. In the inset of Fig. 2.7, we see that for $\gamma = 0.2$, the NMSE of the RC where a second delay has been added is slightly better.

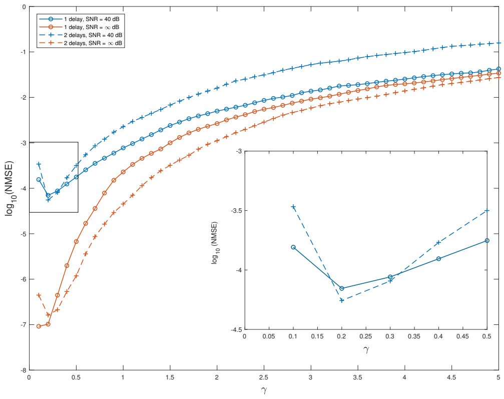
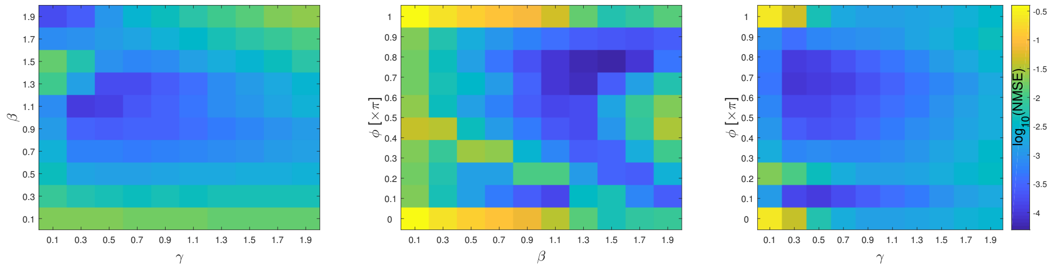


Figure 2.7: One step ahead prediction of the Mackey - Glass chaotic system, the blue lines are for a SNR of 40 dB, red lines are for an infinite SNR. The $\log_{10}(NMSE)$ has been plotted against the parameter γ. In the inset, we zoom-in the region where the best performance appears for this given task. The solid lines with circles correspond to the one delay system, while the dashed lines with cross markers show the prediction for a system with two delays.

However this improvement of the NMSE is not significant. A much lower NMSE, for the same parameters of the reservoir, is obtained when the SNR is increased from 40 dB to the free-noise case.

Now, we extend our analysis to the whole set of parameters. The results are shown in Fig. 2.8. The best combination of parameters is obtained in panel (c) for the combination $\beta = 1.5$, $\gamma = 0.3$ and $\phi = 0.8\pi$ with $\log_{10}(NMSE) = -4.25$. Not only this value is lower than the minimum obtained with one delay but also the parameter region of good operation is significantly extended as it can be clearly seen comparing panel (a) in Figs. 2.6 and 2.8 for the different schemes.



(a) (b) (c)

Figure 2.8: Heat maps for the Mackey - Glass system one step ahead prediction test with two delays. Each map gives the better combination of parameters (a) $\gamma - \beta$ ($\phi = 0.1\pi$), (b) $\phi - \beta$ ($\gamma = 0.3$) and (c) $\phi - \gamma$ ($\beta = 1.5$). The best value of the NMSE is obtained in (c) for the combination $\beta = 1.5$, $\gamma = 0.3$ and $\phi = 0.8\pi$ with $\log_{10}(NMSE) = -4.25$. The color scheme is the same than in Fig. 2.6

For completeness, we also simulate the performance of the system splitting the delay term in two different parts, *i.e.* splitting β into β_1 and β_2. In this case, the equation governing the system can be written as follows:

$$\mathbf{x}_i(k) = \begin{cases} \kappa \sin^2\left(\gamma \mathbf{J}(k) + \beta_1 \mathbf{x}_N(k-2) + \beta_2 \mathbf{x}_{N-6}(k-3) + \phi\right), & i = 1; \\ \kappa \sin^2\left(\gamma \mathbf{J}(k) + \beta_2 \mathbf{x}_{i-1}(k-1) + \beta_1 \mathbf{x}_{N-7+i}(k-3) + \phi\right), & i = 2, \ldots, 7; \\ \kappa \sin^2\left(\gamma \mathbf{J}(k) + \beta_1 \mathbf{x}_{i-1}(k-1) + \beta_2 \mathbf{x}_{i-7}(k-2) + \phi\right), & i = 8, \ldots, N; \end{cases} \tag{2.4}$$

As shown in Fig. 2.9, an anti-diagonal appears such that the increase of one feedback strength can be compensated by reducing the other one. In this case, the best prediction error is $\log_{10}(NMSE) = -4.25$, obtained for $(\beta_1, \beta_2) = (1.3, 1.7)$ and being the rest of parameters the same ones than for the optimum in Fig. 2.8 ($\gamma = 0.3$ and $\phi = 0.8\pi$).

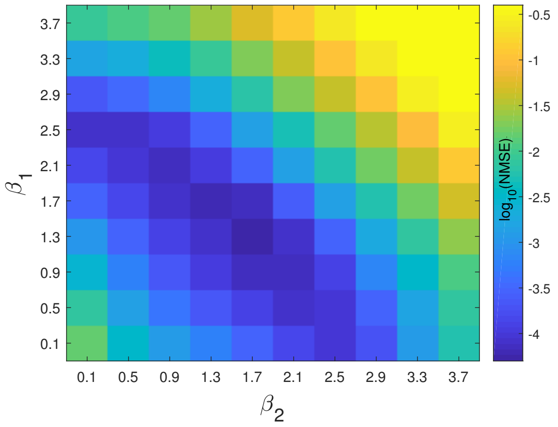


Figure 2.9: MG one step ahead prediction using two different feedback parameters for each delay term. An anti-diagonal appears tracing the best values for β_1 and β_2. The color scheme is the same than in Fig. 2.8.

2.4 NARMA tasks

In this section we will make use of the NARMA task described in Sec. 1.4.3, which will give us a better grasp of the system's dynamics. In the following, we will present the results from numerical simulations demonstrating the computational capabilities of the RC with one single delay for this task.

