Ring topology of an optical phase delayed nonlinear dynamics for neuromorphic photonic computing

Antonio Baylon Fuentes

▶ **To cite this version:**

Antonio Baylon Fuentes. Ring topology of an optical phase delayed nonlinear dynamics for neuromorphic photonic computing. Optics [physics.optics]. Université de Franche-Comté, 2016. English. NNT : 2016BESA2047 . tel-01589681

HAL Id: tel-01589681

https://theses.hal.science/tel-01589681v1

Submitted on 18 Sep 2017

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

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

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UFC UNIVERSITÉ DE FRANCHE-COMTÉ

THÈSE présentée par

ANTONIO BAYLÓN FUENTES

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Optique et Photonique**

Topologie en anneau d'une dynamique non linéaire à retard en phase optique, pour le calcul photonique neuromorphique

Soutenue publiquement le 13 décembre 2016 devant le Jury composé de :

JULIE GROLLIER	Rapporteur	Directeur de recherche à l'Unité Mixte de Physique CNRS/Thales lab
RAUL VICENTE ZAFRA	Rapporteur	Professeur à l'Université de Tartu, Estonie
HERVÉ MAILLOTTE	Examinateur	Directeur de recherche à l'Institut Femto-ST, [UMR CNRS 6174]
FRÉDÉRIC GRILLOT	Examinateur	Professeur de l'université Paris-Saclay
YURI MAISTRENKO	Examinateur	Professeur à l'Académie Nationale des Sciences d'Ukraine
MAXIME JACQUOT	Directeur de thèse	Maître de Conférences à l'Université de Franche-Comté
LAURENT LARGER	Encadrant de thèse	Professeur à l'Université de Franche-Comté
YANNE CHEMBO	Encadrant de thèse	Chargé de recherche à l'Institut Femto-ST, [UMR CNRS 6174]

Acknowledgement

This Doctoral thesis was done at the department of Optical P.M Duffieux of the Institute of investigations FEMTO-ST in Bensancon, the required effort and dedication from the author and the doctoral directors of thesis, it could not have been possible its finalization without the cooperation disinterested of all and every person of which I take the moment to cite and many have been a strong support in moments of anguish and desperation.

First of all, I thank god, for being with me in every step I take, for strengthen my heart and to enlighten my mind, and for putting on my way all those people who have been my support and my company throughout the entire study period.

I thank today and always to my family, despite the fact of not being present physically, I know they always were there looking on my wellness from my country, Mexico, and is clearly that if it was not from the effort done by them, my studies would not have been possible. To my Parents Antonio and Paulina, my siblings Irving, Fatima, Jessica, despite the fact of our distance, their cheers, support and happiness they brought me, it gives me the strength necessary to continue forward.

I would also thank with the sincerest gratitude to the director of the optical department Herve Maillotte for letting me into the lab through this doctoral theses, to my directors of thesis, Dr. Laurent Larger, for his patience, advices and support, but overall the knowledge and the experience that I acquire from him throughout the realization of this thesis, to Maxime Jacquoit for his advices, but mainly for the opportunity of continuing my studies and my investigation (Scholarship Victor Hugo) and to Yanne Chembo for always giving me really good advices and for his availability and his knowledge. Finally, thanks to all of them, to whom I owe the accomplishment of this PhD in a prestigious insti-

tute such as FEMTO-ST.

I thank the jury, and in a particular manner to Julie Grollier and Raul Vicente Zafra for accepting being rapporteurs of this work.

Also, to the entire personal of FEMTO-ST, such as the General Secretary, Principal Faculty, Reception, Administration, Library, Maintenance and Cleaning service, they always received me with a smile.

Likewise, I thank Romain, Souleymane and Bogdan, for giving me their support, encouragement and collaboration in every moment and especially in moments when I needed of them, without giving me impediments or negativities.

To my colleagues and friends, and to the ones who today already are doctors, Jose, Yannick, Abel, Stéphane, Justine and Gloria, for all the encouragement, and patience, for trusting and believing on me, for being like my family and for their valuable friendship. Clem, Batou, Aude, Remi, Ludo, PAM, Caro, Benjamin Irina, Arnaud and Jacques, with whom I share the same experiences and we support each other every time is necessary, for their support and encouragement for every phase that someone goes through and comes with these long years of study.

In general, I would like to thank to all and each person who have lived with me the accomplishment of this doctoral thesis, with their ups and downs and for who I do not need to name because as well as them and I we know that from the most deeply part of my heart I thank them for giving me all their support, collaboration, encouragement and also for their affection and friendship.

Antonio

Acronyms and abbreviations list

For reasons of understanding the meaning of an abbreviation or acronym is displayed in the following list. Due to its meaning is only described in the first appearance in the text of this manuscript.

A

ANN: *Artificial Neural Network*.
AWG: *Arbitrary Waveform Generator*.
AGC: *Automatic Gain Control*.
ASR: *Automatic Speech Recognition*.
ACF: *Auto-Correlation Function*.

C

CAPTCHA: *Completely Automated Public Turing Test to tell Computers and Humans Apart.*
CW: *Continuous Wave.*
CNRS: Centre National de la Recherche Scientifique.

D

DDS: *Delay Dynamical System*.
DDE: *Differential Delay Equation*.
DPSK: *Differential Phase Shift Keying.*
DFT: *Discrete Fourier Transform*.
DCT: *Discrete Cosine Transform*.
DFB: *Distributed Feedback.*

E

ESN: *Echo State Networks*.
ESP: *Echo State Property*.
EO: *Electro-Optic*.

F

FEMTO-ST: Franche-Comté Électronique Mécanique Thermique Optique – Sciences Technologies.
FNN: *Forward Neural Network.*
FT: *Fourier Transform.*
FFT: *Fast Fourier Transform.*

H

HMM: *Hidden Markov Model*.

L

LSM: *Liquid State Machine*.
LCM: *Lyon's Cochlear Model*.

M

MZI: *Mach-Zehnder Interferometer.*
MFCC: *Mell-Frequency Cepstral Coefficient*.

N

NN: *Neural Network.*
NDS: *Nonlinear Dynamical System*.
NTC: *Nonlinear Transient Computing*.

O

ODE: *Ordinary Differential Equation*.
OSA: *Optical Spectrum Analyzer*.

P

PHOCUS: *towards a PHOtonic liquid state machine based on delay-CoUpled Systems*.
PDE: *Partial Differential Equation*.
PM: *Phase Modulator*.

R

RC: *Reservoir Computing*.
RNN: *Recurrent Neural Network*.
RF: *Radio-Frequency.*

S

SDR: *Spoken Digit Recognition*.
SNR: *Signal-to-Noise Ratio*.

T

TDM: *Time-Division Multiplexing*.

U

UFC: Université de Franche-Comté.
UMR: Unité Mixte de Recherche.

V

VNA: *Vector Network Analyzer*.

W

WER: *Word Error Rate*.

X

XOR: *Exclusive or*.

Mathematical and physical notations

Here, we present a summary of mathematical and physical notations, in chronological order of appearance in this manuscript.

Chapiter 1: Artificial neural networks and delay dynamics

f_{NL}: Nonlinear function (referred to as activation function).
$\mathbf{W}^I$, $\mathbf{W}^D$, $\mathbf{W}^R$ and $\mathbf{W}^{FB}$: connectivity matrices of *reservoir computing* (RC).
$\mathbf{u}(n)$: input information of RC.
w^I_{kq}: kq element of the connectivity matrix $\mathbf{W}^I$.
$u_q(n)$: q element of the vector $\mathbf{u}(n)$.
w^N_{kj}: kj element of the connectivity matrix $\mathbf{W}^N$.
$\mathbf{x}(n)$: vector of internal state of RC.
$x_j(n)$: j element of the vector $\mathbf{x}(n)$.
$\mathbf{y}(n)$: vector of output information of RC.
$x(t)$, $y(t)$ and $z(t)$: dynamical variables.
$T(t)$: water temperature.
x_n: discrete dynamical variable.
β: feedback gain of chaotic oscillator (Ikeda model).
ϕ: offset phase of chaotic oscillator (Ikeda model).
Λ: Lyapunov exponent.
K: Kolmogorov entropy.
σ, ρ and ζ: Parameters of Lorenz attractor.
τ_D: time delay (Ikeda and Mackey-Glass model) [s].
α: coupling factor (Mackey-Glass model).
b: coefficient of nonlinearity (Mackey-Glass model).
$h(t)$: impulse response.
$H(j\omega)$: transfer function.
$X(j\omega)$ and $Y(j\omega)$: $x(t)$ and $y(t)$ variables in Fourier space.
ω: cut-off pulsation of a filter [rad $\cdot$ s^{-1}].

τ: time constant of low-pass filter.
θ: time constant of high-pass filter.

Chapter 2: Modeling an EO phase implementation

$\mathbf{x}(n)$: internal state of RC.
$x_k(n)$: k element of vector $\mathbf{x}(n)$.
$\mathbf{u}(n)$: input information of RC.
$u_q(n)$: q element of the vector $\mathbf{u}(n)$.
K: number of virtual nodes within time delay τ_D.
$\mathbf{W}^I$, $\mathbf{W}^N$ and $\mathbf{W}^R$: connectivity matrices of RC.
w^I_{kq}: kq element of the connectivity matrix $\mathbf{W}^I$.
w^N_{kj}: kj element of the connectivity matrix $\mathbf{W}^N$.
f_{NL}: Nonlinear function.
w^R_{mk}: mk element of the connectivity matrix $\mathbf{W}^R$.
$\mathbf{y}(n)$: output information of RC.
$y_m(n)$: m element of the vector $\mathbf{y}(n)$.
$\delta\tau$ elementary delay between two virtual nodes [s].
τ_D : time delay [s].
$h(t)$: impulse response.
$x(t)$: dynamical variable.
σ: virtual spatial dimension.
σ_k: node position.
$\delta(t)$: unit pulse.
$p_\delta\tau$: rectangular temporal window.
u^I_σ: input signal.
ρ: input gain.
$x(t)$: reservoir state of RC.
σ^R_k: node positions for the Read-Out compared with Write-in.
δ^R_τ: Read-Out sampling period.
$\mathbf{M}_x$: response transient matrix corresponding to the dynamical processing performed by the reservoir.
$\mathbf{M}_y$: target matrix.
$\mathbf{W}^R_{opt}$: optimized matrix (post-processing).
λ: regression parameter.
$\mathbf{I}_K$: identity matrix.
$\phi(t)$: input phase modulation.
δT: time imbalancing of Mach-Zehnder interferometer (DPSK telecom demodulator).
β: feedback gain.

Φ_0: offset phase.
λ_0: laser wavelength.
ω_0: angular pulsation.
$x_\sigma(n)$: reservoir response signal.
E_0: light intensity.
δT: time imbalancing for the second arm of DPSK.
G: driver gain.
S: conversion efficiency of photodiode.
$V(j\omega)$: Fourier transform of voltage applied to the PM.
$H(j\omega)$: filter transfer function.
$X(j\omega)$: Fourier transform of dimensionless variable.
f_{NL}: nonlinear function.
$S(k)$: discrete Fourier transform.
$P(k)$: periodogram.
c_{eq}: mel cepstral coefficients.
x_σ^R: sampled response signal.

Chapter 3: Experimental implementation and results

τ: characteristic response time [s].
$\delta\tau$: spacing between two virtual nodes [s].
τ_D: time delay [s].
δT: imbalancing time [s].
$V_{\pi,\mathrm{rf}}$: half-wave voltage.
J_n: Bessel function.
ϕ: instantaneous RF phase.
$E(t)$: electric field.
V_{pp}: voltage peak to peak [V].
V_{pp}: voltage root mean square [V].
ρ: normalized amplitude of input information $u(t)$.
f_c: cut-off frequency [Hz].
$h(t)$: impulse response.
K: number of virtual nodes.
τ_c and τ_f: delay induced respectively by the cables and the fiber [s].
$\Delta\varphi$: phase difference [rad].
ω: frequency [Hz].
k: number of times that a signal phase is shifted by 2π.
a_1, a_2, b_1 and b_2: incident and reflected waves in S-parameters.
τ_G: group delay [s].
σ_x: standard deviation.

C_{xy}: auto-correlation.
E_0: light field.
β: feedback gain.
$x(t)$: dynamical variable.
Φ_0: offset phase.
N_L: number of hidden layers.
$x_k(n)$: k element of vector $\mathbf{x}(n)$.
σ: virtual spatial dimension.
σ_k: node position.
$u_\sigma^I(n)$: q element of the vector $\mathbf{u}(n)$.
$x_\sigma(n)$: reservoir response signal.
ϵ: asynchrony in the Read-Out with respect to the Write-In.

Contents

Introduction

Nowadays there are numerous methods for processing information, but most of them are still based on traditional architectures proposed more than 60 years ago by Alan Mathison Turing and John von Neumann [1]. Such computers have efficient and fast performances when executing tasks that consist of simple mathematical instructions. Nevertheless, for more complex computational tasks, such as classification of characters, atmospheric circulation, evolution of galaxies, problems of artificial intelligence systems and images or speech recognition; they are limited because of their conceptual principle of operation. An example that everyone uses but few people know is a test found on the web that helps to differentiate humans from computers, barely showing us legible words as illustrated in Figure 1. This test is known as CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) was introduced in 2000 by a specialized researcher in cryptography and artificial intelligence named Luis von Anh [2]. But the general idea was proposed in 1952 by Alan Turing, and that is why it is known as Turing test. It is a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human [3].

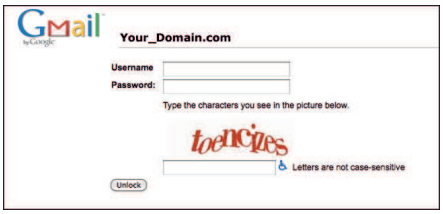


Figure 1: **The Google CAPTCHA.**

Although this test seems a bit absurd, it is considered as a complex one, but for humans this is not considered as a big issue due to the fact that our brain possesses ability of learning and training. These properties enable to change

a complex computational test from the point of view of computer science into something that looks like easier. This is one reason why the computational capabilities of the human brain have been and will be a fascinating topic for researchers worldwide and will continue to attract attention in various research disciplines. Advances in artificial intelligence are indeed increasingly making machines closer to humans, thus tests like the CAPTCHA are easily overcome by computer systems. A second example is the deep learning methods, in which the computational models are composed of hierarchical representations (many processing layers) as the human brain does for communicating, perceiving, acting, etc. The humans can learn with an efficiency that no machine learning method can approach. Nonetheless, E. Bengio and Y. LeCun suggest that in order to learn the kind of complicated functions that can represent high-level abstractions such as visual and speech recognition, it may need deep architectures [4,5]. It surmises that understanding deep learning will not only enable us to build more intelligent machines, but will also help us understand human intelligence and the mechanisms of human learning. For more than two decades computer scientists have also been working with parallel processing as the human brain does. However, not every problem can be parallelized and the fact to add a greater number of processors does not mean that a computer gets more processing power, it implies a great complexity in both hardware and software. There are problems that are inherently sequential and therefore it is difficult to execute them in parallel. Moreover, we might not only need a parallel architecture for executing a particular application, but also an adequate language to express the parallelism of the problem. This point has given rise to various lines of work and research, and one of these lines has naturally been the use of artificial neural networks to develop models of computation and processing based on the principles of biological brains. Throughout history, models of artificial neurons and recurrent neural networks have been developed in the mid-twentieth century, models such as i.e. Perceptron (1957), Adeline and Madeline (1960), Avalanche (1967), Back-propagation (1974), Hopfield (1982), just to name a few; until today they are used in various fields among which we mention the following [6]:

- Electrical engineering (signal or image processing, or control systems).
- Computer science (data mining).
- Artificial intelligence (biometrics or computer vision).
- Mathematics (nonlinear modeling).
- Neuroscience (computational neuroscience).

- Finance (financial modeling and prediction).
- Biology (bioinformatics)

Recently, in the early 2000s two scientists in different fields of research have developed a new approach of artificial neural network, which is known today in the literature as "*Reservoir Computing*" (RC). The first architecture proposed was developed in machine learning by H. Jaeger in 2001 [7], and it is known as *Echo State Network* (ESN). The other one was developed in cognitive neuroscience by W. Maass in 2002 [8], and it is known as *Liquid State Machine* (LSM). These two methods of RC are based on a new paradigm of brain-inspired concept. Although each of these models have been developed in different fields, in both cases the authors introduced a novel way of training and using complex network dynamics of neural nodes. Until 2009, these two models have been experimentally tested as computer algorithms. But today, thanks to the research work of the European PHOCUS project [9] (towards a PHOtonic liquid state machine based on delay-CoUpled Systems) and its contributors, there exist physical implementations of the operating principle of RC. These implementations consist in replacing the dynamical reservoir of neural networks by delayed feedback dynamics. It is precisely this field of research that has motivated the whole work done during this PhD thesis. Therefore, our goal has been to learn and to understand the principles of RC, as well as taking in account the contributions made by the PHOCUS group, in order to make a physical implementation based on optics and optoelectronics, in particular replacing the dynamical network by an electro-optic phase delay dynamics. Indeed, using telecommunication components in our experimental implementation of photonic RC is a strong advantage, because they allow us to achieve bandwidths much higher than those reported in the literature [10–13]. It improves the information processing speed by a factor ~ 3, leading to process around 1 million words per second, establishing state-of-the-art in classification speed.

Thesis plan

This PhD thesis is divided into three chapters, encompassing all work done within the P. M. DUFFIEUX Optics Department of the FEMTO-ST institute.

Chapter one is devoted to the RC presentation, as well as the main domains which led to our photonic realization. First, we will introduce RC models based on artificial neural networks and right after, we will present how a nonlinear delay dynamical system can replace the dynamical reservoir of classic networks. In order to understand these topics, we will introduce their necessary bases,

which will be used for the experimental implementation of a new ultra-fast photonic RC demonstrator based on the principles of RC.

In Chapter two, we will propose a brief overview of the most important principles of a RC system. After that, the theory and modeling of a RC approach using differential delay equations will be presented. Then, our experimental setup based on nonlinear delay photonic system will be introduced and each of its components will be described. We will lay greater emphasis on imbalanced Mach-Zehnder demodulator, because it provides the nonlinear function of our system. Once described our approach of RC, we will make a description of RC information processing: pre-processing (masking procedure and injection of information), internal connectivity (dynamical reservoir) and post-processing (record of transient response for later processing). At last, we will introduce and describe the test used with the aim to evaluate and to quantify the computing power and efficiency of the information processing of our RC demonstrator.

The chapter three is dedicated to the characterization study and to the experimental implementation of our RC demonstrator, which is based on electro-optic phase delay dynamics. We will make the experimental characterization of certain components (laser, phase modulators, etc.), then the time delay of our RC system will be measured through of cross-correlation function. This time delay is used as dynamical reservoir, thus here the information is addressed over virtual nodes, the latter creates spatial dimension virtually. Afterwards, the dynamical characterization will be made showing the different behaviors such as fixed point, periodic oscillations and chaos. At the end of this chapter, two tests based on speech digit recognition (SDR) will be described and detailed. Firstly, the SDR test will be conducted numerically in order to find the parameter values giving the best performance in this classification task. Secondly, the same SDR test will be experimentally performed using the parameters found with the numerical simulation. Finally, the results obtained both numerically and experimentally will be compared with those obtained with approaches based on RC systems, classic approaches (LSM and ESN) and also with Hidden Markov Models (HMM). Furthermore, as our RC system is built with telecommunication devices, which improve the bandwidth of the RC system. This photonic RC setup allows close to 3 times faster information processing speed, therefore it will be compared with the RC approaches.

The PhD thesis will end with a general conclusion and with the perspectives for a future work.

Chapter 1

Artificial neural networks and delay dynamics

Nowadays digital computers based on the so-called Turing-von Neumann architectures [1] are ubiquitous and deeply integrated in our daily lives. Furthermore, they provide many fast and efficient calculation tasks, starting from complex scientific computing through networking and communication systems up to any smartphone device's functionalities and services. The demand for greater computational power is naturally always increasing. As this demand develops, more and more problems are identified as too complex and/or time-consuming, even for the most advanced highly parallelized digital computer farms. Alternative computational paradigms have already been explored long time ago, one obvious direction being naturally suggested by operating principles of the human brain. However, most of the research dedicated to brain-inspired computational paradigm has essentially been performed through computer simulations, i.e. the use of the standard digital Turing-von Neumann computers, which nonetheless are to be replaced.

Among the numerous brain-inspired concepts, RC [14], originally known as echo state network [7] or liquid state machine [8] (see [15] for a review), has recently attracted lots of attention through an unexpected hardware physical implementation [10,11,16] making use of delay dynamical systems (DDS). DDS are here involved as complex dynamical reservoirs, surprisingly and efficiently replacing the traditional spatio-temporal neural network architectures. Moreover, DDS provide a technologically tractable solution to test and design a novel RC system in dedicated hardware. Hence as the present work has been based on artificial neural networks and DDS, this chapter will be dedicated to a brief

review of the main foundations of these two fields, as well as of their respective state-of-the-art.

1.1 Networks

Since the first half of the 20th century researchers have started to develop computer models aimed at emulating the behavior of the human brain [17]. Although many models have been proposed, all of them use a network structure in which the nodes are numerically processed involving the states of other nodes. One class of these computer models is artificial neural networks (ANNs) [18], which will be described in the next section.

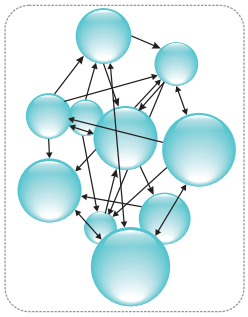


Figure 1.1: **Generic architecture of interacting (arrows indicate coupling direction) dynamical nodes (spheres) forming a network.**

The structure used for these computer models is considered as a complex system that may be decomposed into simple elements in order to be able to understand it, [One efficient way to solve complex problems is following the lemma "*divide and conquer*" (Philip II, 382-336 BC)]. It is also possible to assemble the nodes in order to produce a complex system [19], such as a network. Usually, a network is characterized by the following components: a set of nodes and the weighted connections between them as depicted in Fig. 1.1.

The nodes can be seen as computational units. They process input information and transmit the result to other connected nodes. This processing might be easily achieved by summing the inputs, and by nonlinearly transforming the result of this sum.

The connections can be unidirectional or bidirectional, depending on whether the information flow is in one direction or both. In addition, the way the nodes interact with each other through connections leads to a global behavior of the network, which cannot be observed in single elements. This means that the abilities of the network supersede the ones of its elements. For these reasons they are of great interest in different areas such as computer science, machine learning, cognitive science, mathematics, not only because they can be used as model for a wide range of physical phenomena, but also because many systems can be seen as a network [20].

1.1.1 Artificial neural networks

The ANN is a model inspired by biological neural networks, in particular the brain structure. In this type of network the nodes are seen as "*neurons*". A neural network (NN) can be defined as a set of elements of simple calculation, usually adaptive, massively interconnected in parallel with a hierarchical organization that enables to interact with any system in the same way as biological nervous system does [21]. These ANNs have become very popular because of the ease of use, their implementation and ability to approximate any mathematical functions. Besides, ANNs have a strong ability to get results of complicated and inaccurate data, they can be used to extract patterns and detect frames that are very difficult to appreciate by humans or other computer techniques.

It is noteworthy that ANNs, owing to the massive parallelism of their structure, have a number of advantages :

- **Adaptive learning.** Learning to perform tasks based on training or initial experience.
- **Self-organization.** The ANN may create its own organization or representation of the information received during the learning phase.
- **Fault tolerance**. Thanks to the own distributed information or redundant information coding via, the partial destruction of a network can lead to a degradation of its structure; however, some network capabilities may be retained even suffering of extensive damage.
- **Operation in parallel**. The neural computation can be performed in parallel, either via software or through special machines for this benefit.

These characteristics and their easy implementation into existing digital technology, make it very useful in areas such as biology, finance, industry, environment, military, health, etc. [22]. They are working in applications including

identification of processes [23], fault detection in control systems [24], modeling of nonlinear dynamics [25], nonlinear control systems [26] and optimization of processes [27].

In machine learning and cognitive science the ANNs are generally presented as interconnected systems of inputs/outputs processing neurons, which exchange messages between each other via synapses [28]. These neurons are connected into a network having weighted values that can manipulate the data in the calculation by applying some nonlinear function, commonly referred to as activation function (f_{NL}). The activation function converts a neuron's weighted input to its output activation (see Figure 1.2). The output activation is not only the sum of the activation function of the nodes, it also depends on the way neurons are interconnected within the network. Therefore, the resulting output can be tuned based on experience, making neural nets adaptive to inputs and capable of training. This process is referred to as learning in the neural-network community and it allows to modify the weights in the network to obtain the desired output value for specific inputs [29–31]. There are many methods for training neural networks, but all of them are based upon adapting the weights between interconnected neurons.

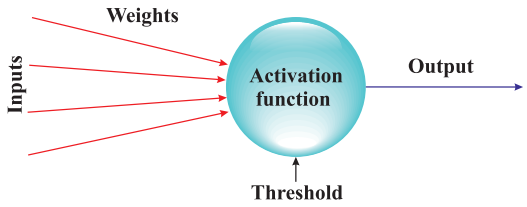


Figure 1.2: **An artificial neuron.**

1.1.2 Main types of neural networks

Depending on the research goal or application there are different models of ANNs in literature. These models of neural networks can be characterized according to different criteria, such as their internal topology, the activation function or the learning method used [32].

The eagerness to simulate the brain (i.e. its structure and its operation) have led researchers to develop numerous models of ANN [33], among them: Perceptron (1957), Adeline and Madeline (1960), Avalanche (1967), Backpropagation (1974), Hopfield (1982), Adaptive Resonance Theory (1986), etc. From the

previous models, it can be seen that this idea is more than 40 years old. Nevertheless, its development has been conceived in recent decades owing to the fact that there exists a technology that allows its efficient application.

Based on the above criteria, the main large families of ANNs are as follow: fully connected network, self-organizing map, feedforward neural networks and recurrent neural networks. In spite of that, in the following sections we only discuss the two latter ones, since this thesis work was based on the operating principles of these networks.

1.1.2.1 Feedforward neural networks

The scheme of a simple feedforward neural network (FNN) is given in Figure 1.3. As the name implies, the FNN starts with the input vector which is equal in size to the number of neurons in the first layer of the network, which processes such vector element by element in parallel. In this type of network, information is modified by the multiplicative factors of weights in each neuron and transmitted forward over the network through the hidden layers, to be finally processed by the output layer. That is why this kind of network are referred to as Feedforward [31].

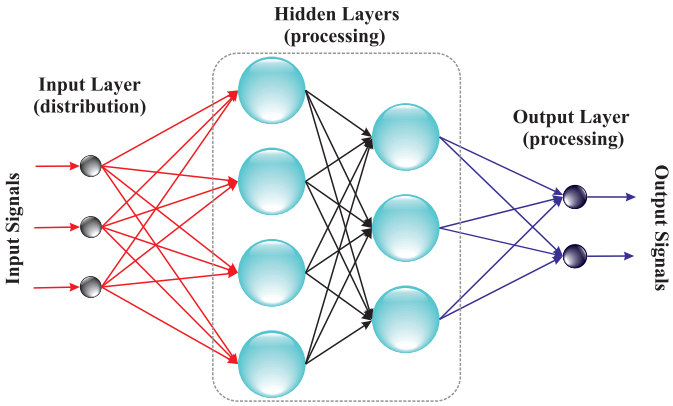


Figure 1.3: **Multilayer feedforward neural network**. On the left side of the network are inputs to the first layer of neurons, followed by interconnected layers of neurons, and finally with outputs from the final layer of neurons.

It is remarkable that this type of architecture is more simple than the others in terms of implementation and simulation. However its performance is good for applications where it is not required that the network retains information from past events to help assess future events, i.e. in spatial data classification or

robotics tasks. In such networks, there are no interconnections between layers (recurrent connections) beyond direct connection forwards to propagate information, layer per layer as a function of discrete time step [34–36]. Moreover, there are no feedback paths to act as network memory, meaning these types of neural networks are not able to process temporal information. These features allow the use of learning algorithms such as backpropagation [37,38]. This method is essentially a gradient descent procedure which searches for the solution of an error-minimization problem along the steepest descent, of the error surface with respect to the connection weights [39].

1.1.2.2 Recurrent neural networks

When working with dynamical patterns, that is, patterns of sequences depending on time, the FNNs are quite limited, due to the fact that their topology does not allow to link neurons creating loops. In recurrent neural networks (RNNs), there are feedback paths between elements in the network, thereby, the information does not only flow in one direction. One neuron is then connected to the subsequent neurons in the next layer, to the previous ones in the past layer, and eventually also to itself, through all variable weights vectors, hence explaining the name of recurrent networks. A schematic view of this topology is shown in Figure 1.4, where all nodes are interconnected with each other and also with previous nodes, through direct connections.

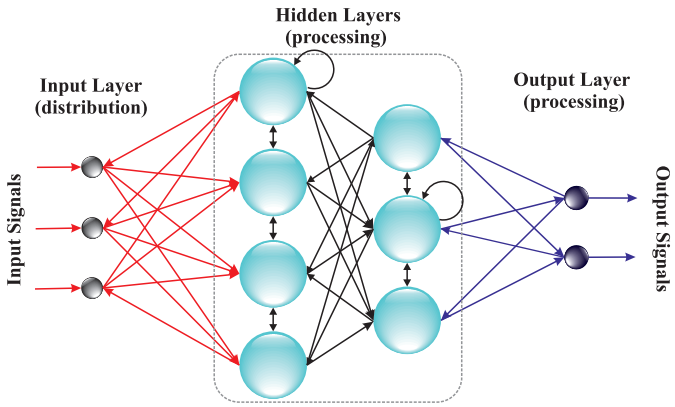


Figure 1.4: **The RNN**. As in the FNNs, it is made up of three layers of neurons (input, hidden and output). However in this type of networks there exists feedback paths.

These recurrences introduced between different layers of neurons cause the information remains circulating inside the network and a form of dynamic mem-

ory is created. In other words, the networks are able of preserving information of previous time states due to the fact that the networks are dynamical systems. Then, the state of neurons does not only depend on the current value of the input, but also on the previous state of the network [40, 41]. This property, when combined with significant temporal nonlinearities, enables to address complex tasks such as recognizing sequential patterns, speech recognition or time series prediction; nevertheless, training is slower and much more complex than the FNNs. The first training algorithm of this kind of networks appeared in 1987, when the backpropagation algorithm of the ANNs was adjusted to the RNNs applied to static patterns ("Recurrent Backpropagation").

Thereafter, the RNNs have been applied to a large number of tasks from speech recognition up to the simulation of finite automata. Nonetheless, the application of RNN addresses a large number of issues. In the case of static patterns, an RNN works presenting a pattern, after making evolve the network to stabilize their outputs. But, it is not guaranteed that oscillatory or chaotic behaviors are not reached. Although there are studies to establish the conditions to prevent it from happening, they are limited to certain very specific architectures, such as Hopfield. In the case of dynamical patterns, it is even more complex because, if the behavior of a recurrent network is a little-known fact, it is difficult to stabilize, meaning less is known about their dynamical behavior. But the real difficulty in this kind of network is the learning procedure, due to the fact that it is necessary to change the network parameters to obtain a desired dynamical behavior and while to obtain the desired output for specific inputs. This process has two targets at the same time, and for this reason it becomes highly nonlinear, more difficult and time-consuming. However, these drawbacks are overcome with recently introduced concept known as *reservoir computing* (RC).

1.2 Reservoir computing

Modern computer systems require methods of information processing more efficient to perform calculations that are increasingly complex. One of the most promising approaches to improve this efficiency is known as RC. This approach relies on using transient states of a nonlinear recurrent system basis, similarly to making a traditional neural network.

1.2.1 Introduction

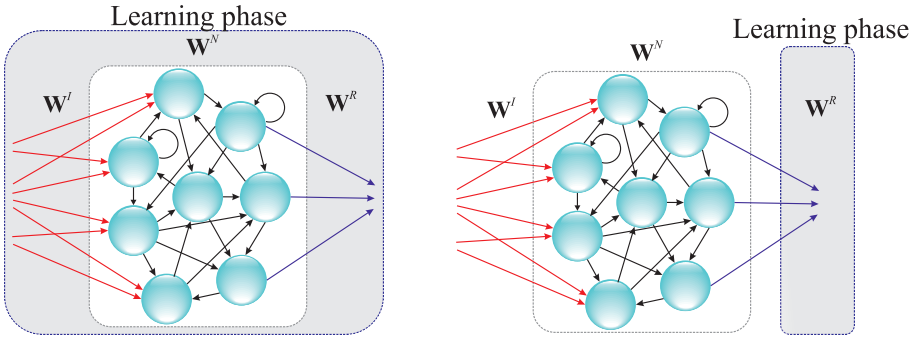
At the beginning of 2000s, a new brain-inspired concept based on family of ANNs derived from RNNs has been introduced. This approach is known as RC [14, 42–44] and it was born after two publications in the context of RNNs and brain-inspired computing [7, 8]. The first one was developed in the machine learning community, its representation of RC is referred to as *echo state network*; the second one was in the field of neuroscience, its representation is named *liquid state machine*. Nonetheless, in both cases the authors introduced a novel way of training and using complex networks of neural nodes. This new approach of RC has been quickly adopted due to its ease of use and its excellent performance. Furthermore, state-of-the-art results have been obtained in tasks that are considered computationally difficult, such as chaotic time series prediction [16, 42], or speech recognition [10–13, 45, 46].

The RC is based on RNN, it consists typically of three layers: (1) an input layer of the injected information, (2) a reservoir of interconnected neurons without any particular organization, and with adaptive mechanisms, (3) and an output layer or readout function, known as "*Read-Out*", upon which learning is applied. The RC structure is similar to the traditional ANNs, but the learning phase is significantly different. Here, the learning phase only consists in adapting the synaptic weights of the Read-Out (matrix $\mathbf{W}^R$). This procedure is depicted in Figure 1.5(b). Contrary to the conventional neural network in which the learning phase is applied to the entire network formed by the elements of the connectivity matrices $\mathbf{W}^I$, $\mathbf{W}^N$ and $\mathbf{W}^R$ (corresponding respectively to the three layers mentioned above), this procedure is depicted in Figure 1.5(a). Accordingly, the training of RC does not affect the dynamics of the reservoir, it simplifies the training and ensures its convergence. This approach makes RC learning fast and reliable, and it was even surprisingly found to be also computationally efficient on many benchmark tasks, being whether comparable or even surpassing RNN techniques. This characteristic is the main advantage of this type of networks.

1.2.2 Reservoir computing methods

1.2.2.1 Echo state networks (ESN)

ESN is one of the two pioneering RC methods [47], its approach is based on RNN. Thereby, an ESN is a special artificial RNN with the following characteristics: a large sparsely interconnected dynamical reservoir (order of 50 to 1000 neurons, previous techniques typically use 5 to 30 neurons) [42], the dynamical



(a) Learning phase for a RNN. (b) Learning phase for a RC.

Figure 1.5: **Comparison between the learning phase of a conventional RNN and a RC.**

reservoir can be excited by inputs and/or feedback of the outputs, the connection weights of the reservoir are not changed by the training, only weights from the reservoir to the output units are adapted (previous techniques tune all synaptic connections, Figure 1.5(a)). Then, the most important point here is that training becomes a linear regression task [15].

The key issue to understand how the ESN principle works is the concept of the so-called *echo states property* (ESP). Indeed, the network should asymptotically wash out its initial state when it is driven by external signal [47,48]. Due to the feedback of RNNs, the reservoir states contain traces of information about the past history of the inputs. This can be seen as a dynamical short-term memory, nonetheless according with ESP, it is known as the *fading memory* effect. Thus, the powerful information processing of these networks is only developed during the transient phase of development of *echos*.

