Spatial recurrent neural networks : from electro-optical towards all-optical

Louis Andreoli

THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE FRANCHE-COMTÉ

École doctorale n° 37

Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Optique et Photonique

par

LOUIS ANDRÉOLI

Spatial Recurrent Neural Networks: From Electro-Optical Towards All-Optical

Réseaux de neurones spatiaux récurrents : de l'électro-optique au tout-optique

Thèse présentée et soutenue à Besançon, le 17 mars 2021

Composition du Jury :

MARIA-PILAR BERNAL	Directrice de Recherche CNRS, Institut Femto-st	Présidente du jury
CLAUDIO CONTI	Professeur, Sapienza University of Rome	Rapporteur
ANTONIO HURTADO	Professeur, University of Strathclyde	Rapporteur
SYLVAIN GIGAN	Professeur, Sorbonne Université	Examinateur
MAXIME JACQUOT	Professeur, Université Bourgogne Franche-Comté	Directeur de thèse
DANIEL BRUNNER	Chargé de Recherche CNRS, Institut Femto-st	Directeur de thèse

N° | X | X | X |

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

Title: Spatial Recurrent Neural Networks: From Electro-Optical Towards All-Optical

Keywords: Optical recurrent neural networks, analogue neural networks, reservoir computing, hardware learning, quantum dots micropillar lasers, micropillar arrays, optical pumping

Abstract:

In recent years, artificial intelligence and more particularly neural networks have played a major role in our technological societies. Nevertheless, neural networks still remain emulated by traditional computers, resulting in challenging problems such as parallelization, energy efficiency and potentially speed. A change of paradigm is desirable but implementing neural networks in hardware is a non-trivial challenge. One promising avenue is optical neural networks, potentially avoiding parallelization bottlenecks. We experimentally implemented the first system where nodes and connections are implemented in parallel hardware, and we exploit our novel system for unique experiments. The electro-optical neural network is composed of up to 961 neurons. Connections between neurons are implemented in fully parallel by optical diffraction, offering highly advantageous scalability. The network's architecture is based on the principle of reservoir computing, therefore the number of readout weights to be trained is simply the nodes number. The trainable output layer is realized by a digital micro-mirror device, which, once trained, has the advantage of being fully parallel, passive, and therefore without bandwidth limitation and energy consumption. Using a *Greedy* evolutionary learning, we find that the learning excellently converges. Furthermore, it appears to possess a conveniently convex-like cost-function and demonstrates exceptional scalability of the learning effort with system size. Finally, the fundamental bases of an all-optical neural network have been studied. Neurons will be implemented in an array of quantum dots micropillar lasers optically coupled. An in-depth study of the optical pumping of micropillar lasers, ultimately allowing a highly scalable system, has revealed a pumping efficiency high enough to realistically operate around a hundred neurons simultaneously.

Titre : Réseaux de neurones spatiaux récurrents : de l'électro-optique au tout-optique

Mots-clés : Réseaux de neurones optiques, micropiliers laser à quantum dots, pompage optique

Résumé :

Depuis les années 2010, l'intelligence artificielle et plus particulièrement les réseaux de neurones jouent un rôle important dans nos sociétés. Cependant, les réseaux neuronaux sont émulés par ordinateurs, entraînant des problèmes complexes tels que la parallélisation, l'efficacité énergétique et potentiellement la vitesse des systèmes. Un changement de paradigme est alors souhaitable, toutefois la mise en œuvre de réseaux de neurones *hardware* reste un réel défi. Les réseaux neuronaux optiques constituent une piste très prometteuse, permettant potentiellement de résoudre la difficulté de la parallélisation. Nous avons expérimentalement conçu le premier réseau spatial où les neurones et les connexions sont matériellement et parallèlement réalisés, nous permettant ainsi d'effectuer des expériences singulières. Ce réseau électro-optique est composé de 961 neurones et leurs connexions neuronales réalisées par diffraction optique offrent au système une grande capacité d'évolution. L'architecture du réseau est basée sur le concept de *reservoir computing*, l'apprentissage consiste alors à optimiser uniquement les poids des connexions de sortie. La couche de sortie, réalisée par un dispositif à micro-miroir numérique, présente, une fois entraînée, les avantages, d'être totalement parallèle, passive, sans limitation de bande passante et par conséquent non-consommatrice d'énergie. En utilisant une version optimisée d'un apprentissage évolutif, nous avons constaté que celui-ci converge parfaitement et son temps d'entraînement est linéaire avec la taille du réseau. Enfin, les principes fondamentaux d'un réseau de neurones tout optique ont été étudiés. Les neurones sont réalisés grâce à une matrice de micropiliers laser à *quantum dots*. Une étude quantitative de leur pompage optique, offrant à terme une grande capacité d'évolution au réseau, a révélé une efficacité suffisante pour faire fonctionner de manière réaliste une centaine de neurones simultanément.

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon, France

"Aucun problème ne peut être résolu sans changer le niveau de conscience qui l'a engendré."

Albert Einstein (1879-1955)

"Tout le chemin de la vie, c'est de passer de l'ignorance à la connaissance, de l'obscurité à la lumière, de l'inaccompli à l'accompli, de l'inconscience à la conscience, de la peur à l'amour."

Frédéric Lenoir (né en 1962)

REMERCIEMENTS

Bien que cela soit une évidence, je préfère l'énoncer explicitement : toute cette production de savoir n'aurait jamais vu le jour sans mon directeur de thèse. C'est pourquoi, Daniel, je te remercie pour la confiance que tu m'as accordée tout au long de cette thèse expérimentale où tout était à construire. Professionnellement tu as été un mentor pour moi, je te remercie de m'avoir enseigné, guidé, éclairé, inspiré, mais également fait grandir scientifiquement en m'apprenant une méthodologie, une rigueur, une exigence, et à croire en mes propres résultats. Enfin, je suis très reconnaissant pour la qualité de ton encadrement, ton écoute réelle et ton humanité sans égo.

J'aimerais mentionner que mes résultats sont également le fruit d'un travail collaboratif, j'ai ainsi bénéficié de l'aide et l'expertise de nombreux doctorants, ingénieurs et chercheurs. Pour cela, je souhaite remercier tous les membres du département Optique et plus particulièrement Luca Furfaro, Cyril Billet, Jean-Marc Merolla, Luc Froehly, Jean-Charles Beugnot, Javier Porte et mon co-directeur de thèse Maxime Jacquot. Un grand merci à Emmanuel Dordor, sans qui un grand nombre de pièces mécaniques essentielles n'aurait jamais vu le jour. Pour terminer, je remercie Valérie Fauvez, Aline Chagrot et Sarah Djaouti pour leur travail de gestion et leur investissement indispensables pour le fonctionnement du département.

Je tiens à remercier Claudio Conti et Antonio Hurtado d'avoir évalué ce manuscrit de thèse en tant que rapporteurs. Merci également à Sylvain Gigan d'avoir accepté d'être examinateur et à Maria-Pilar Bernal d'avoir présidé mon jury de thèse.

Je n'oublie pas tous les compagnons de labo qui ont embelli et enrichi ces années. Je pense notamment aux nombreuses personnes qui ont fait vivre et briller la Pause Café, entité unique, réputée, respectée et pratiquement légendaire dans les esprits. Benoit, Sheler, Coraline, Nicolas, Vincent, Clément, Étienne, Franck, Sébastien, Luc, Alexis et tous les autres, merci à vous. Également un immense merci à Florent, Rémi, Solveig, Cyril et Aurore pour leurs précieuses discussions et échanges qui m'ont spirituellement et humainement grandi.

Je remercie et salue aussi tous les membres de l'équipe, Daniel, Maxime, Sheler, Xavi, Nadya, Vova, Anas, Ria, Adrià, Johnny, mais aussi Stéphane, Luc et les personnes de passage. Merci encore à Daniel, mais aussi Xavi, Maxime et Aurore pour vos relectures et contributions à ce manuscrit.

Un dernier mot pour ma partenaire de vie, Aurore, qui a ensoleillé ces années et qui continue à semer joie et bonheur le long de nos chemins entremêlés. Merci !

ACRONYMS AND ABBREVIATIONS LIST

AND: Logical AND function.

ANN: Analogue Neural Network.

BS: Beam Splitter.

CO2: Carbon dioxide.

CPU: Central Processing Unit.

DET: Detector.

DMD: Digital Micro-mirror Device.

DOE: Diffractive Optical Element.

ESN: Echo State Network.

FALSE: Logical FALSE function.

GPU: Graphics Processing Unit.

GRU: Gated Recurrent Unit.

LED: Light-Emitting Diode.

LP: Long Pass.

LSM: Liquid State Machine.

LSTM: Long Short-Term Memory.

Mag: Magnification.

MO: Microscope Objective.

NA: Numerical Aperture.

ND: Neutral Density.

NMSE: Normalized Mean Squared Error.

NOT: Logical NOT function.

Ops: Operations per second.

OR: Logical OR function.

OSA: Optical Spectrum Analyser.

PBS: Polarizing Beam Splitter.

QD: Quantum Dot.

QDML: Quantum Dots Micropillar Laser.

QDMLA: Quantum Dots Micropillar Lasers Array.

ReLU: Rectified Linear Unit.

RNN: Recurrent Neural Network.

SLM: Spatial Light Modulator.

SNR: Signal to Noise Ratio.

SP: Short Pass.

TPU: Tensor Processing Unit.

TRUE: Logical TRUE function.

USD: United States Dollar.

XOR: Logical XOR function.

CONTENTS

INTRODUCTION

For a few years now, artificial neural networks have deeply penetrated our societies, and artificial intelligence has been used in sciences, transport, finance, games, medicine, agriculture, security, media, management, insurance or engineering [1]. Neural networks have even revolutionized human-machine interaction thanks to speech and hand-writing recognition [2, 3]. However, while studies of the first neural networks began over 70 years ago [4], and theoretical concepts were mostly known since the late 1980s, it is only recently that a sharp acceleration of applications has been observed [5]. In addition to advances in research, this success is the result of two important stages which are closely linked to neural networks. One is the democratization of digital information during the first decade of the 21st century and the advent of social networks which, thanks to human work [6, 7], enabled labelling gigantic databases used to train neural networks. The other is the rise of affordable high performance computing, the medium on which neural networks are mainly emulated. The AlphaGo algorithm's [8] Go game victory against Lee Sedol in 2016 [9] was a major milestone which demonstrated to the world the exceptional advance of neural network research. To reach such a level of expertise in a game renowned for its complexity due to the extraordinarily high number of combinations [10], the algorithm had to train with tens of thousands of games led by expert players, then play millions of games against itself. The training of such neural networks is therefore often carried out using the computing power of data centers.

However, the classical electronic computer is conceptually far from the true topology of neural networks. While the computing power of neural networks is the result of massive and parallel connectivity between neurons, current computers are based on the von Neumann architecture, and operate mostly in series. The size of current transistors is already close to physical limits [11], their number per processor and therefore the computing power increases only with great difficulty [12]. Since fifteen years computing power advances are mainly the result of the increase of the number of processors used [12] and new hardware aimed at improving the parallelization of electronics [13]. One of the main challenges for future hardware artificial neural networks is therefore to achieve highly parallel connectivity.

While electronic computing begins to reach its limits, we will see in Chapter 1 that optical computing research begins to offer serious avenues for future generations of neural networks. Indeed, unlike the electron, the photon has no electrical charge, *i.e.* does not suffer from capacitive and inductive energy dissipation. These intrinsic characteristics of light are already used in telecommunications, where information is optically encoded and exchanged within a single communication channel, such as an optical fiber, with record data rates [14]. The possibility for optically modulated information to be exchanged temporally and spatially in parallel with minimum of crosstalk allows a strong parallelization of the neuron network's connectivity.

Our research goal is to create neural networks with certain similarities to the inspiration from biological neural networks. The basic principle is to use physical phenomena to

avoid emulating a neural network on a computer, but to instead build a physical system that is intrinsically the neural network. It is through this return to physical processes that we leave the digital world and move to an analogue system, arguably as is the case in a brain. To best pursue these ideals and realize neural networks with parallel connectivity, we use optics as means of information propagation. Our neural networks are based on the concept of reservoir computing which is a subclass of recurrent neural networks. This computational framework is well suited for classification tasks and for the processing of temporal information [15]. From a practical point of view, the reservoir computing architecture is simple, which greatly helps to implement it optically.

We will present in Chapter 2 a complete spatial electro-optical recurrent neural network [16]. Its theoretical support, its realization, its adjustment, its learning and its testing will be presented throughout. Neurons are electro-optical, however the entire neural network was designed from the start to serve as a prototype of a more advanced yet equally more delicate system, closer to the previously defined goals, which will be presented in Chapter 4. The spatial distribution of neurons allows increasing the network size without reducing the system bandwidth. The neural network's connectivity is realized through optical diffraction which allows real parallelization and high scalability [17]. The output layer is also implemented using a matrix of digital micro-mirrors, allowing programming of fully parallelized readout weight. The learning algorithm is designed to be an integral part of the neural network, it is not remote learning but operates in real time.