We first create a random uniformly distributed series u_k ranged in [0, 0.5] that will be used as an input for the system. When processed by the system, we obtain the output matrix $\mathbf{X}_i(k)$ that will be used for the training procedure. The targets are the y_{k+1} values of the NARMA task, described by Eqs. (1.6) and (1.7) for NARMA2 and NARMA5 respectively. As usual, we train the system with the first 3000 values and test it with the remaining 1000. To obtain the best values of the parameters that give rise to the lower NMSE, we create heat maps for the parameters γ, β and ϕ as in the previous section. The best values obtained for tasks NARMA2 and NARMA5 are summarized in Tab. 2.1.

	γ	β	ϕ	NMSE
NARMA2	0.9	0.7	0.1π	0.0018
NARMA5	1.1	0.7	0.1π	0.0072

Table 2.1: Sets of parameters γ, β and ϕ for which the best NMSE is obtained for the RC with one delay. The results are averaged over 10 random masks.

For NARMA2 and NARMA5, the best values of the NMSE are 0.0018 and 0.0052, respectively, for the optoelectronic system with a single delay.

Addition of a second delay

In this part we perform the same task but adding a second delay to the reservoir. We follow the same procedure that was carried out with one single delay. The best obtained results are summarized in Tab. 2.2.

	γ	β	ϕ	NMSE
NARMA2	0.5	0.9	0.1π	0.0015
NARMA5	0.8	1.1	0.1π	0.0056

Table 2.2: Sets of parameters γ, β and ϕ for which the best NMSE is obtained for the RC with two delays. In both cases, the NMSE is lower than the one obtained for a single delay in Tab. 2.1. The results are averaged over 10 random masks.

From those results we can confirm that the addition of a second delay systematically decreases the NMSE produced by the delayed optoelectronic reservoir when performing the NARMA2 and NARMA5 tasks.

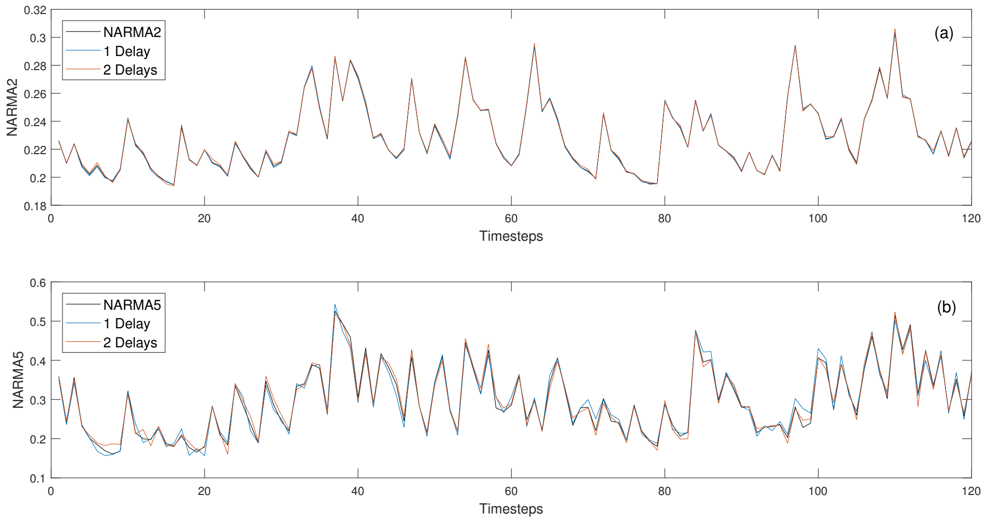


Figure 2.10: Results for one realization of the NARMA task. In (a), 120 timesteps of the test for the NARMA2 task for one (NMSE=0.0013) and two single delays (NMSE=0.0014). In (b), 120 timesteps of the test for the NARMA5 task for one (NMSE=0.021) and two single delays (NMSE=0.013). All the simulations have been done for the optimal combinations of parameters shown in Tabs. 2.1 and 2.2.

2.5 Summary

In this second chapter we have numerically studied the influence of adding a second delay in an optoelectronic reservoir computer. Firstly, we evaluated the memory capacity of the system. The obtained results show that a significant increase on the memory capacity can be obtained through the addition of a second delay. Particularly, we went from a MC of 10 past inputs to a MC of around 16 when a second delay was added to the reservoir for the same value of the SNR. We also introduced the parameter b that allows to obtain different network configurations and studied the different values of the MC resulting from the variation of this parameter. We found that some values of b need to be avoided.

We also tested the one step ahead prediction of the Mackey - Glass chaotic oscillator. In this case by adding a second delay we found a mild improvement on the performance of this task. We also found that by giving a different weight to each of the delays, the strength of one feedback term can be compensated by reducing the other one. However, in order to reduce the NMSE the best strategy that can be carried out is the lowering of the SNR of the experiment.

Finally, we tested the system through the well known NARMA2 and NARMA5 tasks. In both cases, a lower NMSE is obtained by including a second delay to the system and by varying the parameters γ, β and ϕ.

Chapter 3

Experimental setup

The experimental setup for the optoelectronic oscillator with delay has been designed and implemented by Pau Massutti Ballester (detection and amplification stages) and Matthias Häussler (characterization of ADC and DAC adapter cards). The code to program the FPGA for delay-based reservoir computing has been developed by Pau Massutti, Matthias Häussler, Johannes Striebel, Joaquim LLorens Giralt and myself. I (Jesús Yelo Sarrión) have also performed the characterization and the optimization of the dynamics of the optoelectronic oscillator with one and two delays and the optimization of the system's parameters to obtain a good performance for the RC.

3.1 Experimental optoelectronic setup

The experimental optoelectronic reservoir computer is depicted in s Fig. 3.1. It includes several components: a pump semiconductor laser, a Mach-Zehnder modulator, a photodetector, filters, amplifiers, an FPGA (Field-programmable gate array), analog to digital (AD) and digital to analog (DA) converters. The FPGA is a designed network of logic block circuits to be configured by the user with a hardware description language. In the experimental set-up, the FPGA (ALTERA Cyclone IV) is an essential component that takes care of timing operations, the input mask multiplication, the delay-feedback loop, and the trained weights multiplication.