Although, this RC method has obtained excellent performance results on a large number of complex benchmark task, obtaining an optimal dynamical regime for a good ESN is still problematic. It is because of the randomness of dynamical reservoir and because of the fact that it is necessary to find the appropriate global value to scale both the weights from the input to the recurrent network, and of the internal weights of the network. Besides, to the appropriate value of global scaling parameters (spectral radius[1], input scaling[2], output feedback scaling[3] and connectivity) one has to preserve the fading memory prop-

[1]The largest eigenvalue of the reservoir connectivity matrix

[2]The scaling in this case is like a normalization, where the input values are adjusted to fit within smaller range while the neural network is still able to learn information about the inputs

[3]It determines the feedback weights strength from the output layer to the reservoir

erty. Unfortunately at this point, this optimization is performed by numerical simulations generally through brute-force calculations. For further information see the following references [49–52].

1.2.2.2 Liquid state machine

In 2002, Maass, Natschläger and Markram proposed regardless of the ESN, a model of RC called LSM [8]. This model was originally introduced in the domain of neuroscience, with the ambition to contribute to the topic of brain science. Therefore, LSM is based on a biological approach of the operating principle of human brain and nervous system in general.

The model of LSM is based on a rigorous mathematical framework (in practice it usually consists of a recurrent network of biological neurons also known as spiking neurons[4] [53]) and a separate linear readout layer, even if its operating principle is very similar to the ESN. In LSM, the neural network may serve as an unbiased analog fading memory (informally referred to as "*liquid*") about current and past inputs. This "*liquid state*" or state vector is then seen as an analysis of contributions of all neurons in real time. The state vector can be transformed into some target output at any time by a readout function. One advantage of this method is that in the learning phase it is not necessary to take any temporal aspects into account, due to the fact that the temporal processing is done in the recurrent network. Moreover, it is possible to implement several computations in parallel using the same structure, only adjusting the synapses of these readout neurons to produce the desired target output [54, 55].

1.2.3 Operating principle

The RC approach generally uses a recurrent network as a dynamical reservoir. Therefore, its structure is quite similar to the RNN [56, 57], it is most of the time composed of three layers and each is characterized by a connectivity matrix defining the dynamic behavior of the RC, as illustrated in Figure 1.6. This figure shows graphically how information $\mathbf{u}(n)$ (where n corresponds to the discrete time) is encoded and injected through the input layer in red into a nonlinear dynamical system (i.e. a network of firing neurons). The encoding process is performed by the connectivity matrix $\mathbf{W}^I$ whose elements are denoted w^I_{kq}, and this matrix also is responsible to adapt the weights of the input information. Mathematically, the signal injected into the reservoir is represented by the following equation:

[4] It is a mathematical description of the properties of nerve cells, that is designed to accurately describe and predict biological process.

$$\mathbf{W}^I\mathbf{u}(n) = \left[\sum_{q=1}^{Q} w_{kq}^I \cdot u_q(n)\right]. \tag{1.1}$$

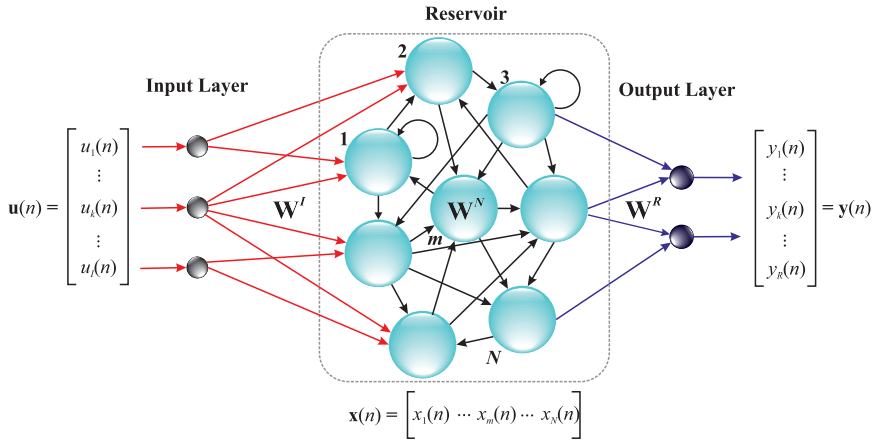


Figure 1.6: **Network topology of a RC.**

The second layer of the RC is the dynamical reservoir, where the input information is projected in a space of high dimension, so that the transient motion is consecutively triggered and it describes a complex motion in the high dimensional phase space of the dynamics of network. The synaptic weights of the network of K neurons are defined by the network coupling matrix $\mathbf{W}^N$, whose elements w_{kj}^N determine the dynamics of the reservoir. Hence, the updated inner state of the reservoir $\mathbf{x}(n+1)$ is defined without input information, i.e. in autonomous operation by:

$$\mathbf{x}(n+1) = \mathbf{W}^N\mathbf{x}(n). \tag{1.2}$$

At last, the output layer in blue is characterized by the connectivity matrix $\mathbf{W}^R$, in which the linear Read-Out operation is performed on this transient motion, whereby the expected solution can be computed. This linear Read-Out remains conceptually into the finding of a hyperplane within the dynamic phase space, which hyperplane locus in the phase space is characteristic of the problem to be calculated. In order to define the optimal location of the hyperplane in the phase space for solving the initial question, a learning procedure is used. As mentioned above in RC section, the learning phase is only performed on the Read-Out layer, while $\mathbf{W}^I$ and $\mathbf{W}^N$ matrices are left fixed. This procedure consists in extracting the optimized parameters of the characteristic hyperplane

equation. These parameters are equivalent to the coefficients of a linear combination of phase space coordinates. Moreover, the output signal $\mathbf{y}(n)$ can be written as a function of the inner state of the reservoir $\mathbf{x}(n)$ as follows:

$$\mathbf{y}(n) = f_{\mathrm{NL}}\left[\mathbf{W}^{R} \cdot \mathbf{x}(n)\right]. \tag{1.3}$$

in which the $\mathbf{x}(n)$ can be written as a function of the input signal $\mathbf{u}(n)$ and the activation function f_{NL} of each neuron as illustrated in the following equation:

$$\mathbf{x}(n) = f_{\mathrm{NL}}\left[\mathbf{W}^{N} \cdot \mathbf{x}(n-1) + \mathbf{W}^{I} \cdot \mathbf{u}(n-1) + \mathbf{W}^{FB} \cdot \mathbf{y}(n-1)\right], \tag{1.4}$$

The term $\mathbf{W}^{FB} \cdot \mathbf{y}(n-1)$ in Eq. (1.4) is the feedback from the output $\mathbf{y}(n)$ to the input $\mathbf{x}(n)$, weighted by the connectivity matrix $\mathbf{W}^{FB}$. This feedback plays an important role in applications that need to know information from past events, and it is also used to maximize performance such as time series prediction test [32].

RC realizes a specific nonlinear transformation of the input signal into a high-dimensional state space in which the nonlinear problem can be solved linearly [57], such as classification [58, 59] and prediction problems [42]. To illustrate this process, consider the values in Table 1.1 for an exclusive disjunction or exclusive-or (XOR) function. It is a logical operation that outputs true only when both inputs differ. This function is an example that cannot be computed by the perceptron [60], because it is not linearly separable. This is geometrically depicted in Figure 1.7(a) for the two dimensional cases and it is known generically as XOR problem. In this problem, the goal is to separate the blue spheres from the red ones by a linear function. Nevertheless, the geometry of Boolean XOR depicted in figure 1.7(a) shows that two straight lines are required for proper class separation.

Input Patterns		Output Patterns
00	$\longrightarrow$	0
01	$\longrightarrow$	1
10	$\longrightarrow$	1
11	$\longrightarrow$	0

Table 1.1: **XOR function.**

There exits two ways in order to solve this problem: the first presents XOR function as a linear combination of Boolean functions $A \oplus B = (A + B) \cdot \overline{AB} = A\overline{B} + \overline{A}B$. In fact XOR function is readily constructed by taking the logical

AND[5] of the OR[6] and NAND[7] functions respectively, that are linearly separable and may result from the network (perceptron) [61]. The second way to solve the problem is to increase its dimension. The main idea consists in expanding a 2D to a 3D mapping, where the problem becomes linearly separable. This map is rather straightforward if we consider carefully the function presented in Table 1.2. This is a three-dimensional problem in which the first two dimensions correspond to the XOR function and the third dimension is the AND function. Figure 1.7(b) illustrates this method, we can also see that adding a third dimension allows a plane to separate the patterns classified in category 0 from those in category 1 [61, 62], which otherwise is not possible due to the XOR problem is not solvable in two dimensions. The nonlinear mapping to high-dimensional space does not construct the hyperplane itself, but increases the probability that the problem becomes linearly separable. This feature is one of the properties of RC, the signal input is nonlinearly mapped within a high-dimensional reservoir state represented by a large number of nodes [29, 32].

Input Patterns		Output Patterns
000	$\longrightarrow$	0
010	$\longrightarrow$	1
100	$\longrightarrow$	1
111	$\longrightarrow$	0

Table 1.2: **Adding an extra input makes it possible to solve XOR problem.**

1.2.4 Properties of reservoir

All RC models use a dynamic reservoir which plays the role of memory. However, a properly functioning reservoir must fulfill several criteria among which the most important are: separation property, approximation property and fading memory [15].

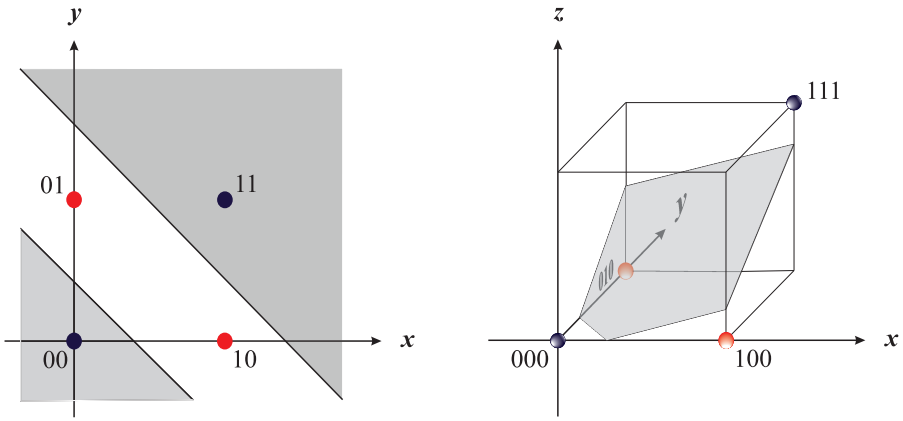
1.2.4.1 Separation property

Separation property is the capacity of the readout function to differentiate the neighboring internal states bound to the different input information $\mathbf{u}(n)$. Meaning the RC has to show a good separation property to map different input

[5] AND function returns true (1) if all its arguments evaluate to true; returns false (0) if one or more arguments evaluate to false.

[6] OR function returns true if any of the conditions are true. Otherwise, it returns false.

[7] NAND function returns false if all its arguments evaluate to false; returns true if one or more arguments evaluate to true.



(a) The XOR represented in two dimensions.

(b) The XOR represented in three dimensions.

Figure 1.7: **The XOR function mapped into three dimensions is linearly separable** [62].

streams into significantly different trajectories in the space of the internal states of the network.

1.2.4.2 Approximation property

Approximation property is the capacity of the readout function to transform the internal states of the reservoir in an output target of learning. In other words, it is the ability to produce a desired output $\mathbf{y}_{\text{target}}$ from specific $\mathbf{x}$.

1.2.4.3 Fading-memory property

Fading-memory property is the capacity to preserve the memory of the recent input (short-term memory), but can forget long term ones. In addition, this property allows systems to return to a relaxed state if no stimulus are present, it means that the memory effect is connected with the transient activation of dynamics of the network [49].

1.2.5 Applications

Nowadays, several successful applications of RC are found in a large variety of fields in the literature. One of these fields is robotics [63–69], where RC systems have been used to control and model a robot arm. Another field is dynamical pattern classification [49] and speech recognition [10, 11, 13, 16, 46, 58, 70, 71]. In the latter task RC systems have done a lot of contributions. Furthermore, RC

have been used for prediction of chaotic time series [12, 49, 72]. RC techniques already outperform state-of-the-art approaches, for example in Mackey-Glass chaotic time series, it is possible to predict with better accuracy than with classical techniques [42] or in isolated digit recognition [13].

1.3 Dynamical systems

1.3.1 Introduction

Dynamical systems are a "*young*" area of mathematics, although its origins date back to Newton with their studies of the laws of motion and the universal gravitation, and much later Jules Henri Poincaré, who proposed a new study of the differential equations, in which the qualitative prediction is emphasized instead of the quantitative one [73]. However, it was just about 40 years ago that dynamical systems were developed as a proper area, thanks to the work of outstanding mathematicians, physicists and engineers such as: James Yorke, Edward Lorenz, Leon Chua, Otto Rössler, Yves Pomeau, Stephen Smale, Yakov Grigorevich Sinai, Aleksandr Mikhailovich Lyapunov, etc.

Actually the dynamical systems are useful to investigate many real life situations [74], which factual consequence is the contribution to the field from many areas other than mathematics, such as physics, chemistry, biology and economics. The term dynamical systems refers to any physical or abstract entity whose configuration at any given time and/or space can be specified by some set of numbers called system variables, and whose configuration at a later time or further position is uniquely determined by its present and past configuration through a set of rules for the transformation of the system variables. In other words, a dynamical system is an approach to understand the nonlinear behavior of complex systems over time using stocks and flows, internal feedback loops and time delays [75].

1.3.2 Delayed feedback systems

Complex dynamical systems are typically marked by interactions and feedbacks of their subsystems. Feedback can often be efficiently described by a time-delayed self-interaction of the subsystems. The resulting delayed feedback systems are a class of dynamical systems, and they are widely used in a variety of areas, among which we can mention: analysis of the complex behavior of financial time series in economy [76], chaos control, the stabilization of unstable periodic orbits of a chaotic system, which is achieved either by com-

bined feedback with the use of a specially designed external oscillator, or by delayed self-controlling feedback without using any external force [77, 78]. In optoelectronics, the dynamics of an electro-optical oscillator subject to a delayed broadband band-pass filtering feedback is modeled by integro-differential delayed equations [79–81], in which both high and low cutoff frequencies of the feedback loop are taken into account. In photonics, particularly in the study of instabilities of semiconductor lasers due to delayed coupling, complex light dynamics are intensively studied with numerous and complex motion that depend on key parameters (feedback delay, pump current, laser nonlinearity, etc.). The latter family of systems may exhibit a variety of dynamical phenomena that can be modeled using delay differential equations (DDEs) [82]. It has been shown that using delayed feedback, an unstable system can be stabilized and effects of external disturbances can be reduced. Furthermore, in the case of a nonlinear feedback, the systems can show complex behavior like bifurcations, several types of oscillations, and chaotic solutions [83]. This behavior has been observed in optical chaos communications [84] or electro-optic (EO) phase oscillators [85].

Dynamical systems can be divided into two general classes: those in which time varies continuously and in which time passes discretely. The dynamical systems in continuous-time are expressed by equations that specify the time derivatives of the system variables in terms of their current (and possible past). These kind of equations may be ordinary differential equations (ODEs), partial differential equations (PDEs) and DDEs. On the other hand, discrete-time systems are described as difference equations giving new values of the system variables as functions of the current (and possibly past). The theory of differential equations with delay is dealt in models where the variation of the state variable x with time at each instant t depends not only on $x(t)$ but also on the values of x in previous moments. The general form of a DDE is given by $\dot{x} = f\,[x(t), x(t-\tau)]$, where f is any given linear or nonlinear function and τ is the delay time. For more general equations and their applications, the reader can refer to the following references [86–89].

Although the general theory of DDE is quite complex, the simplest possible case (a linear scalar equation with constant coefficients dependent on a single parameter of delay) serves to justify the importance of the delay equations. An example illustrating this situation has been proposed by Kolmanovsky and Myshkis in [88]. Consider a person who begins to shower and wants the water to reach optimum temperature T_d by turning a tap. It is assumed that the temperature change ΔT is proportional to the variation of the angle of rotation

$\Delta\alpha$ with coefficient k. We denote by $T(t)$ the water temperature and τ the time required for the water out of the shower. It can be assumed that the variation in the rotation tap is proportional to the difference between the temperature of the water coming out of the shower and the one wanted by the person (T_d) with constant l. This constant depends on the length of tube and temper (or pulse) of the person. It leads to a delay equation for temperature:

$$\dot{T} = -k\dot{\alpha}(t) = -kl\left[T(t-\tau) - T_d\right]. \tag{1.5}$$

The change of variable $x(t) = T(t) - T_d$ converts the Eq. (1.5) in

$$\dot{x} = -ax(t-\tau), \tag{1.6}$$

where $a = kl > 0$. Note that with this change, the approach to variable T_d becomes an approximation to zero. Eq. (1.6) is a linear DDE and, unlike what happens with linear ODE $\dot{x} = -ax(t)$, the behavior of solutions varies with the parameters a and τ. Note that if $\tau = 0$ the solutions of the equation $\dot{x} = -ax(t)$ are of the form $x(t) = x(0)e^{-at}$ and therefore converge exponentially to zero. Returning to our model, water goes cold first then get to the desired temperature gradually, warming increasingly. Probably, it is not exactly what happens in the shower; one indeed might be willing to get more hot water than is desirable before getting an acceptable temperature. Consequently, the model $\tau = 0$ is not very reliable. In the case of positive delay, Eq. (1.6) is not so easy to integrate and the behavior of solutions is given by the roots of the characteristic equation. As in the ordinary case, we seek solutions of the form $e^{\lambda t}$ and substitute directly into the equation for:

$$\lambda e^{\lambda t} = -ae^{\lambda(t-\tau)} \Longleftrightarrow \lambda + ae^{-\lambda\tau} = 0.$$

The roots of this transcendental equation provide the characteristic values that determine the behavior of solutions. In particular, all the characteristic values have negative real part if and only if $a\tau < \pi/2$. We will briefly describe the three behaviors to increase the values of $a\tau$. If $a\tau < 1/e$ (where $e = 2.718...$ is Euler's number), the solutions converge exponentially to zero, as in the ordinary case. It indicates that small delays do not influence the dynamics; corresponding to a very small tube and a quiet person, which is reaching the temperature gradually. See Figure 1.8(a), where values are taken $\tau = 1$ and $a = 0.1$. If $1/e < a\tau < \pi/2$ solutions converge to zero but oscillating. This behavior is characteristic of equations with delay, it is impossible for a first order linear ODE. Here, we are in the case of slightly longer tube and/or a person with more temper. The desired temperature is achieved but through stages of cold and heat, and surely this comportment is more familiar. Figure 1.8(b) shows

one of these solutions for $\tau = 1$ and $a = 1.4$. If $\tau \geq \pi/2$ periodic solutions appear, and even unbounded solutions. The high values of a and τ cause the water to pass successively from hot to cold without stabilizing around the desired value. Figure 1.8(c) shows a periodic solution, obtained by taking $\tau = 1$ and $a = \pi/2$. The existence of periodic solutions in a first order linear differential equation is a phenomenon that distinguishes equations with delay from the ordinary classical ones.

1.3.3 Nonlinear dynamical systems (NDS)

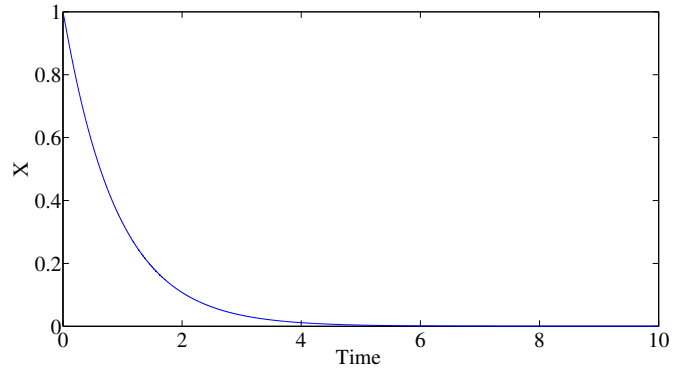
In mathematics, nonlinear systems are those whose behavior is not expressible as the sum of the behaviors of their parts. More formally, a system is nonlinear when the equations of motion and development that govern its behavior are nonlinear. In particular, the behavior of nonlinear systems is not subject to the principle of superposition, as it is the case in linear systems.

From the point of view and evolution of dynamical systems, the RC can be regarded as a complex system, in particular as a nonlinear dynamical system [10,11,13]. This type of systems operate in a transient regime and it is often ousted from the study because its behavior is too complex. Nonetheless, this is precisely the behavior of interest in our approach of RC. Also, the dynamics of nonlinear systems remain finite dimensional in practice [90], but exhibit the properties of high dimensionality and short-term memory [29]. These characteristics have made delay systems suitable candidates for a real-world physical implementation of RC and computational processing. Besides, the three main properties (approximation, separation and fading memory) of a RC are satisfied by the dynamical properties of the network.

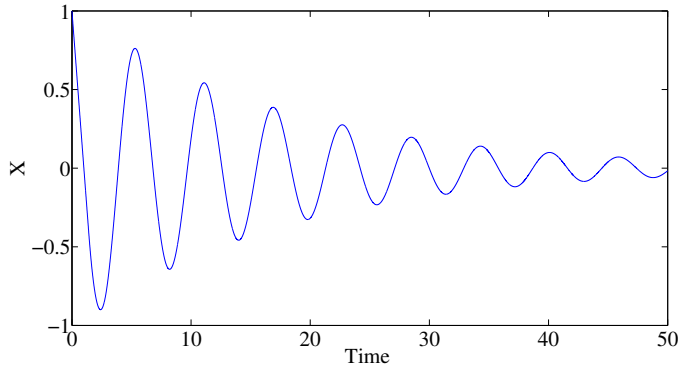
An important point to consider in the study of the behavior of nonlinear dynamical systems is the evolution of the system with respect to the initial conditions, because very small differences in initial conditions can lead to very different solutions. In addition, the initial conditions contribute strongly to the temporal and/or spatial evolution of the dynamical system, as well as its dimensionality (which by the way depends directly on the number of initial conditions).

1.3.4 Characterization of nonlinear dynamics

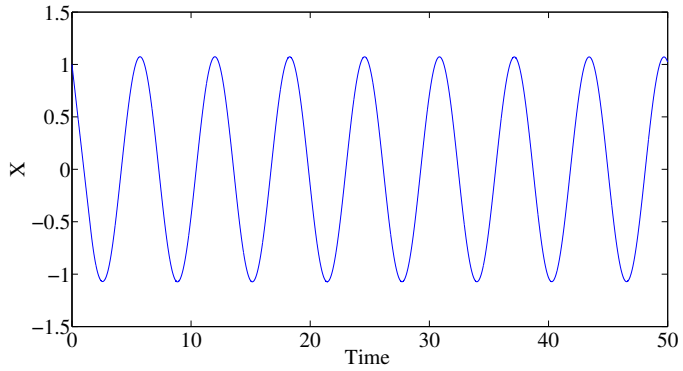
As we discussed in previous sections, the behavior of nonlinear dynamical systems is determined by its initial conditions and by the setting of its internal



(a) Solution of equation $\dot{x} = -0.1x(t-1)$.



(b) Solution of equation $\dot{x} = -1.4x(t-1)$.



(c) Solution of equation $\dot{x} = -\pi/2x(t-1)$.

Figure 1.8: **Delay effect in equation** (1.6).

parameters. Therefore, it could be simple (fixed point, limit cycle, periodic solutions, etc.) or complex (hardly predictable chaotic trajectories). The study of the behavior of these systems is based on the observation and characterization

of their asymptotic solutions. In others words, the study of dynamical system mainly begins after the transient regime has ended. For this purpose indicators are used, which depend mainly on the characteristic of the dynamical behavior of the system that we want to emphasize. The most important indicators are:

- **Lyapunov exponents** evaluate the rate with which grows the separation between neighboring trajectories, or differently, it measures the degree of disorder for the dynamical solution of concern.
- **Entropy** measures the loss of information motivated by the evolution of the trajectories.
- **The time series** or **the graphical representation** of the evolution of the orbits provides information about its regularity.
- **The frequency spectrum** or **Fourier transform**, which in chaotic systems has not defined maxima, indicate the presence of characteristic frequencies.
- **The phase space** of an n-dimensional system is the space where all possible states of a system are represented.
- **Bifurcation diagram** indicates qualitative of the dynamical system solution when a parameter is changed.

In this PhD thesis the dynamical system is studied when it is at rest, where it works with a fixed point that is globally asymptotically stable. It is a special case of asymptotic motion.

In the following sections we will explain in a little more details this case and the indicators used within the current work.

1.3.4.1 Phase space

The phase space is, by definition, a representation where the nature of the nonlinear dynamical system unfolds. The current model is a very common and graphical way to understand the asymptotic or dynamical behavior (strange attractor, singularity, limit cycle, basin of attraction, etc.) of the solutions of a differential equation. The idea to sketch the phase space is to predict the evolution of the asymptotic behavior (for the time tending to infinity) of the solutions depending on where they are initially found. Drawing enough trajectories might

enable to know whether the solutions tend to stabilize at some point, if these oscillate (periodic solutions), if they become infinitely large, if they diverge or converge, etc. Some examples of famous trajectories represented in their tractable 3-dimensional phase space, are: the chaotic attractors of Edward Lorenz [91], Otto Rössler [92] or Leon Chua [93]. Unfortunately, when higher dimensional phase space are concerned (delay dynamics are infinite dimensional phase space), trajectories become difficult to represent graphically in such a way, even when they are not chaotic.

1.3.4.2 Bifurcation diagram

A bifurcation is a typical phenomenon in the study of nonlinear dynamical systems, involving a critical point in the phase space for which a qualitative change occurs in the neighboring of the system while changing one of its parameter. This occurs when a structural instability appears, either by changing the number of solutions, both stationary and periodic, or their stability. The graph representing the solutions and their stability as a function of one or more control parameters is called bifurcation diagram. The stable solutions are usually represented with a continuous line and the unstable ones with a dashed line.

The bifurcation is an essential concept in NDSs because it represents the critical parameter values where something as small as a photon of energy or a slight fluctuation in the external temperature is magnified by iteration to reach such a size that a branch is created. Then, the system adopts a new direction, either fragmenting, falling into chaos, or stabilizing at a new behavior through a series of feedback ripple to couple the new change to its environment and thus to resist until a new critical disturbance [94].

The region of chaos in a bifurcation diagram has a complicated structure, e.g. occurring after several so-called period-doubling bifurcations involving limit cycles trajectories with periods of 2, 4, 8, etc., and then the chaos begins, without regular periods. By further forcing the system, windows with odd periods appear and then the duplication of pair periods resumes: 6, 12, 24, etc. An example of bifurcation diagram is cited in [32], it is based on Ikeda model defined by the Eq. (1.7), where x_n is the discrete dynamical variable [95]. The bifurcation diagram to this equation is shown in Fig. 1.9, it was performed for different values of β, for an initial condition of $x_0 = 0.5$ and $\phi = 0.2$. Accordingly, it is possible to observe the evolution of the static distribution of asymptotic values of x_n as a function of β ranging from 0.5 to 2.7.

$$x_{n+1} = \beta \cdot \sin^2(x_n + \phi). \tag{1.7}$$

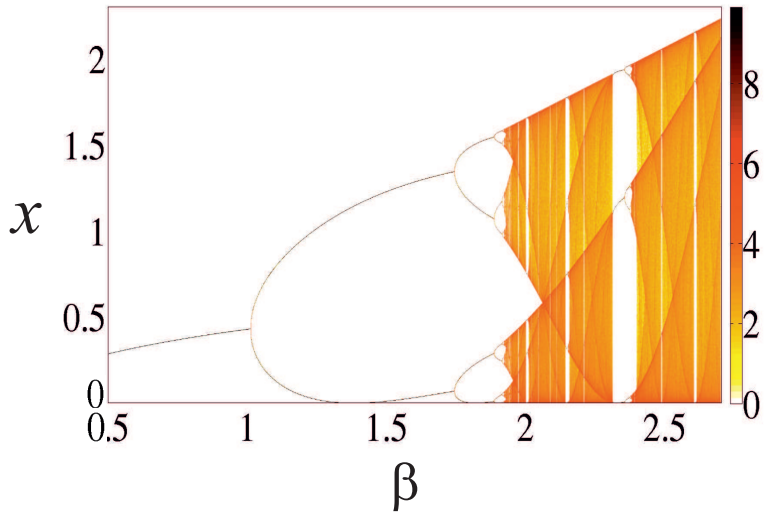


Figure 1.9: **Bifurcation diagram of Ikeda discrete time model**. The parameters used for performing this simulation were: $x_0 = 0.5$, variation of β from 0.5 to 2.7 and $\phi = 0.2$.

We can see in Fig. 1.9 that for values of $\beta < 0.9$ only a single branch of amplitude x_n is obtained (fixed point). For values close to 1 a two branches appear, that is to say, a bifurcation of the solution into an oscillation is observed, which grows in amplitude with β. Then, for values β close to 1.7 each branch is split into two level of oscillations (it is known as period doubling), resulting in periodic oscillations (limit cycles) with different numbers of levels. Finally, for values of $\beta > 1.9$ the amplitude of x_n is continuously spread over one or more amplitude intervals, typically indicating that the chaotic regime is achieved. Moreover, we can see that the use of the bifurcation diagram is a very useful tool from the qualitative point of view to quickly evidence the chaotic behavior of a nonlinear dynamical system.

1.3.4.3 Lyapunov exponents

A general indicator of the presence of chaos in a dynamical system is the maximum Lyapunov exponent. This value gives a measure of the average rate of exponential divergence or convergence of nearby trajectories in phase space. Since close initial conditions correspond to virtually identical initial states, the exponential divergence of the orbits implies the loss of predictability of the system. Any system containing at least one positive Lyapunov exponent is defined as chaotic, with the magnitude of the exponent reflecting the timescale in which

the dynamics are unpredictable.

For a more formal definition, let us start by considering the dynamical system

$$\dot{x} = F(x), \tag{1.8}$$

where $\dot{x}$ represents the time derivative of x and whose solution, we assume to be given by $f'(x)$. Also, consider two close initial conditions in phase space x_0 and $x_0 + \delta x_0$, where δx_0 is a small perturbation of the point x_0. After a time t, the solution for these particular initial conditions, is given by $f'(x_0)$ and $f'(x_0 + \delta x_0)$. If we use the orbit described by the solution with initial condition x_0 as reference, $\Delta f' = f'(x_0 + \delta x_0) - f'(x0)$ the separation between the pair of trajectories will be a function of time and will indicate, for example if the trajectories diverge that the system is unstable. For the particular case of a chaotic orbit, $\Delta f'$ function will erratically vary with time, so it is necessary to enter the exponential average rate of divergence of the pair of trajectories, that is

$$\Lambda = \lim_{t \to \infty} \frac{1}{t} \ln \left| \frac{\Delta f'}{\delta x_0} \right|, \tag{1.9}$$

where Λ is called Lyapunov exponent and it is assumed $\delta x_0 \to 0$.

In general, a dynamical system in a n-dimensional phase space has a spectrum of n Lyapunov exponents, one for each direction of the phase space. In addition, it can be shown that if $F(x)$ in Eq. (1.8) is smooth, the limit in Eq. (1.9) exists and is equal to the maximum Lyapunov exponent [96]. It is important to note that any initial separation x_0, contains, in general, any component in the direction associated with maximum Lyapunov exponent and due to the exponential growth of the latter, the effect of other exponents will be attenuated to fade over time. In this environment, it is important to note that if the exponent is negative, the trajectories get closer; if it is zero, the distance between them is maintained (periodic or quasi-periodic behavior); if it is positive, they are exponentially separated with time, even faster when the Lyapunov exponent is higher. It should be stressed, moreover, that if there is more than one positive exponent, behavior called hyper-chaos would be obtained and, on the contrary, the very negative ones would result in a rapid collapse in the dimensions of a strange attractor [97].

1.3.4.4 Kolmogorov entropy

The most comprehensive measure of the chaotic motion of a dynamical system in phase space is called Kolmogorov entropy, or entropy. Its introduction in-

volves contemplating the movement of the dynamical system from the point of view of information theory of Shannon [98, 99]. According to that theory, the uncertainty we have about the situation in the system in a certain state is defined as:

$$< Incert \cdot (n) >= -\sum_{i=1}^{n}(p_i \cdot \ln p_i), \tag{1.10}$$

where p_i is the probability that the system is in state i.

Kolmogorov used these ideas to establish a measure of the intensity of the chaos of a dynamical system through the loss of information per unit of time of this one: being a dynamical system whose chaotic evolution is represented in a phase space of dimension D following a continuous trajectory, $x(t)$, which is extracted by measuring at equal intervals of time, ς, the following sequence $x_0, x_1, x_2, \ldots, x_n$. We perform a partition of phase space into cells $\kappa, \eta, \xi, \ldots, \nu$ each of which have a volume ε^D ($\varepsilon \approx 0$) and we calculate the probability $P(\kappa, \eta, \xi, ..., \nu)$ of the system to follow the trajectory that involves the following occupations of the phase space:

- For $t = 0$ and $x = x(0)$, the system is in the alpha volume κ.
- For $t = \varsigma$ and $x = x(\varsigma)$, the system is in the alpha volume η.
- For $t = 2 \cdot \varsigma$ and $x = x(2 \cdot \varsigma)$, the system is in the alpha volume ξ.
  ...
- For $t = n \cdot \varsigma$ and $x = x(n \cdot \varsigma)$, the system is in the alpha volume ν.

$$K_n = -\sum_{\kappa,\eta,\xi,...,\nu}[P(\kappa, \eta, \xi, ..., \nu) \cdot \ln P(\kappa, \eta, \xi, ..., \nu)]. \tag{1.11}$$

According to Shannon, the amount should be proportional to the uncertainty to locate the system in a particular neighborhood after n iterations or after a time $n \cdot \varsigma$. In general, we can say that the difference $-(K_{n+1} - K_n)$ is the necessary information so that the system is in the cell $\nu + 1$ at time $(n + 1) \cdot \varsigma$. This implies that the opposite difference, $K_{n+1} - K_n$ represents the loss of information about the system between times $n \cdot \varsigma$ and $(n + 1) \cdot \varsigma$. The Kolmogorov entropy is defined as the average loss of information in any process:

$$K = lim_{\varsigma\to\infty} lim_{\epsilon\to 0} lim_{N\to\infty}\left\{\frac{1}{N \cdot \varsigma}\sum_{n=0}^{N-1}(K_{n+1} - K_n)\right\}. \tag{1.12}$$

The passage to the limit of ϵ is set to eliminate the influence of the partition made in the phase space. As information loss, K, in dynamic systems is due to the existence of positive Lyapunov exponents, there is a strong relationship between entropy K and these exponents.

1.3.5 Nonlinear dynamics modeled by ODEs

In 1963 a meteorologist and mathematician at Massachusetts Institute of Technology Edward Lorenz introduced the following ODE as a simplified model of 3 equations to describe convection flow in the atmosphere [91]

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \zeta z. \end{cases} \tag{1.13}$$

Edward Lorenz said that evolution of meteorological conditions could be predicted if the initial conditions of the system were known. Following this premise, in 1961, he performed numerical simulations and by chance and to accelerate the simulation used the results obtained the previous day. However, when he put these digits in the computer he only used the first three decimal places (0.506 instead of 0.506127) and then the simulation was executed. The results left him perplexed because they diverged with respect to expectations. This discovery puts in evidence the phenomenon of sensitivity to initial conditions (also known as "butterfly effect") of his physical system.