We will discuss in Chapter 3 the features of our hardware embodied neural network. Due to its analogue nature, it is prone to noise. We will study the different noise sources, their specific characteristics and their impact on the system. In particular, we will measure the consistency of our neural network, *i.e.* its ability to similarly process identical information despite the presence of noise. Learning will be investigated in detail, including its convergence and the major role of noise. The results obtained will be derived from general principles which have to be taken into account in all analogue neural networks. Finally, we measured the learning duration which is found to grow linearly with the number of artificial neurons, allowing significant scalability of our neural networks.

We will present in Chapter 4 the foundation of an all-optical neural network. By replacing electro-optical neurons with optical elements, the system presented in Chapter 2 will then become fully optical. The fundamental concepts, the architecture and the neurons coupling, however, will remain the same. Artificial neurons will be an array of quantum dots micropillar lasers [18]. In order to promote a scalable system, semiconductor lasers are optically pumped. Indeed, to create a neural network where each artificial neuron is independent, electrical pumping would require an electrical connection for each laser, which then becomes an extremely difficult challenge for a large-scale system. We will study in depth the optical pumping according to two different mechanisms. In addition to obtaining a pumping efficiency high enough to create a future network of a hundred optical neurons, our quantitative study characterizes laser internal pumping mechanisms in detail and allows us to propose modifications to the lasers promising an energy efficiency improvement between one or two orders of magnitude.

Finally, we will conclude by a summary of the research presented throughout this manuscript. By continuing with our fundamental guidelines and relying on the results obtained, we will propose perspectives targeted in particular to improve the electro-optic neural network but also the next steps necessary to realize the all-optical neural network.

1

ARTIFICIAL NEURAL NETWORKS

Artificial intelligence was initiated by classical philosophers, already Gottfried Wilhelm Leibniz conceived his *calculus ratiocinator*, a philosophical system that tried to describe the human thought process as the mechanical manipulation of symbols. Turing's machine and Church's lambda-calculus suggested that any form of mathematical reasoning could be mechanized [19–21]. This reflection materialized with the invention of the programmable computer in the 1940s [22].

The traditional computer is based on a von Neumann architecture. Schematically described in Fig. 1.1(a), the processing unit which performs the operations is physically separated from the system's memory. The processing unit contains both, the data and the computer program. This type of computer follows a finite series of operations previously defined in order to solve a given problem. In such a context, one generally speaks of the program as an algorithm. As illustrated in Fig. 1.1(b), the programmer completely designs the possible scenarios that the program needs to perform, in other words the entire problem and its solution have to be known. Once programmed, the computer is then an automaton of formidable efficiency.

However many problems cannot be formulated in this form for the simple reason that the way to obtain the required solution is unknown. Another strategy has been developed with expert systems [23]. The principle is no longer to impose a previously defined list of operations on the computer, but to define database and a rule-based system [24]. Solutions will then emerge on their own, without previous definition. This type of system has been effective in formal areas where the rules are explicit enough to be programmed in the inference engine.

Nevertheless, these two approaches are experiencing almost insurmountable challenges in the face of problems which seem trivial to us humans, such as pattern recognition, understanding a language, making a prediction from obscure data. There is no logical sequence of instructions to address these problems. Establishing rules is complicated because they are either too general for a highly specific problem, or too specific and it could approximately take as many rules as there are situations. The traditional computing therefore reaches its limits in the face of problems that are difficult to translate in terms of logical operations.

While traditional computers are very effective in certain situations, their failure to tackle problems that are easily solved by humans rightly suggests that they are powerful machines, but they lack intelligence. Besides, let's go back for a moment to the origin of the word "intelligence". It originates from the Latin "intelligentia" which means "faculty of perceiving, understanding, intelligence". This term is derived from the Greek "intellĕgĕre" (to

discern, understand) composed of the prefix "inter" (between) and the verb "lĕgĕre" (to pick, choose, read). Etymologically, intelligence consists in *making a choice, a selection*. However, a traditional computer just follows instructions or general rules previously defined. Its logic is therefore deductive, *i.e.* going from a general affirmation to a particular conclusion. Human beings seem to be good at inductive reasoning. The principle is to seek general laws from the observation of particular facts, potentially on a probabilistic basis [25]. This ability to generalize allows learning by example. Thus, the recognition of a shape or the prediction of a series of patterns are easy tasks for the human being.

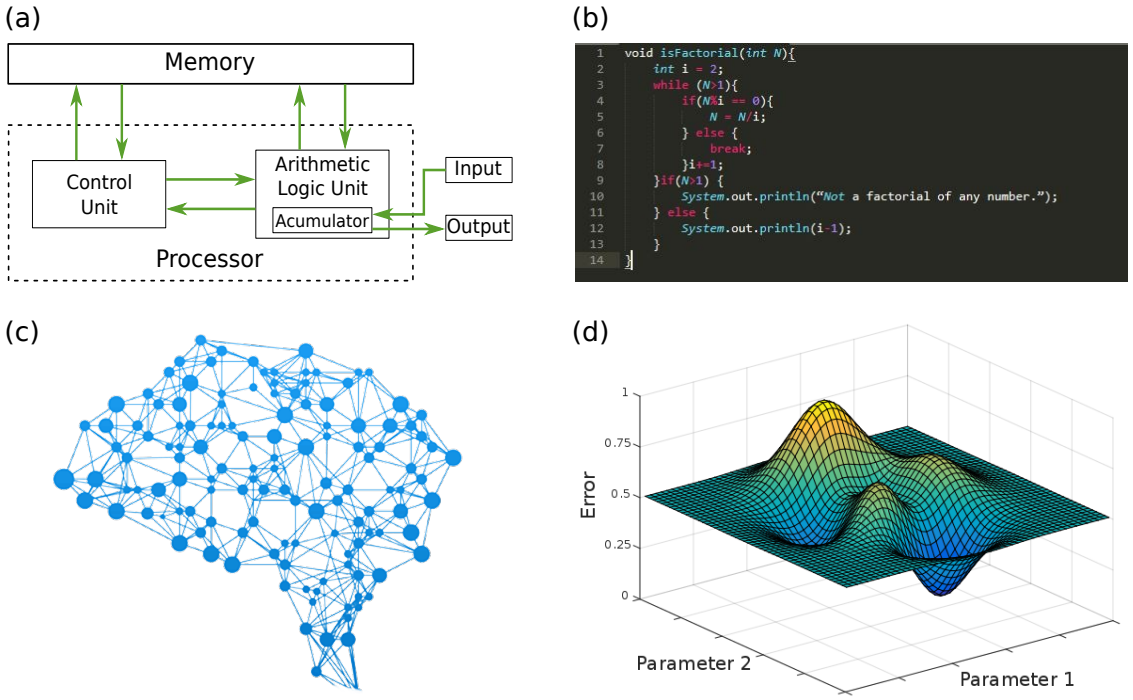

Figure 1.1: (a) Schematic of the von Neumann architecture on which most current computers are based. By sequentially using a pre-established and finite sequence of unambiguous instructions stored in memory, the algorithm shown in panel (b) solves a problem. Panel (c) schematically illustrates a neural network architecture composed of neurons massively interconnected. By optimizing their synaptic weights, the system learns to solve a problem, thereby minimizing the error represented in an error-landscape panel (d).

Taking inspiration from biology, a new type of system emerged. Illustrated in Fig. 1.1(c), its principle is to schematically reproduce a neural network (NN) using a large number of elementary nonlinear functions, called neurons, and to link them together. By adjusting the connections between artificial neurons, it is then possible to solve non-trivial tasks by optimizing connections weights to minimize its error. The system therefore is guided by the topology of an error-landscape, illustrated schematically in Fig. 1.1(d).

During this chapter we will introduce the concept of dynamical systems followed by an overview of different types of neural networks. We will then specify the concept of reservoir computing which is the architecture at the heart of the following chapters. Finally, we will discuss a paradigm shift in the implementation of neural networks. We will study the advantages of building such computing architectures in hardware, once again drawing inspiration from some of a biological neural network's fundamental aspects.

1.1/ INTRODUCTION TO DYNAMICAL SYSTEMS

The study of dynamical systems is a very broad field of research, covering many disciplines such as mathematics, physics [26, 27], chemistry, biology [28] and even economics [29] and engineering [30]. Since many neural networks are governed by the temporal exchange of information, they can also be studied and described from a dynamical point of view [31, 32]. We will limit ourselves here to a small introduction to the field, giving the reader some guidelines. For a much more in-depth study, in particular about the nonlinear dynamics, we refer you to the book *Nonlinear dynamics and chaos* by Steven H. Strogatz [33] or to the excellent non-fiction book *Chaos: Making A New Science* by James Gleick [34].

A dynamic system is a system in which a function describes the evolution of itself. It can be the evolution of a chemical reaction over time, the movement of planets in the solar system or the evolution of a computer's memory under the action of a computer program. A dynamic system has two important properties:

- Causality: the future of a dynamic system depends only on present or past phenomena.
- Determinism: for a given present initial state, there will correspond at each subsequent instant a single future state.

We distinguish continuous time dynamic systems (described by differential equations) from discrete time dynamic systems (described by maps). In the case of neural networks, we are generally more interested in the second type. A discrete dynamic system is generally (but not necessarily) defined by a bijection $\Phi : \Gamma \to \Gamma$ of the phase space Γ on itself. A phase space is an abstract space whose coordinates are the dynamic variables of the system under study and where all possible states of the system are represented.

The bijection operates as follows: given an initial condition x_0 of the system's state, the first and second next states are

$$x_1 = \Phi(x_0), \tag{1.1}$$

then

$$\begin{aligned} x_2 &= \Phi(x_1) \\ &= \Phi(\Phi(x_0)) \\ &= (\Phi \circ \Phi)(x_0) \\ &= \Phi^2(x_0). \end{aligned} \tag{1.2}$$

By continuity, the n-th state is given by

$$\begin{aligned} x_n &= \Phi(x_{n-1}) \\ &= \Phi^n(x_0). \end{aligned} \tag{1.3}$$

To go back in time, simply reverse the function, which is always possible for a bijection. This feature is widely used for learning neural networks. The best approximation function is directly related to the inverse of the correct response to a given input.

We distinguish linear dynamic systems from nonlinear dynamic systems. In the case of a linear system, the sum of two solutions is also a solution (superposition principle). The solutions of a linear equation form a vector space, which allows the use of linear

algebra and considerably simplifies the analysis. Due to the nonlinearity, the analysis of nonlinear dynamical systems is generally more difficult, and it is often in this context that the artificial neural networks take place. However, the nonlinearity gives the opportunity to neurons to potentially be general approximators.

1.2/ OVERVIEW OF ARTIFICIAL NEURAL NETWORKS

1.2.1/ ARTIFICIAL NEURON

An artificial neuron is a mathematical and computational representation of a biological neuron. The first model was proposed by Warren McCulloch and Walter Pitts in 1943 [35]. As illustrated in Fig. 1.2(a), the formal neuron generally has several x_n inputs and one y output which correspond respectively to the dendrites and to the axon starting points of the biological neuron, which certainly is much more complex than the simple mathematical abstraction as illustrated in panel (b). The excitatory and inhibitory actions of biological synapses are represented by numerical coefficients W_n associated with the inputs.

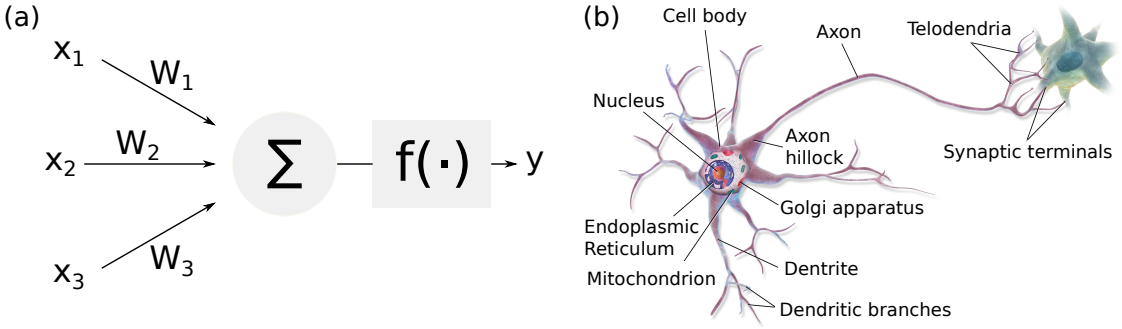

Figure 1.2: Diagram of neurons. (a) An artificial neuron performing a nonlinear transformation $f(\cdot)$ of the sum of the inputs weighted x_n by the synaptic weights W_n. This mathematical representation is a simplified model of biological neurons, illustrated in panel (b). *https://en.wikipedia.org/wiki/Neuron*

The artificial neuron is the elementary unit of artificial neural network in which it is associated with fellow neurons to calculate arbitrarily complex functions, used for various applications in artificial intelligence. Mathematically, the artificial neuron is a function with several variables and real values. In its simplest version, it calculates the weighted sum of the inputs received, then applies to this value an activation function generally nonlinear $f(\cdot)$. The final value obtained is the output of the neuron. Originally for the McCulloch and Pitts neuron [35], an activation function is the Heaviside function $H(\cdot)$. The output of the neuron is therefore

$$y = H(\sum_{j=1}^{m} W_j x_j - W_0) \,, \tag{1.4}$$

with W_0 the activation threshold of the neuron. If the sum $\sum_{j=1}^{m} W_j x_j$ is larger than W_0 then the output is 1, otherwise it is zero.