The Mach-Zehnder (MZ) modulator provides a $\sin^2$ nonlinearity, which can be defined as follows:

$$P_{out} = P_{pl} \sin^2(2\pi \frac{U}{V_\pi} + \phi_b) \tag{3.1}$$

where P_{out} is the output power of the MZ, P_{pl} is the power of the pump laser, U is the radio-frequency input of the MZ (input to be processed and sent by the FPGA), $V_\pi = 2.7$V is the voltage needed to go over one period of the $\sin^2$, and ϕ_b is the operating bias phase of the MZ.

The electronic part of the optoelectronic oscillator includes a photodetector, filters and the AD card. The detection apparatus has an approximate frequency bandwith ranging from $f_{min} = 100$ KHz to $f_{max} = 20$ MHz and an optical to electric conversion factor of 1.6 V/mW. The AD and DA cards have a resolution of 14 bits and a maximum peak-to-peak voltage V_{pp} of 1 V and 600 mV, respectively. We operate the FPGA with a clock frequency of $f_{clock} = 20$ MHz such that we define a temporal spacing between the virtual nodes of $\theta = 1/f_{clock} = 50$ ns. In the following the number of nodes will be set to $N = 500$.

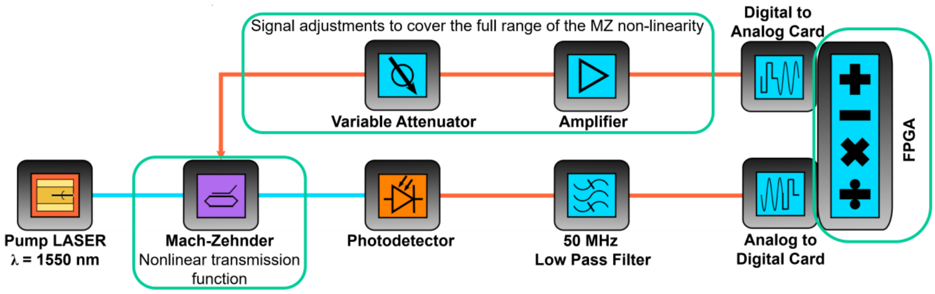


Figure 3.1: Block representation of the optoelectronic reservoir computing with mixed digital and analog hardware.

In the following, we employ a MZ controller to follow long term drifts in the operating point of the MZ. We will operate the MZ at the so-called “Quad-” operational mode, *i.e* $\phi_b = -\pi/4$.

This master thesis contains the first scientific results obtained with such a setup, which has been developed in order to include addittional delays in the delay-based RC approach.

3.2 Characterization of the RC.

The experimental setup is governed by an integro-differential equation [23]:

$$\frac{T}{\theta}\int x(\varepsilon)d\varepsilon + \frac{dx}{d\varepsilon} = -x(\varepsilon) + \kappa \sin^2[z(\varepsilon - \tau') + \phi] \tag{3.2}$$

where θ is the low frequency cut-off, $\varepsilon = t/T$, $\tau' = \tau/T$ where T is the high frequency cut-off and finally, ϕ is the phase of the system. The nonlinearity gain κ can be related to the current pumping the MZ modulator:

$$\kappa = \frac{I_{LD} - I_{Th}}{|I_{LD}|_{\beta=1} - I_{Th}} \tag{3.3}$$

This parameter depends on the power of the pump laser, which can be controlled through the laser diode current above threshold. In eq. (3.3), I_{LD} is the operation current, I_{Th} the threshold current and $|I_{LD}|_{\beta=1}$ is the current at which the first oscillations appear when $\phi_b = -\pi/4$. Setting $\kappa < 1$, the system operates in a steady (fixed point) state. When increasing the laser power, the system becomes unstable, this is the point where $|I_{LD}|_{\beta=1}$ is reached, for this system: $|I_{LD}|_{\beta=1} = 23.4$ mA.

The oscillator’s output strength is linearly controlled by the intensity of the laser as it can be seen in Fig. 3.2. This plot gives us a key information: the laser threshold current, namely I_{Th}. In this case, $I_{Th} \approx 15$ mA. The ADC card saturates for $I > 60$ mA.

There are several sources of noise that affect the system. This intrinsic and extrinsic noise can be quantified. For the complete system, we measure a signal-to-noise ratio (SNR) of 41 dB. It is defined as:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\mathrm{RMS}^2_{signal}}{\mathrm{RMS}^2_{noise}}\right) \tag{3.4}$$

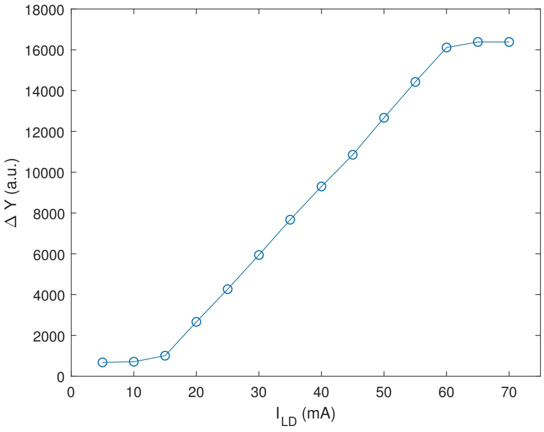


Figure 3.2: Variation of the oscillator's amplitude range as a function of the laser intensity. This plot is obtained in the "Quad -" operational mode.

where RMS is the Root Mean Square value given by

$$\mathrm{RMS}(x_i) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{3.5}$$

Now by introducing an input data, I, the oscillator is transformed into a reservoir computer governed by:

$$\frac{T}{\theta}\int x(\varepsilon)d\varepsilon + \frac{dx}{d\varepsilon} = -x(\varepsilon) + \kappa \sin^2[\gamma I(\varepsilon) + \beta z(\varepsilon - \tau') + \phi] \tag{3.6}$$

where the parameters γ and β control the strength of the input and the delayed feedback, respectively.