In an effort to explain this behavior of his model, he solved Eq. (1.13) with different initial conditions and represented it in phase space of 3D, being the dynamical variables of system $x(t)$, $y(t)$ and $z(t)$. He realized that the evolution of the dynamical variables of his system has a chaotic trajectory, because it goes closer and closer to a certain subspace known as attractor, but it is never intercepted by itself. These chaotic trajectories are illustrated in Fig. 1.10, where can be observed different types of attractors such as fixed points or limit cycles, from simple ones to more complex sets of structures such as strange attractors. Furthermore, if we consider all chaotic trajectories, one will notice that there is a certain order, as they are around two spiral wound into two plans that look like the wings of a butterfly (Lorenz attractor).

The Lorenz model Eq. (1.13) is described by 3 independent variables, thus it has 3 degrees of freedom. Nonetheless, the phase space of some systems can have much higher dimension that presented in Fig. 1.10. Nonetheless, a system defined by a set of ODE will always be limited to a finite phase space dimension, even if it has a large number of degrees of freedom. On the other hand,

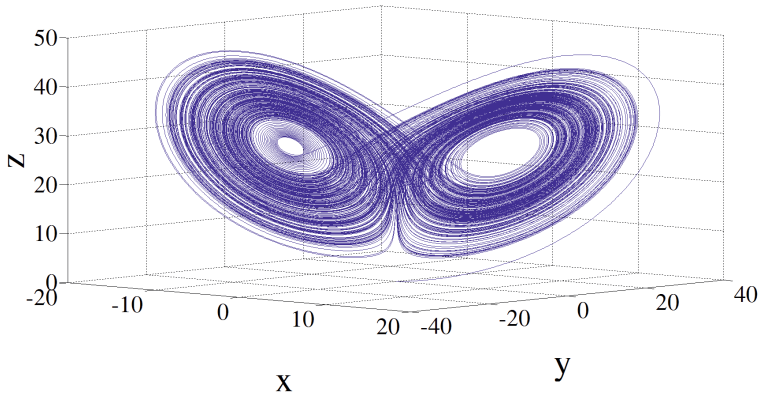


Figure 1.10: **E. Lorenz attractor.**

there exist delayed nonlinear dynamical systems with a single or more dynamical variables with infinite phase space such as: the Ikeda model [100] or the Mackey-Glass model [101], etc. described by Eqs. (1.7) and (1.14) respectively.

The Mackey-Glass model will be presented in more details in the next section.

1.3.6 Delayed nonlinear dynamical system

Delayed nonlinear dynamical systems are often found in different domains of physics (electrical, mechanical, optical, etc.). These systems have been investigated because they exhibit complex chaotic behaviors with high attractor dimension, although their mathematical description can be as simple as a DDE. For example, we can cite the Mackey-Glass model described by Eq. (1.14).

$$\dot{x} = -x(t) + \alpha \left[\frac{x(t-\tau_D)}{1 + bx^p(x-\tau_D)} \right] \tag{1.14}$$

where α is the coupling factor, b is the coefficient of nonlinearity, τ_D is the time delay and $x(t)$ is the dynamical variable delayed by τ_D which appears in the nonlinear term.

This model was proposed by Michael Mackey and Leon Glass in 1977 with the aim of studying complex rhythms observed in the production of blood cells [101]. Moreover, in order to solve Eq. (1.14) we need to know all values of $x(t)$ with $t \in [-\tau_D, 0]$, and one will notice that this interval is constituted of an infinity of values. Consequently, the dynamics of this delayed nonlinear system

is infinite in dimension because an infinite set of independent initial values of numbers is required to specify an initial condition, similarity to spatio-temporal dynamical systems [102]. As mentioned, the time delay τ_D has a direct influence on the dynamical behavior of Eq. (1.14), so that different dynamical regimes are observed for different values of parameters. These regimes are shown in Fig. 1.11, which can be observed at least three different behaviors: fixed point, multi-periodic and chaotic signal.

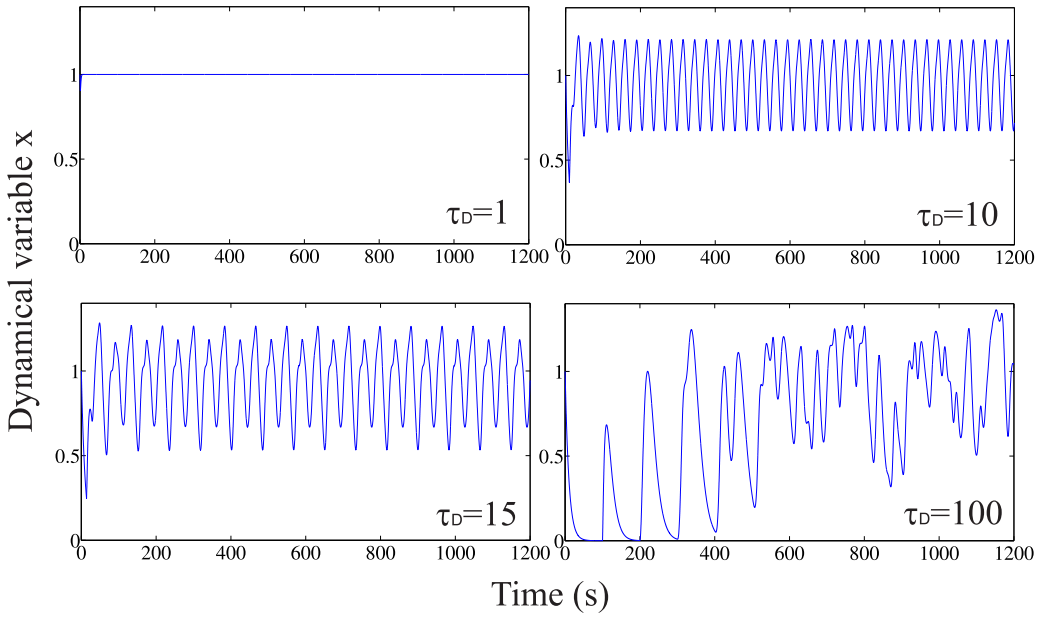


Figure 1.11: **Mackey-Glass dynamics**. Numerical simulation of Eq. (1.14), using the Runge-Kutta method of fourth order with the following parameters : $\alpha = 0.2$, $b = 0.1$, $p = 6.88$ and $\tau_D = 1, 10, 15, 100$.

The Ikeda model is another example of delayed nonlinear dynamical system. This model was proposed by the Japanese physicist Kensuke Ikeda in the 70s [100]. It was based on a nonlinear light absorbing medium placed inside a ring optical cavity. The setup is illustrated in Fig. 1.12. It is comprised of four mirrors, where two of them are partially reflecting and are used to inject a laser light beam into the ring cavity (M_1), and another one to get an output from the cavity (M_2); the other two are 100% reflecting mirrors (M_3 and M_4). The length L of the cavity determines a round-trip time of the light beam through it, which defines the delay time $\tau_D = L/c$ (where c is the speed of the light in vacuum). In addition, the nonlinear medium performs a transformation of the input and delayed light. Thus, the light field to the output is obtained after the interference between the input light beam and the one fed back by the cavity after one

round-trip.

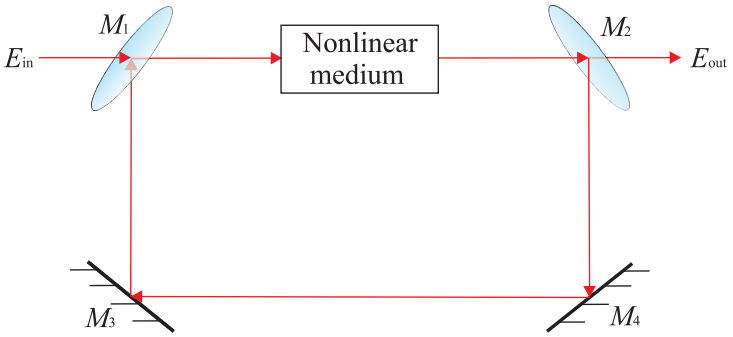


Figure 1.12: **Ikeda setup.**

The schematic setup of Ikeda model can be described by Eq. (1.15), in which the input $x(t)$ with nonlinear function is related (having parameters β and ϕ). Therefore, we can write a mathematical expression of the form :

$$x(t) = \beta \cdot \sin^2(x(t - \tau_D) + \phi) \tag{1.15}$$

where the nonlinear function is applied to the delayed dynamical variable $x(t - \tau_D)$. The importance of the role of the nonlinear transformation in the complexity of a chaotic behavior is determined by two main factors: the amplitude of the magnification factor β for the nonlinear delayed feedback term (usually considered as the bifurcation parameter), and the number of extrema of the nonlinear function. Furthermore to these two factors, the delay time is also a key element in the generation of a high-dimensional chaotic process. A major advantage in this dynamical system is its easy experimental implementation to Electro-Optic (EO) systems.

When associating a time delay (round trip in the cavity) by an iteration, Eq. (1.15) reduces straightforwardly to the Ikeda Map Eq. (1.7). This however neglects an important physical issue that cannot be ignored in a correct modeling of an experiment, the necessarily continuous amplitude motion in time. This important physical constrain related to what is called finite response times, or to limited bandwidth in the frequency domain, for any physical dynamical process. This will be addressed in the next section.

1.3.6.1 Filtering properties

In its broadest definition, a linear filter can be defined as any process that alters the nature of a signal (specifically its spectrum or frequency). The impulse response $h(t)$ is one way to characterize a filter in the time domain, applying an impulse to the input. The Fourier transform of an impulse response of a filter corresponds to the transfer function or frequency representation, characterizing the filter in the frequency domain. This characterization is carried out through its amplitude spectrum and phase spectrum.

$$Amplitude : |H(j\omega)| \tag{1.16}$$
$$Phase : \angle H(j\omega) \tag{1.17}$$

where ω represents a single Fourier component.

Let us consider an input signal $x(t)$ that we process with a filter to generate an output signal $y(t)$, as shown in Fig. 1.13. The spectrum of the output signal $Y(j\omega)$ is obtained by multiplying the input spectrum $X(j\omega)$ by the frequency response of the filter $H(j\omega)$, in other words:

$$Y(j\omega) = X(j\omega) \cdot H(j\omega) \tag{1.18}$$

where $X(j\omega)$, $Y(j\omega)$ and $H(j\omega)$ are the Fourier transforms of $x(t)$, $y(t)$ and $h(t)$, respectively.

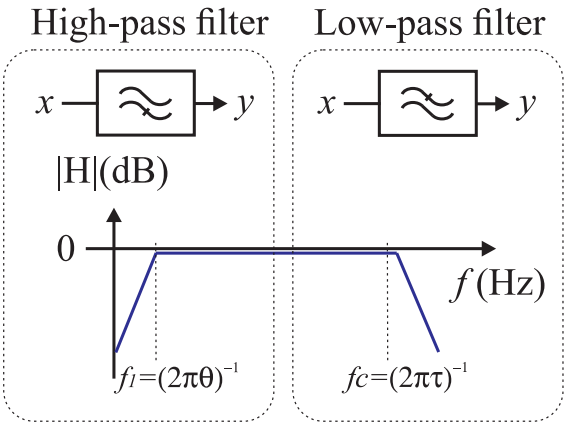


Figure 1.13: **Band-pass filter represented as a cascade of dynamic of a high-pass and low-pass filter, respectively.**

Equation (1.18) is equivalent to the convolution operation ($*$) between the signals in the time domain.

$$y(t) = x(t) * h(t) \tag{1.19}$$

$$y(t) = \int_{-\infty}^{+\infty} x(t-\theta) \cdot h(\theta) d\theta \tag{1.20}$$

From simple product of Eq. (1.18), we can obtain the transfer function of the filter as the ratio between the Fourier transform of the input and the Fourier transform of the output

$$H(j\omega) = \frac{Y(j\omega)}{X(j\omega)} \tag{1.21}$$

Low-pass filter

The Butterworth approximation $H(s)$ to a low-pass filter with cut-off frequency $\omega = \omega_c$ is

$$H(s) = \frac{\omega_c^n}{\left[s - \omega_c e^{j\left(\frac{1}{2}+\frac{1}{2n}\right)\pi}\right]\left[s - \omega_c e^{j\left(\frac{1}{2}+\frac{3}{2n}\right)\pi}\right] \cdots \left[s - \omega_c e^{j\left(\frac{1}{2}+\frac{2n-1}{2n}\right)\pi}\right]} \tag{1.22}$$

with $s = j\omega$.

While this has some complicated form, for the case of $n = 1$, Eq. (1.22) amounts to

$$H(s) = \frac{\omega_c}{s + \omega_c} \tag{1.23}$$

Substituting the filter approximation of Eq. (1.23) in Eq. (1.21), it can be written as

$$H(j\omega) = \frac{1}{1 + j\frac{\omega}{\omega_c}} \tag{1.24}$$

For an expression in the time domain, we apply a property of the Fourier transform defined by:

$$j\omega \cdot Y(j\omega) = FT\left[\frac{dy}{dt}(t)\right] \tag{1.25}$$

If we write Eq. (1.24) as follows:

$$X(j\omega) \cdot \left(\frac{1}{1 + j\frac{\omega}{\omega_c}}\right) = Y(j\omega) \tag{1.26}$$

Developing the left term of Eq. (1.26), and using the property of Eq. (1.25). We can rewrite Eq. (1.18) in temporal domain of the output signal $y(t)$ as a function of the input $x(t)$. Thus, the output $y(t)$ of the low-pass filter is the solution of a first order differential equation given by expression:

$$y(t) + \tau \frac{dy}{dt}(t) = x(t). \tag{1.27}$$

where τ is the time constant of the low-pass dynamics, linked to the cutoff frequency of the filter ω_c by the following relationship: $\tau = \frac{1}{\omega_c}$.

In the case of Ikeda setup τ is physically the non-zero response time on the intensity to phased amplitude conversion in the nonlinear medium (e.g. response time of a Ker effect).

High-pass filter

Low-pass filtering is not the only way limiting actually the bandwidth in a physical system. More subtly, one might also encounter high-pass filtering, removing some of the low frequencies in the delayed feedback process. The study of the high-pass filter is similar to the low-pass filter. Therefore, the output signal $y(t)$ is defined by a convolution product between the input $x(t)$ and the impulse response $h(t)$ as in Eq. (1.20), and in the frequency domain it is described by Eq. (1.21). Consequently, the transfer function of the high-pass filter is defined by:

$$H(j\omega) = \frac{j\frac{\omega}{\omega_1}}{1 + j\frac{\omega}{\omega_1}} = \frac{1}{1 + \frac{\omega_1}{j\omega}}. \tag{1.28}$$

where ω_1 is the cutoff frequency of the filter.

Using the property of the Fourier transform as in the case of the low-pass filter, we can write the high-pass filter as follows:

$$\frac{1}{j\omega} \cdot Y(j\omega) = FT\left[\int_{t_0}^{t} y(s)ds\right]. \tag{1.29}$$

We can write Eq. (1.28) in the following manner

$$X(j\omega) \cdot \left(\frac{1}{1 + \frac{\omega_1}{j\omega}}\right) = Y(j\omega), \tag{1.30}$$

Developing the left term of Eq. (1.29), and using the property of Eq. (1.30), we can rewrite Eq. (1.18) for the case of the high-pass filter in temporal domain,

in which the output signal $y(t)$ is a function of the input $x(t)$. Therefore, the high-pass filter is expressed by the following expression:

$$y(t) + \frac{1}{\theta}\int_{t_0}^{t} y(s)ds = x(t). \tag{1.31}$$

We obtain a new law of evolution governed by an integral involving a new constant of time θ, which corresponds to the characteristic response time of the high-pass filter, linked to the cutoff frequency of the filter ω_1 by the following relationship $\theta = \frac{1}{\omega_1}$.

Band-pass filter

Finally assuming the simplest approximation of a first order high-pass and low-pass filter described by Eqs. (1.31) and (1.27) respectively, considering that Ikeda model is replaced by a band-pass filter profile, then, Eq. (1.15) can be rewritten as simple as a scalar first order differential equation for the variable $x(t)$, where the dynamical law of evolution governing the band-pass dynamic is defined by a nonlinear delayed integro-differential equation as follows:

$$x(t) + \tau\frac{dx}{dt}(t) + \frac{1}{\theta}\int_{t_0}^{t} y(s)ds = \beta \sin^2\left[x(t-\tau_D) + \phi\right]. \tag{1.32}$$

here, the left-hand side is typical of a stable linear second order dynamics, with a characteristic response time τ and an integral (high-pass filtering effect) response time θ; its role is only to limit the feedback oscillation bandwidth. The presence of delay in Eq. (1.32) causes the band-pass dynamic to have complex solutions. Hence, to solve this differential equation, it is necessary to know the initial condition, in other words the value $x(t)$ for $t = 0$, as well as all initial conditions on a time interval $[-\tau_D, 0]$, corresponding to an infinite-dimensional number initial condition $x(t)$ with $t \in [-\tau_D, 0]$. The dimension of a delayed dynamics is then infinite like those of the spatio-temporal dynamics [102]. Precisely this dynamic complexity generated by the delay time is the main reason why they are used in neuromorphic computers, replacing the classic neuron networks (a spatio-temporal dynamics of infinite dimension can be represented as a network of interconnected nodes). Hence, in this manuscript the delayed nonlinear dynamics will be the main basis for the construction of our approach to RC.

1.3.6.2 Reservoir computing based on delayed feedback system approach

RC was developed in the field of ANNs and neuroscience. Nevertheless transposing these concepts in a real-world, experimental implementation was not up till now possible. Since the recent demonstration of its hardware implementation possibility, nonlinear physics and nonlinear dynamics communities have started to also address the topic. Using a delayed feedback system the drawbacks of the experimental implementation of a system with spatio-temporal dynamics can be transposed in a system with only temporal dimension in which the spatial dimension is recreated virtually. Some examples of how RC has been successfully performed using a complex system are: in 2009 Verstraeten explained how the role of reservoir can be seen as a complex nonlinear filter that projects the input signals into a high-dimensional space, where the classification can be done much more accurately [48]. Also in [10] Appeltant *et al.*, introduced in 2011 the general principle about how delay dynamical systems can perform efficient computation. They replaced the entire network of connected nonlinear nodes by one single nonlinear node subjected to delayed feedback. This dramatically reduces the experimental difficulties of the ANN for computing purposes and also demonstrates the extensive computational processing power hidden in even the simplest delay-dynamic system. Another significant demonstration of the use of a complex system like reservoir is given in [13] by Martinenghi *et al.*, in 2012. In this case they reported an experimental demonstration of a hybrid optoelectronic neuromorphic computer based on complex nonlinear wavelength dynamics including multiple delayed feedbacks with randomly defined weights. This implementation allowed to find that using a multiple-delay photonic system with a reduced number of nodes exhibits a comparable computational efficiency (standard benchmark test of spoken digit classification [10,11,16,58,70,71]), with the same order as the best results achieved so far for the same test.

These concepts of the experimental implementation of Martinenghi's demonstrator are used and they are part of the central idea of this doctoral thesis. As described in more detail in the next Chapter, we have taken the general idea of neuromorphic photonic RC, but we have made several and innovative modifications. To begin the realization of a photonic RC can technically shift the intrinsic millisecond time scales of the brain [12, 46, 103], down to pico or even femtosecond ones available in standard Optical Telecommunications, thus basically gaining 9 to 12 orders of magnitude in the processing speed. The present work replaces the dynamical network by an EO phase delay dynamics built with Telecom bandwidth devices, and thus providing ultra-fast information processing while implementing RC concepts with dedicated hardware. The entire process

is detailed in Chapter 2.

1.4 Overview

In this Chapter, we have shown how the properties of ANNs and the dynamical system can be used as a dynamical reservoir in the different models of RC developed along its creation. First, we explained how an ANN can be represented as an interconnected network of neurons. These neurons have weight values that can be manipulated applying a nonlinear function referred to as activation function. Thus, the resulting output can be tuned based on experience or adaptive training. This process is known as learning. Then we explained a particular type of ANN, in which there are feedbacks or recurrences. Thanks to them such networks can generate a memory effect. Therefore the NN can be seen as a complex dynamical system projecting the input signals into a high-dimensional space. The characteristics of this kind of network are used in two methods that gave birth to the concept of RC. These methods (LSM and ESN) are based on the structure of RNNs. However, with them, the drawback of the learning phase of the ANNs is outperform. Here, the learning phase is performed in the readout layer only leaving the other layers fixed, thus resulting in an always converging training. Finally, we presented the delay systems and how their features have made them suitable candidates for a real-word physical implementation. In this case, the RC is seen as a delayed nonlinear dynamical system that may be represented by DDE. On the other hand, we showed three examples (Verstraeten, Appeltant and Martinenghi RC approaches) in which the dynamical reservoir of ANNs has been replaced by a complex system, taking major interest in the experimental demonstration of a hybrid optoelectronic neuromorphic demonstrator based on complex nonlinear wavelength dynamics.

Chapter 2

Modeling an EO phase implementation

In this chapter, we present the theory and modeling of our RC approach, starting with a brief review of the principle of RC systems based on RNN. Then we make a description where the dynamical reservoir of interconnected neurons for the RC has been successfully replaced using a delay dynamical system (DDS). Also, we explain the masking procedure which consists in adapting the signal input information before being injected into the reservoir, as well as the connectivity of the virtual nodes within the delay time τ_D and the separation distance $\delta\tau$ between them. Furthermore in the next section, we describe how the output signal is sampled either following a synchronous or asynchronous mode.

Subsequently, we present the experimental setup which is based on nonlinear delay photonic system. The photonic system used in this work is based on EO phase delay dynamics. We will introduce its physical principles, mathematical model and dynamic characterization. Furthermore, in this chapter we describe two types of tests used in order to evaluate and quantify the efficiency of information processing of our photonic RC.

Finally, the aim of this chapter is to describe how a complex system with only temporal dimension like a nonlinear delayed dynamics can emulate the spatio-temporal dimension of an RNN using a time-division multiplexing (TDM) technique. We introduce an EO phase approach of RC which is built with telecom devices. It allows to increase the bandwidth of our system and consequently faster processing is reached.

2.1 Modeling of RC based on RNN

The RC based on RNN was introduced in Chapter 1, where a NN was used as dynamical reservoir. Besides, its operating principle, as well as its mathematical model were presented in section 1.2.3. Nevertheless in this section, the goal is to showcase the main properties required for reservoir implementation.

The Fig. 2.1 shows the structure of RC. It is generally composed of three parts: an input layer, a dynamical reservoir and an output layer.

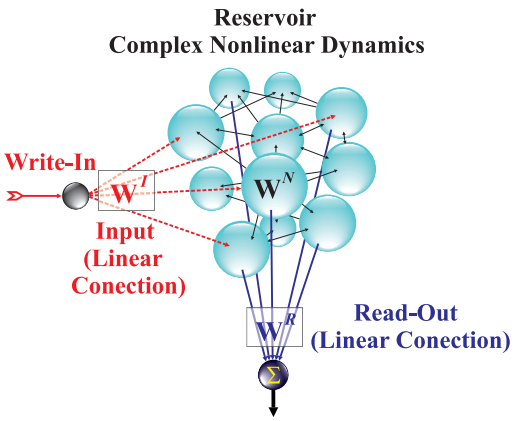


Figure 2.1: **RC scheme based on an approach of type RNN.** Its structure consists of three layers: an input layer, a reservoir and a output layer.

The dynamical reservoir can be considered as a classical network formed by K spatially distributed nodes $\mathbf{x}(n) = [x_1(n)...x_k(n)...x_K(n)]^\intercal$. The discrete time dynamics of the network (as time n is increased) involves the coupling of each node with the other ones according to the network coupling matrix $\mathbf{W}^N$ which is randomly set. Additionally to the internal dynamics, the network also evolves due to the injection of the input information $\mathbf{u}(n)$ to be processed. The information injection is ruled by a input layer connectivity matrix $\mathbf{W}^I$, resulting in a network dynamics reading as:

$$x_k(n) = f_{\mathrm{NL}}\left[\sum_{j=1}^{K} w_{kj}^N\, x_j(n-1) + \sum_{q=1}^{Q} w_{kq}^I\, u_q(n)\right], \tag{2.1}$$

where the f_{NL} term is the nonlinear transformation due to the activation function. It represents the nonlinear node sensitivity to the cumulated stimuli coming from both the network internal connectivity and the input information con-

nectivity. A popular form for f_{NL} is the tanh$-$function also known as sigmoid function. The $\mathbf{W}^I = \left[w^I_{kq}\right]$ term are the input connection weights with $q \in [1, \ldots, Q]$ and the $\mathbf{W}^N = \left[w^N_{kj}\right]$ term are the network internal connections of the reservoir with $j \in [1, \ldots, K]$. According to the principles of RC, the $\left[w^N_{kj}\right]$ elements determine the dynamics of the reservoir and they are not changed in the learning phase. Therefore, the internal state of the reservoir denoted $\mathbf{x}(n)$ in Eq. (2.1) is a vector whose elements depend both on the input information $\mathbf{u}(n)$ but also of connectivity matrices $\mathbf{W}^I$ and $\mathbf{W}^N$, as well as the previous states of the network nodes $\mathbf{x}(n-1)$. Then, the Read-Out operation extracting the computed $M-$dimensional solution consists in a linear combination of each node state according to the output connectivity matrix $\mathbf{W}^R$. Mathematically, we can express the output signal $\mathbf{y}(n)$ in the same way as in Eq. (1.3), by the following relationship:

$$y_m(n) = \sum_{k=1}^{K} w^R_{mk}\, x_k(n). \tag{2.2}$$

In Eq. (2.2) the activation function f_{NL} does not appear anymore, because it is implicit in the nonlinear transformation of the node connections in the network $x_k(n)$, where k is the spatial dimension and n is the discrete time.

2.2 RC based on DDE

This section presents the general principles that explain how a dynamical system was successfully proposed as an unconventional hardware implementation of RC. The main idea is to use a nonlinear delay dynamics instead of the standard spatio-temporal structure consisting in a dynamical network of interconnected nodes (as this was mathematically described in the previous section). One of the most important arguments on which relies this approach is the fact that these systems have a high complexity (introduced by the delay) and at the same time one can have access to it. Moreover, delay dynamical systems are indeed known as having an infinite dimensional phase space and it can be modeled by a DDE, as presented in sections 1.3.2 and 1.3.6 with Eqs. (1.5) and Eq. (1.14) respectively. Consequently, in this approach of RC the NN complexity is replaced by a DDS. It provides a valuable simplification to the study and an attractive hardware solution, especially when photonic implementations are concerned, for which high speed optical telecom device can potentially provide unprecedented processing speed. Unlike the physical realization of a system

with spatio-temporal dimension, which imposes the development of a technology in 3D (as neurons of the human brain), the DDS can be oriented in one dimension (temporal dimension) following an approach of signal processing. It enables the experimental implementation by mean of TDM[1], as illustrated in the scheme of Fig. 2.2. For this, the injection of the input information $\mathbf{u}(n)$ into the reservoir is done serially (so-called pre-processing, detailed in section 2.2.1) and not in parallel as was done in the case of RC based on RNN (see Fig. 2.1). As information has been temporarily addressed through TDM over the virtual nodes within the reservoir, the spatial dimension is virtually created within delay interval τ_D [105]. As a result, the spacing between each successive node $\delta\tau$ is defined by the relationship τ_D/K, K being the number of desired nodes.

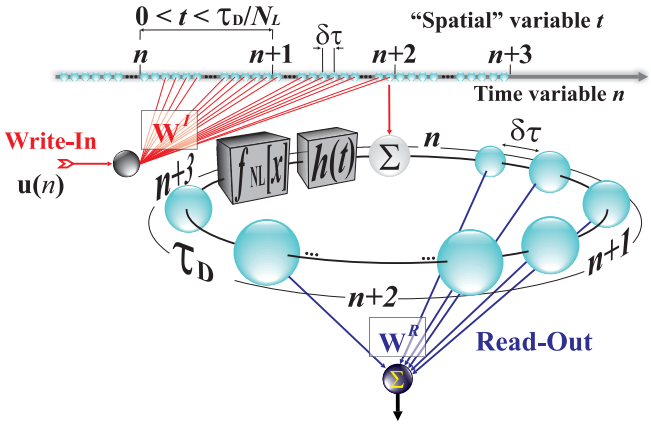


Figure 2.2: **Principles of a RC based on DDE**. A single nonlinear element subject to delayed feedback, where $f_{\mathrm{NL}}(x)$ stands for the nonlinear transformation of the system, and $h(t)$ denotes the impulse response of the system, respectively.

In order to illustrate and justify the use of such systems, let us consider the simplest scalar model of DDE of the following form:

$$\tau\,\dot{x} = -x(t) + f_{\mathrm{NL}}[x(t-\tau_D)], \tag{2.3}$$

where τ_D is the delay time and τ is the characteristic response time limiting the fastest time scales allowed in the feedback loop. The common case, $\tau \ll \tau_D$ has widely demonstrated the emergence of chaotic attractors [106]. The infinite dimensional character can be straightforwardly explained by the kind of initial conditions actually required for the unique definition of a given trajectory: it

[1]TDM is a method of transmitting and receiving independent signals over a common signal path by means of synchronized switches at each end of the transmission line so that each signal appears on the line only a fraction of time in an alternating pattern [104].

consists of a functional[2] $x(t)$ with $t \in [-\tau_D; 0]$ belonging to an infinite dimensional space, as introduced in section 1.3.6.

The spatio-temporal emulation of a delay dynamics has been indeed known for more than 2 decades [105]. In that work, they proposed an experimental system with delay-differential dynamics, consisting in a CO_2 laser driven by a signal proportional to the output laser intensity. In the feedback loop a delay line (τ_D, long time scale) and an amplifier (with bias voltage B) are placed between the modulator and the detector. The case explored was $\tau_D \gg T_c$, with T_c being the correlation time of the laser intensity (short time scale). The results obtained increasing B show different dynamical behaviors with pulse shapes similar to those that are characteristics of the laser without delay. Measuring the time correlation over long times, evidenced the appearance of other features. They imply the existence of two different times scales (multiple scale) in the interplay between the nonlinearity and the delay. It suggests an organization of the data in two-dimensional "space-time" domain, like in the numerical technique for solving DDEs. Following this procedure, the state of Eq. (2.3) is determined by all values x in the interval $[t - \tau_D, t]$ and its evolution consists of a $N-$dimensional discrete mapping taking samples at intervals $\Delta t = \tau_D/(N-1)$. The N-dimensional mapping is defined by a virtual space variable s ranging from 1 to N, and by k being a discrete index counting the delays units. As a result, the space-time representation in $s - k$ domain provides a visual discrimination among the different types of chaotic behavior. For example when $\tau_D \gg T_c$ the points along the s axis decorrelate, and then the correlation revives after one delay τ_D, showing a cellular structure of data as in space-time turbulence, which is much more evident than the one-dimensional representation.

Recently, this approach was studied by our group to demonstrate the existence of Chimera states in delay dynamics [107]. The setup used is based on a modified Ikeda time-delayed equation. It is modeled by a band-pass delay dynamics and analyzed in a virtual, spatial representation. This space-time representation uses a virtual continuous space variable corresponding to a short time scale σ of the dynamics, of the order of τ in Eq. (2.3) or T_c in [105], whereas the long time scale of the order of the delay τ_D defines a discrete time variable with $n \in \mathbb{N}$ and $\sigma \in [0; \tau_D]$. Thus, the evolution of the functional can be defined by the generic term $\{x_\sigma(n) = x(t) \,|\, t = n\tau_D + \sigma\}$, with σ being an index corresponding to a temporal position within a single delay interval and by n

[2] A functional is a function that takes functions as its argument; that is, a function whose domain is a set of functions.

stands for the time steps. With this formalism, in the framework of RC, it appears more convenient to use the form of Eq. (1.20) to rewrite the usual DDE (2.3) into an integral convolution form involving the impulse response $h(t)$, the latter characterizing the linear differential process ruling the delay dynamics:

$$x(t) = \int_{-\infty}^{t} h(t-\xi)\, f_{\mathrm{NL}}[x(\xi-\tau_D)]\, \mathrm{d}\xi, \tag{2.4}$$

where $h(t)$ is the linear impulse response of the delay oscillation loop, which takes the usual form $h(t) = (1/\tau)e^{-t/\tau}$ for $t \geq 0$ (and 0 for $t < 0$ due to causality) for the case of Eq. (2.3). One could notice here that (2.4) is indeed more general than (2.3), since it describes any kind of linear filtering in the delayed feedback loop, not only the first order kind of concern in Eq. (2.3).

In order to model RC, the temporal dimension can be described as the sum of short time scale corresponding to the virtual spatial dimension σ_k, and the long time discrete $n\tau_D$. Therefore, the convolution product in Eq. (2.4) can be described by a more straightforward discrete time map for the functional, thus highlighting the spatio-temporal analogy:

$$x_k(n) = x_k(n-1) + \int_{\sigma_k-\tau_D}^{\sigma_k} h(\sigma_k-\sigma)\, f_{\mathrm{NL}}[x_\sigma(n-1)]\, \mathrm{d}\sigma, \tag{2.5}$$

where the virtual node connections $x_k(n) = x[t(n,\sigma_k)]$ with $t(n,\sigma_k) = n\tau_D + \sigma_k$, provide a writing of the delay dynamics which precisely points at the multiple scale features of such dynamics. Hence, the multiple time scale is used in order to highlight a space-time analogy with the standard discrete time RC model in Eq. (2.1). We recover the discrete time dynamics as the iteration of the functional from $n-1$ to n, whereas the spatial coupling is governed by the integral term where the impulse response $h(t)$ of the dynamics plays the role of a coupling (matrix $\mathbf{W}^R$) between the nodes in a continuous virtual space in σ. Therefore, the state of virtual node x_k at a k position and a n time depends both on the neighboring nodes at the positions σ $k-1$, but also its own previous state at time $(n-1)\tau_D$.

Equations (2.3)-(2.5) correspond to a so-called autonomous dynamics. RC is a computing paradigm which deals with the processing of time dependent information through transient dynamics, thus it is implicitly a non-autonomous system. The typical way the information to be processed is mixed within the delayed feedback loop, is usually to add the encoded input information $\mathbf{u}(n)$ to the dynamics $x(t)$, the added two contributions forming the argument of the nonlinear transformation $f_{\mathrm{NL}}[\cdot]$.

2.2.1 Data pre-processing

We introduced previously how a DDS can successfully emulate the spatio-temporal dynamics like an RNN. It provides a conceptual solution for its experimental implementation as well as similar expansion properties for computational capabilities. We have also explained that the spatial dimension is virtually created within the time delay interval τ_D, for which the input information $\mathbf{u}(n)$ needs a pre-processing in order to adapt the signals before being distributed over the virtual nodes of the reservoir. In this section, we will detail how input data are processed.