Such an artificial neuron can calculate logical functions. Considering the case of a neuron with two inputs x_1 and x_2 provided with weights W_1 and W_2 both positive. The output is

then $y = H(W_1x_1 + W_2x_2 - W_0)$. Simplify further by fixing $x_1 = x_2 = 1$ and we define the symbol $\&$ as the logical AND function. By choosing the value of the activation threshold W_0, it is then possible to calculate all the logic function summarized in Table 1.1.

Functions	Conditions
OR	$0 < W_0 < W_1 \ \& \ W_2$
AND	$W_1 \ \& \ W_2 < W_0$
TRUE	$W_0 < 0$
FALSE	$W_1 \ \& \ W_2 < W_0$
x_1 identity	$0 < W_2 < W_0 < W_1$
x_2 identity	$0 < W_1 < W_0 < W_2$

Table 1.1: Summary of computable logical functions with an artificial neuron having two inputs $x_1 = x_2 = 1$ weighted respectively by W_1, W_2 and a Heaviside function as activation function. The symbol $\&$ is the logical AND function.

The development of artificial neural networks has led to the development of other activation functions. The motivation is not necessarily to get closer to the biological neurons but to improve the performance of the network, reduce the number of neurons or facilitate learning. The activation functions are chosen according to their characteristics, the main ones are as follows:

Non-linearity: When a function is non-linear, a 2-layer neural network can be considered as a universal function approximator [36].

Anywhere differentiable: This property allows you to create optimizations based on gradients [37].

Monotonic: When the function is monotonic, the error surface associated with a monolayer model is always as convex [38].

Density at 0 ($f(x \approx 0) \approx 0$): These functions allow to quickly learn by initializing weights randomly. If the function does not converge to identity at 0, then special care must be taken during the initialization of the weights [39].

Commonly used functions are the sigmoid function, the hyperbolic tangent or the rectified linear unit (ReLU). We will replace the notion of a neuron by the name nonlinear node or simply node of a network.

1.2.2/ FEEDFORWARD NEURAL NETWORKS

A feedforward neural networks is an artificial neural network where the connections between the nodes propagate only from the input to the output. The first architecture, named perceptron, was introduced by Frank Rosenblatt in 1957 [40]. The perceptron is an artificial neuron which is equipped with a learning rule to automatically optimize the synaptic weights in order to solve a linearly separable problem.

Neural networks subsequently became more complex, becoming multi-layered perceptrons. Such systems typically have multiple inputs and outputs which are in general linear functions. Located between these linear in and output layers are the layers comprising

nonlinear nodes, often referred to as hidden layers. Hidden layers are composed of a variable number of nonlinear nodes, and if a neural network contains several hidden layers then one typically refers to a deep neural network [5]. It is important to remember that the nodes are always connected in a feedforward way. Each node in one layer has directed connections to the nodes of the subsequent layer. Connectivity between nodes can be full, in which case each node is connected to all nodes of the next layer [41]. The scheme in Fig. 1.3 depicts a deep neural networks with two inputs and one output. Three layers of four nodes compose the hidden layer.

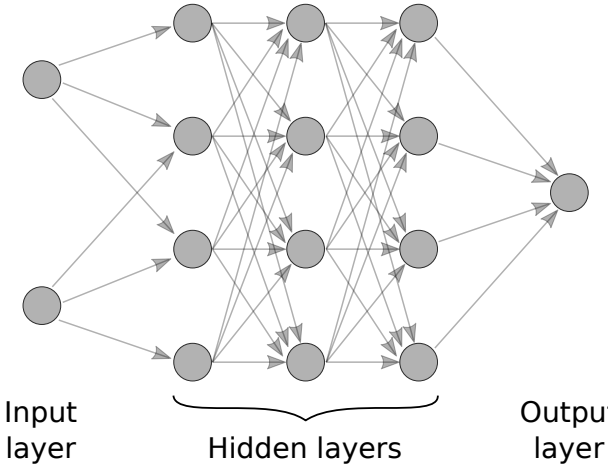

Figure 1.3: Graph representation of a feedforward neural networks with two inputs and one output. The hidden layer is composed of three layers of four nonlinear nodes.

Such an architecture featuring nonlinear nodes can now solve all logical functions. Let us take again the XOR function example which is not monotonic. A classical way to implement an XOR function is its decomposition into more basic logical function. We have $A \oplus B = (A + B)\,\overline{AB} = A\overline{B} + \overline{A}B$, where $+$ and $\oplus$ are the logical OR and XOR functions, respectively. Since a single node can solve the functions AND, OR, TRUE and FALSE, the combination of several nodes can then solve the XOR problem.

Neural network's and comparable systems, however, address such problems in a different manner: their many nonlinear nodes create a high dimensional representation of the original problem. The XOR function can be seen as a 2D problem. As shown in Fig. 1.4(a), the XOR function cannot be linearly solved, *i.e.* separate the solution from the remainder by a single line. However the nonlinear functions of the nodes in the hidden layer can increase the dimension of the problem. Already for passing from 2D to 3D (see panel (b)) allows one to identify a possible plane, therefore a linear function, which discriminates the XOR's two possible solutions. In this illustration the horizontal plane is the original input data, which the additional third dimension would be provided by one or the superposition of several nonlinear nodes of the neural network.

This textbook example illustrates one of the main functions of a neural network. Nonlinearities make it possible to increase the dimension of the original input data inside the representation space of the neural network's nodes, before being separated by a hyperplane. Feedforward neural networks are therefore very efficient at categorizing data.

Other types of NNs exist inside the group of feedforward neural networks, in particular convolutional networks, which are very efficient in image processing [42]. Their nodes of a layer are the convolution of a certain number of nodes of the preceding layer.

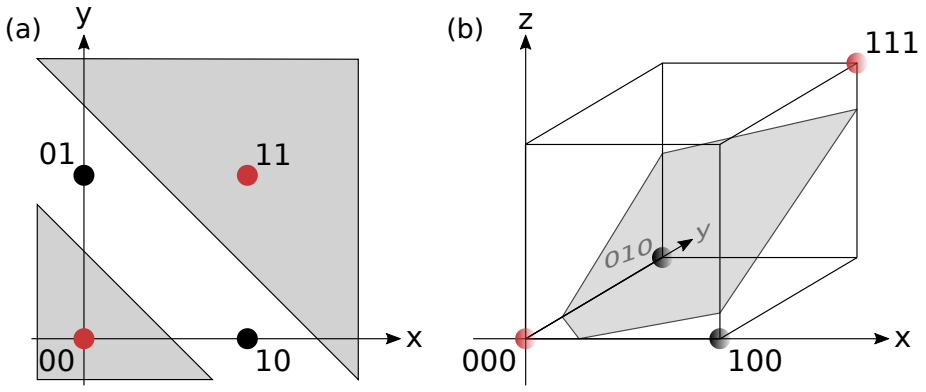

Figure 1.4: Different representation of the logical XOR function problem. (a) The solution is not linearly separate. (b) The XOR problem is extended to a third dimension in which it is now linearly solvable: the red points are above the plane drawn in gray, the black points below. Neural networks with nonlinear nodes can be interpreted in as extremely large dimensionality increasing systems.

1.2.3/ RECURRENT NEURAL NETWORKS

Recurrent neural networks (RNNs) have the particularity of having at least one cyclic connection. Historically two major types of architecture exist: Elman's [43] and Jordan's [44] RNNs. Presented respectively in Fig. 1.5(a) and (b), the first connects the output and the input of the hidden layer, the second connects the output of the network to the input of the hidden layer. This characteristic creates feedback with a certain delay which corresponds to memory within the system. It can be shown that such networks featuring memory are equivalent to a Turing machine [45].

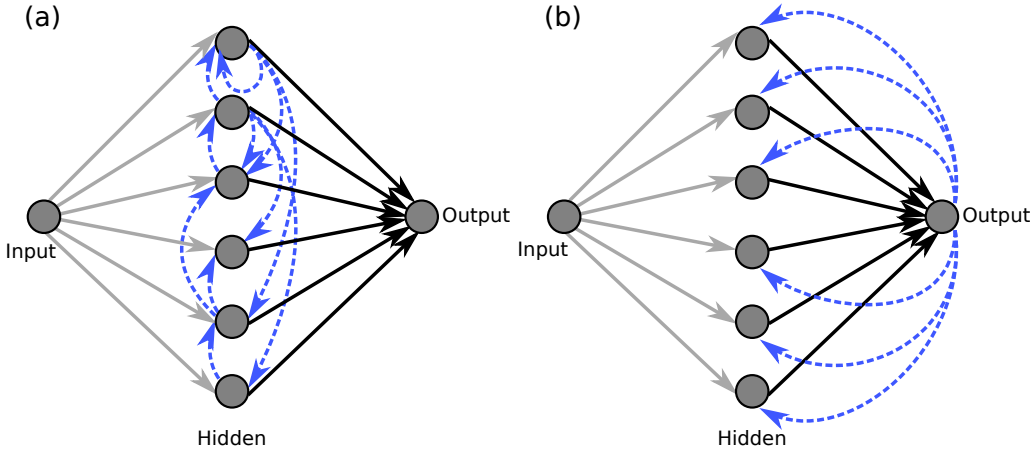

Figure 1.5: Architecture of Elman and Jordan RNN represented in panel (a) and (b), respectively. The output of the hidden nodes or of the neural networks modify the inputs of the nodes via the recurrent connections in blue.

There are multiple variations of RNN architectures, but let's take the simplest as an example: a single hidden recurrent layer h composed of one nonlinear node with a self-feedback [46]. It is architecturally speaking possible to temporally unfold this network to obtain several layers constituting a single node. As illustrated in Fig. 1.6, the layer $h(n)$

processes the information $x(n)$ at each time n but taking into account the past treatment $h(n-1)$. The higher the feedback strength β, the more memory the system has and will depend on older and older events. If the feedback strength is unity or higher, the neural network unfolded is infinite and develops chaotic behaviors.

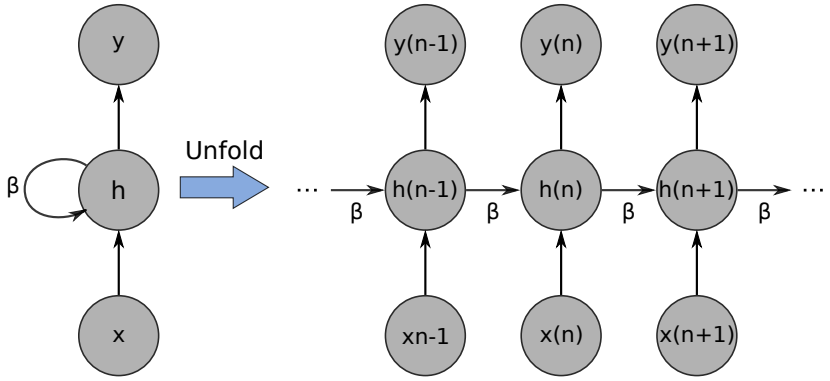

Figure 1.6: Diagram of a RNN. The hidden recurrent layer $h(n)$ is composed of a single node connecting its output to its input with a feedback strength β. On the right, the temporally unfolded version of the structure. The result at time n depends on the input information $x(n)$ but also on the previous results.

RNNs can then use their internal state (memory) to process variable length sequences of inputs [1]. This memory makes RNNs particularly effective in solving temporal problems such as automatic recognition of speech [47, 48] or the future prediction of a sequential information [49].

1.2.4/ LEARNING

Learning in neural networks consists in optimizing the connection weights between nodes to solve a given task with minimum error. The learning is inductive, *i.e.* it is supposed to generalize a characteristic from particular examples. Two general types of learning exist, supervised or unsupervised.

In the first case, labelled data, *i.e.* known data, is sent to the neural network. The principle is to minimize the error between the network output and the correct response. The learning can be translated as a function with multiple parameters. Represented graphically by an error-landscape as illustrated in Fig. 1.7, the training consists in finding the optimal set of parameters, *i.e.* finding an error minimum. Optimization techniques based on gradient descent are typically employed. The principle is to iteratively take steps proportional to the negative of the gradient of the error function at the current point. One of the major difficulties is to find the global minimum and not to get stuck in a local minimum, in particular with a large error.

A method now widely used for its effective results is called backpropagation [50]. This method consists in minimizing the error gradient for each node in a neural network, from the last layer to the first. One of the problems in particular with the deep neural networks [51] or RNNs [52] is the vanishing gradient problem [53]. When the gradient of the error is very small, propagation to an earlier layer is compromised. New architectures such as Long Short-Term Memory (LSTM) [54] or Gated Recurrent Unit (GRU) [55] were subsequently developed to work around this problem.

One of the drawbacks of supervised learning is that the examples must be known. As training usually requires a large number of iterations, large databases must be created for which the right results is labelled for a sufficiently large number of examples.