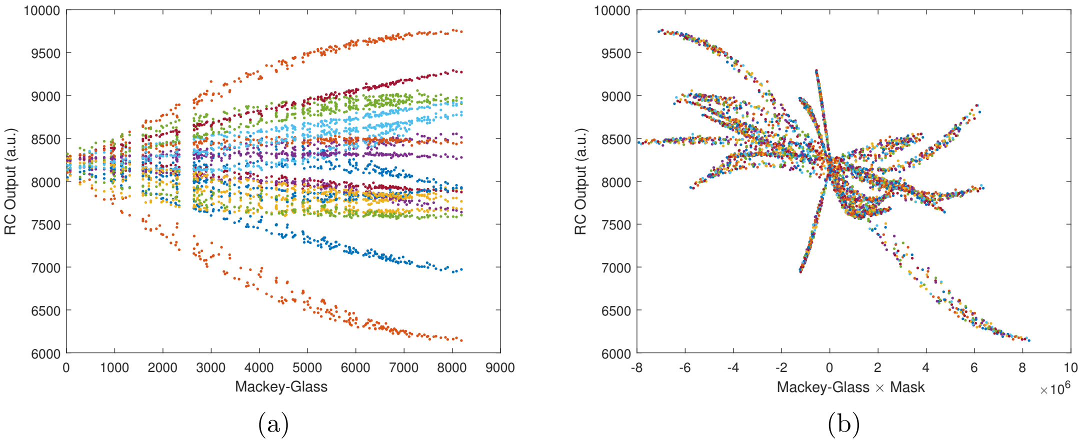


Figure 3.3: Representation of the amplitude of the virtual nodes along the sine squared nonlinearity. The horizontal axis corresponds to (a) the MG time series and (b) the MG time series multiplied by different mask values. The width of the sinusoidal function is related to the feedback's strength and therefore with the value of the parameter β.

In Fig. 3.3 we show the output of the RC when the Mackey - Glass chaotic time-series is used as the input. Initially, we show in Fig. 3.3(a), the RC output function as a function of the MG input values. The different lines in this figure correspond to the different virtual nodes, here $N = 20$.

Actually, the nonlinear responses of the virtual nodes are better visualized in Fig. 3.3(b), where the MG data is scaled by the corresponding input mask value associated with the node. In this figure, it becomes clear that the different virtual nodes have a different effective nonlinearity as a result of the different scaling given by each mask value.

Addition of a second delay

Once the characterization of the RC with a single delay has been done, we proceed with the addition of the second delay line. In order to correctly validate the proposed RC concept with two delays, we set as input a Delta function. When the delta input is processed by the RC, it will give rise to an oscillatory function that will be repeated and damped at each time τ. As the input will be stored in the memory and reproduced by both delays, it must be visible at $(N + 1)$ and $(2N + b)$ time steps later. As the delay echo is rapidly damped, we only detect the first echo of each delay. The equation governing the system in this case would be:

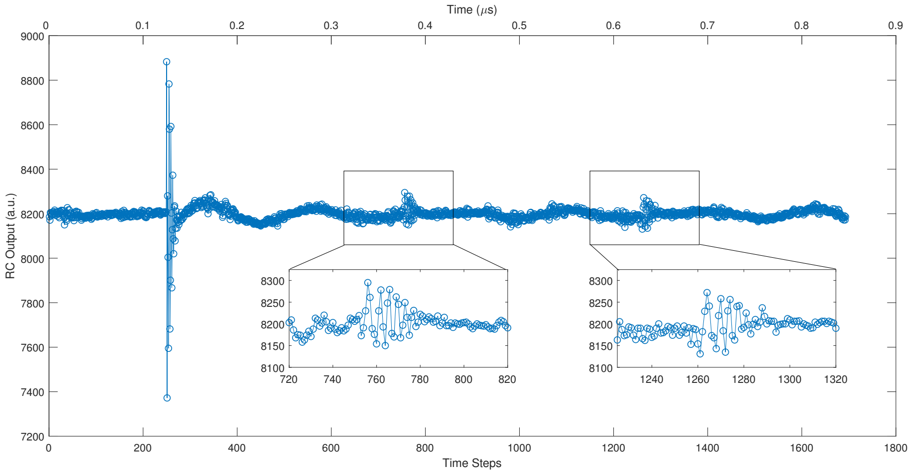


Figure 3.4: Demonstration of the existence of two simultaneous delays. The left hand inset corresponds to the delay set at $N + 1$ while the right hand inset is the result of the second delay at $2N + 7$. The Delta function starts at $t = 250$ and is repeated at $t_1 \approx 753$ and $t_2 \approx 1260$ by the first and second delay respectively. This result was obtained for $\gamma = 1.0$, $\beta = 1.875$, $I = 29.7$ mA and $N = 500$.

$$\frac{T}{\theta}\int x(\varepsilon)d\varepsilon + \frac{dx}{d\varepsilon} = -x(\varepsilon) + \kappa\sin^2[\gamma I(\varepsilon) + \beta(z(\varepsilon - \tau_1) + z(\varepsilon - 2\tau_2)) + \phi] \quad (3.7)$$

Figure 3.4 shows how the oscillation created by the second delay has the same amplitude than the one from the first delay.

3.3 Memory Capacity

In this section, the objective is to find the set of parameters β and γ that gives rise to the best memory capacity in the experimental implementation. The data used for the memory capacity are a set of 4000 values from a random uniform distribution in the range [-1,1] and we used a uniformly distributed mask where each value is repeated twice. As in the numerical simulations, we train the output layer with 3000 values and test it with the other 1000. In this case the aim is to test how many values the system is able to remember.

The memory function $m(i)$, as defined in Sec. 1.4.1, is evaluated for different combinations of parameters γ and β with a laser intensity of 60 mA. The area under the curve $m(i)$ corresponds to the memory capacity. The results for the MC are shown in Tab. 3.1. Those parameters have been chosen after checking different combinations of them. Afterwards, we explored with a higher accuracy around the place where the largest MC was obtained until completing a table where all the neighbours of the center value had a lower MC.