In our approach the information injection into the virtual space σ of a delay dynamics is achieved through TDM. It defines the node position $k \in [1; K]$ as time position $\sigma_k = (k-1)\delta\tau$ within time delay. Similarly to the information contribution in Eq. (2.1), a given virtual node $x_k(n)$ receives an input information signal contribution (Write-In) in order to increase variability in the network. Nonetheless in this case, the input information is weighted by the synaptic weights corresponding to the elements w^I_{kq} of a new function (so-called mask). This mask function is defined by $(q-1)\delta\tau \leq t < q\delta\tau$, with $q \in [1, \ldots, Q]$, which redefines the connectivity $(K \times Q)$ matrix $\mathbf{W}^I$ (we recall that K is the number of virtual nodes and Q are the amplitudes of the input). The mask output is then a piecewise constant function, constant over an interval $\delta\tau$ and periodic with period τ_D. The Q amplitude values of the mask during each interval of length $\delta\tau$ are randomly chosen and these values weight the serial input information. In this way, the information contribution is the product of the input $\mathbf{u}(t)$ and the mask function. This masking procedure is depicted in Fig. 2.3, where values $w^I_{kq} = \begin{cases} \pm 1 \\ 0 \end{cases}$ and it can be defined mathematically by Eq. (2.6):

$$\sum_q w^I_{kq}\, u_q(n)\, \delta(t - n\tau_D - \sigma_k), \tag{2.6}$$

where $\delta(t) = 1$ at $t = 0$ and is equal to zero at any other time. In the particular case of a photonic delay dynamics as reported in this thesis, the unit pulse $\delta(t)$ is replaced by a physically more realistic rectangular temporal window also known as an "order zero" sample and hold operation:

$$p_{\delta\tau}(t) = \begin{cases} 1 \text{ if } t \in [0; \delta\tau] \\ 0 \text{ otherwise.} \end{cases} \tag{2.7}$$

Substituting the $\delta(t)$ by the temporal window of Eq. (2.7) in the information contribution Eq. (2.6), it can be rewritten as follows :

$$u_{\sigma}^{I}(n) = \sum_{k=1}^{K} \left[\sum_{q=1}^{Q} w_{kq}^{I} \, u_q(n) \right] \, p_{\delta\tau}(\sigma - \sigma_k), \tag{2.8}$$

where the input signal $u_{\sigma}^{I}(n)$ is a stepwise constant function defined according to the TDM principle performing the distribution of each input information sample $\mathbf{u}(n)$ over each of the virtual nodes defined within a time delay interval. Considering the input signal, the dynamical law for the transient can be identified with the generic Eq. (2.1) modeling initially the ESN concept. In this way, DDS allows a formalization of a signal processing much like to this one used in ESN and thus the dynamical law can be described in relation to the following equation:

$$x_k(n) = x_k(n-1) + \int_{\sigma_k - \tau_D}^{\sigma_k} h(\sigma - \sigma_k) \times f_{\mathrm{NL}} \left[x_{\sigma}(n-1) + \rho \cdot u_{\sigma}^{I}(n-1) \right] \, \mathrm{d}\sigma. \tag{2.9}$$

where ρ is an adjustable parameter (usually referred to as input gain).

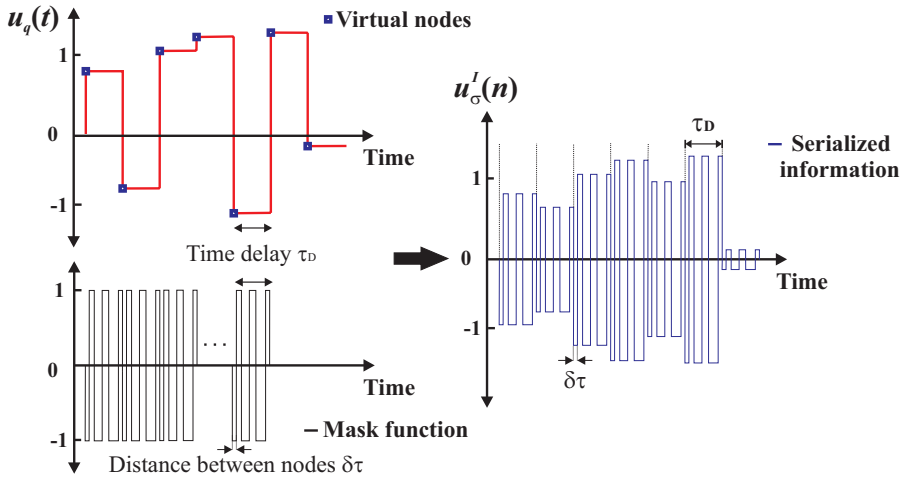
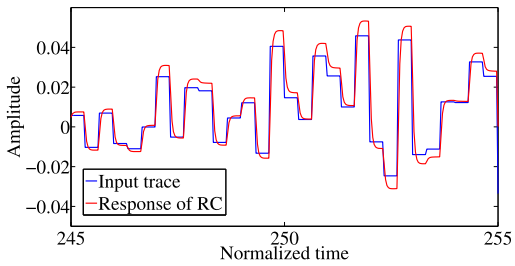


Figure 2.3: **Description of pre-processing of input information**. The virtual nodes corresponding to the input information $\mathbf{u}(n)$ are in blue (for simplicity $u(n)$ is considered here as a scalar and consequently $\mathbf{W}^{I}$ is a vector and not a matrix anymore), which have been held up during a time delay τ_D, then they are multiplied by the elements of the mask function whit amplitudes ± 1, and 0 (dashed lines in black). The role of the mask is to distribute information throughout τ_D. Finally, the blue line is the information adapted and ready to be injected into the RC system.

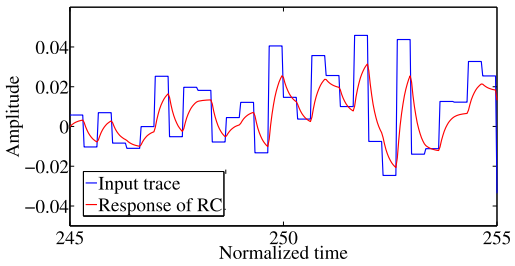
2.2.2 Internal connectivity of the reservoir

In the same way as in a traditional NNs, the interconnections of the virtual nodes within the reservoir are governed by several parameters: the separation distance between each successive node $\delta\tau$, the delay time due to the feedback loop τ_D, the integral response time θ and the response time τ of the RC system set by the profile of the dynamic low-pass or band-pass filter. However, for a good performance of the system it is necessary to find a balance between these constants, because it will determine the different types of operation, as illustrated in Fig. 2.4. The value of this balance was determined empirically (numerical simulations) by the consortium of the PHOCUS project [108]. They found that good performance occurs when the RC works under transient conditions, wherein the distance between two successive virtual nodes is small enough with respect to the response time of the system, thus the time scales are related $\delta\tau \leq \tau \ll \tau_D$.

If $\tau \ll \delta\tau$, the response of RC system very rapidly reaches its steady state, as depicted in Fig. 2.4(a). In this case, there is no coupling between the virtual nodes and each of them is then decorrelated from the rest of the network nodes, as well as, those of their successive states. This means that, the reservoir state $x(t)$ is only determined by the instantaneous value of the input $u(t)$ and the delayed reservoir state $x(t - \tau_D)$. Under this condition the dynamics are not sufficient for good performance of RC system.



(a) Response of RC very close to steady state working in asymptotic regime for $\tau \ll \delta\tau$.



(b) Transient regime of RC for $\tau > \delta\tau$, in this case the reservoir is a network of virtual nodes strongly interconnected.

Figure 2.4: **Internal connectivity of the reservoir**. In blue input signal $u(t)$ and in red response output.

Considering $\tau > \delta\tau$, the response of RC operates in transient regime and this one never reaches its steady state, as depicted in Fig. 2.4(b). In this case, the interconnections of the virtual nodes are more complicated, because the distance between two neighboring nodes is much smaller than the response time

of the RC system. For this reason, the reservoir state $x(t)$ at time t depends on its previous state and those of their neighbors. In this transient dynamics the best performance of RC system is obtained. For example in a test of speech recognition (isolated digits) the best results were found when the system worked in transient regime. This regime can be mathematically defined by the following relationship $\delta\tau/\tau \simeq 0.2$ [29,108]. Using the total number of virtual nodes K and the time constants of the system τ, τ_D, the relationship $\delta\tau/\tau$ can be written as follows :

$$\frac{\tau_D}{K} \simeq \frac{1}{5}\tau. \tag{2.10}$$

This empirical rule will be used throughout this thesis in numerical and experimental tests.

2.2.3 Post-processing of the output information

Practically the output signal of reservoir $x(t)$ is a linear combination of the states of the virtual nodes and it depends on one dimension (time) as the input signal $u(t)$. For this reason, a post-processing of the output signals or temporal traces is applied, and it is realized in two steps. The first step consists in sampling and formatting the temporal traces recorded at the output of our reservoir. The second step consists in applying the learning phase in order to find the good coefficients of Read-out function.

2.2.3.1 Read-Out

The output signal of the reservoir is sampled following values $\delta\tau$ and τ_D (corresponding to the virtual distance between two successive nodes and the value of the time delay respectively) and then the Read-Out $\mathbf{y}(n)$ can be expressed in the same way as in Eq. (2.2). If a synchronous sampling (synchronous mode) is adopted at the Read-Out compared to the Write-in, the sampling rate between the input signal and the output signal is the same and thus the virtual distance $\delta\tau$ within in a time delay interval τ_D is equal in both signals.

$$y_m(n) = \sum_{k=1}^{K} w_{mk}^R \, x_k(n), \tag{2.11}$$

On the contrary if the asynchronous sampling (asynchronous mode) at the Read-Out is adopted. Then one can play with the degrees of freedom brought by a continuous time dynamics, here defining different node positions σ_k^R for the Read-Out compared with Write-In:

$$y_m(n) = \sum_{k=1}^{K} w_{mk}^R \, x\left(n\tau_D + \sigma_k^R\right), \text{ with } \sigma_k^R \in [0; \tau_D[\,. \tag{2.12}$$

This issue makes use of degrees of freedom brought by the continuous flow of information at the output of the Reservoir dynamics, thus changing the way this information flow is actually sampled with respect to the Write-In. The Read-Out sampling period $\delta\tau^R$ was chosen slightly different from the Write-In sampling period $\delta\tau$, $\delta\tau^R = (1+\varepsilon)\delta\tau$, which resulted in a radically improved RC processing efficiency by more than one order of magnitude. This improvement was confirmed numerically and experimentally. This effect of asynchronism will be described in more detail in Chapter 3.

The output signal $y_m(n)$ in either synchronous or asynchronous mode Eqs. (2.12) and (2.11) is put in a 2D format with spatio-temporal dimension, resulting a matrix $\mathbf{M}_x$. This procedure is shown in Fig. 2.5. A given response matrix $\mathbf{M}_x$ is representative of the dynamical processing performed by the dynamical reservoir and it consists in the sampling of the transient trajectory triggered by one input signal. This trajectory is taking place in the infinite dimensional phase space of the delay dynamics.

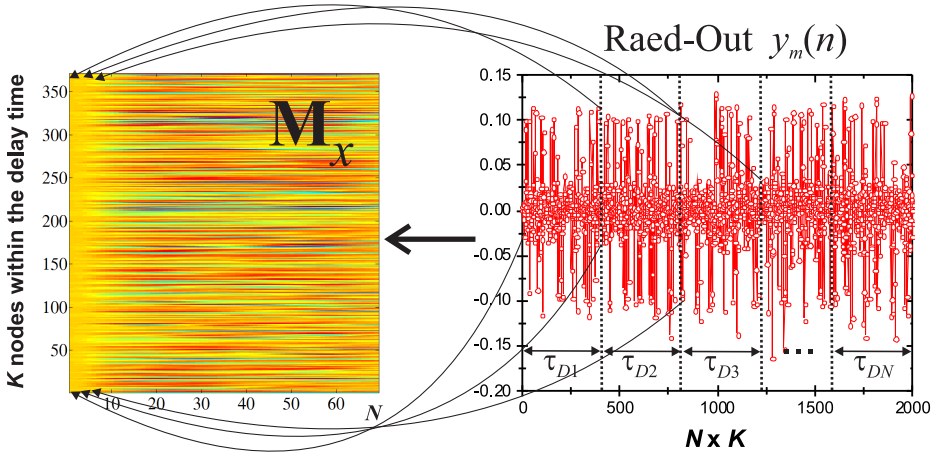


Figure 2.5: **Procedure for the formatting of the output signal $\mathbf{y}(n)$ in order to put the transient response in a spatio-temporal form (matrix $\mathbf{M}_x$)**. The vertical values K are virtual nodes distributed along the time delay τ_D and horizontally it is a sequence of columns N corresponding to the temporal length of the information being processed.

2.2.3.2 Learning procedure

The learning procedure is based on the same learning method proposed by H. Jaeger and W. Maass [7, 8]. Then, according to the concept of RC, one needs to obtain or to find the optimal coefficients $[\omega^R_{mk}]$ of a hyperplane in this phase space. The hyperplane is expected to have a characteristic role for the problem presented to the RC processor. We propose a description of the learning procedure without taking into account any particular problem. The implementation for a particular case will be detailed later in section 2.4.2.1.

As we have mentioned in chapter 1, the learning process of RC is different from the traditional one (RNN method), where the conceptual separation of the dynamical reservoir and Read-Out function does not exist, and train both the inner reservoir and the coefficients $[\omega^R_{mk}]$ in technically the same fashion. RC methods only train the last layer (Read-Out) and the other layers are fixed (Write-In and dynamical reservoir). According to it, the learning process assumes that the Read-Out function produces the desired output from the state expansion $\mathbf{x}(n)$. Therefore, the output weight learning of Read-Out matrix $\mathbf{W}^R$ can be phrased as solving a system of linear equations

$$\mathbf{W}^R \times \mathbf{M}_x = \mathbf{M}_y, \tag{2.13}$$

where $\mathbf{M}_x$ are all neuron signals $\mathbf{x}(n)$ produced by presenting the reservoir with the input $\mathbf{u}(n)$ and $\mathbf{M}_y$ is the so-called target matrix with dimension $M \times N$. It is also the set of coefficients to learn. The correct answer is then expected to be easily identified from the content of $\mathbf{M}_y$.

The learning requires practically the use of a subset from the whole number of data points in the task. The complementary part of this subset is later used for the testing of untaught data points. This is basically performed through a simple ridge regression minimizing the error for all possible $\mathbf{W}^R$ matrices in order to obtain the proper target for each data point of the training subset:

$$\mathbf{W}^R_{\text{opt}} = \text{argmin}\|\mathbf{W}^R \cdot \mathbf{M}_x - \mathbf{M}_y\|^2 + \lambda\|\mathbf{W}^R\|^2, \tag{2.14}$$

where λ is the regression parameter to compensate for a necessarily ill-posed problem. The matrices $\mathbf{M}_x$ and $\mathbf{M}_y$ are horizontally concatenated matrices gathering all experimental response matrices of the training subset and all their corresponding targets, respectively.

The optimal Read-Out matrix is calculated off-line using the recorded transient experimental response matrices $\mathbf{M}_x$. This calculation is done by a Matlab

routine executing the ridge regression through a standard Moore-Penrose matrix inversion, according to the following expression:

$$\mathbf{W}^R = \mathbf{M}_y \cdot \mathbf{M}_x^\intercal (\mathbf{M}_x \mathbf{M}_x^\intercal - \lambda \mathbf{I}_K)^{-1}, \tag{2.15}$$

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

2.3 Modeling the setup performing a photonic delay-based Reservoir

In this section, we present the physical implementation of our photonic RC based on a nonlinear DDS. The idea to use such systems intended for signal or information processing has been already studied few years ago, even-though it was for totally different applications such as ultra-high spectral purity microwave generation [109] or physical layer optical encryption using broadband chaotic waveforms [110]. In this work, the design of our experimental setup is built taking as model a version of a optoelectronic circuit. Among the different setups reported in the literature, our group contributed to several versions of the so-called optoelectronic chaos oscillator [111]. The latest version was especially designed in the framework of 10 Gb/s optical chaos communication, and it succeeded in establishing the state-of-the-art speed and transmission quality thanks to highly controllable differential phase shift keying (DPSK) optical communication techniques [112]. Beyond the operational chaos communication demonstration, this so-called "phase chaos setup" is characterized by temporally non-local delayed feedback which provided additional virtual space-time coupling features with novel bifurcation phenomena [84,85]. Consequently, the phase chaos architecture appeared to us of high potential interest in the framework of photonic RC processing as well.

2.3.1 EO phase delay dynamics

The setup performing RC through the transient motion of an electro-optically modulated phase of a laser beam, is depicted in Figure 2.6. It is implemented in a closed-loop configuration based on the Ikeda setup idea. The delayed feedback loop is inspired by the standard optoelectronic delayed feedback architecture [111] which provides the "recurrent" character of the dynamics. This oscillator makes use of two EO phase modulators (PM) (20 Gbps, EOspace) which electrical to optical phase conversion efficiency is denoted V_π (also technically referred to as half-wave voltage). One can see in Figure 2.6 that the first

EO PM allows to imprint the initial electrical domain information $u_\sigma^I(n)$ onto the optical phase of the laser light beam carrier (PureSpectrum -FSL, TeraXion), whereas the second EO PM combines additively to this input information the nonlinear delayed self feedback $x_\sigma(n)$. The resulting total light phase modulation is $\varphi(t) = x_\sigma(n) + \rho\, u_\sigma^I(n)$, where ρ is an amplification factor scaling the information signal compared to the feedback one. The recurrent dynamics of the RC concept is provided by the delayed optoelectronic feedback loop. This loop consists of:

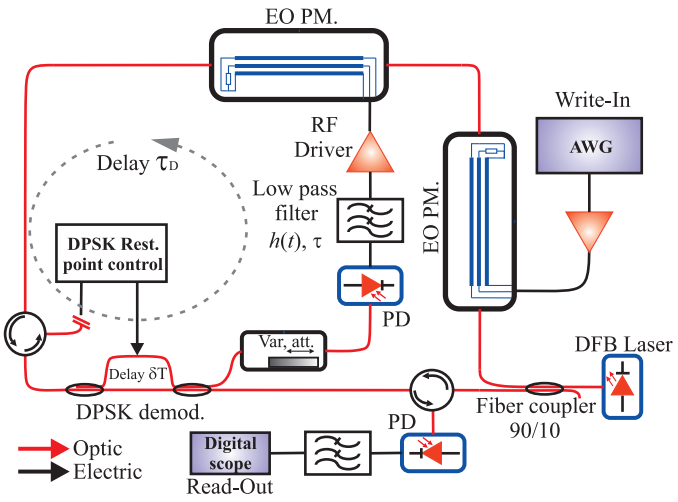


Figure 2.6: **EO phase setup involving two integrated optic PM followed by an imbalanced Mach-Zehnder DPSK demodulator providing a temporally nonlocal nonlinear phase-to-intensity conversion**. The information to be processed by this delay photonic Reservoir is provided by a high speed arbitrary waveform generator. The response signal from the delay dynamics is recorded by an ultra-fast real time digital oscilloscope at the bottom of the setup, after the circulator followed by an amplified photodiode and a filter.

- A significant time delay τ_D issued from the serial combination of optical fiber pigtails and small electrical links connecting the different electronic, optoelectronic, and EO devices of the loop.
- A fiber-based passive imbalanced Mach-Zehnder interferometer (MZI), which actually consists in a conventional DPSK telecom demodulator (2.5 GHz free spectral range, ITF technologies), with a time imbalancing $\delta T \simeq$ 402.68 ps. This demodulator converts nonlinearly the input phase modulation $\varphi(t)$ into an output intensity modulation, according to the stan-

dard two-wave intensity interference function (with an offset shift ensuring $f_{\mathrm{NL}}[0] = 0$, and a gain factor β):

$$f_{\mathrm{NL}}[\varphi] = \beta \cdot \{\cos^2[\varphi(t) - \varphi(t - \delta T) + \Phi_0] - \cos^2 \Phi_0\}, \tag{2.16}$$

in which the imbalancing $\delta T \gg \tau$ reveals a strong time non-locality of the nonlinear transformation contrarily to most other delay dynamics [79,109, 111,113]. Such a non locality becomes dynamically significant, as soon as the phase modulation occurs at time scales comparable or faster than δT. This is obviously the case in our work, as e.g. the input information occurs at more than 10 GHz sampling frequency, the PMs being 10 Gb/s telecom devices. It is worth noticing the role of the offset phase $\Phi_0 = \omega_0 \delta T$, which is to fix the shape of the nonlinear transformation profile. This actual shape can be set to linear or quadratic for $\Phi_0 = \pi/4$ or 0, modulo $\pi/2$, respectively. For $\Phi_0 = \pi/4$ and a strong enough phase modulation span (parameter ρ), the actual shape can even be of cubic type. The parameter Φ_0 is practically adjusted whether through the central wavelength of the laser $\lambda_0 = 2\pi c/\omega_0$, or through a fine tuning of the time imbalancing. Fiber-based DPSK are indeed usually equipped with a heating wire rolled around one interferometer arm, thus allowing to finely and actively control the absolute differential phase shift, or DPSK rest point; such an active control is implemented in our setup, making use of the backward optical path in the DPSK demodulator, through the backward injection of the same but unmodulated laser beam through a 10/90 fiber coupler as shown in Figure 2.6; two optical circulators are used to separate or combine the forward modulated light beam -counter clockwise propagation in the oscillator loop in- and the backward unmodulated one (see [114] for details).

The parameter β is a normalized weighting factor for the nonlinear delayed feedback, it scales linearly with the different EO and optoelectronic conversion efficiency, as well as with the electronic gain, and the forward path optical power κP_0. The latter power is practically tuned through a variable optical attenuator (M420, Eigenlight with attenuation factor κ) in order to adjust β in the experiment, to the optimal value determined from numerical simulations. It is important to note that the power can be controlled by tun ing the laser injection current, but varying this current the wavelength is also changed. This wavelength shift affects the offset phase of the interplay of waves and hence changes the nonlinear transform.

- Telecom amplified photodiodes (PDs) with sensitivity S are used to con-

vert the intensity fluctuations at the DPSK demodulator outputs into electronic signals. In the delayed optoelectronic feedback the signal applied to the second EO PM is coming from a broadband PD (30 KHz-12.5 GHz, model DR-125G-A, Miteq) and in the output of oscillator a PD (10 GHz bandwidth, model DSC-R402 from Discovery Semiconductor, Inc) performs the extraction of the signal $x_\sigma(n)$. The latter signal represents the reservoir response to the transient triggered by the input information, and it is recorded by a digital scope for further off-line processing corresponding to learning and testing of the Read-Out layer.

- A low-pass filter is used to "slow-down" the internal delay dynamics time scale, resulting in a characteristic response time τ or differently speaking it results in the impulse response $h(t)$ involved in equations (2.4)-(2.9). The motivation for this filter is to slow down the actually available 10 GHz bandwidth in the feedback dynamics, which is governed on one side by empirically determined [10] optimal distance between the virtual nodes ($\delta\tau$) of a delay reservoir as described in Eq. (2.10), and on the other side by the maximum sampling frequency of our AWG (Tektronix 7122C) up to 24 GS/s. Input sampling rate is deduced from one fifth of the low-pass filter characteristic time scale $(\delta\tau)^{-1}$, indeed fulfilled these two technical requirements.

- A broadband Telecom RF driver with gain G is used for amplification (SHF100CP, SHF), before applying the nonlinear delayed filtered feedback to the second EO PM. A similar driver is also used between the AWG and the first EO PM, to ensure a strong enough modulation span [weight ρ in equation (2.9)]

The schematic of the feedback loop is shown in Figure 2.7 and a physical demonstration of the function of certain components (particularly the DPSK) will be held in the next section.

2.3.2 Modeling EO phase setup

The feedback loop in Figure 2.7 is fed by a semiconductor CW laser injecting a light wave in the form of $E_0 e^{j\omega_0 t}$. The optical phase of the injected light is then modulated by a value $\varphi(t) = \pi v(t)/V_\pi$ according to the voltage $v(t)$ applied to the electrical RF input of the PM, while the intensity of the output light remains constant. Then the signal is delayed by a serial combination of the few meters of fiber pigtails and the electrical connections, resulting in a total delay in the closed-loop oscillation we denoted as τ_D. Then, the output light leaving the PM

is tuned by a variable optical attenuator, the resulting light intensity is linearly influencing the feedback loop gain β. The phase of the attenuated light is then nonlinearly converted into intensity one by means of the DPSK demodulator, which can be detected by an amplified PD. The electrical signal at the PD output is filtered according to the overall bandwidth of the electronic feedback and is determined by a low and high cut-off time scales that we denoted θ and τ (from filter). The electrical signal leaving the filter is then amplified by a linear amplifier with gain G which drives the input of the PM, thus closing the loop.

2.3.2.1 Derivation of the nonlinear transformation by the DPSK, i.e. the activation function

The optical output at the PM ($E_0 e^{j[\omega_0 t+\varphi(t)]}$) is split in two optical waves delayed by τ_D at one arm and by $\tau_D + \delta T$ at another. Interference of re-coupled waves results in the following output electrical field:

$$\begin{aligned} E(t) &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)+j\varphi(t-\tau_D)} + \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D-\delta T)+j\varphi(t-\tau_D-\delta T)} \\ &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)} \cdot \left[e^{j\varphi(t-\tau_D)} + e^{-j\omega_0\delta T+j\varphi(t-\tau_D-\delta T)}\right] \\ &= E_0 e^{j\Psi} \cos\left[\frac{\omega_0\delta T + \varphi(t-\tau_D) - \varphi(t-\tau_D-\delta T)}{2}\right], \end{aligned} \tag{2.17}$$

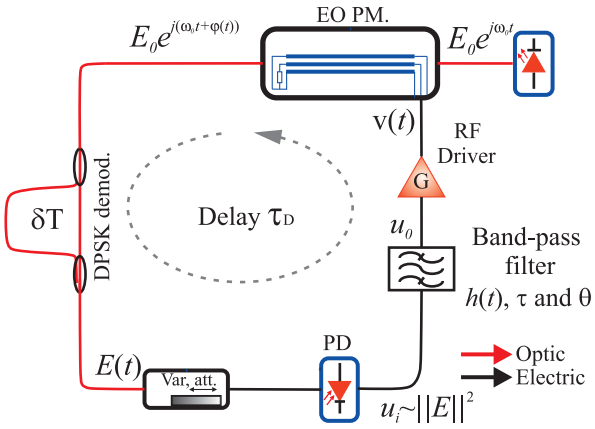


Figure 2.7: **EO phase delay model.**

$$\Psi = \omega_0(t-\tau_D-\delta T) + \frac{[\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)]}{2}. \tag{2.18}$$

Therefore, the light intensity at the output of PD can be described as follows:

$$\begin{aligned} I(t) &= \|E(t)\|^2 = P_0 \cos^2\left[\frac{\omega_0 \delta T + \varphi(t-\tau_D-\delta T)+\varphi(t-\tau_D)}{2}\right] \quad (2.19)\\ P_0 &= \|E_0\|^2 \end{aligned}$$

The bandwidth of the system is limited (usually by the response time of the electronic devices), and it is modeled by two cascaded linear first-order high-pass $(2\pi\theta)^{-1}$ and low-pass $(2\pi\tau)^{-1}$ filter, otherwise by a band-pass filter. If we consider the voltage at the input and at the output of the filter as follows:

$$u_i(t) = S \cdot I(t) \Rightarrow U_i(j\omega) = SI(j\omega) \quad (2.20)$$
$$u_0(t) = v(t)/G \Rightarrow U_0(j\omega) = V(j\omega)/G \quad (2.21)$$

where S is the conversion efficiency of PD, G is the gain of the amplifier and $V(j\omega) = \mathrm{FT}[v(t)]$ is defined as the Fourier transform of the voltage applied to the PM. Using a simple definition of filter transfer function Eq. (1.21) introduced in section 1.3.6.1

$$H(j\omega) = \frac{U_0(j\omega)}{U_i(j\omega)}, \quad (2.22)$$

using $\mathrm{FT}[v(t)]$ we can rewrite the Eq. (2.22) as:

$$\mathrm{FT}^{-1}\left[\frac{V(j\omega)}{H(j\omega)}\right] = SG \cdot I(t) \quad (2.23)$$

Substituting Eq. (2.20) into Eq. (2.23), taking into account additional losses γ and also including delay (δT) due to the DPSK into overall feedback delay time τ_D, then Eq. (2.23) can be rewritten as:

$$\mathrm{FT}^{-1}\left[\frac{V(j\omega)}{H(j\omega)}\right] = \gamma SGP_0 \cos^2\left[\frac{\omega_0 \delta T + \varphi(t-\tau_D-\delta T)+\varphi(t-\tau_D)}{2}\right] \quad (2.24)$$

We consider $x(t) = \varphi(t)/2 = v(t) \cdot \pi/(2C_\pi)$ as the dimensionless variable describing the system, $\beta = \pi\gamma SGP_0/2$ and $\Phi_0 = \omega_0 \delta T/2$ as parameters into Eq. (2.24), it can be written in the following way:

$$\mathrm{FT}^{-1}\left[\frac{X(j\omega)}{H(j\omega)}\right] = \beta \cos^2\left[x(t-\tau_D) - x(t-\tau_D-\delta T) + \Phi_0\right] \quad (2.25)$$

where $X(j\omega) = \mathrm{FT}[x(t)]$. The right hand side of the equation reflects its nonlinear and delayed nature, which usually is denoted as $f_{NL}[x(t-\tau_D), x(t-\tau_D-\delta T)]$. Whereas in our case, the dynamical model has a dual-delay, thus the function not only depends on a single delay, but also on the delay value introduced by the interferometer $x(t-\tau_D-\delta T)$. We have also introduced a normalized parameter β defining the feedback gain in the system, that is practically adjusted through a variable optical attenuator. Finally, we introduced an offset interference phase condition Φ_0, that can be tuned by changing the wavelength of the laser or the static shift in the imbalanced fiber interferometer (through an electrical current flowing in a heating wire). The entire description of the mathematical model is further explained in Appendix A.

According to Eq. (2.25), the nonlinear transformation is a function of a single parameter (Φ_0), however the parameter β controls the amplitude of such a nonlinear transformation, while Φ_0 controls the operating point of the nonlinear function horizontally.

$$\psi = x(t-\tau_D) - x(t-\tau_D-\delta T) + \Phi_0 \tag{2.26}$$

$$f_{NL} = \beta \cos^2(\psi) \tag{2.27}$$

Figure 2.8 illustrates the static nonlinear transformation function of Eq. (2.27) with $\beta = 1$, and it is possible to observe the offset point (Φ_0) corresponding to the case $x(t) \equiv x_0$. Moreover, the argument of the $\cos^2$ function (ψ) is deviated from the offset point depending on the value of $x(t)$ at two moments delayed in time [114].

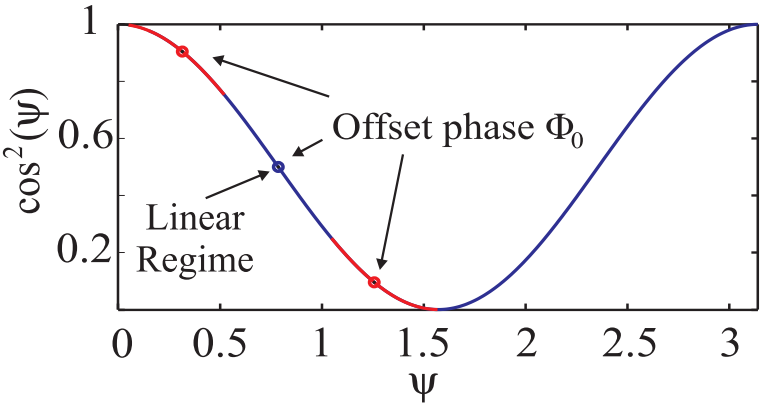


Figure 2.8: **The $\cos^2$ static nonlinear transformation function, with three different values of offset.**

Returning to the Eq. (2.25), the left hand side is a standard definition of filter transfer function $H(j\omega)$. This transfer function can be substituted with a specific approximation of an actual filter. The simplest description in the frequency domain for such filtering profile corresponds to a second order band-pass filter. Nevertheless, it has been notified in [113] that using a simplest approximation does not allow to detect certain relevant characteristics, as when a high-order approximation is used. To explain the profile of the band-pass filter, we consider the dynamics of this filter as a succession in cascade of a high-pass filter and a low-pass one [79, 85], it is illustrated in Figure 1.13 in section 1.3.6.1. Thus, assuming the simplest band-pass filter approximation (firts-order low-pass and high-pass) described by Eqs. (1.27) and (1.31), the dynamics of the system is defined by a nonlinear delayed integro-differential equation

$$x(t) + \tau \frac{dx}{dt}(t) + \frac{1}{\theta} \int_{t_0}^{t} x(s)ds = \beta \cos^2 \left[x(t - \tau_D) - x(t - \tau_D - \delta T) + \Phi_0 \right] \quad (2.28)$$

where τ and θ correspond to high and low cut-off times of the electronics.

2.4 Benchmark test: spoken digit recognition (SDR)

There are many standard tests to evaluate the performance of RC systems. They are used in neuroscience or learning machine communities. In this work, two classification tests are used to quantify and to evaluate the computing power of our RC system and these are based on speech recognition . The objective of these tests is to classify a series of pronounced digits. The first test is a subset of a standard database **NIST TI-**46 **corpus** [115], made available by Texas Instrument. This database is composed of 500 isolated digits from 0 to 9, each being pronounced 5 times by 10 different women and sampled at 12.5 KHz.

The second one is **distributed speech reconigtion, like AURORA-2**. It is a subset based on a version of original TIDigits LDC1993s10. The original test contains a corpus from 326 speakers (111 men, 114 women, 50 boys and 51 girls) each speaking approximately 77 digit sequences and has been divided into test and training subsets. The digit sequences were made up of the digits: zero, oh, one, two, three, four, five, six, seven, eight and nine. Furthermore, these digit sequences spoken by each speaker can be broken down as follows: 22 isolated digits (2 productions of each of 11 digits), 11 sequences of 2-digits, 3-digits, 4-digits, 5-digits and 7-digits [116]. On the other hand, the AURORA-2 test is composed of two training modes **clean** and **multi-condition**. The advantage of training on clean is only the modeling of speech without distortion by any type

of noise. However, these models contain no information about possible distortion. This aspect can be considered as an advantage of multi-condition training where distorted speech signals are taken as training data. For the first mode, 8440 utterances are selected from the training part of the TIDigits containing the recordings of 55 male and 55 female adults. These signals are filtered with the G.712 characteristic[3]. The same 8440 utterances are taken for the second mode too. They are equally split into 20 subsets with 422 utterances, each subset contains a few utterances of all training speakers. The 20 subsets represent 4 different noise scenarios at 5 different SNRs. The noises are subway, babble, car and exhibition. The SNRs are 20 dB, 15dB, 10dB, 5dB, 0, -5dB and the clean condition. Three tests of data exist, the first one **test set A** contains 4004 utterances from 52 male and 52 female speakers. This set is split into 4 subsets with 1001 utterances and all speakers are present in each subset. One noise signal is added at SNRs 20 dB, 15dB, 10dB, 5dB, 0dB, -5dB and the clean case without noise is taken as seventh condition. Moreover, speech and noise are filtered with the G.712 characteristic. Therefore test set A consists of 4 subsets with 7 conditions of the 1001 utterances each one ($4 \times 7 \times 1001 = 28028$ utterances). The second test called **set B** is created in exactly the same way, but using 4 different noises (restaurant, street, airport and train-station). Third test called **set C** contains 2 of the 4 subsets, but the speech and noise are filtered with MIRS characteristic, thus $2 \times 7 \times 1001 = 14014$ utterances [117,118].

In the speech recognition tests, it is common to use a pre-treatment of the information (mathematical models of human auditory systems). These models can significantly improve the sound signal processing and system performances. In addition, ear models can contribute to the understanding of experimental hearing data. This kind of representation subdues the dynamic characteristics and the nonlinearity of the signal. They are based on mechanics and neural phenomenology of the inner ear (cochlea) [119]. There are many different auditory models and they are being continually upgraded. The principal models are:

- Flanagan's Model
- Lyon's Cochlear Model (LCM)
- Meddis' Inner Hair Cell

[3]It deals with the transmission performance characteristics of pulse-code modulation channels of digital transmission equipment. Requirements to be met between 4-wire and 2-wire analogue ports are given as well as requirements for analogue-to-digital and digital-to-analogue connections.

- Auditory Image Model
- Mel-Frequency Cepstral Coefficients (MFC)
- Patterson's Ear Model

LCM and MFCC are the two models used for the speech recognition in this work, hence they will be briefly detailed below.

2.4.1 Ear model

Before addressing the LCM and the MFCC, it is useful to give a simple description of the hearing process. The human ear can be divided in three different parts. The first is the *outer ear*, which consists of the pinna, or auricle, and the ear canal (external auditory meatus). After we have the *middle ear*, which includes the tympanum and mastoid cavities. The tympanic cavity is composed of three auditory ossicles (malleus, incus and stapes) and the last part is the *inner ear*, which contains the sensory organs for hearing and balance (the cochlea and the auditory nerve). Each of the parts of the ear has embryological origins, roles and different physiology, as depicted in Fig. 2.9.