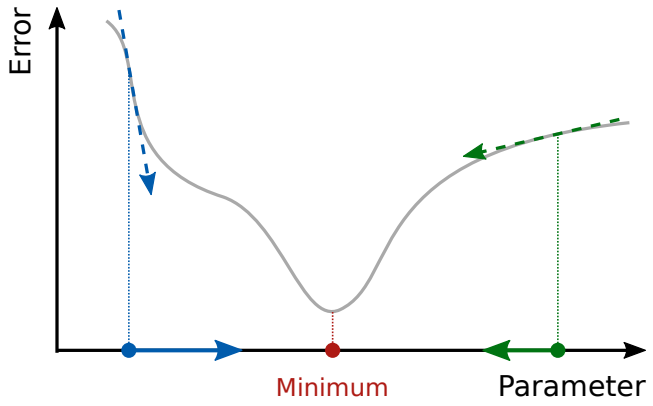

Figure 1.7: One-dimensional error-landscape. To find a local minimum of a function using gradient descent, we take steps proportional to the negative of the gradient of the function at the current point.

Unsupervised learning refers to the case where data is not labelled. It is therefore a matter of discovering the structures underlying this unlabelled data. Many algorithms exist based on different methods such as principal component, cluster analysis or anomaly detection. One of the first techniques used is the so called Hebb's rule [56]. This theory is used both as a hypothesis in neuroscience and as a concept in neural networks. This rule suggests that when two neurons are excited together, they create or reinforce the connection between them. To prevent the connection weight from becoming infinite, the method was subsequently modified [57]. However, since the data is not labelled, it is impossible for the algorithm to calculate a success score with certainty. In general, systems based on unsupervised learning allow more complex tasks to be performed than supervised learning systems, but they can also be more unpredictable.

An important problem in system optimization and learning is overfitting. Since learning is inductive, it is essential that the result remains generally valid for the entire class of the particular problem, and not only for the input examples. During a supervised learning process, the system will increasingly optimize system parameters in order to maximally reduce the error. However, care must be taken that the system is not too specialized in the data series used. An overfitted model is a statistical model that contains more parameters than can be justified by the data [58]. The black line in Fig. 1.8 is a general model which differentiates the red and blue data. The green line is an overfitted model, its error for this particular case is extremely low but it loses its general characteristics. To limit this problem, it is necessary to use an adequate number of nodes and hidden layers. However, these parameters are difficult to determine in advance. Several pragmatic techniques exist. The best known is cross validation. This method consists of separating the data into two categories. The learning is performed on the first data, called learning data, then the system is tested with the other data called testing data. Typically test and training data are reshuffled, until the entire data has once served as test and training data. The error obtained during learning must never be lower than the result during the test. Otherwise it means that the ability to generalize characteristics is lost.

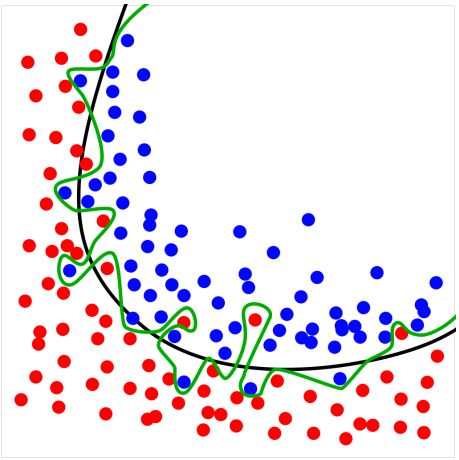

Figure 1.8: While the black line represents a general model to differentiate between red and blue data, the green line is an overfitted model. *https://en.wikipedia.org/wiki/Overfitting*.

1.3/ RESERVOIR COMPUTING

Reservoir computing is a unified computing concept that has emerged from several independent RNN concepts, in particular the Echo State Networks (ESNs) [59] and Liquid State Machines (LSMs) [60]. Reservoir computing is above all well suited to the processing of temporal or sequential information [15].

1.3.1/ ARCHITECTURE

As illustrated in Fig. 1.9, the architecture of a reservoir computer is composed of three parts. The first is the input layer, the second the recurrent hidden layer is called reservoir and finally an output layer.

The input layer is used to distribute information inside the reservoir. Input connections can be random but always remain constant in time. The reservoir is a network of nonlinear nodes connected to each other. The connectivity can be random, total or partial, but also remains constant in time. Such a reservoir projects the input information into a large dimensional space. Importantly, the dimensions of a dynamical system can be seen as the number of its initial condition. These depend on the number of nodes in the reservoir. The output layer, called readout, performs a linear transformation of the reservoir state.

The main feature of reservoir computing is that the input and reservoir weights remain constant throughout the computation. The learning will be performed by optimizing only the output weights. While in a conventional RNN the number of weights involved in the learning process increases nonlinearly with the network's size, in reservoir computing this increases linearly with the number of nodes. In comparison, this simplicity of training makes it possible to drastically reduce the cost of computing training, which is a major advantage of reservoir computing [61] .

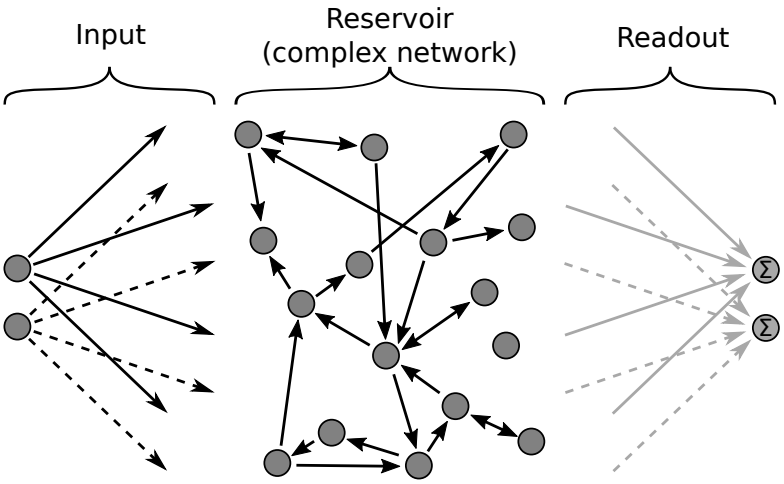

Figure 1.9: Reservoir computing architecture. The input layer injects information into the reservoir, which expands the dimensionality of the input's representation. The weights of input layer and reservoir can be random but stay constant in time. The readout weights are optimize to best approximate the required output and hence computation.

1.3.2/ OPERATING PRINCIPLE

The reservoir computing model uses a reservoir based on a discrete-time n RNN. The input layer composed of Q linear nodes denoted $q = 1, 2, \cdots, Q \in \mathbb{N}$, receives the input information defined by the vector $\mathbf{u}$. The input signal is injected inside the reservoir thanks to the injection matrix $\mathbf{W}^{\text{inj}}$ composed of Q rows and N columns, with N the number of nodes of the reservoir. Each node i of the reservoir receives at each instant n the input signal $W^{\text{inj}}_{q,i} u_q(n)$.

The weights of the reservoir are defined by the coupling matrix $\mathbf{W}^{\text{R}}$ of dimension $N \times N$. In the case where the output of the network is not reconnected to the system, the temporal evolution of the neuronal state $\mathbf{x}$ within the reservoir is then

$$\mathbf{x}(n+1) = \mathbf{f}\left(\mathbf{W}^{\text{inj}}\mathbf{u}(n+1) + \mathbf{W}^{\text{R}}\mathbf{x}(n)\right), \tag{1.5}$$

where the function $\mathbf{f}(\cdot)$ represents an element-wise activation function.

Equation 1.5 represents a non-autonomous dynamical system forced by the input signal $\mathbf{u}(n)$. The output $\mathbf{y}^{\text{out}}$ of the reservoir is a linear combination of the state of the reservoir as

$$\mathbf{y}^{\text{out}}(n+1) = \mathbf{W}^{\text{out}} \cdot \mathbf{x}(n+1), \tag{1.6}$$

where $\mathbf{W}^{\text{out}}$ is the readout matrix of dimension N.

Only the readout weights are optimized during learning. By knowing the desired output of the system and by calculating the state of the reservoir, we therefore obtain

$$\mathbf{W}^{\text{out}} = \mathbf{y}^{\text{out}}(n+1)(\mathbf{x}(n+1))^{-1} \tag{1.7}$$

for all n in the training data. Learning is therefore a linear matrix inversion problem, which is convex and hence admits a unique optimal solution [62]. Remember that the reservoir is unchanged during learning which optimizes the system to solve a specific problem. It is therefore possible to train several outputs to solve several problems simultaneously with a single reservoir.

1.3.3/ RESERVOIR PROPERTIES

The role of the reservoir is to map the sequential input information into a large space so that the output can linearly approximate the desired result. The reservoir must also have the following properties:

Separation property: Separation property is the capacity of the reservoir to distinguish two neighboring internal states from two different input information.

Approximation property: Approximation property is the capacity to distinguish and transform different internal states in order to approximate any function [60].

Fading-memory property: The recurrences within the reservoir give memory capacity to the system. However, the system must be able to return to a steady state in the presence of no external input. This means that the system memory is linked to the transient state of the network dynamics [63]. This property therefore implies that the reservoir must have a finite memory in order to process the information locally.

1.4/ NEURAL NETWORK IMPLEMENTATIONS

We discussed in the introduction some limitations of traditional computing. Thus, neural networks have emerged and are now effective in solving problems previously known to be difficult. However, the bio inspiration has mainly focused on the software aspect, the majority of neural networks are emulated on traditional computers. In this part we will explore the physical implementation of neural network systems by mimicking once again the advantageous characteristics found in biology.

1.4.1/ HUMAN BRAIN AS BENCHMARK

The human brain is able to solve complex and varied problems, has short and very long term memories and is able to make choices depending on the situation. The biological brain, while widely studied, is far from having exposed all its mysteries [64]. Nevertheless let us examine some main physical characteristics which we can use as humble inspiration for concepts.

On average the human brain weighs 1.3 kg and has a volume of 1200 cm^3 [65]. It is composed of around 170 billion cells, of which 86 billion are neurons [66]. Each neuron can form between five and sixty thousand synapses [67]. The nerve impulse which propagate with 1 m/s can be accelerated to 100 m/s in the presence of myelin around the axons. The thickness of the cerebral cortex is between 1 and 4.5 mm and its surface area is around 2000 cm^2 [68]. In order to find space inside the human skull, the cortex is folded by furrows or fissures, sinuous and of varying depth. Neurons are connected by physical connections, fast, very largely parallelized and their functioning is autonomous, not clocked. Neurons are arranged in 2D while their connections evolve in a three-dimensional space. In addition, to optimize space, the connections have a fractal architecture [69]. The white matter of the brain (mainly composed of myelinated axons) makes up about half of the total brain volume [70]. The connections linking neurons to each other and thus creating

an information exchange network, therefore seem to occupy a very important part in the functioning of the brain. Measurements in a small cube of brain tissue revealed a storage capacity of about 4.7 bits of information per synapse [71], suggesting memory on the order of 10^{15} bytes at the whole brain scale. The brain represents about 20 % of the energy consumption of the human body [72], or about 25 W, which makes the system extremely efficient.

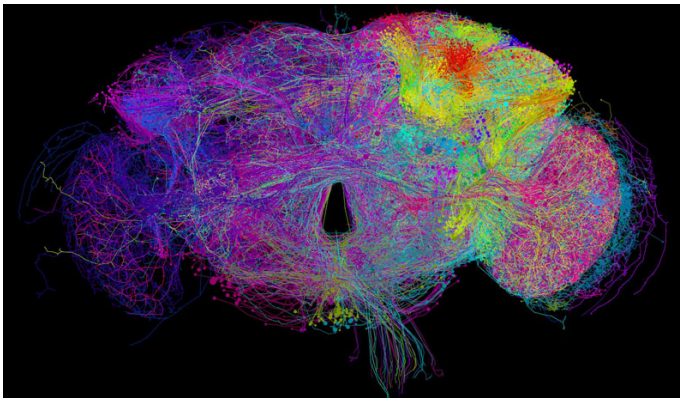

Figure 1.10: Volume reconstruction of a complete adult fly brain [73]. Neighboring neurons were marked with identical color.

1.4.2/ DIGITAL NEURAL NETWORK: REVOLUTION AND LIMITATIONS

Neural networks are now performing tasks that were only a dream ten years earlier. Their shape detection performance is equivalent to that of a human, images are described automatically [74], the translation of foreign languages has also been automatized [75], there are early beginnings of autonomous cars [76], the computer surpasses humans in many games such as the game of go, chess, shogi [77] but also games with hidden information like poker [78] or Starcraft II [79]. However, this revolution is not only the result of recent advances in neural networks. The constitution of large databases were also decisive for the learning phases. But more importantly, neural networks are mostly implemented on traditional computers. The major advances are based in part on other computer technology and electronics. Indeed, research in computer science, now over seventy years old, allowed developments which massively democratized this technology. The standardization of manufacturing processes in particular by CMOS technology, the accumulated know-how and the growing number of people working in this sector have made it possible to produce extremely high-performance systems. It is therefore in part the mastery of computer science that allowed the field of neural networking to make a leap forward. However, the improvement of the current neural networks accuracy depends on the availability of exceptionally large computational resources that require equally great power consumption. As illustrated in Fig. 1.11, neural networks training in data centers is expensive, requires a lot of energy and has a large carbon footprint.