γ \ β	0.25	**0.3125**	0.375
0.375	9.74	9.35	9.18
0.4375	9.13	**9.79**	9.08
0.5	8.54	9.58	9.25

Table 3.1: Memory capacity for different values of β, γ and $I = 60$ mA.

The best obtained value is MC $= 9.79$ for the set of values $\gamma = 0.4375$ and $\beta = 0.3125$ for $I = 60$ mA. Introducing this intensity into Eq. (3.3), we find $\kappa = 5.36$, and therefore, the products $\kappa \cdot \beta$ and $\kappa \cdot \gamma$ give 1.675 and 2.345, respectively.

Addition of a second delay

To test the influence of the network connectivity, we evaluate the MC after the addition of a second delay in the experiment. The program used to control the RC behavior only allows discrete changes of the parameters (addition or subtraction). As β is set the same for the first and second delays, a slight modification of this parameter gives rise to a huge change of the results. For this reason, instead of modifying the parameters γ and β of the system, we play with the intensity of the laser. By varying I_{LD}, we change κ and to some extent, this is equivalent to changing simultaneously and proportionally γ and β.

Starting from a set of parameters close to the optimal ones obtained with a single delay, we made several measurements of the memory capacity of the system with two delays. The first and second delays are set as in the previous section, at $N + 1$ and $2N + 7$ respectively ($N = 500$). The values of the parameters used are $\gamma = 0.4375$ and $\beta = 0.2815$. The parameter β is the same for the two delayed feedback terms but it is not divided in half as happened in the simulations.

I (mA)	25	30	31	32.5	**35**	40
MC	6.90	9.53	9.62	10.26	**10.64**	9.93

Table 3.2: Memory capacity for different values of I(mA), with $\gamma = 0.4375$ and $\beta = 0.2815$.

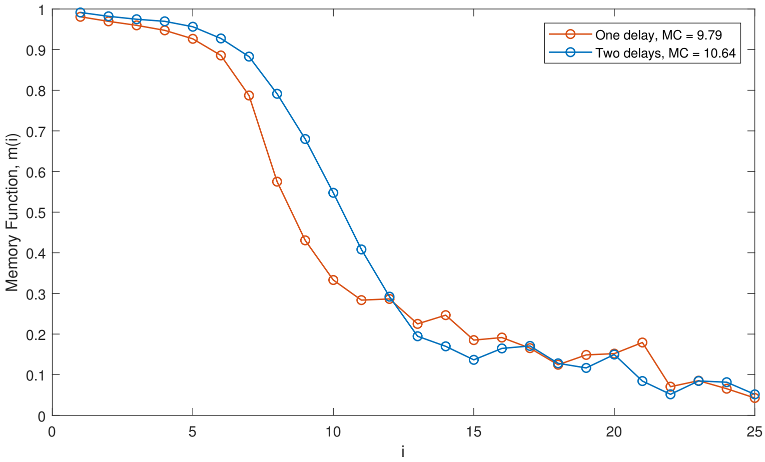


Figure 3.5: Memory function $m(i)$ for systems with one and two delays for optimized conditions.

The results are displayed in Ta. 3.2. The best value found for the memory capacity using a RC with two delays is $MC = 10.64$. Experimentally, a higher value than with one delay has been found for different intensities and therefore for different combinations of γ and β. In order to compare the obtained results in this configuration with the one single delay configuration we introduce the optimal value of the intensity in Eq. (3.3), which gives $\kappa = 2.38$. Now, the new values for the product of the nonlinearity gain κ by the feedback gain and the input gain are $2(\kappa \cdot \beta) = 1.34$ and $\kappa \cdot \gamma = 1.04$. Interestingly, the feedback strength is similar to the one with one single delay and the input scaling is a half of the one obtained for the first configuration.

In Fig. 3.5 we show the different shapes of the memory function $m(i)$ for one (blue) and two (red) delays. The curves that are represented in this figure correspond to the ones with a higher MC for each configuration of the reservoir reported in Tabs. 3.1 and 3.2.

3.4 Chaotic time series prediction

The second task that evaluates the performance of the RC under different configurations of the reservoir is the chaotic time series prediction.

The input time series originates a Mackey - Glass delayed oscillator (see Sec. 1.4.2), where 3000 samples are used for training the output weights and 1000 for the test. The numerical simulations have shown that, for certain parameters, the performance of the RC with two delays is better than with one single delay but the difference was relatively small.

Here, we evaluate the system with the parameters that have been shown to give a better memory capacity, for one delay, $\gamma = 0.4375$, $\beta = 0.3125$ and $I = 60$ mA. In this case the NMSE obtained is 0.0027 for the train and 0.0052 for the test. This low prediction error is comparable to a previous optoelectronic RC implementation (NMSE=0.0036 [12]) and one order of magnitude lower than an all-optical RC implementations (NMSE=0.019 [24]) and NMSE=0.042 [25]), validating

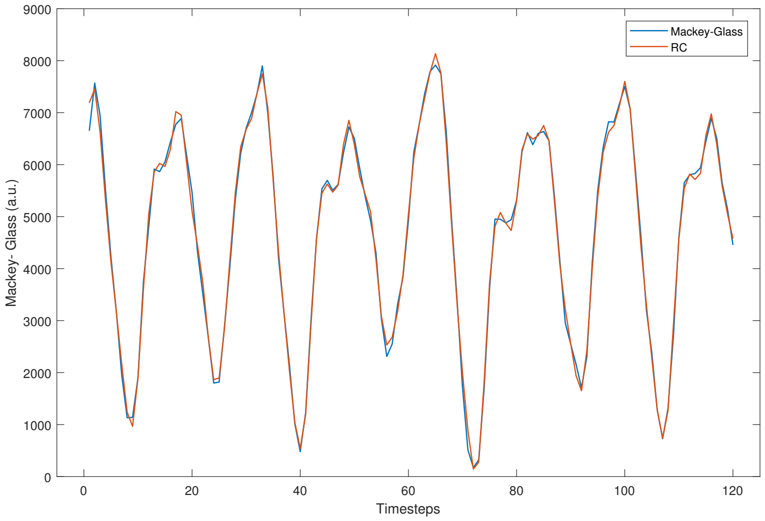


Figure 3.6: Mackey - Glass test prediction for a system with two delays. The figure was obtained for parameters: $\gamma = 0.4375$, $\beta = 0.2815$ and a laser intensity of 35 mA. The result is $NMSE = 0.0056$ for the test with 1000 samples.

the current setup for RC.