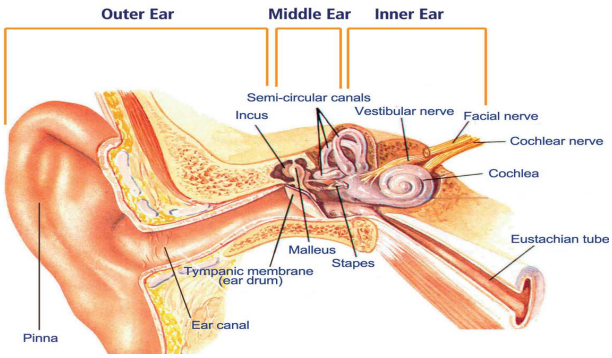


Figure 2.9: **Illustration of the human ear**. Anatomically it is divided in: the outer ear that acts as screening and channeling sound stimuli, the middle ear which leads and amplifies the vibrations of the tympanic membrane and the inner ear where vibration is collected and the mechanical signal is then transformed into an electrical signal by stimulating hair cells. This stimulation performed via the auditory nerve and through the central nervous system reaches the cerebral cortex, where auditory perception becomes conscious and the brain then interprets these electrical signals as sound [120].

2.4.1.1 LCM

The model of analog electronic cochlea used in our thesis's work was developed by Richard F. Lyon [121] and it is based on the knowledge of biological functioning of the cochlea. This model describes the propagation of the 1D sound waveform (pressure variations over time) in the inner ear and the conversion of the acoustical energy into neural representation. This process is performed by a cascade of selective filters and the action of the active outer hair cells is modeled by a set of Automatic Gain Controls (AGC) which simulated the dynamic compression of the intensity range on the basiliar membrane, as depicted in Figure 2.10. Also this figure shows how the sampled acoustic waveform is converted into a 2D frequency-time representation so-called "*cochleagram*". This representation corresponds to the time evolution all along the pronunciation of the Fourier spectral power distribution for $Q = 86$ frequency channels.

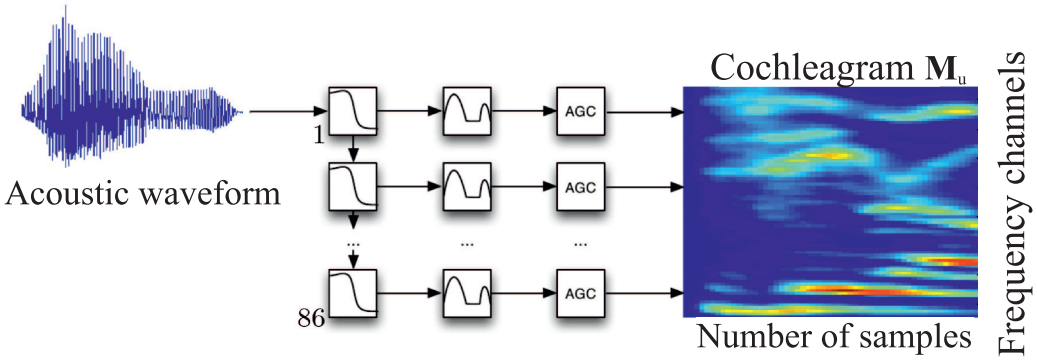


Figure 2.10: **Illustration of the LCM**. The acoustic waveform is decomposed in 86 frequency channels before being represented as cochleagram matrix $\mathbf{M}_u$.

To perform this model, we use a toolbox of Matlab called auditory model developed by Malcolm Slaney. This implementation is freely accessible in the following link: "*https://engineering.purdue.edu/ malcolm/interval/1998-010/*".

Considering the isolated spoken digit recognition: TI-46 test described in section 2.4, the cochleagram is then a $(Q \times N)$ matrix whose dimensions are $Q = 86$ frequency channels (rows) and the number of time-samples (columns), corresponding to the input information signal $\mathbf{u}(n) = [u_1(n)...u_q(n)...u_Q(n)]^\intercal$. The successive columns of the matrix $\mathbf{M}_u$ is thus indexed by the discrete time n. Furthermore, in this test the number of columns N can vary from 32 to 132 depending on the duration of pronunciation for each digit. The resulting matrix

$\mathbf{M}_u$ is the input information, but before being injected in the dynamical reservoir, a pre-processing is applied. This is detailed in section 2.4.2.

2.4.1.2 MFCC implementation

MFCCs are the coefficients for the representation of speech based on human auditory perception. These arise from the need in the area of automatic audio recognition of extracting characteristics of components of an audio signal which are suitable for identification of relevant content, as well as ignore all those who have little valuable information such as background noise, volume, pitch, etc., and they added nothing to the recognition process.

MFCCs were introduced by Davis and Mermelstein in the 80s [122] and have been state-of-the-art ever since (in automatic speech recognition). They are a feature widely used in automatic speech and speaker recognition or environmental sound events, and because of their simplicity and efficiency they have become standard in such domains [123,124]. These are commonly calculated as follows:

1. The acoustic signal is divided into short frames of 20-40 ms with frame step usually of 10 ms, which allows some overlap between themselves.

2. For each frame a periodogram estimate of the power spectrum is calculated in order to adapt the frequency resolution to the properties of the human ear. It is computed applying a discrete Fourier transform (DFT) to every single frame for the discrete-time signal $s(n)$ with length N, described by:

$$S(k) = \sum_{n=1}^{N-1} w(n)s(n)e^{(-j2\pi kn/N)} \qquad 1 \leq k \leq K \tag{2.29}$$

   where K is the length of the DFT and $w(n)$ is a time-window (e.g. Hamming window). Consequently, the periodogram-based power spectral estimate for speech frame $S(k)$ is given by:

$$P(k) = \frac{1}{N} \mid S(k) \mid^2 \tag{2.30}$$

3. The next step is to apply the Mel[4] filterbank to the periodogram power spectral, segmenting into a number of critical bands. Then they are summed

[4]The Mel scale proposed by Stevens, Volkmann and Newmann in 1937, is a scale of pitches judged by listeners to be equal in distance one from another. It relates perceived frequency, or pitch to its actual measured frequency.

to get an idea of how much energy exists in various frequency regions, it is performed by a set of 20-40 overlapping triangular filters.

4. Once the filterbank energies are obtained, their logarithm is considered. It allows to use cepstral mean subtraction.

5. The final step is to compute the discrete cosine transform (DCT) of the log filterbank energies. The resulting features is a raw MFCC vector, in which the highest cepstral coefficients are omitted because they represent fast changes in the filterbank energies and it turns that these fast changes actually degrade automatic speech recognition performance (ASR). For ASR only the lower 12-13 of all coefficients are kept. Mathematically it can be described by:

$$\mathbf{c_{eq}} = [c_{eq}(1,t), c_{eq}(2,t)...c_{eq}(12,t)]^{\mathsf{T}} \tag{2.31}$$

where t represents the frame index.

Finally the MFCC vector only describes the power spectral envelope of a single frame, but this can be modified by adding the deltas Δ and delta-deltas $\Delta\Delta$ (time derivatives), also known as velocity and acceleration coefficients. These coefficients describe the dynamic information, calculating the trajectories of the MFCC over time. Considering the AURORA-2 test, the MFCC feature vector consists of 13 static MFCCs, 13 Δ coefficients and 13 $\Delta\Delta$ coefficients, and these can be represented as matrix $\mathbf{M}_u$ in the same way as was done with LCM. Figure 2.11 shows the MFCCs.

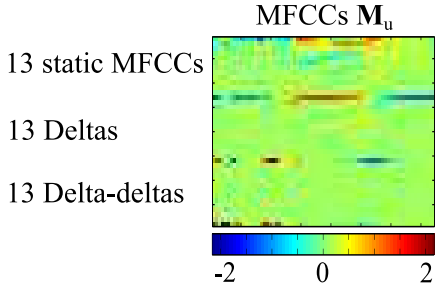
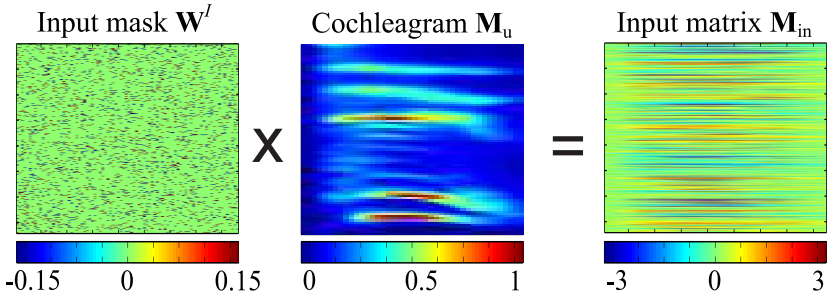


Figure 2.11: **Representation of the MFCCs**. The acoustic waveform is decomposed in 39 coefficients (static, velocity and acceleration).

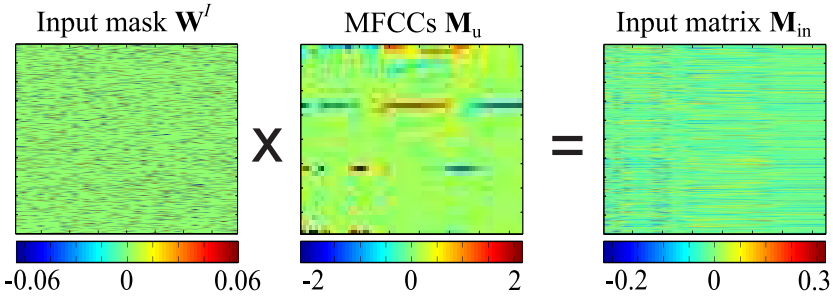
To perform this model, we use the distributed speech recognition front-end algorithm based on the mel-cepstral feature extraction technique. The specification covers the computation of feature vectors from speech waveform sampled at different rates (8 kHz, 11 kHz and 16 kHz) [125].

2.4.2 RC processing

For RC processing the cochleagram or MFCCS matrix $\mathbf{M}_{\mathrm{u}}$ must follow the pre-processing described in section 2.2.1 (masking procedure). It is intended to ensure a kind of relevant expansion over all nodes of the dynamical system, of many possible Fourier or Ceptrums components, at every time of the duration of pronunciation. For this a connectivity matrix $\mathbf{W}^I$ is used to spread each of these input samples $\mathbf{u}(n)$ onto each of virtual nodes (position σ_k) defined within one layer of the delay dynamics. The scalar input waveform signal $u^I_\sigma(n)$ appears as the queued columns of the matrix product $\mathbf{M}_{\mathrm{in}} = \mathbf{W}^I \times \mathbf{M}_{\mathrm{u}}$, as illustrated in Figure 2.12.



(a) Pre-proccesing for TI46 test.



(b) Pre-proccesing for AURORA-2 test.

Figure 2.12: **Illustration of the masking procedure**. The sparse and random $K \times Q$ connectivity matrix $\mathbf{W}^I$ performs a spreading of the input information represented as a $Q \times N$ cochleagram or MFCCs matrix $\mathbf{M}_{\mathrm{u}}$. The resulting $K \times N$ matrix $\mathbf{M}_{\mathrm{in}}$ defines a scalar temporal waveform $u^I_\sigma(n)$ obtained after queuing horizontally the N columns, each of them being formed by the K amplitudes addressing the virtual nodes in one layer.

According to the matrix multiplication and consistently with Eq. (2.8), the single-layer signal $u^I_{\sigma_k}(n) \,|\, k \in [1;K]$ can be interpreted as a set of K linear combinations of the Q different Fourier frequency components or Cepstrums coefficients (depending on the test) occurring at time n, each node k of the network

receiving a different linear combination whose coefficients are defined by the non zero elements in the rows of the connectivity matrix $\mathbf{W}^I = [w^I_{kq}]$.

In order to inject physically the input matrix $\mathbf{M}_{\text{in}}$ into the dynamical reservoir, it has to be converted into a scalar signal $u(t)$ depending of the time. To do this, we have used the reverse process to this one performed in the post-processing of information (described by Figure 2.5). Here, the N columns of the matrix $\mathbf{M}_{\text{in}}$ are unfolded and then they are put in a horizontal sequence, thus forming the temporal signal $u(t)$ that is injected into the reservoir, as depicted in Figure 2.13.

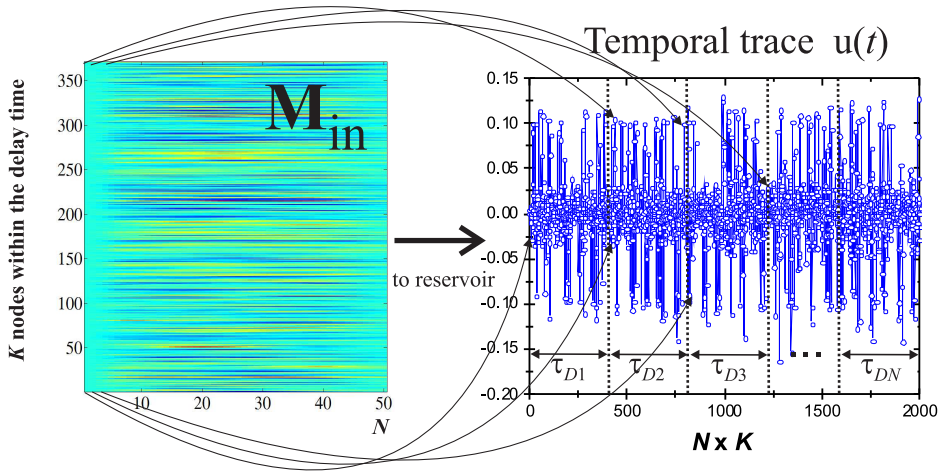


Figure 2.13: **Procedure to obtain a scalar signal $u(t)$ from ($K \times N$) input matrix $\mathbf{M}_{\text{in}}$.** The vertical values K are virtual nodes distributed along the time delay τ_D and horizontally it is a sequence of columns N corresponding to the length of the spoken digit.

2.4.2.1 Learning and testing

The signal $\mathbf{u}(t)$ after being dynamically treated by the reservoir follows a post-process in order to measure the performance of our RC system. The post-process consists of a phase of learning and testing, and the results obtained are a measure of word error rate (WER) which can be interpreted as a reference of the computation power of our RC.

This is developed as follows. First the encoded digital input signal $u^I_\sigma(n) \mid n \in [1; N]$ corresponding to the spoken digit (matrix $\mathbf{M}_{\text{in}}$ unfolded), is stored in

an arbitrary waveform generator (AWG) allowing to "'play" each scalar input signal as a voltage driving the signal input of our photonic RC. The playing rate is set at 17.6 GS/s, so that each sample duration matches the virtual node spacing calculated from the photonic RC response time τ, $\delta\tau = \tau/5$. The response of this drive, at the output of the DPSK demodulator, is detected by a broadband photodiode (DSC R402), which output voltage is recorded by a real time ultrafast digital scope (Lecroy WaveMaster 845Zi-A, 30 GHz analogue bandwidth, 80 GS/s) with an over-sampling e.g. averaging, or selection of interpolated samples according to the already mentioned asynchronous Read-Out sampling in Eq. (2.12). Each sampled response signal $x_\sigma^R(n) \mid n \in [1; N]$ is then conditioned in the form of a matrix $(K \times N)$ $\mathbf{M}_x$ (see Figure 2.5), thus in a similar form compared to the previously introduced input signal matrix $\mathbf{M}_{\text{in}}$.

TI-46 test

Considering the 500 spoken digits of TI-46 test, the learning phase introduced in section 2.2.3.2 is applied to a subset of digits in order to calculate the Read-Out matrix $\mathbf{W}^R$ and then the testing is applied to the subset of un-taugh digits. For the particular case of a classification problem as for the spoken digit recognition task, we expect to learn $M = 10$ such set of coefficients, i.e. one for each of the 10 possible modalities (the 10 digits from zero to nine). The $M \times K$ Read-Out matrix can be constructed solving Eq. (2.13), in which the transient response matrix $\mathbf{M}_x$ is a spatio-temporal representation of the amplitude of virtual nodes of reservoir and the target matrix $\mathbf{M}_y$ corresponds to the desired response as represented in Figure 2.14.

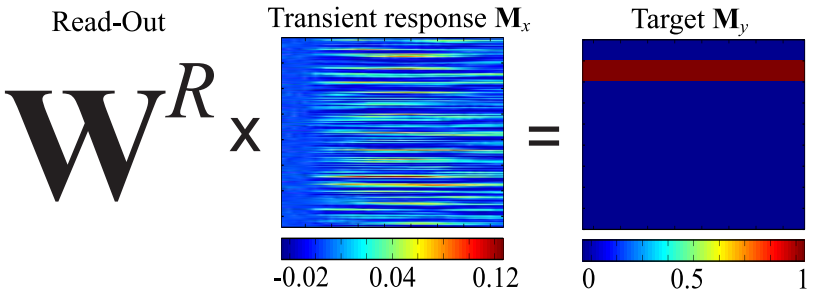


Figure 2.14: **Illustration of the expected optimized Read-Out processing**, by the product between the $(M \times K)$ matrix $\mathbf{W}^R$ and the transient response $(K \times N)$ matrix $\mathbf{M}_x$, thus resulting in an easy-to-interpret target $(M \times N)$ matrix $\mathbf{M}_y$. The latter matrix is aimed at designating the right answer for the digit to be identified (here, the digit "1").

Hence, the aim in the leaning phase is to learn the coefficients $[w^R_{mk}]$ of a hyperplane in this phase space. This hyperplane is expected to have a characteristic role for the problem presented to the RC processor. Once the Read-Out matrix is calculated, an optimization of this one is performed through a linear regression described by Eq. (2.14) in order to obtain the proper target for each digit of the training subset, resulting in a matrix $\mathbf{W}^R_{opt}$. The regression parameter λ empirically gives good results for values around of 0.001.

An improvement of this test consists of using the principle of cross-validation. We practically chose randomly 475 digits for the training subset, the 25 remaining being left for testing. This partitioning is essentially guided by the implementations of the same test, consequently allowing for a more confident performance evaluation for the RC processing efficiency. Cross-validation was performed through the repetition of 20 learning procedure, according to the choice of 20 non-intersecting subsets of 25 testing digits, and their corresponding training subsets, as illustrated in Figure 2.15.

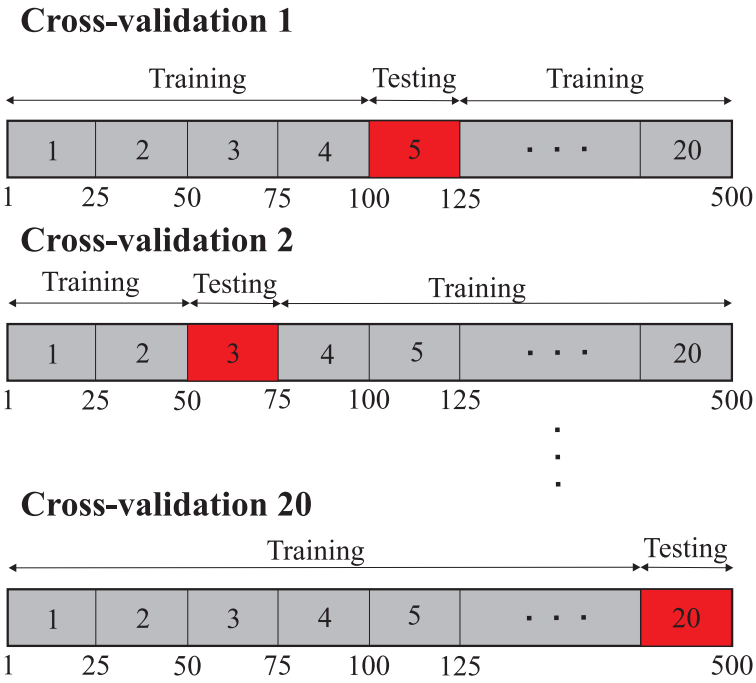


Figure 2.15: **Schematic principle of cross-validation**. The cross validations are used to increase the precision with which we calculate the power of our RC systems [32].

Cross-validation allowed to obtain an average computational performance

measure in which each of the 500 digits had been used once for testing. The testing phase consists of evaluating the result of the RC processing on each of the testing digits by calculating $\tilde{\mathbf{M}}_y = \mathbf{W}_{\text{opt}}^R \mathbf{M}_x$, which is expected to resemble the target matrix $\mathbf{M}_y$. One naturally does not obtain a perfect right line with ones, and wrong lines with zeros, but integrating (summing) the elements of each line of $\tilde{\mathbf{M}}_y$ over n leads to a score for each modality, as shown in Figure 2.16.

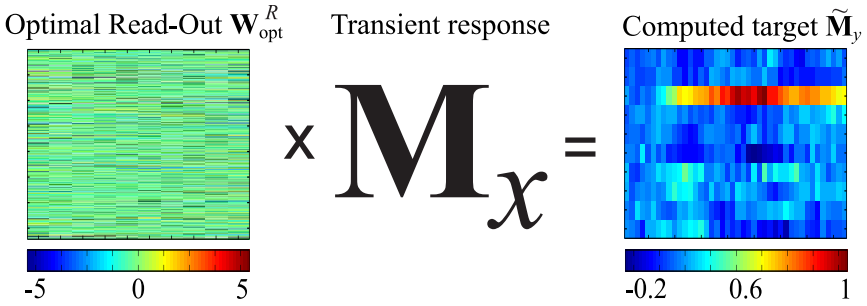


Figure 2.16: **Example of an imperfect "reservoir-computed" target answer while testing the optimal Read-Out $\mathbf{W}_{\text{opt}}^R$ on an untrained digit of response $\mathbf{M}_x$.** The digit "2" however clearly appears as the most obvious answer for this untrained tested digit.

The RC processing is then simply determined by the modality showing the best score; this procedure is depicted in Fig. 2.17, which can be compared with the correct answer in order to count the number of errors.

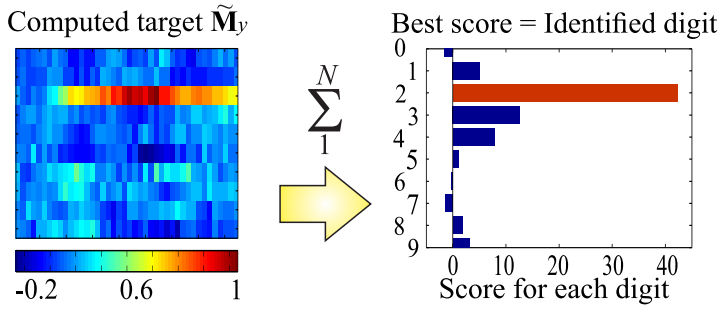


Figure 2.17: **Illustration of the decision procedure for the computed answer**. The temporal amplitudes of the actual target are summed over time for each line (or modality), i.e. for each of the 10 possible digits. The right modality is then declared as the one with the highest sum.

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

AURORA-2 test

Now considering the AURORA-2 test for the case where isolated digits are to be classified, the learning and testing are performed in the same way as TI-46 test. The MFCCs corresponding to the spoken digits are processed forming a matrix $\mathbf{M}_{\text{in}}$ (see Figure 2.12(b)) that is injected into the dynamical reservoir. Right after, in the output of reservoir, this signal is recorded and is then conditioned forming the transient response matrix $\mathbf{M}_x$. As mentioned above the learning phase is applied to a subset of digits in order to calculate the elements of Read-Out matrix $\mathbf{W}^R$ and the testing is applied to the left subset of un-taught digits. Here we expect to learn $M = 11$ of different modalities, corresponding to the 10 possibles digits (zero to nine) and to the word (oh). Then, The $M \times K$ Read-Out matrix can be constructed solving Eq. (2.13) as was performed for the TI-46 test (see Figure 2.18). Once the Read-Out matrix is computed, it is optimized by a linear regression Eq. (2.14).

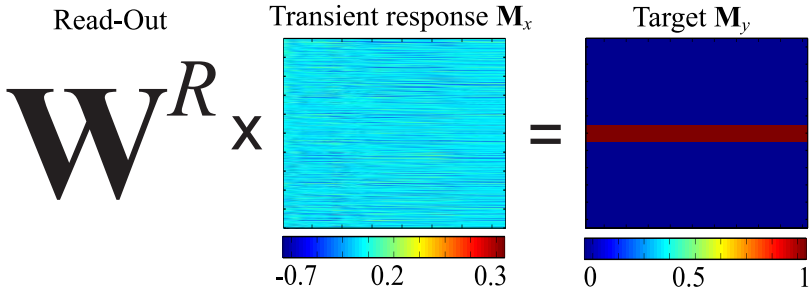


Figure 2.18: **Illustration of the expected optimized Read-Out processing for the AURORA-2 test**. This illustration is to be compared with Figure 2.14 (TI-46 test). The latter matrix is aimed at designating the right answer for the digit to be identified (here, the digit "5").

As in TI-46 test the cross-validation is used in order to have more confident evaluation. For training we use a set belonging to the folder called *clean1* (**multi-condition** mode), with a total of 1001 sequences of digits, among which 289 are isolated digits. We use 265 digits for the training subset chosen randomly and 24 digits for the testing subset (see Figure 2.15). The testing is then performed following the same protocol used in the other test of classification (TI-46). It consists in evaluating the result obtained by RC processing on the testing digits

by calculating $\tilde{\mathbf{M}}_y = \mathbf{W}^R_{\text{opt}} \mathbf{M}_x$, as illustrated in Figure 2.19. The RC processing is thus determined by the modality showing the best score, as illustrated in Figure 2.20. Finally, the WER is determined as the ratio of the sum of errors on the total number of spoken digits in the test.

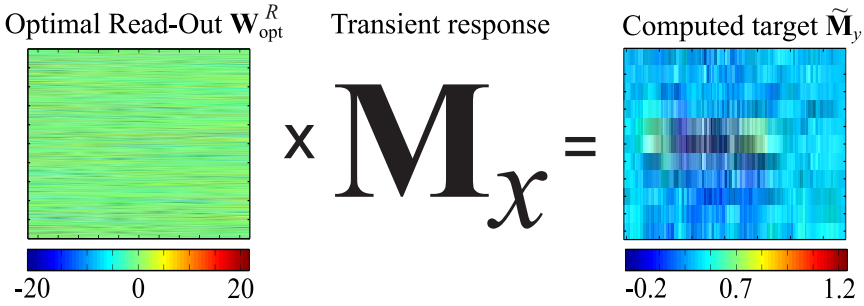


Figure 2.19: **Example of an imperfect "reservoir-computed" target answer while testing the optimal Read-Out $\mathbf{W}^R_{\text{opt}}$ for the AURORA-2 test (to be compared to Fig. 2.16 for the TI-46 test)**. The answer appears here to be the digit "5".

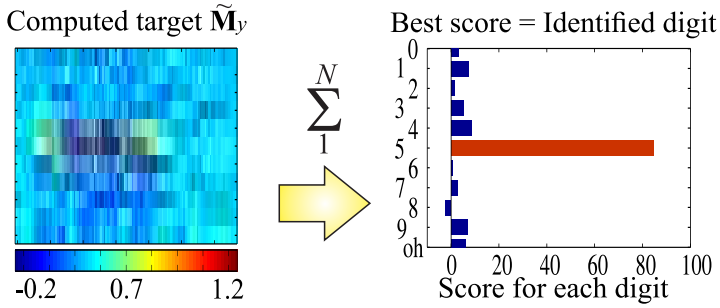


Figure 2.20: **Decision procedure for the computed answer with the AURORA-2 test (to be compared to Fig. 2.17 for the TI-46 test)**. The outcome of the test is the digit "5".

2.5 Overview

In this chapter, we described the main principles of RC system based on RNN. Then, we did a description of how a nonlinear DDS can be used to reproduce the

functionality of a reservoir. Replacing the standard RC approach with spatio-temporal structure of interconnected nodes involves an adaptation of the signals to be processed. This is performed by the steps of pre-processing and post-processing of information: the spatial dimension of NN is virtually recreated in the delay (τ_D), which implies the use of a mask procedure and TDM technique in order to address each virtual node. We presented the schematic setup of DDS based on an EO phase delay dynamic approach, which is built with telecom bandwidth devices. The use of telecom devices provide ultra-fast information processing capability and at the same time allows for a hardware implementation.

We introduced the asynchronous mode at the Read-Out. It enables the possibility to play with an additional degree of freedom. The use of asynchronous sampling allows us to define different node positions σ_k^R for the Read-Out compared to Write-in. This issue makes use of degrees of freedom brought by the continuous flow of information at the output of the reservoir dynamics, changing the way of information flow, which is usually sampled with respect to the Write-in (synchronous mode).

Finally after having described the injection of information (Write-In) and reading of transient signals reservoir (Read-Out), we introduced two standard test of speech recognition, which allow us to quantify the computing power of our photonic RC. These tests of classification will be implemented experimentally in Chapter 3.

Chapter 3

Experimental implementation and results

The experimental implementation and evaluation of the performances of our RC system through the standard tests of classification correspond to the most important part of the work realized during this PhD thesis. Therefore, this chapter is devoted to the experimental study using our photonic RC system.

This chapter characterizes experimentally the photonic RC system presented in Chapter 2 section 2.3. The first part details the experimental setup, which consists in an time-delayed EO oscillator with a feedback loop. Since the delay line serves as reservoir, it has been measured by different methods. On the other hand, the EO phase delay dynamics is characterized temporally and dynamically. Afterwards, we make a quick comparison between numerical and experimental time traces.

The last part of this chapter is dedicated to the quantitative evaluation of the performances of our RC system. For this purpose, we used two standard tests of speech recognition, that have been introduced in Chapter 2 section 2.4. Right after, a comparison between the experimental and numerical results using synchronous and asynchronous sampling is performed. Furthermore in this work, we report on the possibility to introduce for single delay reservoir the analogue of so-called *hidden layers* in standard RNN.

3.1 EO time-delayed oscillator

The EO phase delay dynamics has been used for the realization of a RC system (see Figure 2.6). At the beginning of this work the EO oscillator was mainly linked to the work on optical chaos telecommunications [112], as well as to the study of chaotic regimes [84,85]. Because of the existing setup, we have changed and adapted some components of the original setup (EO oscillator). Nevertheless, these modifications change the behavior of the oscillator, requiring the related characterization (measurement of delay and dynamics characterization).

3.1.1 Experimental implementation

The Figure 3.5 shows the physical realization of the photonic RC system. The operating principle and its structure is detailed in Chapter 2 section 2.3. As it may be difficult to identify each device from the following picture, it is possible to refer to the schematic setup shown in Figure 2.6.

Figure 3.1: **Picture of the photonic RC setup based on an EO phase delay dynamics.**

In this experimental implementation, the time scales (response time τ, node spacing $\delta\tau$, time delay τ_D and imbalancing time δT) as well as the feedback gain β and the offset phase Φ_0 are of paramount importance. These parameters

play a crucial role to calculate the number of virtual nodes K, the reservoir size, the internal connectivity but also the dynamical behavior of the RC system. Besides, they can be adjusted to optimize the performance of RC processor. The following sections will detail the optimization of all these parameters.

3.1.2 Open loop measurements

A way to characterize the RC system is to work in open-loop configuration, as shown in Figure 3.2. This method consists in disconnecting the feedback output of the line delay. It allows us to characterize each element without the feedback influence.

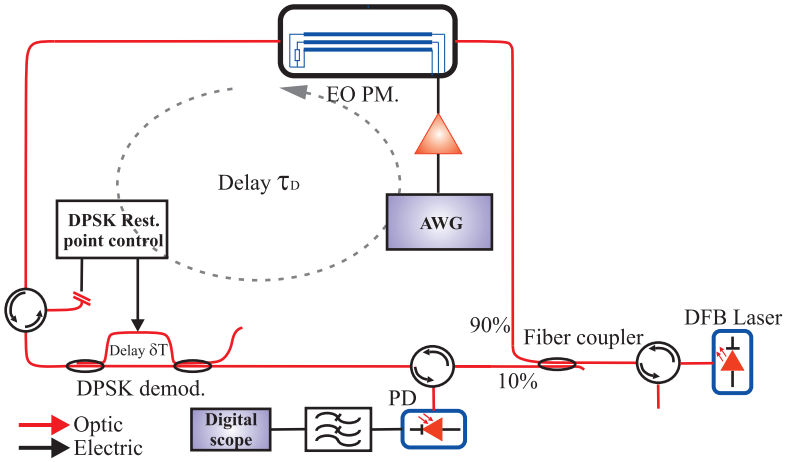
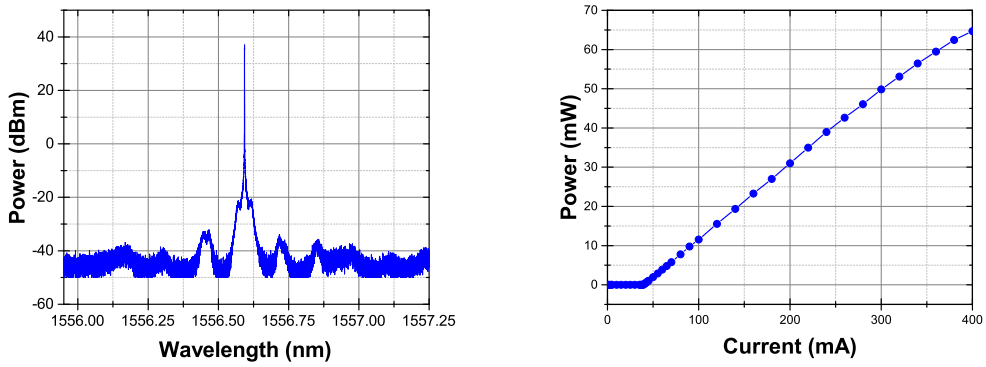


Figure 3.2: **RC system in open-loop configuration.**

3.1.2.1 Characterization of the laser source

The source used is a DFB semiconductor laser module with a very high stability and frequency accuracy (Teraxion, PureSpectrum™-FSL Frequency Stabilized Laser). The module integrates a semiconductor DFB laser, a low noise laser driver, a temperature controller and frequency control electronics in a compact package for use with the original manufacture instrumentation.

We have characterized the optical power as a function of the current at a temperature of 25°C. It was achieved by using an optical fiber detector (Newport, 918D-IS/818-IS Series, S/N 2710, 400-1650 nm) and a dual channel 3.3(b) the power-current graph for current values from 1 to 400 mA, from which the threshold current can be determined close to 40 mA.



(a) Optical spectrum of the laser source. (b) Power-current graph of the laser source.

Figure 3.3: **Experimental characterization of Teraxion Pure Spectrum source**.

For all experiments, the injection current of our laser is set at 140 mA to work with an optical power $\sim$ 19.36 mW. This optical power was calculated through measurements of power loss and is enough in order to observe all different dynamical regimes.