To understand this, let's go back to the von Neumann architecture, which is still widely used in computing and far from an artificial neural network's (see Figures 1.1(a) and 1.10). Computation is performed synchronously and the communications between memory and the Central Processing Unit (CPU) are mainly in series. This limiting factor has been fully understood and the parallelization of calculations is now taken into account

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$\text{Transformer}_{base}$	P100x8	1415.78	12	27	26	\$41–\$140
Transformer_{big}	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT_{base}	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT_{base}	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

Figure 1.11: Estimated cost in terms of CO2 emissions (lbs) and cloud compute cost (USD) for training neural networks models. Power and carbon footprint are omitted for TPUs due to lack of public data [80].

through the multiplication of CPU cores or the use of Graphics Processing Unit (GPU) as well as moving memory closer to the processor. Recent research aimed at optimizing the parallelization of digital calculations is working on new hardware chips, such as the development of a Tensor Processing Unit (TPU) [13], which is an artificial intelligence accelerator application-specific integrated circuit.

Nevertheless, intrinsic constraints to electronics limit the development of neural networks. It is now accepted that the semiconductor industry is abandoning the pursuit of Moore's Law [11]. In 2020 the semiconductor industry will undertake the transition to the 5 nm manufacturing process, notably based on Fin Field-Effect Transistor and Gate-All-Around Field-Effect Transistor technologies which are non-planar transistors. Even though the number of transistors continues to increase, as indicated by the orange data in Fig. 1.12, the physical limits of computation based on the current architectures are apparently reached, see blue green and red data in the same Figure. The computing power increase over the past fifteen years is mainly due to the increase in the number of cores in processors (black data).

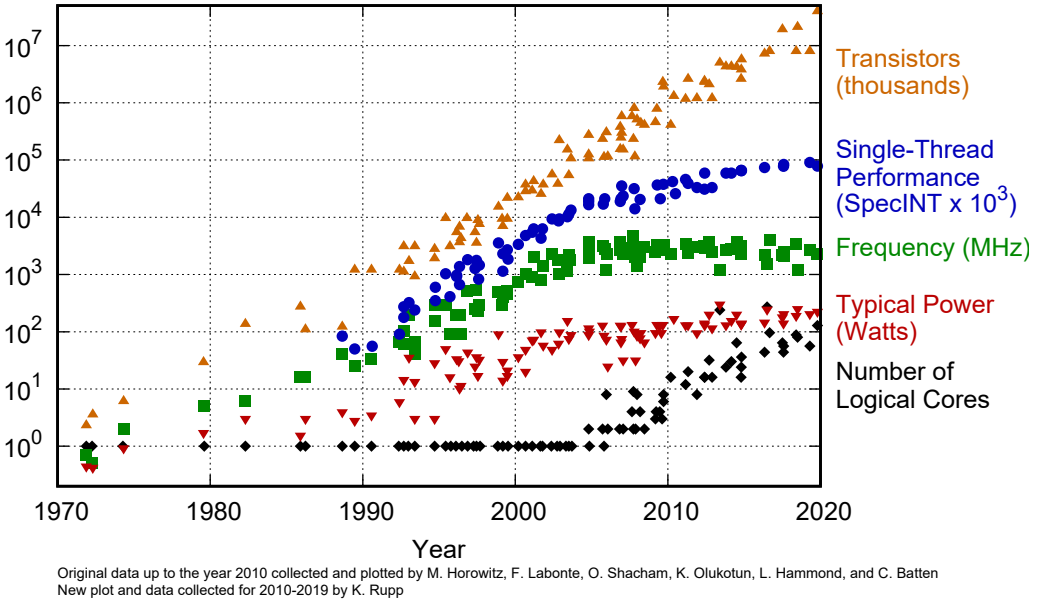

Figure 1.12: 42 years of microprocessor trend data [12].
https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/

The development of a large connectivity between electronic artificial neurons faces an intrinsic problem of electrical connections. The manufacture of semiconductors is mainly based on thin film deposits, and even though multilayer strategies have appeared, the overall architecture remains mostly planar. In addition, the electrical connections cannot cross each other or be too close in order not to lose or alter the information. Therefore, increasing the number of electrical connections necessarily implies reducing the diameter and increasing the length of the wires. However, the flow of information (bits/s) along an electrical connection is limited by the aspect ratio of the wire, *i.e.* the ratio between its section and its length [81]. This characteristic is therefore a strong physical limitation to a significant increase in the number of electrical connections. The article entitled "Design challenges of technology scaling" concluded already in 1999 that the theory of scaling in processors, whether it be of power, density or performance, was going to be stopped by the consumption and dissipation of energy [82]. The analysis is completed in [83] and depicted in Fig. 1.13 which shows that the number of connections in a processor decreases exponentially with the increase in their length. Indeed, the heat dissipation of a electrical wire increases with its length [83], therefore at constant energy dissipation, the increase in the number of connections can only be achieved by reducing their length. The exponential trend visible in Fig. 1.13 clearly shows that achieving parallelized connectivity of an electronic neural network requires extremely short connections, which is in contradiction with a planar architecture and wires that cannot cross each other.

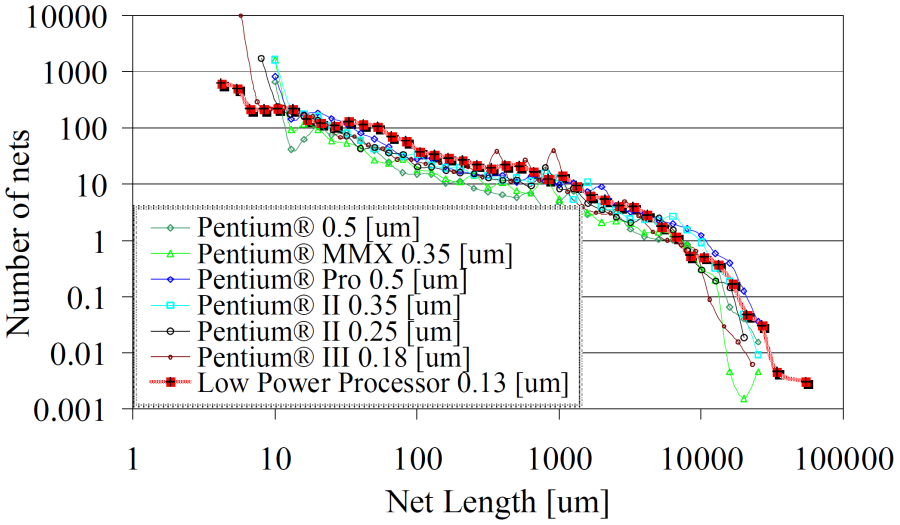

Figure 1.13: Number of interconnections of seven processors according to the connection length [83].

The physical limit for scaling the connections is the fundamental constraint which prevents the massive parallelization of an electronic system. By studying a large number of machine learning accelerators in Fig. 1.14, it appears that the power consumed increases linearly with the number of operations per second (Ops) [84]. On average, an emulated neural networks compute 1 TeraOps/W, and the most powerful NNs running on data center system can currently consume up to 10 KW. Most importantly, even the most recent electronic artificial neural network special purpose chips do not overcome this limitation.

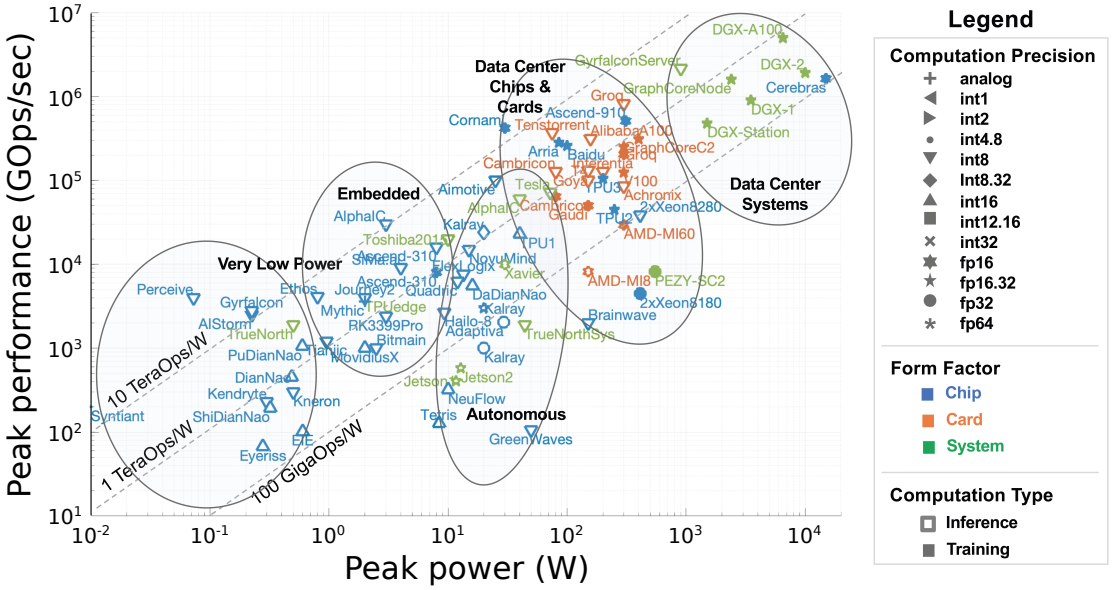

Figure 1.14: Peak performance as function of power of publicly announced artificial intelligence accelerators and processors [84].

1.4.3/ ANALOGUE NEURAL NETWORK: A PARADIGM SHIFT

Since the intrinsic characteristics of digital computing are far from the architecture observed in biological neural networks, research has turned to alternative hardware implementations. As the weights' variation in an analogue neural network (ANN) is not trivial to be realized in hardware, the concept of reservoir computing is widely used. Neural networks have appeared in very different fields, whether with analogue electronic circuits [85, 86], in Field-Programmable Gate Array (FPGA) [87], with memristors [88], in spintronic [89] and optics [90–92]. The principle of the Liquid State Machine has even been physically implemented by calculating with a water bucket [93]. A large review of the field can be found in [94].

The principle of physical neural networks is to use physical phenomena to calculate. By changing the paradigm, the idea is to create networks of nodes which have intrinsic characteristics tending towards those of an ANN. However, work on this topic is very recent and faces many challenges. The technologies used are not yet mature or even remain at the state of research. Using analogue phenomena ultimately results in ANN-circuits exhibiting noise and drifts, which are a priori not beneficial for computation. Studying this impact is therefore essential to understand the underlying mechanisms and potentially how to take advantage of them. Moreover, if the reservoir is entirely analogue, the learning rule for reservoir computing is no longer applicable because it requires knowledge of the entire reservoir state at all times during the learning period. Measuring that state would be to fall back into the pitfalls of digital computing, requiring a consequent instrumentation which can quickly become predominant compared to the network itself. Therefore, it will be necessary to design novel types of learning. Many challenges are still to be overcome, and current research is still in proof of concepts or demonstrators, results comparable to the performance of digitally emulated neural networks remain scarce.

1.4.4/ OPTICAL HARDWARE COMPUTING

Widely used in the telecommunications field, information can be encoded with light. The large-scale commercialization of optical fibers and semiconductor lasers has revolutionized long-distance information transport. The many types of existing light modulation for information encoding (amplitude, frequency, phase, polarization, spatial) have made it possible to increase the communication data rate, recently reaching 178 Tbit/s [14]. In addition, optics have long been a solution considered for working with electronics [95], in particular to address interconnection problems whether inside digital processors [96] or on electronic chips [97] . Indeed, light has an intrinsic characteristic which radically differentiates it from electronics. The photon is a boson, therefore it is not subject to the Pauli exclusion principle. Roughly summarized, this means that the photons can be at the same time in the same place. Furthermore, a photon has no electric charge, hence does not suffer from capacitive and inductive energy dissipation. This fundamentally changes the transport of information due to the ability of communication channels to be able to cross or even be superimposed, strongly enabling parallelization. In addition to its high speed, optics or optoelectronics can perform calculations and communications at very low energies (~10 fJ/bit), which is approaching a competitive level when compared to electronics [98, 99].

As light has great advantages for the transport of information, optics therefore seems an excellent candidate for the implementation of a neural network where connections are a fundamentally important part of the system. Since the beginning of the 2010s, many scientific publications relating to this field emerged, a recent review can be found in [94, 100]. Optics is even considered through the manipulation of single photons or entangled photons for the implementation of quantum neural networks [101, 102]. Still, with the idea to roughly mimic biology, optical neural spikes are achievable, for example using excitable semiconductor lasers. The biological neurons features such excitability, total and partial refractory periods, which are properties recently reproduced in semiconductor lasers [103–105].