Addition of a second delay

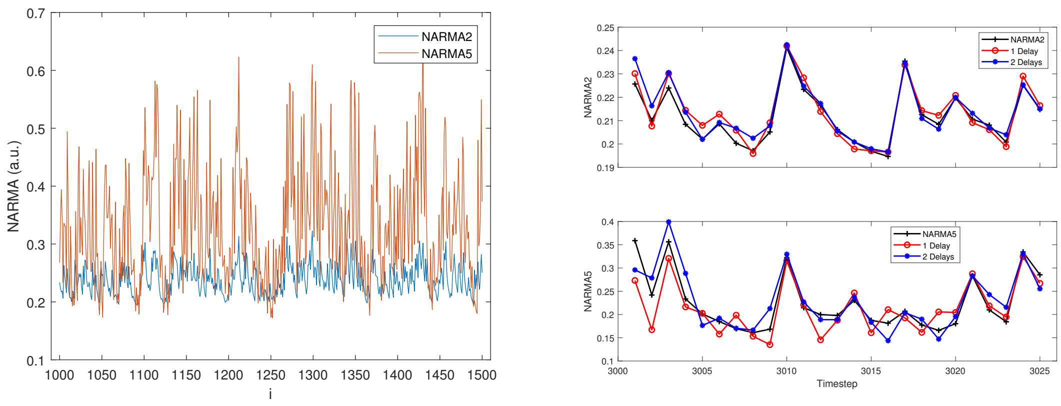
Once the system's best parameters were found for the memory capacity, we proceed with the one step ahead prediction of the Mackey-Glass delayed differential equation for a configuration of the reservoir with two delays. For an intensity of $I = 35$ mA, a feeback gain $\beta = 0.2815$ and an input gain $\gamma = 0.4375$, the NMSE is 0.0027 for the training and 0.0056 for the test. A These results are similar but slightly worse than the prediction error obtained in the case of a single delay for this particular task. As shown in the numerics, the NMSE is bounded by the SNR of the system and does not improve much with the addition of the second delay.

3.5 NARMA tasks

In this section, the RC performance is assessed through two NARMA tasks. For this kind of tasks, the memory capacity of the system is fundamental. Particularly, the first $n-$values of the memory capacity $m(i)$ (see *e.g.* Fig. 3.5) play a key role. Not only the MC must be greater than the $n-$order of the NARMA task but those n first values of the $m(i)$ should be close to 1 to obtain a good performance.

The input $u(i)$ of the NARMA task is the same collection of uniform random numbers that was used in section 3.3 but normalized to fit in the range $[0, 0.5]$ that is required for this task. Introducing those values into:

$$y_{k+1} = 0.4y_k + 0.4y_k y_{k-1} + 0.6u_k^3 + 0.1 \tag{3.8}$$



(a) (b)

Figure 3.7: In panel (a) typical NARMA2 and NARMA5 output time series. In (b) we show a segment of the prediction made by the RC for NARMA2 (top) and NARMA5 (bottom) with one (red) and two (blue) delays, respectively. Black line corresponds to the target.

for NARMA2 and, respectively for NARMA5, into:

$$y_{k+1} = 0.3y_k + 0.05y_k\left(\sum_{j=0}^{4} y_{k-j}\right) + 1.5u_{k-4}u_k + 0.1 \tag{3.9}$$

we get a numerical series of 4000 data where the first 3000 are used for the training and the remaining 1000 for the testing of the output layer. The results for the test are illustrated in Fig. 3.7 and summarized in Fig. 3.8. for the one single delay reservoir.

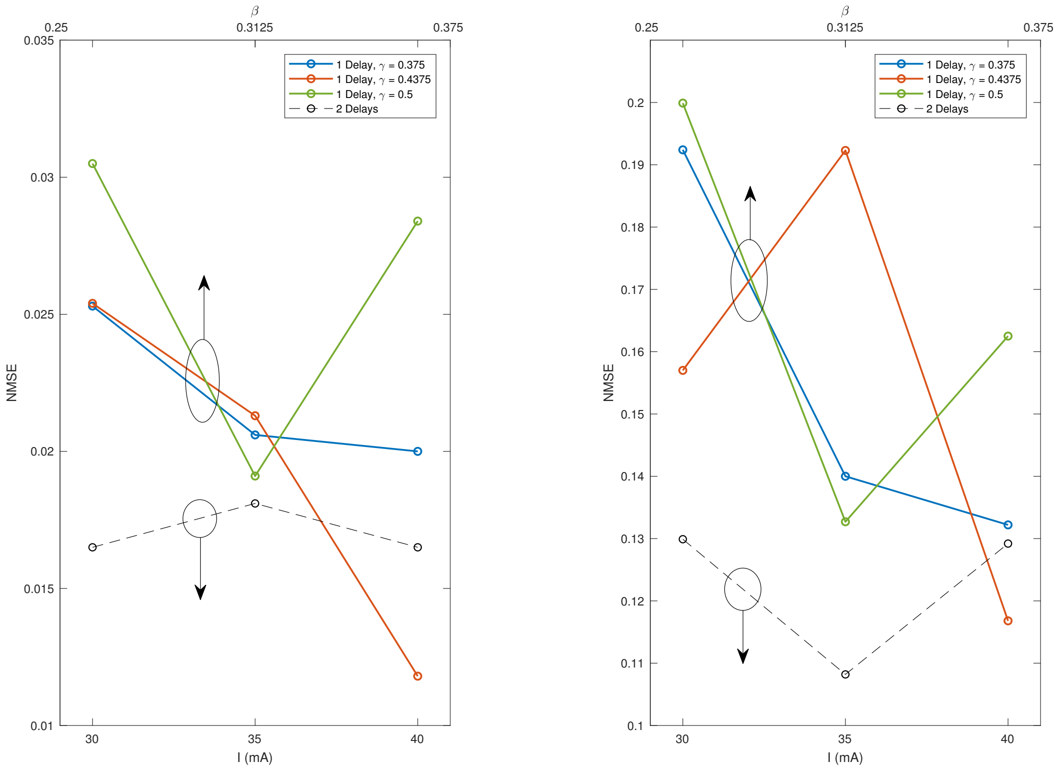
The best value for the prediction of the NARMA2 series is obtained for parameters $\gamma = 0.4375$ and $\beta = 0.375$ where the NMSE is 0.0118 and the corresponding memory capacity was shown to be $MC = 9.08$. It is also this set of parameters that gives rise to the best prediction of the NARMA5 random series with an NMSE of 0.1168.