3.1.2.2 Phase Modulator (PM)

Two EO PM (EOspace Lithium Niobate $LiNbO_3$ phase modulator) are used in the RC setup and their roles have been mathematically described in Chapter 2 section 2.3. Here we determine experimentally the value of half-wave voltage $V_{\pi,\mathrm{rf}}$. First of all, the optical spectrum in the C-band is measured with a high-resolution optical spectrum analyzer (OSA, APEX model APE2040A). The amplitude of optical phase modulation peaks are then given by the Jacobi-Anger identity:

$$e^{i\delta\cos\phi}\sum_{-\infty}^{+\infty} = i^n J_n(\delta)e^{in\phi}, \tag{3.1}$$

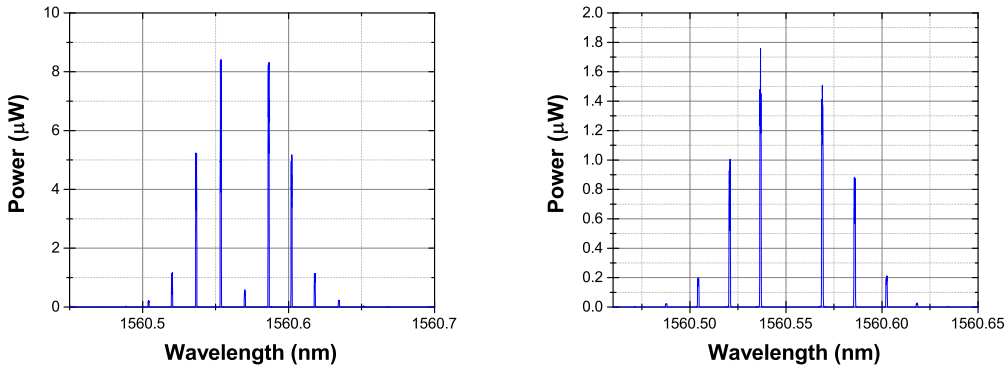
where $\delta = \pi U_0/V_\pi$ is the optical phase amplitude, U_0 being the amplitude of voltage applied to the RF electrodes and V_π the characteristic voltage of the PM that induces an optical phase shift of π. The phase $\phi = \Omega t$ is the instantaneous RF phase of the sinusoidal electrical modulation applied to the electrodes and J_n is the Bessel function of the first kind of order n.

The complex amplitude of the electric field of the optical output at the PM is written as:

$$E(t) = E_0 e^{i[\omega_0 t + \varphi(t)]}, \tag{3.2}$$

where the factor of instantaneous optical phase is due to the voltage applied to the electrodes is $\varphi(t) = \pi(U_0/V_\pi)\cos\Omega t = \delta\cos\phi$, which can be identified with the left argument of Eq. (3.1). The right arguments in the same equation allows us to find the harmonics components owing to phase modulation: for a lag of $n\Omega$ from ω_0 a peak amplitude $J_n(\delta)$ will be in the optical spectrum.

The experimental approach of cancellation of the central component $J_0(\delta)$ allows us to determine the sensitivity of phase modulation as illustrated in Figure 3.4 for each of the PMs. From the results obtained previously we have then determined the $V_{\pi,\mathrm{rf}}$ for the two PM with respect to different frequency modulation. This characterization was achieved by using an OSA with 10 MHz resolution, an RF synthesizer (HP 83751A) and a driver SHF 100CP. The latter component is used in order to cancel the central optical peak (i.e. $V_\pi = 4$ V, the RF amplitude modulation on the phase modulator should be 3 V or 6 V_{pp}, in other words 2.12 V_{rms} or 19.5 dBm, this amplitude is quite large and cannot be reached by our synthesizer). Figure 3.5 shows the $V_{\pi,\mathrm{rf}}$ values for each modulator with respect to the frequency.



(a) (b)

Figure 3.4: **Experimental cancellation of the central component $J_0(\delta)$ for each PM**.

This characterization allows us to get a physical relationship between the normalized amplitude parameter ρ with the input information $u(t)$ in Eq. (2.9). Then, the amplitude $\rho = 1$ corresponds to $\rho = V_{\pi,\mathrm{rf}}$.

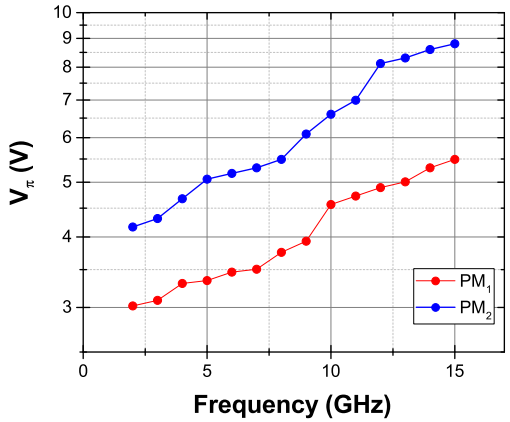


Figure 3.5: **Determination of the half-wave voltage $V_{\pi,\mathbf{rf}}$ values** .

3.1.3 Time scales in photonic RC

The first timescale to calculate is the characteristic response time τ of the filter, which is used in RC setup in order to slow-down the bandwidth in the feedback dynamics (see Figure 2.6). This is a low-pass filter (Picosecond Pulse Labs) having a cutoff frequency $f_c = 560$ MHz and a similar attenuation and group delay response to those of Bessel-Thomson filter. Its characteristic response time τ can be calculated as follows $\tau = 1/(f_c \cdot 2\pi) = 284.2$ ps. In substituting this previous value in Eq. (2.10) the node spacing can be also obtained as $\delta_\tau = 56.84$ ps. The node distance $\delta\tau$ within a layer, according to the discrete time dynamics described in Eq.(2.5), can be viewed as directly related to the local span of the impulse response $h(t)$ defined by the linear filter, and thus it is ruled by the characteristic time scale τ of this filter. This makes the connection between node separation and filter bandwidth, as already mentioned through an empirically determined optimal node distance of ca. $\tau/5$. One could however notice here, similarly to what was introduced in [13] with a multiple delayed feedback, that the non-local character of the nonlinear transformation adds another contribution to the virtual network connectivity with a coupling distance related to δT.

Once $\delta\tau$ calculated the other value necessary in order to calculate the number of virtual nodes K in the reservoir is the time delay τ_D brought by the series of optical and electrical connections in the feedback loop.

3.1.3.1 Time delay measurements

We know that the time delay τ_D has one of the main role in our RC processor (the spatial dimension will be virtually created inside it). Hence we propose to use an experimental methods to measure it.

Time delay τ_D value from cross-correlation function

The cross-correlation of two signals $x(t)$ and $y(t)$ is defined as:

$$C_{xy} = \int_{-\infty}^{+\infty} x(t)y(t-\tau)dt. \tag{3.3}$$

Eq. (3.3) reflects the similarity between $x(t)$ and $y(t)$ shifted in time by τ. When the signals are the same ($y \equiv x$), $C_{xx}(\tau)$ is known as the auto-correlation function (ACF), which allows to detect patterns and repeated profiles between a signal with itself delayed in time.

The cross-correlation theorem (Wiener-Khinchin) defines the relation between the cross-correlation and the FT as follows :

$$FT[C_{xy}] = FT[x(t)]^* \cdot FT[y(t)] = X^*(j\omega) \cdot Y(j\omega), \tag{3.4}$$

where * denotes complex conjugation.

In case of $y(t)$ is obtained from $x(t)$ by shifting in time and deforming, C_{xy} shows a peak at the corresponding position, which allows to extract the time lag. This is exclusively adequate to the evolution ruled by a DDE, as the signal is practically defined by a transformation applied to its own delayed signal (except the filtering effect). Figure 3.6 illustrates the ACF applied to a temporal trace in chaotic regime of our photonic RC, this ACF shows a peak linked to the time delay τ_D and a second one linked to the time imbalancing owing to the DPSK demodulator @$\tau_D + \delta T$.

Finally, in Figure 3.6, it can be observed that the separation between two harmonics corresponds to the time delay value $\tau_D = 63.35$ ns, and performing a zoom on the peak of the harmonic the time imbalancing of DPSK can be also obtained $\delta T =\sim 0.402$ ns. One way to evaluate the accuracy of this measurement is zooming on the maximum point of a peak and then using these values around it. This value is ~ 12 ps, so that the time delay measured with the ACF is $\tau_D = 63.35 \pm 0.01$ ns.

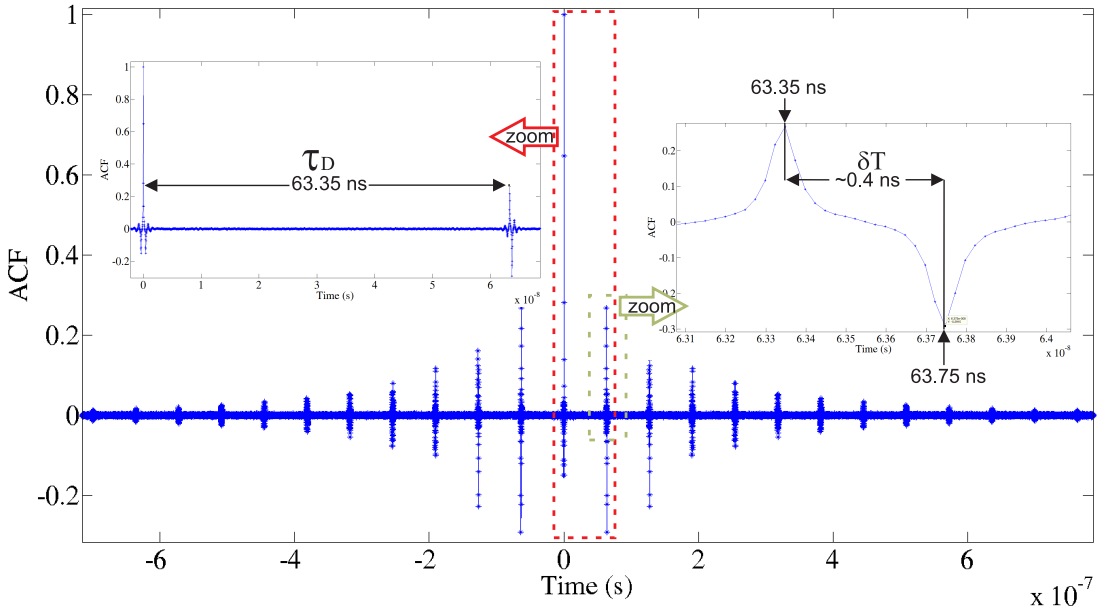


Figure 3.6: **ACF applied to a chaotic trace of our RC system.**

Then with the methods used we have obtained three slightly different values of time delay. But, for the future application we will use the τ_D value obtained by the ACF, because this has been measured directly in one of the chaotic trace recorded at the output of our RC photonic.

The time constants and the typical values of the parameters present in our photonic RC are presented in Table 3.1.

Variable	Meaning	Value	Information
τ	response time	284.2ps	low-pass filter $f_c = 560$ MHz
θ	low frequency cutoff	5.3 ms	RF driver limit $f_{lf} \simeq 30$ KHz
τ_D	time delay	63.35 ns	
$\delta\tau$	spacing node	56.8 ps	
δT	DPSK imbalance	0.4 ns	
β	normalized gain	0 - 5	
Φ_0	DPSK rest point	$\pi/4$	maximum slope

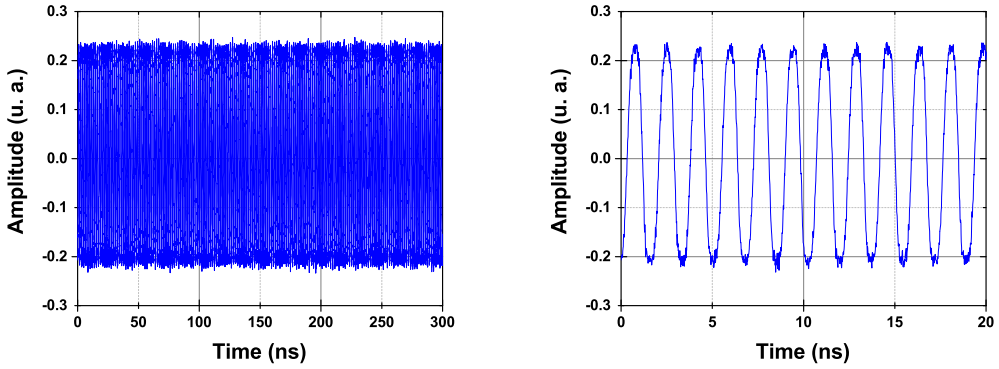
Table 3.1: **Typical parameters values.**

3.2 Dynamical characterization

In this section the autonomous dynamical properties of the RC system is investigated while tuning the parameters (particularly, the feedback loop gain β and offset phase Φ_0). The parameter β can be tuned by changing the in-loop attenuator settings (see Figure 3.5), with a maximum accessible value defined by saturation of RF driver (26 dBm). The offset interference phase $\Phi_0 = \omega_0 \delta T$, as already mentioned in section 2.3, can be tuned through the central wavelength of the laser or by applying a heating around one interferometer arm in the DPSK control. Usually, any Φ_0 operating point can be set to observe the dynamics of the system. However, in order to observe it at the lowest values of β, the operating point should be set $\pi/4$ corresponding to the maximum slope of the DPSK nonlinear function (see Figure 2.8).

3.2.1 Dynamical regimes observed experimentally

Figure 3.7 shows the time series for a low value of the feedback gain $\beta \simeq 0.5$. This time trace has been obtained from an electrical output proportional to the optical phase shift and the normalized variable $x(t)$. It was recorded with an 80 GS Lecroy WaveMaster 845Zi. When the feedback gain is increased gradually from zero, fast oscillations are first observed. A bifurcation point is reached at $\beta \simeq 0.5$ and right after this Hopf bifurcation, the oscillations are weakly stable. The first Hopf bifurcation serves as a reference for the experimental calibration of the bifurcation parameter β obtained from the optical power in the RC system feedback loop.



(a) (b)

Figure 3.7: **Experimental time trace for** $\beta \simeq 0.5$. (b) is a zoom of (a).

Applying the fast Fourier transform (FFT) to the time traces recorded at a 125 ns sampling length, the regular oscillation frequency $f_1 = 1.15$ GHz ($T_1 = 869$ ps) can be identified in the RF spectrum, as illustrated in Figure 3.8.

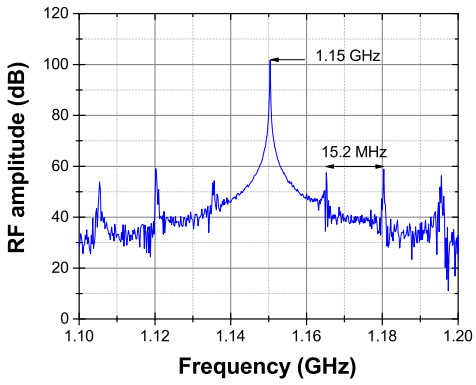
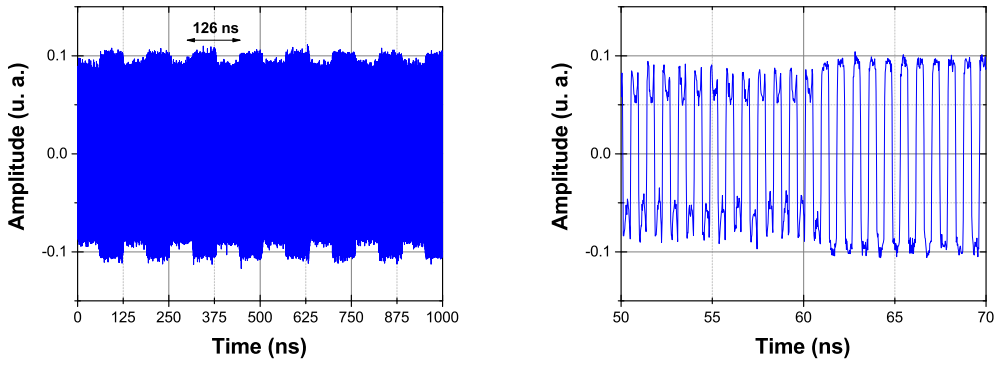


Figure 3.8: **Experimental RF spectrum from time trace ($\beta \gtrsim 0.5$).**

The dynamical behavior is then related to the DPSK demodulator time imbalancing, and is approximately equal to $2\delta T = 0.805$ ns. At this point the large delay τ_D does not influence the dynamical behavior, suggesting that only δT plays a role for the Hopf bifurcation. From the RF spectrum it is also possible to notice that the spacing between the noisy peak modes are originated by the large time delay $1/\tau_D = 15.79$ ns. Their noisy character exhibits amplitude fluctuations in time from the FFT, while the deterministic character of dominant peak leads to a much more stable amplitude peak. From values of $\beta \simeq 0.6$ to 0.9 the noisy character disappears and the shape of the oscillations becomes square waves with the same fundamental frequency [85]. In Ikeda-like systems, the first oscillations typically appear at $\beta = 1$. In our case, this threshold is reduced by a factor of 2, because instead of a single delayed combination parameter $x(t - \tau_D)$ of the nonlinear transformation there is a difference between $x(t - \tau_D)$ and $x(t - \tau_D - \delta T)$. Hence, if the solution $x(t)$ changes its sign when going from t to $t + \delta T$, the amplitude of a difference $x(t - \tau_D) - x(t - \tau_D - \delta T)$ can be twice as high as in case of a single delayed term $x(t - T)$, thus resulting in a lower threshold for β. A change of sign each δT period results practically in oscillations with period of $2\delta T$, which are indeed observed [114].

From further values of β a secondary bifurcation is obtained at $\beta = 1.14$ and other oscillations appear: fast oscillations modulated by a slowly varying envelope square, as shown in time traces in Figure 3.9. The period of the envelope is

close to $T_2 \simeq 122.99$ ns, which is close to $2\tau_D = 126.7$ ns, whereas the fast oscillations within the envelope still exhibit the faster period T_1. Therefore, both the large time delay τ_D and small delay δT play a role for the secondary bifurcation.



(a) (b)

Figure 3.9: **Crenelated oscillations with period related to large scale time by $2\tau_D$ modulating faster oscillations with period close to $2\delta T$.** (b) is a zoom of (a).

Applying the FFT to this time trace, the period of the slow-time envelope is observed, with sideband peaks in the RF spectrum at $\pm(2\tau_D)^{-1} = 7.89$ MHz, as illustrated in Figure 3.10.

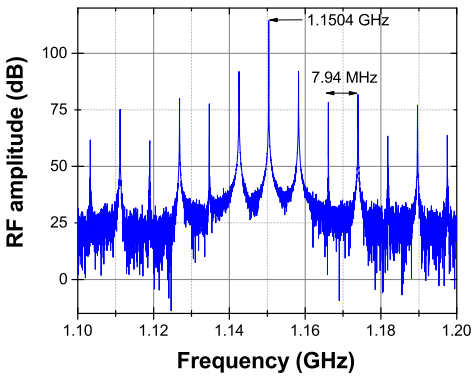
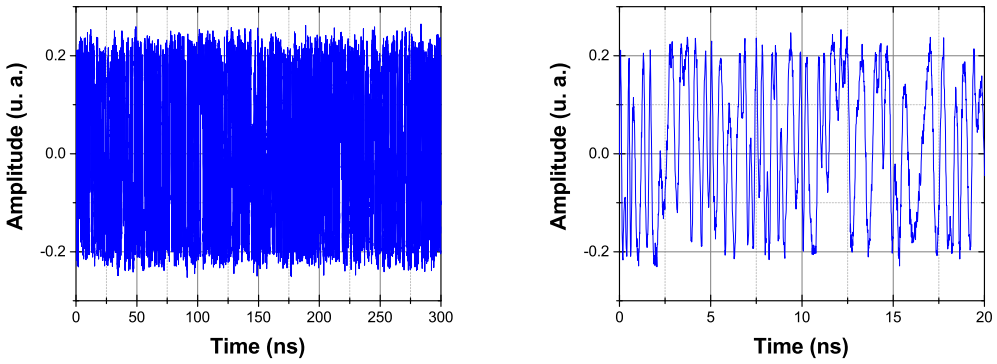


Figure 3.10: **Experimental RF spectrum measured right after the secondary bifurcation $\beta > 1.14$.**

These sideband peaks are consistent with the frequencies of the envelope

and the carrier signals in the amplitude modulation observed in the time series of Figure 3.9.

For much larger values of β, the time series exhibit chaotic oscillations and its RF spectrum is very flat as illustrated in Figures 3.11 and 3.12. This dynamical regime can be used in optical chaos telecommunication applications [112].



(a) (b)

Figure 3.11: **Chaotic dynamics**. In this case no characteristic time scale is directly observable in the time traces. (b) is a zoom of (a).

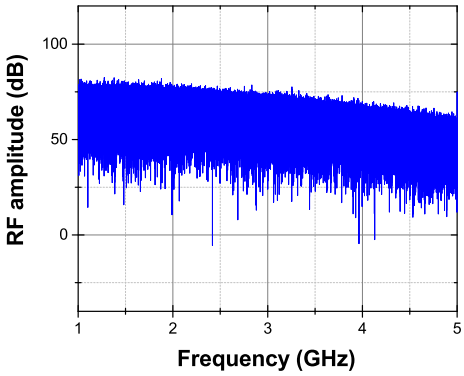


Figure 3.12: **RF spectrum very flat without any characteristic frequency.**

We can summarize from the results previously obtained, that the dynamical behavior of our RC system depends on a large time delay τ_D and a relatively small delay δT. These delays are responsible for a cascade of period-doubling

bifurcations to multiperiodic regimes that depend on two basic periods only, T_1 and T_2.

3.2.1.1 Bifurcations

From the dynamical model described by Eq (2.28), we have $\beta \propto \kappa \mid E_0 \mid^2$ which is adjusted through a variable optical attenuator with factor (κ). To facilitate the mathematical treatment, we fix the offset of phase $\Phi_0 = \pm\pi/4$. Eq. (2.28) then simplifies as:

$$x(t) + \tau\frac{dx}{dt}(t) + \frac{1}{\theta}\int_{t_0}^{t} x(s)ds = \pm\frac{\beta}{2}\sin^2\left[x(t-\tau_D) - x(t-\tau_D-\delta T)\right], \quad (3.5)$$

where the plus and minus signs correspond to $\Phi_0 = \pi/4$ and $\Phi_0 = -\pi/4$. With these values and increasing β from zero, the first Hopf bifurcation is reached at its lowest value. We note that the term $\omega_0 T$ only appears by a factor multiplying β. The minimum value of $\mid \beta \mid$ for a Hopf bifurcation occurs if $\mid \sin(\omega_0\tau) \mid = 1$, meaning $\Phi_0 = \pm\pi/4$ [85]. In order to demonstrate the latter, we introduce a new variable $u \equiv \int_0^t x(\xi)d\xi$, which leads Eq. (3.5) to two coupled first-order DDEs:

$$\begin{cases} \tau\frac{dx}{dt} = -x - \frac{1}{\theta}u - \frac{\beta}{2}\sin[2(x_{\tau_D} - x_{\tau_D+\delta T})] \\ \frac{du}{dt} = x, \end{cases} \quad (3.6)$$

where $x_{\tau_D} = (t-\tau_D)$ and $x_{\tau_D+\delta T} = x(t-\tau_D-\delta T)$.

Figure 3.13 shows the numerical simulation of Eq. (3.6) performed with Matlab. In the bifurcation diagram from Figure 3.13, three regimes can be observed: steady state, periodic oscillation and chaos. Each of them is used for different applications, i.e. in steady state regime the nonlinear transient computer is performed [11–13], in periodic oscillations a high spectral purity micro-wave can be generated [84, 85] and chaos can be used for optical communications [110–112]. In the present work we will set in steady state with not too small feedback strength β, allowing nonlinearity mixing however not needed too close from the instability threshold. In addition, one can note that the β values to reach the bifurcations are matched with those obtained experimentally.

For the eigenvalues a mathematical analysis is performed in Appendix B.

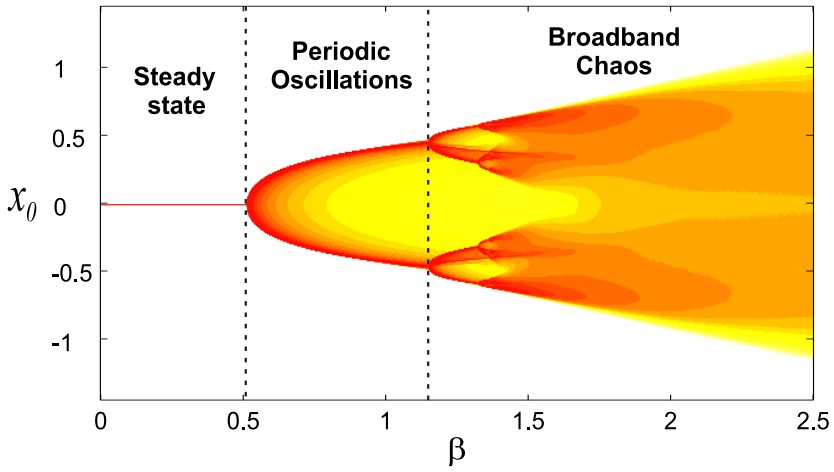


Figure 3.13: **Numerical bifurcation diagram, showing the β values for which the Hopf bifurcation is reached, as well as the secondary bifurcation.**

3.3 Performances of RC system

The nonlinear dynamical properties of time-delayed ultra-fast photonic oscillator need to be studied as a RC processor. In order to compare this RC efficiency with respect to those performed by computer simulations and to the studies previously reported [10–13, 16, 126], we decided to realize two types of classification tests. Both tests are about the SDR, the first is called TI-46 and the second one is the AURORA-2, which have been introduced in Chapter 2.

3.3.1 Classification tests: NIST TI-46 and AURORA-2

The operating principle of these tests including the model used by the pronounced digits (LCM and MFCCs), the masking procedure, the process to physically inject the signal into the dynamical reservoir and the learning and the testing steps have been described in Chapter 2 section section 2.4. The SDR task for both tests has been implemented in our RC system, following an identical experimental development. Although experimentation is performed with the same setup, the total number of virtual nodes, the reservoir size, the internal connectivity and settings of the system need to be adapted. For this reason, we propose a global view of each test for experimental implementation and right after, we present and discuss the results that are numerically and experimentally obtained.

3.3.1.1 Experimental implementation of TI-46 test

This test can be developed into 4 steps:

1. $\mathbf{u}(t)$ signal injection corresponding to the input information to be processed in the reservoir.
2. Dynamical processing of information performed with EO oscillator within the feedback loop RC.
3. Recording the output signal $x(t)$ within the reservoir via a digital scope.
4. The experimental traces recorded follow a learning phase aimed at calculating the Read-Out matrix $\mathbf{W}^R$, which is then used for the testing on untrained data.

The first step consists of building 1D input signal $\mathbf{u}(t)$, also called Write-In as described in section 2.2.1. Before, the input signal is pre-processed following the LCM (see section 2.4.1.1), as well as the masking procedure. During this pre-processing, the input information $\mathbf{u}(t)$ is saved in (.txt) format, which is then programmed into the internal memory of an AWG. In order to get all the 500 digits processed in the physical experiment, a sequence of digits are actually gathered in 10 to 20 digits typically (depending on the internal memory of the AWG used to generate $\mathbf{u}(t)$), leading to 20 to 50 sequences (in our case we have 33 sequences of 15 digits and 1 of 5 digits). The digits are randomly permuted to avoid any bias potentially introduced through the originally ordered arrangement. Moreover, time delays of zeros are added at the end of each digit in order to have the same length for each digit and to avoid any influence of one digit onto the next. In this step there are two parameters to adjust, the amplitude of $\mathbf{u}(t)$, which is adjusted by the amplification coefficient ρ of AWG and the sampling rate (S/s) so that information is distributed correctly on each virtual node within of the time delay τ_D.

Injection data using *"hidden layers"*.

In the following, we report on the possibility to introduce for delay reservoir, a similar conceptwith respect to the analogue of the so-called *hidden layers* in standard RNN. Unlike, the delay-based RC demonstrations proposed up to now, it was always assumed that each input sample $\mathbf{u}(n)$ is spread through the input layer over the full time delay duration τ_D. Here, hidden layers can be emulated through a N_L:1 ratio between the number N_L of input samples spanned over one time delay. In standard methods, a single layer i used corresponding to 1:1 spanning for each input sample over all the K nodes of a time delay.

Considering the case where both the separation between virtual nodes $\delta\tau$ and the numbers of nodes are fixed, it is then necessary to redefine or to increase the time delay by a factor N_L, thus leading to the temporal parameter scaling Eq. (3.7), ruling the node spacing, the number of internal layers, the number of emulated nodes and the time delay.

$$\tau_D = N_L K \delta\tau. \tag{3.7}$$

Compared to the single layer always adopted so far and for which the former theoretical description was detailed in Chapter 2, one needs to slightly redefine some of the equations of the RC processing model:

- The number of virtual nodes (per layer) is still defined as K, which imposes the number of columns of the input connectivity matrix $\mathbf{W}^I$.
- The recurrence time scale defined by the time delay τ_D is now covering N_L successive reservoir vectors $\{x_k(n), x_k(n+1), \ldots, x_k(n+N_L-1) \,|\, k = 1 \ldots K\}$, instead of a single reservoir vector $\{x_k(n) \,|\, k = 1 \ldots K\}$. The reservoir vector with K nodes corresponds then to the following definition:

$$x_k(n) = x(t) \quad \text{with} \quad t = n(\tau_D/N_L) + \sigma_k \tag{3.8}$$

  where the node position $\sigma_k = (k-1)\tau_D/(N_L K) \in [0; \tau_D/N_L]$.

It implies that the discrete time dynamics previously derived in (2.9) will develop as a result of the delayed feedback loop between input samples $\mathbf{u}(n)$ and $\mathbf{u}(n-N_L)$ instead of $\mathbf{u}(n)$ and $\mathbf{u}(n-1)$. Consequently, the delay-Reservoir dynamics has to be re-written as

$$x_k(n) = x_k(n-N_L) + \int_{\sigma_k-\tau_D}^{\sigma_k} h(\sigma-\sigma_k) \times f_{\mathrm{NL}}\left[x_\sigma(n-N_L) + \rho \cdot u_\sigma^I(n-N_L)\right] \mathrm{d}\sigma. \tag{3.9}$$

Motivated by a qualitative analysis of performances compared to the existing literature, we tried to adjust the number of hidden layers within the time delay, so that our experiment also involves a number of nodes per layer close to the 400 ones most often used in the literature [10–12]. The time delay $\tau_D = 63.35$ ns resulting in a total number of nodes of $\tau_D/\delta\tau = 5\tau_D/\tau \simeq 1114$, we chose $L = 3$ hidden layers, this implies that three successive input samples $u_\sigma^I(n)$ indexed from n to $n+2$ are needed to fill the delay line, instead of a single one as practiced in previous experiments. Each of these samples for a given n is also spanned over $\tau_D/3$ according to the input connectivity, leading to a TDM of the input information over $K = 371$ virtual nodes, according to Eq. (2.8). The

node separation is here imposed by the AWG sampling rate set to 17.6 GHz, defining theoretically the node spacing $\delta\tau = 56.8$ ps. Here, there actually exists a difference between the stepwise constant input as described in Eq. (2.8) and the physical light phase modulation. Indeed, the AWG has a 24 GS/s maximum sampling rate, but it also has a limited analogue bandwidth of 9 GHz. Thus it prevents from generating a stepwise constant amplitude, since the harmonics, and even the fundamental, are necessary filtered out by the AWG analogue bandwidth. The actual phase modulation ρu_σ^I is thus a smooth waveform of time. Nevertheless, there is no theoretical requirement for the stepwise constant profile for the input signal, the motivation being here essentially to keep the dynamics in permanent transient regime all along the duration of the input information. The latter condition is indeed also fulfilled even with a smooth input waveform, as long as its frequency content is fast enough compared to the characteristic time scale of the delay dynamics.

In a second step, the photonic delay dynamics provides an output signal or nonlinear transient response $x(t)$, which is ruled by the non-autonomous DDE (2.28). At this point two parameters are adjusted (β and Φ_0) before each experiment. β that represents the weight in the feedback delay loop, and its setting has to fulfill stability conditions of the dynamics in autonomous regime (referred as to spectral radius in RNN). Φ_0 is an offset phase setting the average rest point along the nonlinear function when no information is present.

In the third step, each transient response $\{x_{\sigma_k^R}(n) \,|\, n \in [1;N]\}$ of the dynamics to each of the 500 encoded input digital signal $\{u_\sigma^I(n) \,|\, n \in [1;N]\}$ is recorded by a real time ultra-fast digital scope with an over-sampling (80 GS/s) allowing for further post-processing.

The post-processing is the last step, in which the learning and testing phase described in section 2.4.2.1 are applied in order to quantify the results of SDR test. Once the answers of both learning and testing sets are known, the counting, the identification and the analysis of errors occurring under different processing conditions (ρ, β, Φ, and N_L) are easy to make. Figure 3.14 shows SDR test step by step.

3.3.1.2 RC: numerical simulation performances

The previously described physical parameters are directly accessible in the experiment so that they are adjusted to obtain the minimum WER. In this context, intensive numerical simulations of Eq. (2.28) have been performed in order to identify the optimal parameter conditions for best RC performances. The WER

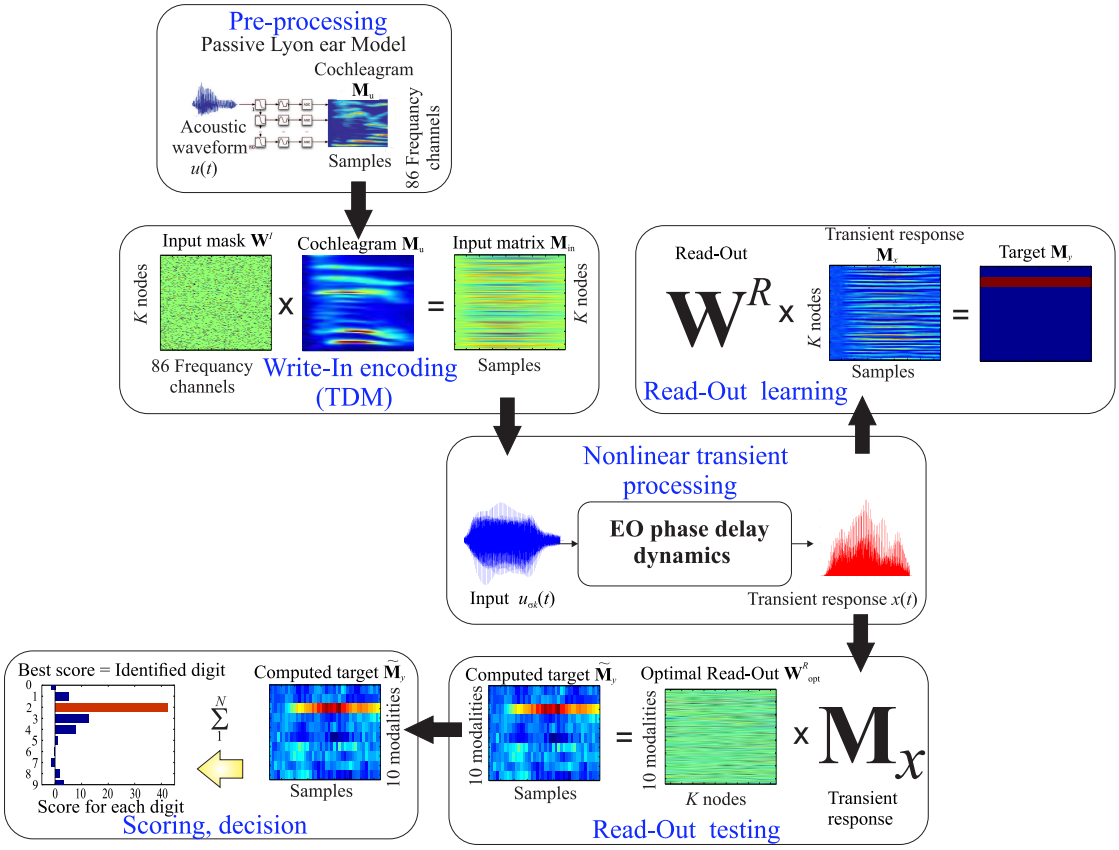
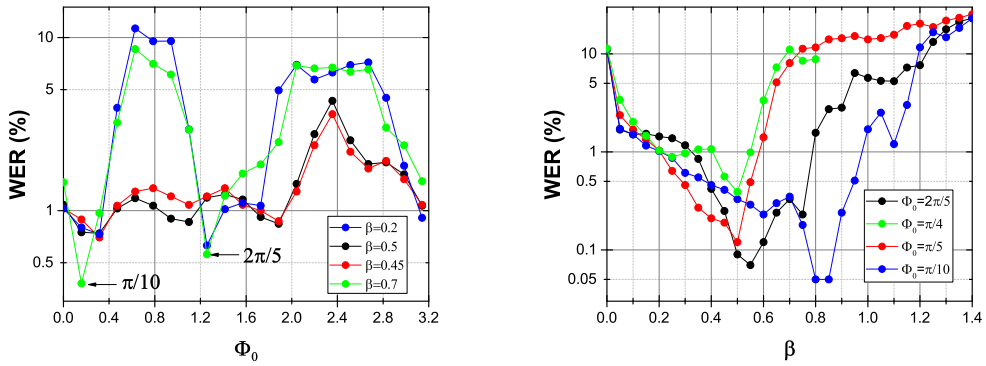


Figure 3.14: **Processing steps for the SDR task.**

obtained allows us to set the values of Φ_0, β and ρ. The other parameters τ, τ_D, $\delta\tau$ and N_L are set by the operating principle in transient regime of the RC processor.