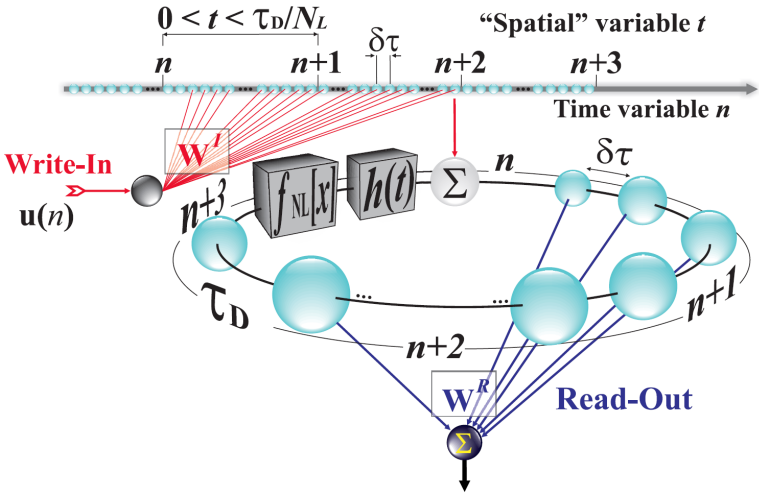

Figure 1.15: Principles of a RC based on delay system. A nonlinear delayed feedback dynamics emulating virtual nodes which are connected via time multiplexing [106].

The field of hardware neural networks has made a major advance with the use of delay systems. As illustrated by Fig. 1.15, the principle is to create a network with a single physical nonlinear node but mix the information using time multiplexing, thus producing coupling between a large number of virtual nodes [85]. The implementation of delay systems in optoelectronics is now well mastered and gives excellent results with for example the classification of a million words per second [106]. However, this type of network is difficult to scale since increasing the number of nodes means increasing the delay and therefore reducing the global data rate of the system.

A spatio-temporal architecture, where nodes are spatially distributed, avoids this drawback. Such a parallel reservoir computing architecture was proposed in 2008 [107], then enhanced in 2011 [90], and it potentially offers fundamentally improved computation speed and low power consumption. The optical reservoir, whose architecture is illustrated in Fig. 1.16, was composed of 4×4 semiconductor optical amplifiers (SOAs) assembled in a array. Each nonlinear node is connected to a maximum of four neighbors in a swirl configuration.

Based on similar spatio-temporal topologies, other parallel optical reservoir were investigated. A building blocks for an all-optical high-speed reservoir-computing composed of 6×6 micro-ring resonators array was numerically demonstrated [91]. An experimental demonstration of cascaded array of 56 programmable Mach–Zehnder interferometers in a silicon photonic integrated circuit was successfully performed [108]. A passive integrated photonic reservoir computing platform based on low multimodal loss Y-junctions has been designed to be CMOS compatible [109]. However, a planar architecture makes it difficult to connect a large neural network, the scalability of 2D systems is therefore delicate.

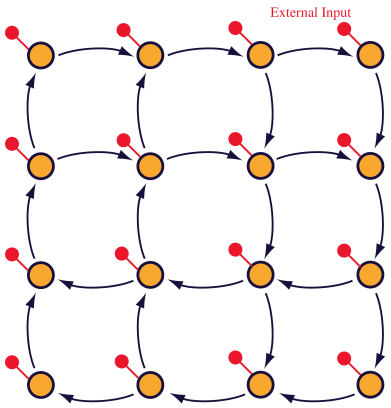

Figure 1.16: Spatio-temporal neural network architecture with a 2D topology [90].

Another family of ANNs with photonic nodes is based on free-space optics. As illustrated in Fig. 1.17, the first implementation of a Hopfield network (RNN with discrete time) in optics dates back to 1985 [110] where information was processed spatially [111, 112]. The main idea is to build artificial neurons on a surface and couple them by optical connections which takes place in the third dimension, which is the propagation direction of the photons [113]. Information can be transported without significant crosstalk in overlapping optical beams, the connections between the nodes are realized by a spatiotemporal optical feedback. Intrinsically this architecture then presents a high potential scalability.

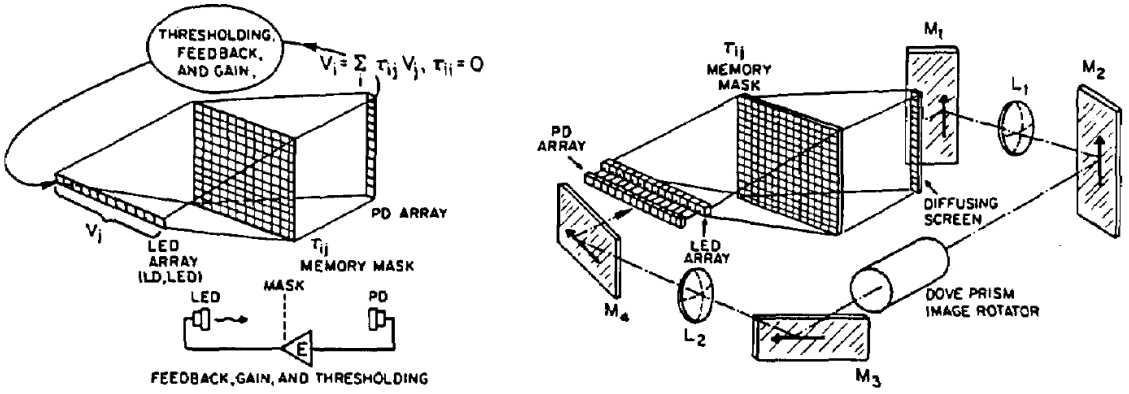

Figure 1.17: Schematic illustration of a Hopfield network implemented electro-optically and all-optically in (a) and (b), respectively [111].

Several recent works use optical diffraction to create an ANNs connections. The implementation of this new architecture was introduced with diffractive coupling of an 8×8 semiconductor lasers array [113]. Another system using diffraction grating and Fourier imaging was designed to connect micro-ring resonators [114]. The scalability of this coupling technique is promising for realizing large neural networks [17]. Thus, a network using optical diffraction to couple 900 nodes created via a spatial light modulator (SLM), and a digital micro-mirrors device (DMD) to implement the readout weights has recently been realized [16]. Diffraction is a technique that allows interconnecting many types of photonic artificial neurons, in particular semiconductor lasers which have high potential to create an ultra-fast system with low energy consumption [115].

Keeping the advantage of optical coupling leveraging all three dimensions, other methods of connections have been realized. In particular, neural network using speckle imaging technique [116, 117] and random connection matrix consisting of a scattering media also show a high scalability [118, 119]. A recent and promising 3D coupling technique is based on optical waveguides 3D printed with a fractal architecture [120]. Such structures are scalable and high density integration of optical interconnect based on photonic waveguides.

1.5/ SUMMARY

We have introduced and discussed during this chapter the basics of neural networks, starting with an introduction of dynamic systems since it is a fundamental feature of these systems. We have defined the role of artificial neurons and then described different types of network architecture. Feedforward neural networks are now widely used, their learning is well mastered in particular thanks to the error gradient backpropagation technique. RNNs are more complex. It is possible to topologically unfold these networks in time, and, we then observed that nodes process the present information taking into account past situations. This property corresponds to memory, which makes RNNs efficient to process sequential information.

Neural network training generally consists in modifying the network's connection weights

in order to optimize the projection of the data in a large dimension space and to optimize the linear separation allowing to classify the information. Such learning can be unsupervised, *i.e.* it is based on structure recognition in the data. In the case of supervised learning, the network's weights are optimized to reduce the error between the neural network result and the input data which is previously known for a typically large number of examples. The result of the neural network should be a generalization of the training data, and if its computation is too specifically optimized for the training data we speak of overfitting.

Current neural networks are mainly computer emulated. Thus, neural networks deployed within our societies are based on nowadays highly developed tools of software and the maturity of the semiconductor industry. Nevertheless, the intrinsic architecture of traditional computers is very far from the one of ANNs. Since the miniaturization of components has practically reached its physical limits, the issue has turned to parallelization to further increase the computing power. However, the integration of electronic circuits is performed by surface deposits, thus the electrical connections between transistors are also mostly in 2D and can neither cross nor be too close. In addition, we have seen that the maximum flow of information in an electrical connection is limited by its aspect ratio. As a consequence, the implementation of fully parallel and integrated electronic ANN processors is currently facing major challenges.

We discussed a paradigm shift by raising the issue of hardware computing. This development has been strongly advanced thanks to the concept of reservoir computing. The architecture is based on RNN but learning only consists in optimizing the output weights. In addition to the highly simplifying learning, the concept leaves the possibility of using many types of reservoir. The reservoir can therefore be analogue and exploit the features of physical phenomena.

Finally, we have seen that light is an exceptional vector of information. Since connections are an important part of neural networks, optics is an excellent candidate for their implementation. More specifically, we introduce the idea of a size scalable and parallel ANN architecture based on the advantageous properties of optics. The main idea is to create optical artificial neurons on a 2D surface but to connect them optically in the third dimension. In practice, the information from all nodes can be transported by non-interacting, hence parallel, optical beams and coupling can be implemented physically by exploiting the diffraction of light. This strategy will then be implemented for the design of an electro-optical neural network presented in chapter 2 and will also be the architecture targeting for the construction of an all-optical network whose fundamental steps will be presented in chapter 4.

2

ELECTRO-OPTICAL RECURRENT NEURAL NETWORKS

Inspired by the reservoir computing concept illustrated in Fig. 1.9 of the previous chapter, we create a recurrent neural network where internal and readout connections as well as nodes are realized in hardware. Once the system has been trained, all connections are passive and due to the intrinsic characteristics of the optics they are fully parallel.

It is generally difficult to separate the role of hardware from the software part of such a neural network since the two are merged, one interacting with the other. Therefore, for increasing the clarity of our discussion this chapter will be divided into two parts. In the first part, we will discuss how to build such an electro-optical neural network and characterize the components constituting the system. We will first study the photonic neurons, then the recurrent connections. The method of coupling based on diffraction forms a principle element of this network and will therefore be investigated in detail, particularly with regard to the limits of this technique [17]. Finally, we will show how the network readout has been implemented physically.

The second part concerns then the dynamics of the neural network. The network architecture and the physical phenomena used are ultimately the result of a process aimed at processing information using the reservoir computing concept. We will describe the temporal evolution of the network's state and study the impact of the feedback strength within the reservoir.

A learning algorithm well suited for a hardware implementation will be presented in the third part. Learning is a major challenge in a complete hardware implemented neural. In addition, a strategy to mitigate the unipolarity of optical intensity will be presented. The learning ability of the system will be tested with a complex time-series prediction task: one-step-ahead prediction of a chaotic Mackey-Glass sequence.

2.1/ NEURAL NETWORKS: ENGINEERING PERSPECTIVE

Since the neural network is an electro-optical experiment, in this part we will characterize the system with a focus on engineering. The experiment will first be introduced as the entire setup, then each component constituting the neural network will be described in detail.

2.1.1/ EXPERIMENTAL SCHEME

Figure 2.1 schematically describes the photonic neural network. The pixels of the spatial light modulator (SLM) are illuminated by a plane wave coming from a laser s-polarized (661.2 nm). The pixels are used to physically implement the network nodes, the illuminating light provides the maximum signal amplitude of each pixel's nonlinear response. The light is reflected by the SLM and the optical beam's p-polarization passes through the polarizing beam splitter (PBS), realizing a $\cos^2$ non-linearity in the transmitted intensity. Thus, these illuminated SLM pixels combined with the nonlinearity due to the PBS create the nonlinear nodes and hence defines the neural network states $\mathbf{x}(n)$ at integer time n. The light then passes through a diffractive optical element (DOE) twice due to reflection by a mirror. The DOE establishes the recurrent connections using diffractive side-orders, creating the connection matrix $\mathbf{W}^{\mathrm{DOE}}$. After the double pass through the quarter wave plate, the PBS directs the entire signal towards the camera, where each pixel of the SLM is imaged. The image recorded is used to drive the SLM by electronic feedback implemented via the control computer.

Taking advantage of the RC concept, we restrict learning to the modification of the readout weights. We build the readout system using the polarization-filtering of the PBS, whose reflection constitutes the polarization orthogonal to the network state. According to $\cos^2\Theta + \sin^2\Theta = 1$, this signal contains the same amount of information as the network state. The readout connections are created by imaging the network's orthogonal state onto the digital micro-mirror device (DMD). The DMD's mirrors can be configured in two positions with an angle of $\pm 12^\circ$ from normal. A photodiode, which is positioned such that it only detects the optical field of mirrors oriented at $+12^\circ$, provides the network's output $y^{\mathrm{out}}(n)$.

The input layer is not yet implemented in hardware but is currently realized digitally based on the control computer. Input information $u(n)$ is injected into the system according to constant random connections W^{inj}.

2.1.2/ NONLINEAR NEURONS SUBSTRATE

The nodes' hardware substrate is a SLM (Hamamatsu X13267-01) whose pixels give a physical and spatial entity to nodes. The neural network is composed of 2025 nodes (45×45 pixels) for its characterization and 961 (31×31 pixels) during its operation. The pixel pitch is $\mathrm{p}^{\mathrm{SLM}} = 12.5$ µm, the total size of the network on the SLM's surface is a square of 0.316 mm^2 or 0.150 mm^2 for 45×45 or 31×31 nodes, respectively.