Addition of a second delay

In NARMA2, the RC with one single delay performs better than after adding a second delay, the lower NMSE at NARMA2 is 0.0165 obtained for $I = 30$ and 45 mA in the case of two delays, as shown in Fig. 3.7. However for the NARMA5 task it is the RC with two delays that obtains the lower NMSE, for the set $\gamma = 0.4375$, $\beta = 0.2815$ and $I = 35$ mA. The error between the prediction and the target is in this case $NMSE = 0.1082$.

3.6 Other types of memory

The memory capacity can also be evaluated in nonlinear scenarios. In this case, we have asked the RC to remember the values $y_i(n) = u^2(n-i)$, $y_i(n) = \sin(u(n-i))$ and $y_i(n) = \sin(3u(n-i))$. The same procedure than for the linear memory capacity



(a) NARMA2

(b) NARMA5

Figure 3.8: Results for the NARMA task with the RC configurations with one (solid lines) and two delays (dashed lines). With one delay, the NMSE is plotted as a function of the variable β (upper axis) for three different combinations ($\gamma = (0.375, 4375, 0.5)$, $I = 60$ mA), while with two delays the NMSE is plotted as a function of the intensity (lower axis).

is followed, however the targets have been modified to the non linear cases referred before. In all cases, the memory capacity exhibited by the RC with two delays is higher than the one showed by the one delay combination. This shows that the addition of the second delay increases both the linear and the nonlinear memory capacities. Interestingly, this systematic improvement in the memory turns out to have a rather neutral influence for the prediction tasks.

3.7 Summary

In this third chapter, we studied and discussed the influence of the additional delay on the experimental setup of the optoelectronic reservoir computer. Firstly, we characterized the behavior of the RC which is governed by an integro-differential equation. We saw how the parameters $\kappa(I)$, β and γ in this equation influence the dynamics. The effects of the addition of a second delay were illustrated for different benchmark tasks. A second delay appears at $2N + b$, being b a parameter that we can choose. It is the addition of this second delay that was proven to be useful in order to increase the memory. We found some values of the system parameters that

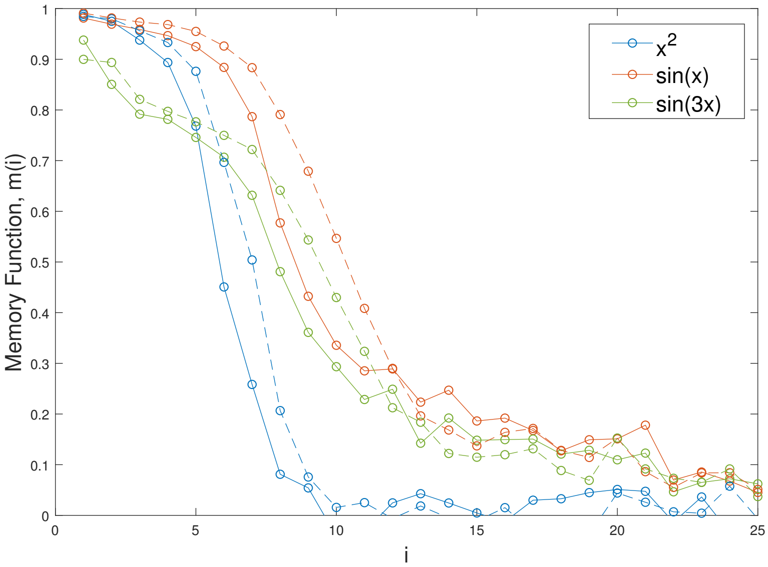


Figure 3.9: Memory capacity for different non linear functions. Solid lines corresponds to the system with one single delay, while dashed lines are for the RC system with two delays.

give rise to a larger memory capacity in the RC's version with two delays than in the version with a single delay.

The increase in the memory capacity does not seem to have an effect on the accuracy for the prediction of chaotic time series prediction. The results obtained for one single delay and two delays were similar for the prediction of the Mackey - Glass oscillator. The NARMA2 task is performed better by the RC with one single delay. However, it is for the NARMA5 task where a slight increase of the prediction capacity appears when it is performed by the optoelectronic system with two delays.

Finally, it was also shown that not only the linear memory is increased by the addition of a second delay but also the non linear memory.

Conclusions and Outlook

Nowadays, traditional information processing techniques are reaching its physical limits and new unconventional methods are emerging. In this way, alternative computational paradigms like machine learning techniques and hardware substrates like optical systems are called to be key part to overcome this challenge. In particular, reservoir computing is emerging and obtaining state-of-the-art results [10, 13].

In this MSc Thesis we have demonstrated, both in simulations and experiments, that the addition of a second delay to an optoelectronic reservoir computer can efficiently enhance the memory capacity and the performance of the device for different benchmark tasks.

Firstly, in numerical simulations, we found the combination of parameters that gives rise to a better performance of the optoelectronic reservoir computer when adding a second delay. This includes the feedback gain β, the input gain γ, and the phase ϕ. A significant increase of the memory capacity was found with this new configuration. For the one step ahead Mackey - Glass time series prediction, we found that not only a slightly lower NMSE occurs for two delays but also that the parameter region of good operation is significantly extended by the addition of a second delay. In parallel, we have studied the influence of a changing network connectivity by adding a new parameter b. When a combination of two delays is used, the best performance is obtained for $b \geq 3$ when setting the first delay to $N+1$. For the system with two delays we found evidences that the increase of one feedback strength can be compensated by reducing the other one. Finally, the NARMA task was performed better by the reservoir with two delays than the reservoir with one single delay.