The weight ρ for the input signal injected into the first EO PM allows for a more or less wide scan of the nonlinear function profile, thus controlling the actual nonlinear expansion of the information into the delay dynamics phase space. A value close to unity confirmed that a large amplitude injected information is in favor of optimal performances. This unit peak-to-peak amplitude is to be compared to the delay feedback signal peak-to-peak amplitude, $x_\sigma(n)$, which is of the order of $|2\beta \sin(2\Phi_0)| \ll 1$ (for $\Phi_0 \simeq \pi/4$). Excessive values of ρ are detrimental to RC processing because they can trigger transient self-oscillations if β is above the minimum Hopf threshold (at $\beta = 0.5$ if $\Phi = \pm\pi/4$).

Figure 3.15 shows the numerical simulations for SDR task. The WER has been calculated with a fixed value of $\rho = 1$ corresponding to V_π, rf (see section 3.1.2.2) and for different values of β, scanning the offset phase value $\Phi_0 \in [0; \pi]$. For finding the best value of β the previously procedure was performed, ρ remains equal to 1, but Φ_0 takes different values, while the feedback gain is scanned $\beta \in [0.01; 1.4]$ as represented in Figure 3.15(b).



(a) WER vs. Φ_0 parameter.

(b) WER vs. β parameter.

Figure 3.15: **Numerical simulations of SDR task (TI-46 test) in order to find the best values (a) for the offset phase and (b) for the feedback gain under synchronous conditions**.

In Figure 3.15(a), it can be observed that the optimal performance for this classification task is found for nearly quadratic nonlinearity, i.e. for an offset phase Φ_0 close to 0 modulo $\pi/2$. It is worth noticing that operations close to a minimum or a maximum lead to comparable performance optimizations compared to the linear operating point ($\pi/4$ modulo $\pi/2$), with a WER improvement factor of the order of 10. We also notice that the minimum WER is not due to a purely parabolic operating condition, but always slightly shifted from the minimum or the maximum, e.g. $\pi/10$ or $2\pi/5$ (see Figure 3.15(a) for $\beta = 0.7$) instead of 0 and $\pi/2$. This indicates that a linear contribution might be needed for an optimal processing of this classification task. Figure 3.15(b) shows an optimal value for the feedback gain slightly below from the instability threshold for the fixed point, as described by $|2\beta \sin(2\Phi_0)| < 1$ in our setup. The optimal values of $\beta \simeq 0.55$ and $\beta \simeq 0.8$ were obtained for the optimal nearly quadratic offset phase $\Phi_0 \simeq 2\pi/5$ and for slightly shifted of maximum operating point $\Phi_0 \simeq \pi/10$, respectively.

As mentioned before, the hidden layers are set by the operating principle in

transient regime, being $N_L = 3$ in our case. This concept of hidden layers for delay-based reservoir was also found to be of practical interest for the improvement of RC processing efficiency. Its motivation was initially inspired by the long time delays compared to the fast response times of photonic telecom devices, essentially due to the too long fiber pigtails in these devices (63 ns$\gg$284 ps). It leads to an initial scaling of the delay dynamics with more than 1000 virtual nodes. Introducing three hidden layers enabled us to process information with a comparable number of nodes (around 400) with respect to some of the already existing literature. Simulating the processing with an arbitrary number of hidden layers indicated that the concept indeed improved the performances compared to the single layer topologies reported up to now. Additional numerical simulations performed a scanning of the hidden layers length $N_L \in [1, \ldots, 17]$ with $\rho = 1$, $\beta = 0.6$ and $\Phi_0 = 2\pi/5$ (see Figure 3.16), even showed us that an optimal scaling can be obtained for around 3 to 5 hidden layers. This might be related to some internal correlation of human speech, which could be more efficiently revealed by the use of hidden layers. Nevertheless in experimental practice, we continue to use the value of $N_L = 3$ due to limitations of the internal memory capacity imposed by AWG.

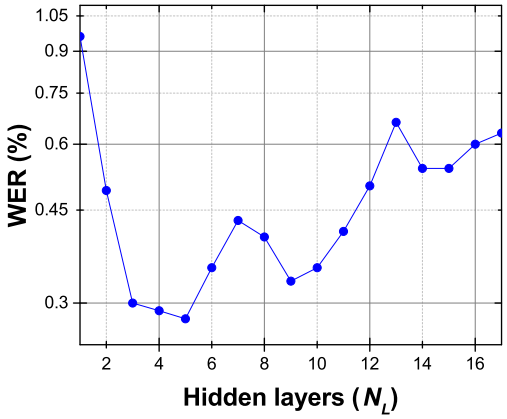


Figure 3.16: **Numerical simulation of hidden layers with values fixed** $\rho = 1$ $\beta = 0.6$ **and** $\Phi_0 = 2\pi/5$.

3.3.1.3 Numerical and experimental results

The numerical simulations performed in the previous section allow us to set the experimental parameters and also to find qualitative criteria that are fully con-

sistent with previously obtained results [10–12]. The experimental parameters used for SDR TI-46 are presented in Table 3.2.

Parameter	Meaning	Value
β	normalized gain	0.1-1.5
ρ	information amplitude	1
Φ_0	offset phase	$2\pi/5$
$\delta\tau$	spacing node	56.8 ps
$(\delta\tau)^{-1}$	sampling rate (AWG)	17.6 GS/s
N_L	hidden layers	3
K	number of nodes	371
f_s	sampling frequency (Lecroy scope)	80 GS/s

Table 3.2: **Experimental parameters values.**

One important point to highlight is that the sampling frequency f_s was only used in experiments. The temporal traces are recorded by a real-time scope at a $f_s = 80$ GS/s, thus allowing to choose among 5 samples in the Read-Out processing. The experimental results of WER obtained with the parameter values of Table 3.2 are presented in Figure 3.17 and also the results numerically obtained under the same conditions are presented.

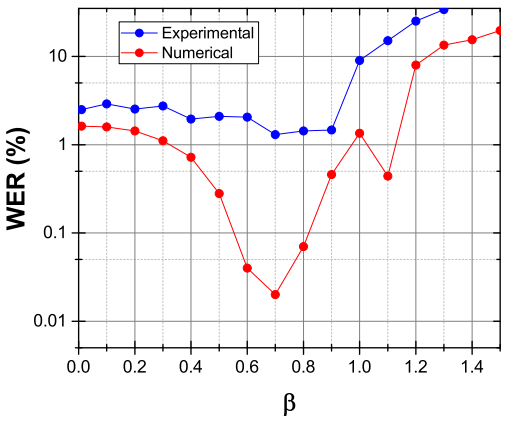


Figure 3.17: **WER vs. β parameter, under synchronous conditions** $\delta\tau = \delta\tau^R$. Red line is numerical and blue line is experimental.

In Figure 3.17 it can be noted that the experimental WER is degraded compared to numerical results, however it remained at a very good level with a best operating point slightly above 1%. On the other hand, it is important to say that numerical simulations are performed in ideal case. They do not take into account the noise added by the devices, the cables, neither the bandwidths imposed for the instrumentation. Therefore, we can conclude that the experimental result can be compared with this one obtained numerically.

3.3.1.4 Results: synchronous and asynchronous mode

Up to this point all simulations and experiments have been performed in synchronous conditions between Write-in and Read-Out according to Eq. (2.11). In order to enrich the dynamics, we proposed a similar desynchronization, but not equal, to the one described in [16, 45]. In Massar, *et al.*, paper, the input information is desynchronized with respect to the total delay of the feedback loop, dividing the reservoir into different independent subsystems. In our case we have worked with the reservoir in synchronous mode and in transient regime ($\tau > \delta\tau$) as [10, 11, 13]. Nevertheless, we have desynchronized the response record of the virtual nodes amplitudes (matrix $\mathbf{M}_x$) with respect to the node separation ($\delta\tau$) forced at the input information Write-In. This effect of an asynchronous Read-Out (sampling of the transient response) with respect to the Write-In (frequency imposed by the AWG) has been finely explored, both numerically and experimentally. In Figure 3.18 a simulation of this desynchronization can be observed and the asynchrony is quantified as $\varepsilon = \delta\tau^R/\delta\tau - 1$.

We discovered a significant RC processing enhancement by a factor of more than 10, for a very small relative ε deviation between $\delta\tau$ and $\delta\tau^R$, of the order of 10^{-4}. Such an improvement was moreover obtained for whether positive or negative relative asynchrony, the exact synchrony leading actually to a deterioration of the RC classification performance when considering the large asynchrony ($|\varepsilon| > 5 \times 10^{-4}$) asymptotes as shown in Fig. 3.18. The origin of this phenomenon is yet unexplained, and was not discovered earlier, as far as we know. Due to the extremely small time scale shifts of concern, this phenomenon is necessarily related to fast response times, and thus to the continuous time information mixing through the impulse response $h(t)$. We anticipate that efficient information extraction from the input signal actually requires asynchrony, possibly related to globally non-uniform sampling issues.

An illustration of this approach is shown in Fig. 3.19. Note that these results have been obtained with same parameters used in Fig. 3.17, the only difference is that in this case the asynchronous Read-Out is applied ($\varepsilon = \pm 10^{-4}$),

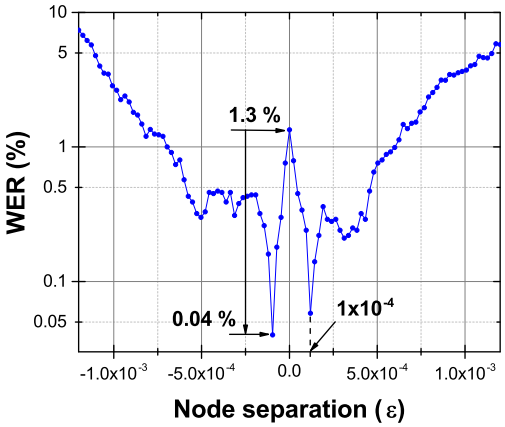


Figure 3.18: **Numerical simulation of WER as function of the relative Read-Out vs. Write-In asynchrony** ε.

improving both the numerical and experimental results. Numerically WER ranging from $0.02 \pm 0.02\%$ down to 0% and experimentally WER ranging from $1.3 \pm 0.2\%$ down to only $0.042 \pm 0.03\%$, thereby implying that the asynchronous Read-Out enable us to achieve performance comparable to state-of-the-art systems. This results validates the computational efficiency of our ultra-fast photonic RC setup.

Quantitatively, excellent RC classification performances have been obtained both numerically and experimentally with WER close to 0%, thus confirming previous results from other photonic setups [10–13,16,45,46,58,126,127]. For the results obtained with our RC system, it is worth emphasizing the high speed of our setup presents a state-of-the-art improvement by a factor ~ 3, which is the most important contribution of this PhD work. We introduced a novel RC setup based on an EO phase delay dynamics setup. The ultra-fast RC setup is built with telecom components, which allows a relatively faster information processing, reaching state-of-the-art classification speed: for an average duration of 60 samples for one spoken digit, the analogue processing can be performed within 1.3 μs, thus resulting in a processing speed of the order of 1 million words per second. In the following section both WER and information processing speed will be compared with the approaches of RC.

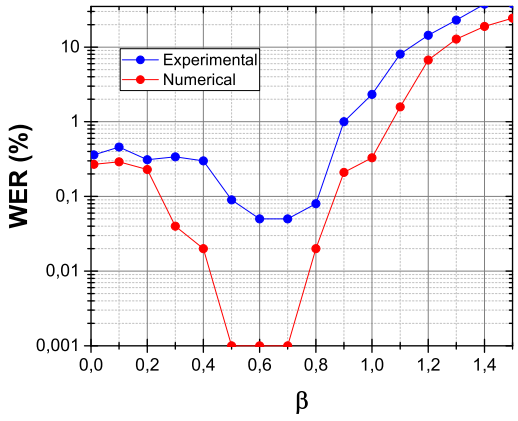


Figure 3.19: **WER as function of the relative Read-Out vs. Write-In asynchrony quantified as** $\varepsilon = \pm 10^{-4}$. Experimental and numerical results.

3.3.1.5 Comparison with state-of-the-art methods

Finally, the WER results obtained experimentally with our RC processor are compared with those obtained with approaches based on RC, with those obtained by classic approaches (LSM and ESN), and also with *Hidden Markov Models* (HMM). This comparison is described in table 3.3:

3.3.2 Development of AURORA-2 test

This test is broken down into the same number of steps used in TI-46 test. The only differences are the contribution of Write-In and the Read-Out function. In the Write-In the input information $\mathbf{u}(t)$ follows a pre-processing based on MFCCs described in section 2.4.1.2 and in Read-Out there are 11 modalities to learn. The following steps (i.e. masking procedure, data injection, training and testing) are performed in the same way than in the previous test.

3.3.2.1 Numerical performance obtained

In order to evaluate the efficiency of our RC processor, we performed numerical simulations using only a part of this test (see Chapter 2 section 2.4). The part used corresponds to the folder called *clean1* (*multi-condition* mode), with a total of 1001 sequences of digits, among which 289 are isolated digits. We have randomly chosen 265 digits for the training subset, and 24 digits remaining were left for testing. The setting of the parameters has been based on the best op-

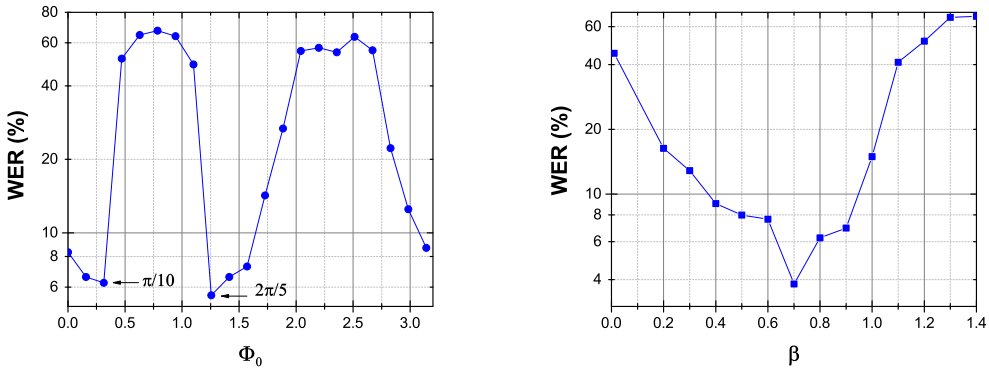
Approach	**WER (%)**	**Information**	**Processing speed (S/s)**
Photonic RC	0.042 ± 0.03	This PhD thesis	17.6 G
Photonic NTC	0.06 ± 0.02	[13, 126]	628.3 K & 10 M
Silicon RC	close to 0	[46]	
Parallel photonic information processing	0.014	[12]	5 G
All-optical RC	0.3	[45]	200 M
Optoelectronic implementation of RC	0.04 ± 0.017	[11]	19.1 M
RC using delay systems	0.2	[10]	40 K
Optoelectronic RC	0.4	[16]	23.5 M
RC-based in techniques SDR	0.2	[58]	
Information processing using transient dynamics	0.14	[127]	5 G
LSM	4.3	[59]	
ESN	1.3	[128]	
HMM	0.55	[129]	

Table 3.3: **Comparison of WER obtained with different approaches.**

erating point found for the TI-46 test, being $\rho = 1$, $\beta = 0.7$, $\Phi_0 = 2\pi/5$, the number of virtual nodes and the hidden layers are the same. With these setting the WER found was around $5 \pm 1\%$, suggesting that perhaps the parameters obtained giving the best efficiency for the previous SDR task (TI-46 test) are not suitable to generate the same performances. For this reason, numerical simulations were performed and they are represented in Figure 3.20.

From the results of Figure 3.20, we confirmed that the value of the parameters found in the TI-46 test are similar to those giving the best results for the AURORA-2 test. Nonetheless, the WER obtained in this simulation is at least degraded by a factor 10 with respect to those obtained numerically in TI-46 test. We then performed other simulations to investigate the influence of the reservoir size (number of nodes) as proposed in [117]. This simulation shows that the WER decreases with the reservoir size, but it starts to saturate as soon as the number of trainable parameters reaches 1000 as shown in Figure 3.21.

Increasing the number of virtual nodes, the WER decreases down close to 1%, motivating to perform once again the previous simulation with the same



(a) WER vs. Φ_0 parameter.

(b) WER vs. β parameter.

Figure 3.20: **Numerical simulations of AURORA-2 test in order to find the best values (a) for the offset phase and (b) for the feedback gain under synchronous conditions**. This is performed using a folder called *clean1*, which is only a part of the test.

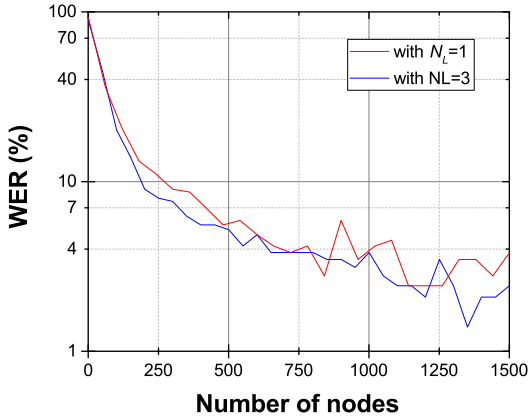


Figure 3.21: **WER as function of reservoir size**. Red line shows the WER without hidden layers, and blue line is using $N_L = 3$.

parameters except for the size of the reservoir ($K = 1000$). We used three hidden layers, so that we have indeed $1000 \times 3 = 3000$ effective nodes. With this protocol we have obtained a WER $1.21 \pm 0.5\%$, which allowed us to find qualitative criteria that are fully consistent with previously obtained results [117,118].

Another test was conducted with all sequences of digits (1001) of *clean1* from multi-condition mode. These sequences have been treated with Matlab in order to obtain isolated digits, so that the number of digits isolated for this test is 3257. The learning set has been completed using the same process, but the digits used correspond to the training for clean1 condition in the test AURORA-2. This subset consists in 422 digit sequences, which were trimmed with Matlab in order to obtain a total of 1397 isolated digits. The learning phase was then conducted with the same parameters used in the previous test, with a set of 1397 digits for the learning phase, and 3257 digits for testing. Under these conditions, we have obtained a WER around 6%.

At last, a final test was conducted with all isolated digits of test for the *clean* mode. This mode has 8440 digit sequences among which 2412 are isolated digits. The parameters of this test are exactly the same as those used in the previous cases, except the set of digits for the training and testing. For the training phase we used 75% of all digits, being 1809 digits and the remaining 25% has been used for the testing phase, being 603 digits. With these settings, the WER obtained is 0.69 ± 0.21. The latter value is even better than the one obtained in [117] with RC techniques, and it is also comparable to those obtained with HMM [118].

We were unable to test this experimentally, but good correspondence between theory and experiment were found in TI-46 task, leading us to conclusion that our system can also reach a good processing efficiency in the experiment.

3.3.2.2 Comparison with state-of-the-art methods

Finally, the WER results obtained numerically are compared with with these obtained with approaches based on RC and with HMM. This comparison is described in table 3.4.

Approach	**WER (%)**	**Information**	Number of nodes
Photonic RC	0.69 ± 0.21	This PhD thesis	1000
RC	1.21	[117]	4000
HMM	0.65	[118]	

Table 3.4: **Comparison of WER obtained with different approaches.**

3.4 Overview

In this chapter, we reported on the experimental implementation of an ultra-fast photonic RC processor. A complete characterization of each element has been proposed. We measured the time delay by ACF method. Then a dynamical study was performed, finding at least three different types of regimes (steady state, oscillations and chaos). For low values of feedback gain $\beta \gtrsim 0.5$ fast oscillations are observed and a bifurcation point is reached at $\beta = 0.5$. Applying the FTT to the time traces other dynamics are revealed, these dynamics are related to the time imbalancing δT and the time delay τ_D. Increasing $\beta \simeq 0.6$ to 0.9 crenelated oscillations are obtained, corresponding to oscillations that are modulated by a slow envelope with a period close to $2\tau_D$ (the fast oscillations inside remaining with a period close to $2\delta T$). Computing the FFT the slow-time envelope is observed, being determined by the sideband peaks. Further increasing β, a secondary bifurcation is reached at $\beta = 1.14$ and for values much larger than the chaotic oscillations appear. The bifurcations have been confirmed numerically in Matlab, leading to a good correspondence.

One part of this chapter was dedicated to the validation and to the quantification of the performances of our ultra-fast photonic RC, using two types of classification tests based on SDR. Before, reporting their results, we introduced the possibility to use a hidden layers concept instead of using a single as layer proposed up to now. Then, the SDR tests were numerically performed with the purpose to find the best operating point for the RC information processing. The found parameters giving the best results are $\rho = 1$, $\beta = 0.7$ and $\Phi_0 = 2\pi/5$ and $\Phi_0 = \pi/10$, the number of nodes and the hidden layers are fixed by the experimental setup ($K = 371$ and $N_L = 3$). Nevertheless, the numerical simulation of the hidden layers showed that using around 3 to 5 such layers the efficiency of RC processing can be improved. With these setting, we found numerically a WER=$0.02 \pm 0.02\%$ and experimentally $1.3 \pm 0.2\%$, showing that the experimental WER is degraded compared with numerical WER, but it is worth to mention that numerical simulations are made in the ideal case. Furthermore, in order to enrich the dynamics, we proposed an asynchronous sampling for the the Read-Out, consisting in desynchronizing recording rhyme of the transient response signals with respect to the virtual node separation imposed by the Write-in. This asynchrony enhances the RC processing by a factor of more than 10, lowering the experimental WER down to only $0.042 \pm 0.03\%$, thereby implying that the asynchronous Read-Out enable us to achieve performances comparable to state-of-the-art systems. Therefore, in this PhD thesis we proposed two novel techniques with which the efficiency of our RC processor is

improved (hidden layers and asynchronous Read-Out), but the main contribution in this work is the realization of a photonic RC using telecommunication devices. The experimental implementation of ultra-fast RC is based on an EO phase delay dynamics, enabling the possibility to achieve bandwidths much higher than those reported in previous RC systems [10–13]. Indeed, the new RC setup allowed us to achieve a record of information processing speed at 17 GS/s, improving the classification speed by a factor of ~ 3. Formulated in another way, this RC system can process around of 1 million of words per second, establishing the state-of-the-art in classification speed.

At last, the AURORA-2 test was conducted in the same way as TI-46 test, except for the number of virtual nodes (1000 nodes) and the number of spoken digits to classify. The test was performed with all isolated digits of the AURORA-2 test, being 2412 digits. With these settings, we found numerically a WER=0.69 ± 0.21. However we were unable to test this experimentally. A globally good correspondence between theory and experiment as found in TI-46 task, allows leads us to speculate that our system can also obtain a good efficiency even for experimental information processing.

Conclusions and perspectives

Conclusions

This PhD thesis focused on the study and implementation of RC as an unit for information processing. Our approach of photonic RC enabled us to achieve a faster processing by a factor of 3, establishing state-of-the-art classification speed. The analogue processing speed was evaluated close to 1 million words per second.

At the beginning of this work, the study of RC was based on a new brain-inspired concept using an RNN in the field of neuroscience and in machine learning. The new approach was promptly adopted due to its performances in complex tasks such as time series prediction or speech recognition. However, in order to realize an experimental implementation the dynamical reservoir of interconnected neurons is replaced by an DDS. The idea is to use a DDS instead of standard spatio-temporal structure of RNN. It simplifies the design, providing an innovative hardware solution, as well as similar expansion properties for computational capabilities. The RC system consists typically in three layers: an input layer where the information is injected, a second layer corresponding to a dynamical reservoir of virtual nodes interconnected randomly with synaptic weights, and an output layer, known as the Read-Out function that is responsible for extracting information from the reservoir. Also in this layer the learning phase is applied. In the RC system the main advantage is the fact that the training does not change the dynamics of the reservoir, hence the learning phase is faster, reliable and easier than classical methods.

The proposed RC system is based on an EO phase delay dynamics (DDS), which was developed for applications in the areas of optical chaos telecommunications. The dynamical properties of DDS, e.g. the nonlinear function, the fading memory and the infinite dimensional phase space are the arguments on

which relies our RC approach. Nevertheless, in this system the spatial dimension does not exist physically, but it is virtually created within the temporal dimension, implying that the information is serially injected following a technique of signal processing known as TDM. As a result, the input information needs a pre-processing, which consists in the adaptation of the signal before being addressed onto each virtual node in the dynamical reservoir. In the pre-processing, the input information is weighted by the synaptic points in accordance to the elements of a mask function. The mask output is then a piecewise constant function, constant over an interval covering the separation between two successive nodes $\delta\tau$ and periodic with a period corresponding to the time delay τ_D.

The internal connectivity of the reservoir is ruled by several parameters: the time delay, the response time and the most important of them is the separation between two successive virtual nodes. Within the PHOCUS project, it was determined empirically that the good performances of RC are obtained when it operates in transient regime. In this PhD thesis, we confirmed numerically that the good efficiency of RC system occurs when it works under transient conditions, meaning that the separation between two successive virtual nodes is small enough with respect to the characteristic response time of the system.

A corresponding mathematical model of the EO oscillator was proposed. It was particularly centered in the DPSK component because it is the one that provides the nonlinear function. Afterwards, a dynamical characterization was performed showing at least three different behaviors (steady state, period oscillation and chaos). On the other hand, we introduced the possibility to use hidden layers instead of assuming always that each input sample is spread through a single input layer as it has been proposed up to now. In order to evaluate and to quantify the efficiency of our ultra-fast RC system, two standard tests based on SDR are used. The goals of these tests are to classify a series of spoken digits, one of them is named TI-46 and the other one is AURORA-2. The WER obtained for the TI-46 test in synchronous mode is 1.3%, but if asynchronous sampling in Read-Out is adopted (the transient response signals are desynchronized with respect to the virtual node separation imposed by the Write-In) the WER is then improved by a factor of more than 10. Therefore, this asynchrony highlights the efficiency of information processing and enables us to achieve performances comparable to state-of-the-art systems, experimentally and numerically: we found a WER of $0.042\pm0.03\%$ and $0.02\pm0.02\%$ respectively. These results were obtained with the following parameter settings: 3 hidden layers, 371 virtual nodes and the best operating point for the RC infor-

mation processing ($\beta = 0.7$, $\phi = 2\pi/5$ and $\rho = 1$). At last, the AURORA-2 test was conducted with the same settings used in TI-46 test, however consisting of a more important set of digits to be tested 2412 digits. In this test, the WER result was too degraded with respect to this one obtained in TI-46 test. The solution proposed in order to improve the efficiency for RC information processing was to explore the performances of the system by changing the reservoir size. It was performed numerically from 371 virtual nodes to 1500 nodes and it was noticed that a reservoir with around 1000 virtual nodes gives a good efficiency with the AURORA-2 test. Accordingly, a reservoir of 1000 virtual nodes allowed to find numerically a WER=0.69$\pm$0.21%, which is even better than the one obtained in [117] with RC techniques, and it is also comparable to those obtained with HMM [118].

Perspectives

In this PhD thesis we introduced an ultra-fast photonic RC based on EO phase delay dynamics. The new RC setup allowed us to process 1 million words per second, the possibility to use hidden layers, as well as an asynchronous sampling mode in the Read-Out. These characteristics enabled our RC approach to establish the state-of-the-art in classification speed and also to obtain a WER comparable and even better to those found in the literature. Although good results have been obtained, many are perspectives for the photonic RC system.

- The topology of the proposed RC approach does not allow to process digit-sequences. Then a modified architecture might be interesting for future study. Furthermore, this modification enables us to process all digits in AURORA-2 test and also to validate the results obtained numerically. On the other hand, in the current topology the testing phase is performed offline, however using integrated circuit such as FPGA (Field-Programmable Gate Array) board would enable a hardware solution for full experimental implementation and at the same time would be possible to use multi-delay instead of a single delay.

- Another important point would be to use the MFCC model in the TI-46 test and while in AURORA-2 test the Lyon model will be adopted. It would provide a measurable point of comparison, which makes it easier to know which model is more efficient.

- The results obtained with the photonic RC suggest to continue in this line,

using the dynamics based on Chimera states for information processing.

- The tests performed in this work were based on SDR or classification task, and might be interesting to introduce other tests based on prediction of time series.
- We introduced the hidden layers as a solution in order to use a number of virtual nodes similar to this one generally used in RC approaches. It means that the information within each hidden layer is exactly the same and in Read-Out function only one layer is used in order to extract the transient response. That is why, a study that allows to explore the possibility to learn each layer individually as carried out in deep learning methods would be very suitable.

Papers and conferences related to this PhD thesis

Articles in international reviewed publication

Larger, L., Baylón-Fuentes, A., Martinenghi, R., Udaltsov, V. S., Chembo, Y. K., & Jacquot, M. High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification. Physical Review X, 7(1), 011015 (2017).

Proceedings

R. Martinenghi, A. Baylon-Fuentes, F. Xiaole, M. Jacquot, Y. Chembo and L. Larger. "Optoelectronic nonlinear transient computing with multiple delays". In Conference on Lasers and Electro-Optics - International Quantum Electronics Conference (CLEO/EUROPE - IQEC), Munich, Germany, may 2013.

A. Baylon-Fuentes, R. Martinenghi, M. Jacquot, Y. Chembo and L. Larger. "10 GHz bandwidth nonlinear delay electro-optic phase dynamics for ultra-fast nonlinear transient computing". In Conference on Lasers and Electro-Optics - International Quantum Electronics Conference (CLEO/EUROPE - IQEC) - IEEE/OSA, page 1530, Munich, Germany, may 2013.

R. Martinenghi, A. Baylon Fuentes, M. Jacquot, Y. Kouomou Chembo, and L. Larger. Towards optoelectronic architectures for integrated neuromorphic computers. In Proc. SPIE 8989, Smart Photonic and Optoelectronic Integrated Circuits XVI, number 89890K, San Francisco, California, United State, feb 2014.

L. Larger, A. Baylon Fuentes, R. Martinenghi, M. Jacquot and Y. Kouomou Chembo. Delayed electro-optic phase dynamics for ultra-fast nonlinear transient computing. In 2014 International Symposium on Nonlinear Theory and its Applications (NOLTA 2014), page 553, Luzern, Switzerland., sep 2014.

L. Larger, A. Baylon-Fuentes, R. Martinenghi, M. Jacquot, Y. Kouomou Chembo, and V. Udaltsov. Collective motion in delay dynamics photonic brain inspired processors. In WIAS workshop on Collective dynamics in coupled oscillator systems., Berlin, Germany., nov 2014.

L. Larger, B. Marquez, A. Baylon Fuentes, R. Martinenghi, M. Jacquot, Y. Kouomou Chembo, and V. Udaltsov. Photonic nonlinear delay dynamics for advanced information processing: from secure chaos communications to brain-inspired computing. In NOLTA 2015, International Symposium on Nonlinear Theory and its Applications, number Paper ID : 6263, pages 538–541, Hong Kong, China, dec 2015.

Conferences

R. Martinenghi, A. Baylon Fuentes, M. Jacquot, Y. Kouomou Chembo, and L. Larger. Expérience électro-optique générique de calculateur neuromorphique utilisant les transitoires complexes des dynamiquesa retard. In 16èmes Rencontres du Non Linéaire, pages 98 – 103, Université Paris-Diderot, Paris, France, mar 2013.

R. Martinenghi, A. Baylon-Fuentes, M. Jacquot, Y. Chembo and L. Larger. Introduction au "Reservoir Computing" basé sur une dynamique non-linéaire à retards multiples photonique. Journée de la Photonique 2013, Dijon, Novembre 2013.

A. Baylon-Fuentes. 10 GHz bandwidth nonlinear delay electro-optic phase dynamics for ultra-fast nonlinear transient computing. SMart sYstems for a better LifE (SMYLE) 2013, Besançon, France, 12 – 13 sep 2013.

L. Larger, B. Penkovskyi, A. Baylon-Fuentes, M. Jacquot, Y. Kouomou Chembo, and Y. Maistrenko. New paradigm in delay dynamics: Virtual chimera states and brain-inspired processing. In The 4th International Conference on Complex Systems and Applications - ICCSA., Le Havre, France., jun 2014.

L. Larger, R. Lavrov, A. Baylon-Fuentes, M. Jacquot, Y. Kouomou Chembo, and Vladimir Udaltsov. Dual delay electrooptic phase oscillator: A nonlinear non-local dynamics for high performance secure optical chaos communications. In The 4th International Conference on Complex Systems and Applications., Le Havre, France., jun 2014.

L. Larger, A. Baylon-Fuentes, B. Penkovskyi, M. Jacquot, Y. Kouomou Chembo, and V. Udaltsov. High-speed photonic demonstration of reservoir computing. In GdR ISIS Journ´ees NeuroStic., Univ. de Cergy-Pontoise, France., jul 2014.

A. Baylón-Fuentes, R. Martinenghi, I. Zaldıvar-Huerta, B. Márquez, V. Udaltsov, M. Jacquot, and L. Larger. Reservoir Computing ultra-rapide basé sur une dynamique non-linéaire électro-optique en phase. Résumés des exposés de la 18e Rencontre du Non-Linéaire Paris 2015, 75.

A. Baylon-Fuentes, M. Jacquot, Y. K. Chembo, B. Marquez, R. Martinenghi, I. Zaldivar-Huerta, and L. Larger. Calculateur photonique neuromorphique réalisant un test de reconnaissance vocale à 1 million de mots/s. Journée du Club IETR, Nantes, France, 12 juin 2015.