Pixels i of the SLM are illuminated by a plane wave $|\mathrm{E}_i^0|^2$. To achieve this, the colimated diode laser (Thorlabs LP660 SF20, λ = 662.1 nm), aligned to s-polarization by paddle polarization controllers (Thorlabs FPC030) is focused in the back focal plane of the microscope objective MO1 (Nikon CFI Plan Achro 10x) by the lens L1 (Thorlabs AC254-200-AB-ML). As illustrated in Fig.2.2(a), the pixels of the SLM receive an illumination with a Gaussian profile due to the spatial mode of the single mode fiber-coupled laser. To obtain non-linearity the SLM has to be in intensity modulation configuration. For this, the laser is linearly polarized and the optical beam is reflected by the PBS towards the SLM. The half-wave plate is adjusted such that the linear polarization of the incident beam is at 45 degrees from the ordinary and extraordinary axis of the SLM. The extraordinary axis of the pixels is modulated by a voltage resulting in a refractive index modification controlled

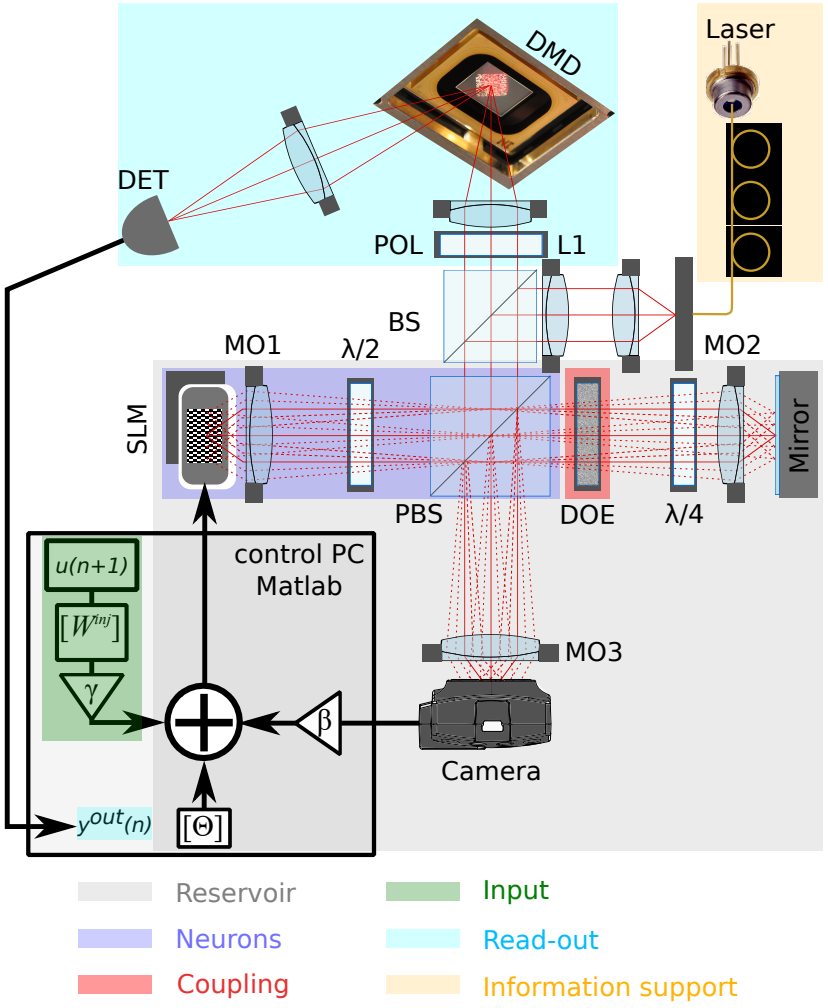

Figure 2.1: Photonic implementation of a spatio-temporal neural network with 961 nodes. An optical plane-wave illuminates the spatial light modulator (SLM), the neural network states are encoded by the SLM pixels. They are imaged on the camera, passing through the polarizing beam splitter (PBS) and the diffracting optical element (DOE) creating non-linearity and recurrent coupling between network states. The information detected by the camera is used to drive the SLM. We image the nodes on a micro-mirrors array (DMD), which allows to select pixels imaged onto the detector and hence implements Boolean readout weights.

by the gray scale value (GS) of the SLM state vector $\mathbf{x}^{\mathrm{SLM}} \in \{0,1,....,255\}$. This modulation introduces a phase-shift of the optical field along the extraordinary axis, which results in a rotation of the polarization angle of the reflected light. Therefore, the optical beam reflected by SLM pixel i is linearly polarized with a specific angle depending on $\mathbf{x}_i^{\mathrm{SLM}}$. The resulting optical field of pixel i after polarization filtering by the PBS is then

$$\mathrm{E}_i = \mathrm{E}_i^0 \cos\left(\frac{2\pi}{K_{\mathrm{SLM}}} \cdot (\mathbf{x}_i^{\mathrm{SLM}} + \varphi_i^0)\right), \tag{2.1}$$

where $K_{\mathrm{SLM}} = 244.6 \pm 1.6$ is the conversion between pixel gray scale and polarization angle in radians, and $\varphi_i^0 = 11.1 \pm 1.1$ the gray scale offset which is a device related constant.

These two parameters have been characterized and their uncertainty is the standard deviation over all measured SLM pixels.

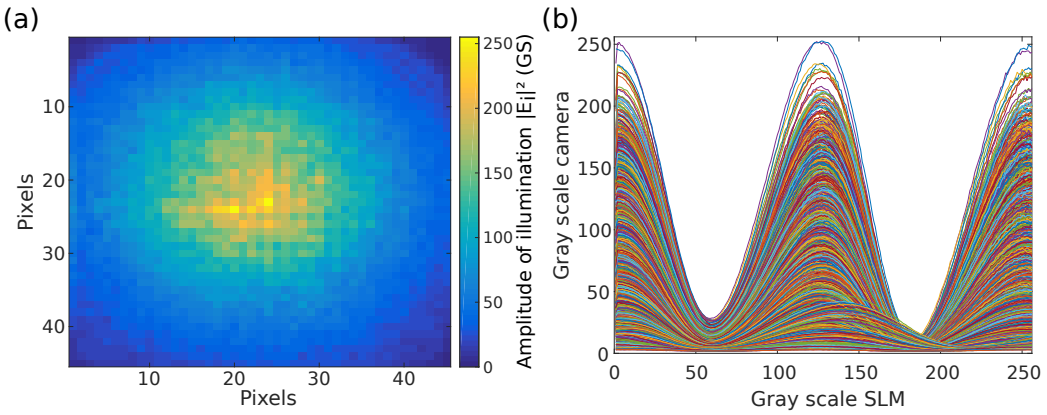

Figure 2.2: (a) Gaussien illumination of the SLM pixels. The amplitude in gray scale (GS) is encoded on 8-bits. (b) Averaging of twenty measurements of the nonlinearity of the 2025 nonlinear nodes which composed the network. For the low amplitude detected by the camera, we can clearly see a phase shift.

Figure 2.2(b) illustrates the intensity modulation of each pixel, thus realizing the nonlinearity of the nodes. For nodes with a low maximal intensity, the maximum modulation is slightly shifted. The intensity difference between pixels is the consequence of the illumination by a Gaussian profile. The low intensities of a Gaussian beam are by definition at the borders of the Gaussian. Focusing the laser in MO1 back focal plane, the wavefront can no longer be approximated by a plane wave (paraxial approximation) beyond a radius. The half-wave plate adjusted for a plane-wave, *i.e.* the Gaussian beam center, is no longer precisely set for the pixels at the edges of the illuminated SLM surface. The polarization angle of the incident beam is therefore slightly different for low intensities.

2.1.3/ RECURRENT NEURAL NETWORKS

To obtain a recurrent neural networks, the nonlinear nodes must be self-coupled through feedback. Since the control of nodes is electronic, their optical information must be digitally converted using a camera (Thorlabs DCC1545M)). SLM and the camera are in the focal plane of MO1 and MO3 (Nikon CFI Plan Fluor 4x), respectively. However, the camera is not placed behind the PBS, otherwise an inverted image would be formed ($2f$ imaging). To reverse the image, a mirror is placed in the focal plane of MO2 (identical to MO1), which results in a 4f imaging configuration. A quarter-wave plate is added on the optical path between the PBS and the mirror, transforming the linear polarization of the incident beam into circular polarization. When the beam returns, the polarization is again transformed into linear polarization, but at 90 degrees from the incident beam. Therefore, the light reflected by the mirror is fully reflected by the PBS towards the camera. This 4f-system is not essential for this electro-optic neural network since the image of SLM could be digitally inverted. However, this configuration is relevant for the self-coupling in an all-optical neural network, which requires self-imaging and hence a 4f configuration.

The camera adjustment is a critical step of the optical alignment. The camera is positioned on a tilting micrometric platform (Thorlabs PY003/M) in a way that its axes are

aligned with those of the SLM. Using a non-trivial intensity modulation pattern displayed on the SLM, the area of interest of the camera is specified and the orientation of the camera is mechanically adjusted with a precision exceeding 4 µrad. In the absence of recurrent connections between the nonlinear nodes, the signal recorded on the camera is then

$$x_j^C = \alpha |\mathrm{E}_j|^2, \tag{2.2}$$

where $x^C \in \{0, 1, ..., GS^{\max} = 255\}$ is the 8-bit camera-state. To use the full dynamic range of the camera and avoid over-exposure, the optical intensity $|\mathrm{E}|^2$ is converted by $\alpha = \mathrm{ND} \cdot GS^{\max} / I^{\mathrm{sat}}$, where I^{sat} is the saturation intensity of the camera and ND the transmission efficiency of the optical path including added neutral density filters.

Due to an optical imaging magnification ($\mathrm{Mag} = 2.5$), and different pixel pitch of SLM ($p^{\mathrm{SLM}} = 12.5$ µm) and camera ($p^{\mathrm{C}} = 5.2$ µm), the number of pixels in the area of interest is not the same between the camera and the SLM. Camera-state $\mathbf{x}^{\mathrm{C}}$ is therefore linearly rescaled in size to match the number of active SLM pixels, resulting in $\tilde{\mathbf{x}}^{\mathrm{C}}$. Each active pixel i of the SLM is therefore imaged on a group of pixels i of the camera, and the normalized and re-scaled camera-state $\tilde{x}_i^{\mathrm{C}} = \alpha |E_i|^2$ drives the SLM's state vector $\mathbf{x}^{\mathrm{SLM}}$. This self-coupling is measured and represented in Fig. 2.3, where the pixels of the SLM are all turned off ($x^{\mathrm{SLM}} = 55$) and sequentially a single pixel's intensity is turned on ($x^{\mathrm{SLM}} = 125$) and the image is recorded by the camera. The resulting camera images are reshaped into single-column vectors, and the individual images for each pixel "switched on" is appended to an array. The result is a diagonal appear, proof that each node is self-coupled. The low visibility of the diagonal is due to a single pixel width, and the variation of the intensity detected by the camera is the consequence of the Gaussian illumination.

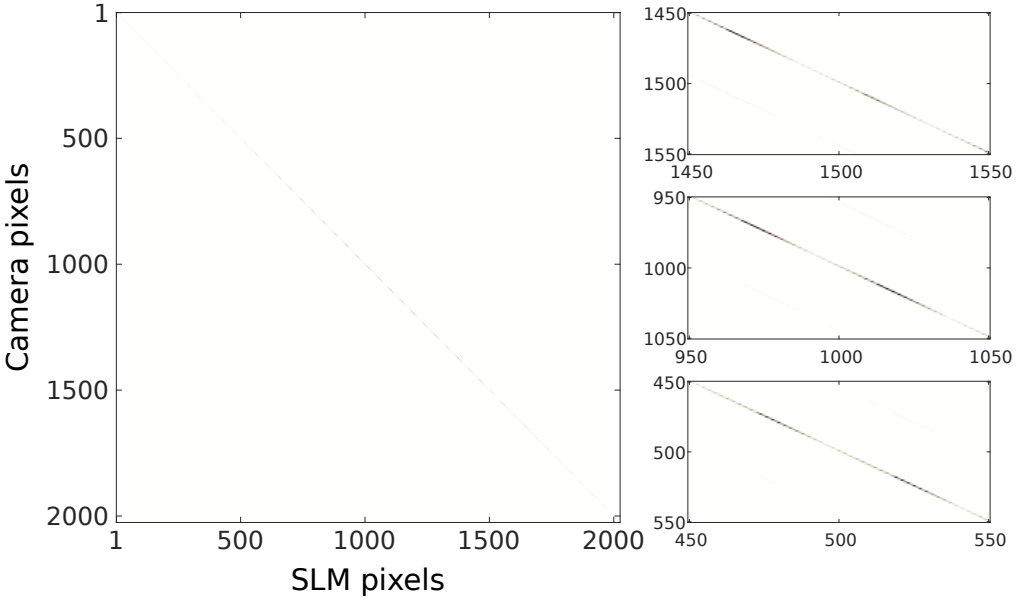

Figure 2.3: Recurrent neural network self-coupling. Each pixel of the SLM is linked to a single pixels array of the camera. The three right insets are zooms into smaller regions. The intensity is normalized but is identical for the four panels.