In the experimental setup, we characterized the reservoir computer and successfully tested the addition of the second delay. The influence of this novel configuration has been studied through the different benchmark tasks. We also discussed the dynamical effects of varying the parameters $\kappa(I)$, β and γ. The memory capacity of the experimental optoelectronic reservoir computer was increased around a 9% by the addition of the second delay. In the chaotic time series prediction and the NARMA2 task, we did not found any improvement in the NMSE but we found it in the NARMA5 task. Finally, an increase of the the non linear memory capacity was also found.

Time restrictions prevented us to go in-depth in the exploration of the optoelectronic reservoir computer, specially in the experimental setup. Nevertheless, through this MSc thesis we went a step further in the intersection between machine learning, optics and information processing. The challenge is to extract the best of each of those fields. In this way, many different strategies can still be addressed in order to optimize the optoelectronic reservoir computer: reduction of the SNR, tailoring the random mask or look for better combinations of parameters.

Bibliography

[1] J. P. Crutchfield, W. L. Ditto, and S. Sinha. “Introduction to focus issue: Intrinsic and designed computation: Information processing in dynamical systems-beyond the digital hegemony”. In: *Chaos* 20.3 (2010), p. 2. DOI: 10.1063/1.3492712.

[2] D. A. B. Miller, M. H. Mozolowski, and A. Miller. “Non-Linear Optical Effects in InSb with a C.W. CO Laser”. In: *Optics Communications* 27.1 (1978), pp. 133–136.

[3] E. Abraham and S. D. Smith. “Optical bistability and related devices”. In: *Reports on Progress in Physics* 45.8 (1982), pp. 815–885. DOI: 10.1088/0034-4885/45/8/001.

[4] J. L. O’Brien. “Optical Quantum Computing”. In: *Science* 318.5856 (2007), pp. 1567–1570. DOI: 10.1126/science.1142892.

[5] R. S. Tucker. “The role of optics in computing”. In: *Nature Photonics* 4.7 (2010), pp. 406–407. DOI: 10.1038/nphoton.2010.164.

[6] H. J. Caulfield and S. Dolev. “Why future supercomputing requires optics”. In: *Nature Photonics* 4.5 (2010), pp. 261–263. DOI: 10.1038/nphoton.2010.94.

[7] D. Woods and T. J. Naughton. “Photonic neural networks”. In: *Nature Physics* 8.4 (2012), pp. 257–259. DOI: 10.1038/nphys2283.

[8] H. Haas and H. Jaeger. “Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication”. In: *Science* 304.5667 (2004), pp. 78–80. DOI: 10.1126/science.1091277.

[9] D. Verstraeten et al. “An experimental unification of reservoir computing methods”. In: *Neural Networks* 20.3 (2007), pp. 391–403. DOI: 10.1016/j.neunet.2007.04.003.

[10] L. Larger et al. “Photonic information processing beyond Turing : an optoelectronic implementation of reservoir computing”. In: *Optics Express* 20.3 (2012), pp. 3241–3249.

[11] Y. Paquot et al. “Optoelectronic Reservoir Computing”. In: *Scientific Reports* 2 (2012), p. 287. DOI: 10.1038/srep00287.

[12] S. Ortín et al. “A Unified Framework for Reservoir Computing and Extreme Learning Machines based on a Single Time-delayed Neuron”. In: *Scientific Reports* 5 (2015), p. 14945. DOI: 10.1038/srep14945.

[13] L. Appeltant et al. “Information processing using a single dynamical node as complex system”. In: *Nature Communications* 2 (2011), p. 468. DOI: 10.1038/ncomms1476.

[14] M. Rabinovich. “Transient Dynamics for Neural Processing”. In: *Science* 321.2008 (2011), p. 48. DOI: 10.1126/science.1155564.

[15] H. Jaeger. “Short term memory in echo state networks”. In: *GMD Report 152* Vol. 5 (2002).

[16] M. C. Mackey and L. Glass. “Oscillation and chaos in physiological control systems”. In: *Science* 197.4300 (1977), pp. 287–289. DOI: 10.1126/science.267326.

[17] H. Jaeger. “Adaptive Nonlinear System Identification with Echo State Networks”. In: *Advances in neural information processing systems* 4 (2002), pp. 593–600.

[18] A. Rodan and P. Tino. “Minimum complexity echo state network”. In: *IEEE Transactions on Neural Networks* 22.1 (2011), pp. 131–144. DOI: 10.1109/TNN.2010.2089641.

[19] K. Nakajima et al. “Boosting Computational Power through Spatial Multiplexing in Quantum Reservoir Computing”. In: *Physical Review Applied* 11.3 (2019), p. 034021. DOI: 10.1103/PhysRevApplied.11.034021.

[20] A. F. Atiya and A. G. Parlos. “New results on recurrent network training: unifying the algorithms and accelerating convergence”. In: *IEEE Transactions on Neural Networks* 11.3 (2000), pp. 697–709. DOI: 10.1109/72.846741.

[21] J. Nakayama, K. Kanno, and A. Uchida. “Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal”. In: *Optics Express* 24.8 (2016), p. 8679. DOI: 10.1364/oe.24.008679.

[22] L. Appeltant. “Reservoir computing based on delay-dynamical systems.” PhD thesis. UIB, VUB, 2012.

[23] M. Peil et al. “Routes to chaos and multiple time scale dynamics in broadband bandpass nonlinear delay electro-optic oscillators”. In: *Physical Review E* 79.2 (2009), p. 026208. DOI: 10.1103/PhysRevE.79.026208.

[24] J. Bueno et al. “Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback”. In: *Optics Express* 25.3 (2017), p. 2401. DOI: 10.1364/oe.25.002401.

[25] J. Bueno et al. “Reinforcement Learning in a large scale photonic Recurrent Neural Network”. In: *Optica* 5.6 (2017), p. 756. DOI: 10.1364/OPTICA.5.000756.