List of Figures

List of Tables

Appendices

Appendix A: Mathematical model of EO phase delay

Equation (2.17) in chapter two shows the model for the interference between the two arms of the interferometer DPSK.

$$\begin{aligned} E(t) &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)+j\varphi(t-\tau_D)} + \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D-\delta T)+j\varphi(t-\tau_D-\delta T)} \\ &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)+j\varphi(t-\tau_D)} + \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)-j\omega_0\delta T+j\varphi(t-\tau_D-\delta T)} \\ &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)} \cdot \left[e^{j\varphi(t-\tau_D)} + e^{-j\omega_0\delta T+j\varphi(t-\tau_D-\delta T)}\right], \end{aligned} \tag{10}$$

If we multiply the Eq. 10 by $e^{j\omega_0\delta T}$ obtain

$$\begin{aligned} E(t) &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)} e^{j\omega_0\delta T} \cdot \left[e^{j\varphi(t-\tau_D)} e^{j\omega_0\delta T} + e^{j\omega_0\delta T} e^{-j\omega_0\delta T+j\varphi(t-\tau_D-\delta T)}\right] \\ &= \frac{1}{2}E_0 e^{j\omega_0(t-\tau_D-\delta T)} \cdot \left[e^{j\omega_0\delta T+j\varphi(t-\tau_D)} + e^{j\varphi(t-\tau_D-\delta T)}\right], \end{aligned} \tag{11}$$

Now the equation (11) is multiplied by $\frac{e^{\left[-j\omega_0\delta T-j\varphi(t-\tau_D)-j\varphi(t-\tau_D-\delta T)\right]/2}}{e^{\left[-j\omega_0\delta T-j\varphi(t-\tau_D)-j\varphi(t-\tau_D-\delta T)\right]/2}} = 1$,

$$
\begin{aligned}
E(t) &= \frac{\frac{1}{2}E_0 e^{j\omega_0(t-\tau_D)}}{e^{[-j\omega_0\delta T - j\varphi(t-\tau_D) - j\varphi(t-\tau_D-\delta T)]/2}} \left[e^{j\omega_0\delta T - j\omega_0\delta T/2 + j\varphi(t-\tau_D) - j\varphi(t-\tau_D)/2 \cdot} \right. \\
&\cdot \left. e^{-j\varphi(t-\tau_D-\delta T)} + e^{j\varphi(t-\tau_D-\delta T) - j\omega_0\delta T)/2 - j\varphi(t-\tau_D)/2 - j\varphi(t-\tau_D-\delta T)/2} \right], \\
&= \frac{1}{2}E_0 e^{j\left[\omega_0(t-\tau_D-\delta T + \frac{\omega_0\delta T + \varphi(t-\tau_D) + \varphi(t-\tau_D-\delta T)}{2})\right]} \left[e^{j\omega_0\delta T/2 + j\varphi(t-\tau_D)/2 - j\varphi(t-\tau_D-\delta T)/2} + \right. \\
&+ \left. e^{-j\omega_0\delta T/2 - j\varphi(t-\tau_D)/2 - j\varphi(t-\tau_D-\delta T)/2 + j\varphi(t-\tau_D-\delta T)} \right], \qquad (12) \\
& \qquad (13)
\end{aligned}
$$

If we consider $\Psi = \omega_0(t - \tau_D - \delta T) + \frac{[\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)]}{2}$, then the equation (13) can be written as:

$$
E(t) = \frac{E_0 e^{j\Psi}}{2} \left[\frac{e^{j\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)}}{2} + \frac{e^{-j\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)}}{2} \right], \quad (14)
$$

Finally using the following trigonometric identity $e^{j\theta} + e^{-j\theta} = 2\cos\theta$, the equation (14) is written as follows

$$
E(t) = E_0 e^{j\Psi} \cos\left[\frac{j\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)}{2} \right]. \qquad (15)
$$

Thus, the intensity of light at the output of PD can be described as follows:

$$
\begin{aligned}
I(t) &= \|E(t)\|^2 = P_0 \cos^2\left[\frac{\omega_0\delta T + \varphi(t-\tau_D-\delta T) + \varphi(t-\tau_D)}{2} \right]. \quad (16) \\
P_0 &= \|E_0\|^2
\end{aligned}
$$

Appendix B: Dual-delay dynamical model

To facilitate the mathematical treatment of the Eq. (2.28), we fix the offset of phase $\Phi_0 = \pi/2$ and introduce a new variable $u \equiv \int_0^t x(\xi)d\xi$. It leads to two coupled first-order DDEs:

$$\begin{cases} \tau\frac{dx}{dt} = -x - \frac{1}{\theta}u - \frac{\beta}{2}\sin[2(x_{\tau_D} - x_{\tau_D+\delta T})] \\ \frac{du}{dt} = x, \end{cases} \tag{17}$$

where $x_{\tau_D} = (t - \tau_D)$, $x_{\tau_D+\delta T} = x(t - \tau_D - \delta T)$ and $\beta \propto \kappa \mid E_0 \mid^2$ is adjusted through a variable optical attenuator with factor (κ).

Firstly we determine the Hopf bifurcation of the basic state $(x, u) = (0, 0)$, it allows to find pulsating instabilities if they exist. Then, from (17) we formulate the linearized equations, for $x \to 0$, we have to

$$\tau\frac{d^2(dx)}{(dt)^2} = -\frac{dx}{dt} - \frac{1}{\theta}\frac{du}{dt} - \beta(x_{\tau_D} - x_{\tau_D+\delta T}), \qquad \frac{du}{dt} = x, \tag{18}$$

resulting

$$\tau\frac{d^2(dx)}{(dt)^2} + \frac{dx}{dt} + \frac{1}{\theta}\frac{du}{dt} + \beta(x_{\tau_D} - x_{\tau_D+\delta T}) = 0. \tag{19}$$

We look a periodic solution of form $x = a\exp(i\sigma s)$, where a is a constant and $s = t/\delta T$ is time normalized to the small delay. Then Eq. (19) can be rewrite

$$-\tau a\frac{\sigma^2}{\delta T^2}e^{i\sigma t/\delta T} + \frac{ai\sigma}{\delta T}e^{i\sigma t/\delta T} - \frac{1}{\theta}ae^{i\sigma t/\delta T} + \beta\left[\frac{i\sigma}{\delta T}e^{i\sigma x_{\tau_D}/\delta T} - a\frac{i\sigma}{\delta T}e^{i\sigma x_{\tau_D+\delta T}/\delta T}\right] = 0,$$

$$-\frac{\tau\sigma^2}{\delta T^2} + \frac{i\sigma}{\delta T} + \frac{1}{\theta} + \beta\frac{i\sigma}{\delta T}e^{-i\sigma\tau_D/\delta T\left(1-e^{-i\sigma}\right)} = 0. \tag{20}$$

Proposing new variables as follows:

$$\varepsilon_1 \equiv \frac{\delta T}{\theta}, \qquad \varepsilon_2 \equiv \frac{\tau}{\delta T}, \quad \text{and} \quad \zeta \equiv \frac{\tau_D}{\delta T}. \tag{21}$$

Substituting the variables of Eq. (21) in Eq. (20), the characteristic equation can be described by:

$$(1 + i\varepsilon_2\sigma)i\sigma + \varepsilon_1 + i\sigma\beta[e^{-i\sigma\zeta} - e^{-i\sigma(\zeta+1)}] = 0. \tag{22}$$

If σ satisfies the characteristic equation, leading a nontrivial solution. For this we use the values of experimental parameters θ, τ, T and δT to determine $\varepsilon_1 \simeq 7.5 \times 10^{-8}$, $\varepsilon_2 \simeq 706 \times 10^{-3}$, and $\zeta \pm 157.58$. The small values of ε_1 and ε_2 suggest the elimination of the terms multiplying in Eq. (22).

$$\begin{aligned}
i\sigma + i\sigma\beta\left[e^{-i\sigma\zeta} - e^{-i\sigma(\zeta+1)}\right] &= 0 \\
\beta\left[e^{-i\sigma(\zeta+1/2-1/2)} - e^{-i\sigma(\zeta+1/2+1/2)}\right] &= -1 \\
\beta\left[e^{-i\sigma(\zeta+1/2)}e^{i\sigma/2} - e^{-i\sigma(\zeta+1/2)}e^{-i\sigma/2}\right] &= -1 \\
\beta e^{-i\sigma(\zeta+1/2)}\left[e^{i\sigma/2} - e^{-i\sigma/2}\right] &= -1 \\
2i\beta\left(\frac{e^{i\sigma/2}e^{-i\sigma/2}}{2i}\right) &= -e^{i\sigma(\zeta+1/2)} \\
-2i\beta\sin\left(\frac{\sigma}{2}\right) &= e^{i\sigma(\zeta+1/2)}
\end{aligned} \tag{23}$$

Applying the formula of Euler to Eq. (23)

$$-2i\beta\sin\left(\frac{\sigma}{2}\right) = \cos\left[\sigma\left(\zeta + \frac{1}{2}\right)\right] + i\sin\left[\sigma\left(\zeta + \frac{1}{2}\right)\right]. \tag{24}$$

Separating the real part and the imaginary part of Eq. (24), we obtain two conditions for β and ρ

$$\cos\left[\sigma\left(\zeta + \frac{1}{2}\right)\right] = 0 \tag{25}$$

$$\sin\left[\sigma\left(\zeta + \frac{1}{2}\right)\right] = -2\beta\sin\left(\frac{\sigma}{2}\right) \tag{26}$$

We concentrate only in the case $\rho > 0$ (the eigenvalue is greater than zero) for analysis of stability of fixed point of Eq. (25), which is satisfied if

$$\sigma_p = \frac{\pi}{2\zeta + 1} + \frac{2p\pi}{2\zeta + 1} = \frac{1}{\zeta + 1/2}\left(\frac{\pi + 2p\pi}{2}\right), \tag{27}$$

replacing $\zeta \equiv \tau_D/\delta T$ in Eq. (27)

$$\sigma_p = \frac{\pi}{\frac{2\tau_D}{\delta} + 1} + \frac{2p\pi}{\frac{2\tau_D}{\delta T} + 1}. \tag{28}$$

We found that Eq. (28) determines the p-discrete frequencies, for τ_D period in units of δT. Integrating this equation in imaginary part described by Eq. (26), we obtain the feedback gain β as

$$\begin{aligned}
\sin\left(\frac{\pi+2p\pi}{2}\right) &= -\beta\sin\left[\left(\frac{\pi}{2\zeta+1}+\frac{2p\pi}{2\zeta+1}\right)\frac{1}{2}\right] \\
(-1)^p &= (-1)2\beta p\sin\left(\frac{\sigma_p}{2}\right) \\
(-1)^{p+1} &= 2\beta p\sin\left(\frac{\sigma_p}{2}\right) \\
\beta_p &= \frac{(-1)^{p+1}}{2\beta p\sin\left(\frac{\sigma_p}{2}\right)},
\end{aligned} \tag{29}$$

if p is a even number, then $p = 2n$, and we obtain from Eq. (26) that $\sin(\sigma/2) < 0$; and if $p = 2n+1$ is a odd number, it leads to $\sin(\sigma/2) > 0$, so

$$\beta_p = \frac{(-1)^{p+1}}{2\beta p\sin\left(\frac{\sigma_p}{2}\right)} > 0. \tag{30}$$

Bibliography

[1] Tanguy Chouard. Legacy of a universal mind. *Nature*, 482(7386):455–455, 2012.

[2] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *Advances in Cryptology—EUROCRYPT 2003*, pages 294–311. Springer, 2003.

[3] Alan Turing, Richard Braithwaite, Geoffrey Jefferson, and Max Newman. Can automatic calculating machines be said to think?(1952). *B. Jack Copeland*, page 487, 2004.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[5] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pretraining help deep learning? *Journal of Machine Learning Research*, 11(Feb):625–660, 2010.

[6] Martin T Hagan, Howard B Demuth, Mark H Beale, et al. *Neural network design*. Pws Pub. Boston, 1996.

[7] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148:34, 2001.

[8] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[9] *PHOCUS*, 2010 (accessed April, 2013). `http://ifisc.uib-csic.es/phocus/`.

[10] Lennert Appeltant, Miguel Cornelles Soriano, Guy Van der Sande, Jan Danckaert, Serge Massar, Joni Dambre, Benjamin Schrauwen, Claudio R Mirasso, and Ingo Fischer. Information processing using a single dynamical node as complex system. *Nature communications*, 2:468, 2011.

[11] Laurent Larger, Miguel C Soriano, Daniel Brunner, Lennert Appeltant, Jose M Gutiérrez, Luis Pesquera, Claudio R Mirasso, and Ingo Fischer. Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Optics express*, 20(3):3241–3249, 2012.

[12] Daniel Brunner, Miguel C Soriano, Claudio R Mirasso, and Ingo Fischer. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature communications*, 4:1364, 2013.

[13] Romain Martinenghi, Sergei Rybalko, Maxime Jacquot, Yanne Kouomou Chembo, and Laurent Larger. Photonic nonlinear transient computing with multiple-delay wavelength dynamics. *Physical review letters*, 108:244101, 2012.

[14] David Verstraeten, Benjamin Schrauwen, Michiel d'Haene, and Dirk Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391–403, 2007.

[15] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[16] Yvan Paquot, Francois Duport, Antoneo Smerieri, Joni Dambre, Benjamin Schrauwen, Marc Haelterman, and Serge Massar. Optoelectronic reservoir computing. *Scientific reports*, 2, 2012.

[17] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[18] Simon Haykin and Neural Network. A comprehensive foundation. *Neural Networks*, 2(2004), 2004.

[19] Y. Bar-Yam. *Dynamics of Complex Systems*. Studies in Nonlinearity. Westview Press, 1997.

[20] Carlos Gershenson. Artificial neural networks for beginners. *arXiv preprint cs/0308031*, 2003.

[21] Teuvo Kohonen. *Self-organization and associative memory*, volume 8. Springer, 2012.

[22] José Ramón Hilera González and VJ Martínez Hernando. Redes neuronales artificiales: fundamentos modelos y aplicaciones. *Madrid: Editorial Alfaomega Ra-Ma*, 2000.

[23] R Gonzalez-Garcia, R Rico-Martinez, and IG Kevrekidis. Identification of distributed parameter systems: A neural net based approach. *Computers & chemical engineering*, 22:S965–S968, 1998.

[24] Timo Sorsa and Heikki N Koivo. Application of artificial neural networks in process fault diagnosis. *Automatica*, 29(4):843–849, 1993.

[25] Kumpati S Narendra and Kannan Parthasarathy. Identification and control of dynamical systems using neural networks. *Neural Networks, IEEE Transactions on*, 1(1):4–27, 1990.

[26] Anthony J Calise, Naira Hovakimyan, and Moshe Idan. Adaptive output feedback control of nonlinear systems using neural networks. *Automatica*, 37(8):1201–1211, 2001.

[27] Jyh-Shing Roger Jang, Chuen-Tsai Sun, and Eiji Mizutani. Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence. 1997.

[28] J Pearce. Sir charles scott sherrington (1857–1952) and the synapse. *Journal of neurology, neurosurgery, and psychiatry*, 75(4):544, 2004.

[29] Lennert Appeltant. *Reservoir Computing based on Delay-dynamical Systems*. PhD thesis, Vrije Universiteit Brussel and Universitat de les Illes Balears, May 2012.

[30] Jan Mycielski. Review: Marvin minsky and seymour papert, perceptrons, an introduction to computational geometry. *Bull. Amer. Math. Soc.*, 78(1):12–15, 01 1972.

[31] Kevin L Priddy and Paul E Keller. *Artificial neural networks: an introduction*, volume 68. SPIE Press, 2005.

[32] Romain Martinenghi. *Démostration opto-électronique du concept de calculateur neuromorphique par reservoir computing*. PhD thesis, École doctorale sciences pour l'ingénieur et microtechniques, Université de Franche–comté, Décember 2013.

[33] James A Freeman and David M Skapura. *Algorithms, Applications, and Programming Techniques*. Citeseer, 1991.

[34] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

[35] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[36] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088–3092, 1984.

[37] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5:3, 1988.

[38] Frank Rosenblatt. The perceptron-a perceiving and recognizing automaton. *Cornell Aeronautical Lab*, 1957.

[39] Manfred M Fischer and Arthur Getis. *Recent developments in spatial analysis: Spatial statistics, behavioural modelling and computational intelligence*. Springer Science & Business Media, 1997.

[40] Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.

[41] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[42] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

[43] Wolfgang Maass, Prashant Joshi, and Eduardo D Sontag. Computational aspects of feedback in neural circuits. *PLoS Comput Biol*, 3(1):e165, 2007.

[44] Dean V Buonomano and Wolfgang Maass. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2):113–125, 2009.

[45] François Duport, Bendix Schneider, Anteo Smerieri, Marc Haelterman, and Serge Massar. All-optical reservoir computing. *Optics express*, 20(20):22783–22795, 2012.

[46] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nature communications*, 5, 2014.

[47] Herbert Jaeger. Echo state network. *Scholarpedia*, 2(9):2330, 2007.

[48] David Verstraeten. *Reservoir Computing: computation with dynamical systems*. PhD thesis, Ghent University, 2009.

[49] Herbert Jaeger. *Short term memory in echo state networks*. GMD-Forschungszentrum Informationstechnik, 2001.

[50] Herbert Jaeger. *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach*. GMD-Forschungszentrum Informationstechnik, 2002.

[51] Herbert Jaeger. Reservoir riddles: Suggestions for echo state network research. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pages 1460–1462. IEEE, 2005.

[52] Herbert Jaeger. Discovering multiscale dynamical features with hierarchical echo state networks. *Jacobs University Bremen, Tech. Rep*, 2007.

[53] Thomas Natschläger and Wolfgang Maass. Spiking neurons and the induction of finite state machines. *Theoretical Computer Science*, 287(1):251–265, 2002.

[54] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Computational models for generic cortical microcircuits. *Computational neuroscience: A comprehensive approach*, 18:575, 2004.

[55] Thomas Natschläger, Wolfgang Maass, and Henry Markram. The" liquid computer": A novel strategy for real-time computing on time series. *Special issue on Foundations of Information Processing of TELEMATIK*, 8(LNMC-ARTICLE-2002-005):39–43, 2002.

[56] Misha Rabinovich, Ramon Huerta, and Gilles Laurent. Transient dynamics for neural processing. *Science*, 321(5885):48–50, 2008.

[57] Barbara Hammer, Benjamin Schrauwen, and Jochen J Steil. Recent advances in efficient learning of recurrent networks. In *ESANN*, 2009.

[58] David Verstraeten, Benjamin Schrauwen, and Dirk Stroobandt. Reservoir-based techniques for speech recognition. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 1050–1053. IEEE, 2006.

[59] David Verstraeten, Benjamin Schrauwen, Dirk Stroobandt, and Jan Van Campenhout. Isolated word recognition with the liquid state machine: a case study. *Information Processing Letters*, 95(6):521–528, 2005.

[60] M.L. Minsky and S. Papert. *Perceptrons; an Introduction to Computational Geometry*. MIT Press, 1969.

[61] James L McClelland and David E Rumelhart. *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press, 1989.

[62] Satish Kumar. *Neural networks: a classroom approach*. Tata McGraw-Hill Education, 2004.

[63] Prashant Joshi and Wolfgang Maass. Movement generation and control with generic neural microcircuits. In *Biologically inspired approaches to advanced information technology*, pages 258–273. Springer, 2004.

[64] Harald Burgsteiner. Training networks of biological realistic spiking neurons for real-time robot control. In *Proceedings of the 9th international conference on engineering applications of neural networks, Lille, France*, pages 129–136, 2005.

[65] Wolfgang Maass, Robert Legenstein, and Henry Markram. A new approach towards vision suggested by biologically realistic neural microcircuit models. In *Biologically Motivated Computer Vision*, pages 282–293. Springer, 2002.

[66] Joachim Hertzberg, Herbert Jaeger, and Frank Schönherr. Learning to ground fact symbols in behavior-based robots. In *ECAI*, pages 708–712, 2002.

[67] Eric A Antonelo, Benjamin Schrauwen, and Dirk Stroobandt. Event detection and localization for small mobile robots using reservoir computing. *Neural Networks*, 21(6):862–871, 2008.

[68] Eric Antonelo, Benjamin Schrauwen, and Dirk Stroobandt. Modeling multiple autonomous robot behaviors and behavior switching with a single reservoir computing network. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 1843–1848. IEEE, 2008.

[69] Eric A Antonelo and Benjamin Schrauwen. Unsupervised learning in reservoir computing: Modeling hippocampal place cells for small mobile robots. In *Artificial Neural Networks–ICANN 2009*, pages 747–756. Springer, 2009.

[70] Fabian Triefenbach, Azarakhsh Jalalvand, Benjamin Schrauwen, and Jean-Pierre Martens. Phoneme recognition with large hierarchical reservoirs. In *Advances in neural information processing systems*, pages 2307–2315, 2010.

[71] Mark D Skowronski and John G Harris. Automatic speech recognition using a predictive echo state network classifier. *Neural networks*, 20(3):414–423, 2007.

[72] Miguel C Soriano, S Ortín, Daniel Brunner, Laurent Larger, Claudio R Mirasso, Ingo Fischer, and Luıs Pesquera. Optoelectronic reservoir computing: tackling noise-induced performance degradation. *Optics express*, 21(1):12–20, 2013.

[73] Henri Poincaré and R Magini. Les méthodes nouvelles de la mécanique céleste. *Il Nuovo Cimento (1895-1900)*, 10(1):128–130, 1899.

[74] Thomas Erneux. *Applied delay differential equations*, volume 3. Springer Science & Business Media, 2009.

[75] MIT. *System Dynamics*, 1997 (accessed February 3, 2014). `http://web.mit.edu/sd-intro/`.

[76] Henning U Voss and Jürgen Kurths. Analysis of economic delayed-feedback dynamics. In *Modelling and Forecasting Financial Data*, pages 327–349. Springer, 2002.

[77] Kestutis Pyragas. Continuous control of chaos by self-controlling feedback. *Physics letters A*, 170(6):421–428, 1992.

[78] Guanrong Chen and Xiaoning Dong. From chaos to order—perspectives and methodologies in controlling chaotic nonlinear dynamical systems. *International Journal of Bifurcation and Chaos*, 3(06):1363–1409, 1993.

[79] Y Chembo Kouomou, Pere Colet, Laurent Larger, and Nicolas Gastaud. Chaotic breathers in delayed electro-optical systems. *Physical review letters*, 95(20):203903, 2005.

[80] Laurent Larger. Complexity in electro-optic delay dynamics: modelling, design and applications. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(1999):20120464, 2013.

[81] Laurent Larger, Romain Martinenghi, Maxime Jacquot, and Yanne Kouomou Chembo. Complex photonic nonlinear delay dynamics for high performance signal and information processing. In *International conference Optics & Photonics Taïwan (OPTIC 2013)*, 2013.

[82] Miguel C Soriano, Jordi García-Ojalvo, Claudio R Mirasso, and Ingo Fischer. Complex photonics: Dynamics and applications of delay-coupled semiconductors lasers. *Reviews of Modern Physics*, 85(1):421, 2013.

[83] Eckehard Schöll and Heinz Georg Schuster. *Handbook of chaos control*. John Wiley & Sons, 2008.

[84] Roman Lavrov, Michael Peil, Maxime Jacquot, Laurent Larger, Vladimir Udaltsov, and John Dudley. Electro-optic delay oscillator with nonlocal nonlinearity: Optical phase dynamics, chaos, and synchronization. *Physical Review E*, 80(2):026207, 2009.

[85] Lionel Weicker, Thomas Erneux, Maxime Jacquot, Yanne Chembo, and Laurent Larger. Crenelated fast oscillatory outputs of a two-delay electro-optic oscillator. *Physical Review E*, 85(2):026206, 2012.

[86] Richard Ernest Bellman and Kenneth L Cooke. Differential-difference equations. 1963.

[87] Jack K Hale and Sjoerd M Verduyn Lunel. *Introduction to functional differential equations*, volume 99. Springer Science & Business Media, 2013.

[88] V Kolmanovskii and A Myshkis. *Introduction to the theory and applications of functional differential equations*, volume 463. Springer Science & Business Media, 2013.

[89] Yang Kuang. *Delay differential equations: with applications in population dynamics*. Academic Press, 1993.

[90] M Le Berre, E Ressayre, A Tallet, HM Gibbs, DL Kaplan, and MH Rose. Conjecture on the dimensions of chaotic attractors of delayed-feedback dynamical systems. *Physical Review A*, 35(9):4020, 1987.

[91] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

[92] Otto E Rössler. An equation for continuous chaos. *Physics Letters A*, 57(5):397–398, 1976.

[93] Takashi Matsumoto, Leon O Chua, and Motomasa Komuro. The double scroll. *Circuits and Systems, IEEE Transactions on*, 32(8):797–818, 1985.

[94] John Briggs and F David Peat. *Turbulent mirror: An illustrated guide to chaos theory and the science of wholeness*. HarperCollins Publishers, 1989.

[95] Kensuke Ikeda and Kenji Matsumoto. High-dimensional chaotic behavior in systems with time-delayed feedback. *Physica D: Nonlinear Phenomena*, 29(1):223–235, 1987.

[96] Valery Iustinovich Oseledec. A multiplicative ergodic theorem. lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc*, 19(2):197–231, 1968.

[97] Miguel A Andrade, Juan C Nuño, Federico Morán, Francisco Montero, and George J Mpitsos. Complex dynamics of a catalytic network having faulty replication into error-species. *Physica D: Nonlinear Phenomena*, 63(1):21–40, 1993.

[98] G. Jumarie. *Maximum Entropy, Information Without Probability and Complex Fractals: Classical and Quantum Approach*. Fundamental Theories of Physics. Springer Netherlands, 2013.

[99] M. Girvan. *Relating the Kolmogorov Entropy of Chaotic Dynamics with the Physical Entropy of Statistical Mechanics*. Massachusetts Institute of Technology, Department of Physics, 1999.

[100] Kensuke Ikeda. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Optics communications*, 30(2):257–261, 1979.

[101] Michael C Mackey and Leon Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, 1977.

[102] J Doyne Farmer. Chaotic attractors of an infinite-dimensional dynamical system. *Physica D: Nonlinear Phenomena*, 4(3):366–393, 1982.

[103] Martin Andre Agnes Fiers, Thomas Van Vaerenbergh, Francis Wyffels, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(2):344–355, 2014.

[104] R.E. Ziemer and W.H. Tranter. *Principles of Communications, 7th Edition:*. John Wiley & Sons, 2014.

[105] FT Arecchi, G Giacomelli, A Lapucci, and R Meucci. Two-dimensional representation of a delayed dynamical system. *Physical Review A*, 45(7):R4225, 1992.

[106] Martine Le Berre, Elisabeth Ressayre, Andrée Tallet, and Hyatt M Gibbs. High-dimension chaotic attractors of a nonlinear ring cavity. *Physical review letters*, 56(4):274, 1986.

[107] Laurent Larger, Bogdan Penkovsky, and Yuri Maistrenko. Virtual chimera states for delayed-feedback systems. *Physical review letters*, 111(5):054103, 2013.

[108] L. Appeltant *et al.* Information processing using a single dynamical node as complex system. *Internal reporting of activities: PHOCUS project*, 2010.

[109] X Steve Yao and Lute Maleki. High frequency optical subcarrier generator. *Electronics Letters*, 30(18):1525–1526, 1994.

[110] Apostolos Argyris, Dimitris Syvridis, Laurent Larger, Valerio Annovazzi-Lodi, Pere Colet, Ingo Fischer, Jordi García-Ojalvo, Claudio R Mirasso, Luis Pesquera, and K Alan Shore. Chaos-based communications at high bit rates using commercial fibre-optic links. *Nature*, 438(7066):343–346, 2005.

[111] Laurent Larger and John M Dudley. Nonlinear dynamics: Optoelectronic chaos. *Nature*, 465(7294):41–42, 2010.

[112] Roman Lavrov, Maxime Jacquot, and Laurent Larger. Nonlocal nonlinear electro-optic phase dynamics demonstrating 10 gb/s chaos communications. *Quantum Electronics, IEEE Journal of*, 46(10):1430–1435, 2010.

[113] Michael Peil, Maxime Jacquot, Yanne Kouomou Chembo, Laurent Larger, and Thomas Erneux. Routes to chaos and multiple time scale dynamics in broadband bandpass nonlinear delay electro-optic oscillators. *Physical Review E*, 79(2):026208, 2009.

[114] Roman Lavrov. *Chaos électro-optique par modulation differéntielle de phase: application à la sécurisation de la couche physique des transmissions optiques à 10 Gb/s*. PhD thesis, École doctorale sciences pour l'ingénieur et microtechniques, Université de Franche–comté, Août 2010.

[115] Texas Instruments-Developed. 46-word speaker-dependent isolated word corpus (ti46). *NIST Speech Disc*, pages 7–1, 1991.

[116] R Gary Leonard and George Doddington. Tidigits. *Linguistic Data Consortium, Philadelphia*, 1993.

[117] Azarakhsh Jalalvand, Fabian Triefenbach, David Verstraeten, and Jean-Pierre Martens. Connected digit recognition by means of reservoir computing. In *INTERSPEECH*, pages 1725–1728, 2011.

[118] Hans-Günter Hirsch and David Pearce. The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*, 2000.

[119] Egbert De Boer. Mechanics of the cochlea: modeling efforts. In *The cochlea*, pages 258–317. Springer, 1996.

[120] Cephalicvein. *Anatomy Of Ear*, july 22, 2016 (accessed September 5, 2016). `http://http://cephalicvein.com/2016/07/anatomy-of-ear/`.

[121] Richard F Lyon and Carver Mead. An analog electronic cochlea. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7):1119–1134, 1988.

[122] Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980.

[123] X Valero, P Farré, and F Alías. Comparison of machine learning techniques for the automatic recognition of soundscapes. In *Proceedings of Forum Acusticum*, pages 2037–2042, 2011.

[124] Jonathan T Foote. Content-based retrieval of music and audio. In *Voice, Video, and Data Communications*, pages 138–147. International Society for Optics and Photonics, 1997.

[125] ETSI Speech Processing. Transmission and quality aspects (stq); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms. *ETSI ES*, 202:050, 2002.

[126] Romain Martinenghi, Antonio Baylon Fuentes, Maxime Jacquot, Yanne K Chembo, and Laurent Larger. Towards optoelectronic architectures for integrated neuromorphic computers. In *SPIE OPTO*, pages 89890K–89890K. International Society for Optics and Photonics, 2014.

[127] Konstantin Hicke, Miguel Angel Escalona-Morán, Daniel Brunner, Miguel Cornelles Soriano, Ingo Fischer, and Claudio R Mirasso. Information processing using transient dynamics of semiconductor lasers subject to delayed feedback. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(4):1501610–1501610, 2013.

[128] Ali Rodan and Peter Tino. Minimum complexity echo state network. *IEEE transactions on neural networks*, 22(1):131–144, 2011.

[129] Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, and Joe Woelfel. Sphinx-4: A flexible open source framework for speech recognition. 2004.

[130] Romain Martinenghi, A. Baylon-Fuentes, Xiaole Fang, Maxime Jacquot, Yanne Kouomou Chembo, and Laurent Larger. Optoelectronic nonlinear transient computing with multiple delays. In *Conference on lasers and electro-optics - International quantum electronics conference (CLEO/EUROPE - IQEC)*, Munich, Germany., may 2013.

Document réalisé avec LaTeX et :
le style LaTeX pour Thèse de Doctorat créé par S. Galland — http://www.multiagent.fr/ThesisStyle
la collection de paquets `tex-upmethodology` — http://www.arakhne.org/tex-upmethodology/

Résumé :

Aujourd'hui, la plupart des ordinateurs sont encore basés sur des concepts développés il y a plus de 60 ans par Alan Turing et John von Neumann. Cependant, ces ordinateurs numériques ont déjà commencé à atteindre certaines limites physiques via la technologie de la microélectronique au silicium (dissipation, vitesse, limites d'intégration, consommation d'énergie). Des approches alternatives, plus puissantes, plus efficaces et moins consommatrices d'énergie, constituent depuis plusieurs années un enjeu scientifique majeur. Beaucoup de ces approches s'inspirent naturellement du cerveau humain, dont les principes opérationnels sont encore loin d'être compris. Au début des années 2000, la communauté scientifique s'est aperçue qu'une modification du réseau neuronal récurrent (RNN), plus simple et maintenant appelée Reservoir Computing (RC), est parfois plus efficace pour certaines fonctionnalités, et est un nouveau paradigme de calcul qui s'inspire du cerveau. Sa structure est assez semblable aux concepts classiques de RNN, présentant généralement trois parties: une couche d'entrée pour injecter l'information dans un système dynamique non-linéaire (Write-In), une seconde couche où l'information d'entrée est projetée dans un espace de grande dimension (appelé réservoir dynamique) et une couche de sortie à partir de laquelle les informations traitées sont extraites par une fonction dite de lecture-sortie. Dans l'approche RC, la procédure d'apprentissage est effectuée uniquement dans la couche de sortie, tandis que la couche d'entrée et la couche réservoir sont fixées de manière aléatoire, ce qui constitue l'originalité principale du RC par rapport aux méthodes RNN. Cette fonctionnalité permet d'obtenir plus d'efficacité, de rapidité, de convergence d'apprentissage, et permet une mise en œuvre expérimentale. Cette thèse de doctorat a pour objectifs d'implémenter pour la première fois le RC photoniques en utilisant des dispositifs de télécommunication. Notre mise en œuvre expérimentale est basée sur un système dynamique non linéaire à retard, qui repose sur un oscillateur électro-optique (EO) avec une modulation de phase différentielle. Cet oscillateur EO a été largement étudié dans le contexte de la cryptographie optique du chaos. La dynamique présentée par de tels systèmes est en effet exploitée pour développer des comportements complexes dans un espace de phase à dimension infinie, et des analogies avec la dynamique spatio-temporelle (tels que les réseaux neuronaux) sont également trouvés dans la littérature. De telles particularités des systèmes à retard ont conforté l'idée de remplacer le RNN traditionnel (généralement difficile à concevoir technologiquement) par une architecture à retard d'EO non linéaire. Afin d'évaluer la puissance de calcul de notre approche RC, nous avons mis en œuvre deux tests de reconnaissance de chiffres parlés (tests de classification) à partir d'une base de données standard en intelligence artificielle (TI-46 et AURORA-2), et nous avons obtenu des performances très proches de l'état de l'art tout en établissant un nouvel état de l'art en ce qui concerne la vitesse de classification. Notre approche RC photonique nous a en effet permis de traiter environ 1 million de mots par seconde, améliorant la vitesse de traitement de l'information d'un facteur supérieur à ~3.

Mots-clés : Reservoir computing, Réseau neuronal récurrent, Dynamique non linéaire à retard, Oscillateur electro-optique

Abstract:

Nowadays most of computers are still based on concepts developed more than 60 years ago by Alan Turing and John von Neumann. However, these digital computers have already begun to reach certain physical limits of their implementation via silicon microelectronics technology (dissipation, speed, integration limits, energy consumption). Alternative approaches, more powerful, more efficient and with less consume of energy, have constituted a major scientific issue for several years. Many of these approaches naturally attempt to get inspiration for the human brain, whose operating principles are still far from being understood. In this line of research, a surprising variation of recurrent neural network (RNN), simpler, and also even sometimes more efficient for features or processing cases, has appeared in the early 2000s, now known as Reservoir Computing (RC), which is currently emerging new brain-inspired computational paradigm. Its structure is quite similar to the classical RNN computing concepts, exhibiting generally three parts: an input layer to inject the information into a nonlinear dynamical system (Write-In), a second layer where the input information is projected in a space of high dimension called dynamical reservoir and an output layer from which the processed information is extracted through a so-called Read-Out function. In RC approach the learning procedure is performed in the output layer only, while the input and reservoir layer are randomly fixed, being the main originality of RC compared to the RNN methods. This feature allows to get more efficiency, rapidity and a learning convergence, as well as to provide an experimental implementation solution. This PhD thesis is dedicated to one of the first photonic RC implementation using telecommunication devices. Our experimental implementation is based on a nonlinear delayed dynamical system, which relies on an electro-optic (EO) oscillator with a differential phase modulation. This EO oscillator was extensively studied in the context of the optical chaos cryptography. Dynamics exhibited by such systems are indeed known to develop complex behaviors in an infinite dimensional phase space, and analogies with space-time dynamics (as neural network ones are a kind of) are also found in the literature. Such peculiarities of delay systems supported the idea of replacing the traditional RNN (usually difficult to design technologically) by a nonlinear EO delay architecture. In order to evaluate the computational power of our RC approach, we implement two spoken digit recognition tests (classification tests) taken from a standard databases in artificial intelligence TI-46 and AURORA- 2, obtaining results very close to state-of-the-art performances and establishing state-of-the-art in classification speed. Our photonic RC approach allowed us to process around of 1 million of words per second, improving the information processing speed by a factor ~3.

Keywords: Reservoir computing, Recurrent neural network, Delay dynamics, Nonlinear delay electro-optic oscillator

■ École doctorale SPIM 1 rue Claude Goudimel F - 25030 Besançon cedex

■ tél. +33 (0)3 81 66 66 02 ■ **ed-spim@univ-fcomte.fr** ■ www.ed-spim.univ-fcomte.fr

UFC UNIVERSITÉ DE FRANCHE-COMTÉ