2.1.4/ PHOTONIC NETWORK CONNECTIVITY

Following the principle of reservoir computing [15], the coupling between nonlinear nodes can be random but needs to remain constant in topological terms over time. The neural connectivity obtained by using the diffraction property of light perfectly meets this condition. The idea is to locally connect each nonlinear node to its neighbours by adding a DOE (HOLOOR MS-443-650-Y-X) to the optical path.

2.1.4.1/ DIFFRACTIVE COUPLING

Our DOE is comparable to an optical transmission grating with a phase-modulation along the x and y-axis. Such a DOE divides a single spatially coherent beam into several beams by diffraction. In our case, the DOE splits the incident optical beam into nine, each with the original beam except for its power and angle of propagation. The different diffractive orders separated by angle φ^{diff} are imaged onto the camera with a distance d^{diff} between them. The DOE is located in the collimated space of our imaging system, and we can treat the system as infinite diffraction by considering only plane waves. The diffractive angle $\varphi^{\text{diff}}_{i,m}$ for diffractive order m of pixel i can be obtained by the grating equation

$$\sin\left(\varphi^{\text{diff}}_{i,m}\right) = \sin\left(\varphi^{\text{im}}_{i}\right) + m\frac{\lambda}{p^{\text{DOE}}}, \tag{2.3}$$

where $\varphi^{\text{im}}_{i} = \arctan\left(i \cdot p^{\text{SLM}}/f_1\right)$ is the incident angle of the collimated beam reflected from pixel i, $p^{\text{DOE}} = 1006.55$ µm the grating pitch of the DOE and λ the wavelength of the illuminating laser. Figure 2.4(a) illustrates the coupling mechanism by diffractive imaging for the simpler case of a 2f imaging system. Importantly, the working principle is identical for 4f imaging. For pixels (nodes) i and $i+1$ the DOE creates three orders of diffraction in this planar projection. In order to obtain a spatial superposition of their optical fields, which is the mechanism we use for creating coupling, the propagation angle of the diffracted orders must be identical. The criteria for coupling between the diffractive order $m = -1$ of emitter $i = 0$ and order $m = 0$ of emitter $i = 1$ is then given by

$$\varphi^{\text{diff}}_{0,-1} = \varphi^{\text{im}}_{1}. \tag{2.4}$$

The angles can be written as

$$\varphi^{\text{diff}}_{0,1} = \arcsin\left(\frac{\lambda}{p^{\text{DOE}}}\right) = \arctan\left(\frac{d^{\text{diff}}}{f_2}\right), \qquad \varphi^{\text{im}}_{1} = \arctan\left(\frac{p^{\text{SLM}}}{f_1}\right). \tag{2.5}$$

The parameter λ, p^{SLM} and p^{DOE} should therefore be optimized to satisfy Eq. (2.4). However, it is clear that these parameters are only exact for coupling the particular pixels selected for optimising Eq. (2.5), usually the SLM's center pixels. For pairs of transmitters located far from the optimal position, the overlap between neighboring pixels is degraded due to the different trigonometric relationships given by Eq. (2.5).

According to our experimental parameters, we obtain $d^{\text{diff}} = 33.105$ µm. Comparing to the pixel pitch after imaging, the distance between diffractive orders has an absolute error $\Delta = |\text{Mag} \cdot \text{p}^{\text{cam}} - \text{d}^{\text{diff}}| = 1.85$ µm and a relative error $\delta = \Delta/\text{Mag} \cdot \text{p}^{\text{cam}} = 5.6 \cdot 10^{-2}$. This mismatch represents a compromise between performance and the use of products off the shelf such as the laser, the SLM and the DOE.

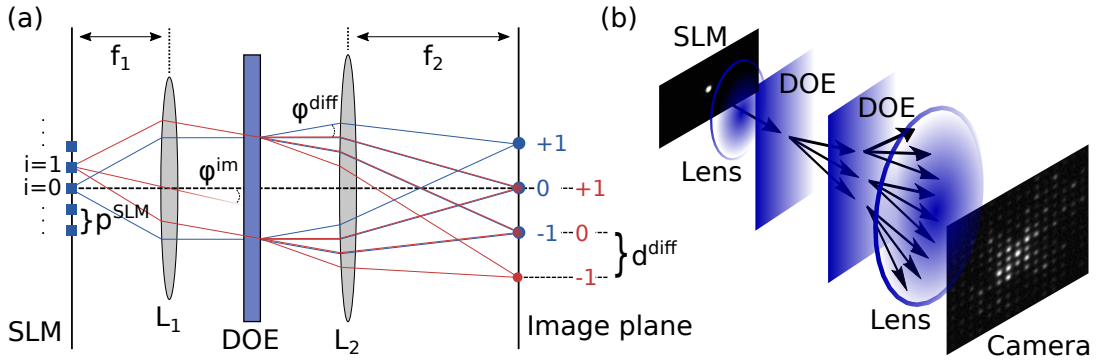

Figure 2.4: (a) 2D schematic illustration of coupling of the SLM pixels i by a diffractive process. Two pixels are imaged via a 2f system including a DOE creating three orders of diffraction. f_1 and f_2 are the focal distance of lenses L_1 and L_2, respectively. φ^{diff} is the angle between diffraction orders and φ^{im} is the angle between the principle rays of neihboring emitters. If both angles are equal, the distance between diffractive orders in the image plane d^{diff} is identical to the distance between neihboring emitters p^{SLM} considering the magnification : $d^{\text{diff}} = f_1 p^{\text{SLM}}/f_2$. (b) Illustration of (only along the vertical axis) the diffraction of an optical beam during its double passage through the DOE. The image of the SLM and the camera are both experimental data measured before and after the DOE, respectively.

Returning to the experiment, the optical beam passes a second time through the DOE after being reflected by the mirror, therefore, as schematically illustrated in Fig. 2.4(b), the diffraction orders are again diffracted. Ultimately, this DOE double-pass transforms the beam of each pixel of the SLM into 25 beams propagating at a discrete set of angles. Each pixel is directly connected to all its close neighbours within a coupling radius of two pixels. The signal recorded by the camera is then

$$\tilde{\mathbf{x}}_i^C = \alpha \left| \sum_j^N W_{i,j}^{\text{DOE}} \mathrm{E}_j \right|^2, \tag{2.6}$$

where $\mathbf{W}^{\text{DOE}}$ is the coupling matrix and N the total number of nonlinear nodes. In the same way as explained for Fig. 2.3, except that the DOE is introduced into the beam path, we measure qualitatively the coupling matrix. As illustrated by Fig. 2.5, now there is not only a single diagonal of a single pixel width as previously, but five lines of five pixels width. As before the real 45×45 2D matrices are unfolded into a 2025×1 vector in 1D in order to obtain a representation of Fig. 2.5 in 2D. Each pixel of the SLM is therefore coupled to 24 of its close neighbours and itself. As better visible in the right insets, the width of the diagonals corresponds to the coupling along the vertical axis between the nodes. The vertical distance of 45 pixels between the diagonals is the result of the transformation of the 45×45 matrix into a 2025×1 vector, but actually corresponds to the coupling of neighboring pixels along the horizontal axis. The coupling of neighboring pixels along the vertical axis is for its part illustrated by diagonals five pixels wide.

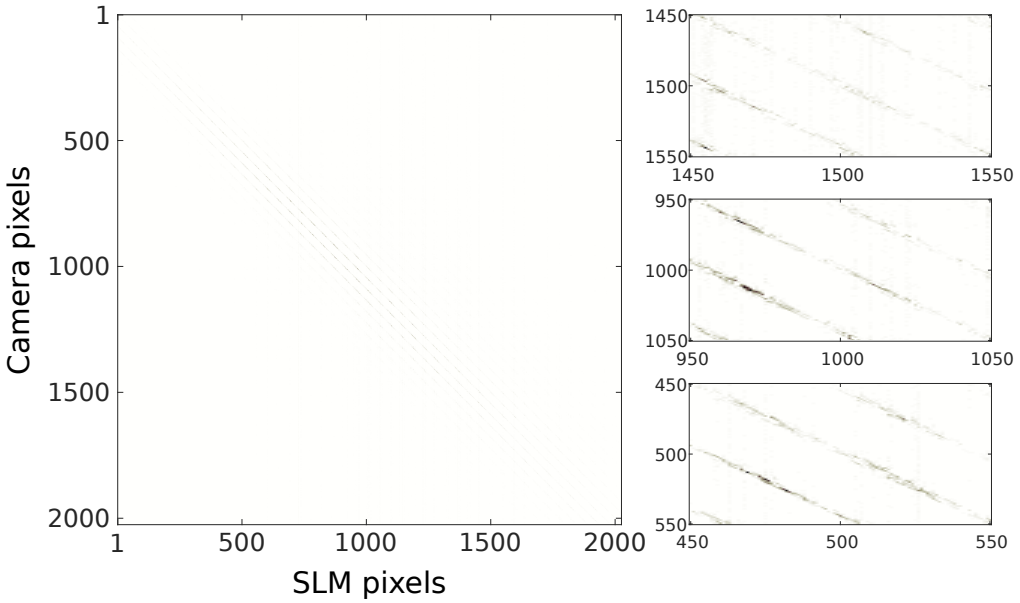

Figure 2.5: Recurrent neural network's coupling matrix $\mathbf{W}^{\mathrm{DOE}}$ established by the DOE. Each pixels of the SLM is coupled to its close neighbours. Due to the unfolding of 2D matrices of pixels into 1D vectors, the width of the diagonals corresponds to the coupling of nodes along the vertical axis, the vertical distance between the diagonals corresponds to the coupling along the horizontal axis. The three right insets are zooms into smaller regions. The intensity is normalized but is identical for the four panels.

2.1.4.2/ COUPLING STRENGTH

Studying the insets on the right in Fig. 2.5, we can see a strong variation in local connectivity strength. This is the result of the spatial distribution of the pixels of the SLM. The different positions of the pixels lead to collimated beams with different propagation angles, thus illuminating a slightly different area of the DOE. Therefore, the distribution intensity between the diffractive orders varies. This phenomenon, intrinsically linked to the coupling method, gives heterogeneous connectivity to the neural network, which is beneficial for computation according to the reservoir computing concept [15].

We continue to study in detail the profile of the coupling strength. During the first passage of the optical beam through the DOE, the pattern produced is a 3×3 configuration of diffractive orders. The cumulative intensity of the diffracted orders represents approximately 70% (72% according to the manufacturer) of the entire optical intensity. We could then approximate this coupling strength distribution as the result of a convolution of a single delta peak with a 2D-step function which has the value 1 for each of its 3×3 inputs and zero otherwise. This convolution results in the same 2D step function. The second path through DOE could be viewed as the self-convolution of the 3×3 step function. The result is a 2D pyramidal intensity distribution with 5x5 non-zero entries.

Figure 2.6 presents the averaged coupling properties obtained from our neural network. To achieve this, we use the data measured by the camera during the characterization presented in Fig. 2.5. We select an area of 9×9 pixels centered around the activated SLM pixel i, hence the subarray center is the position of the zero diffractive order. Subarrays are normalized to unity and are consequently summed together, and, the image obtained

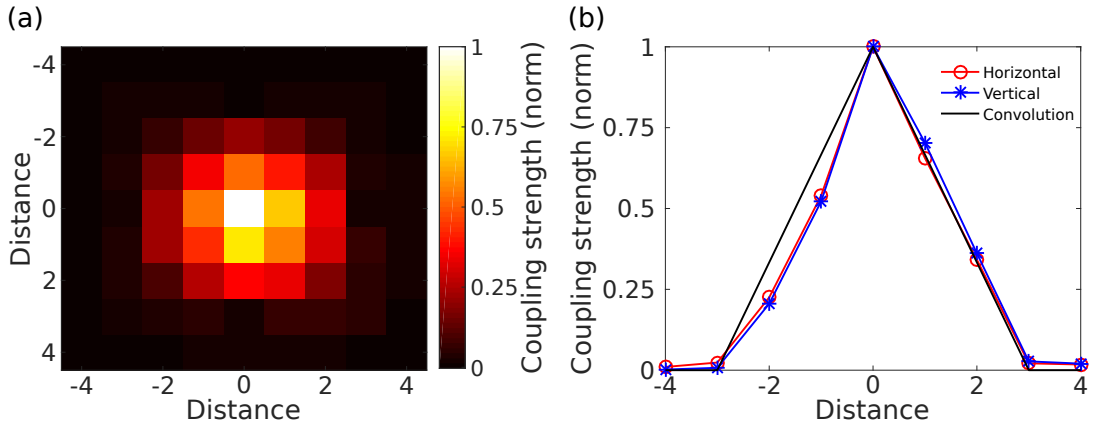

Figure 2.6: (a) Normalized, average coupling strength against coupling distance. The dominating coupling term corresponds to self-coupling. (b) Horizontal and vertical profiles through the center position of panel (a). The triangle topology obtained is the result of double passing the DOE. The black line is the theoretical result of the self-convolutions of 3×3 step function.

is shown in Fig. 2.6(a). As expected, the coupling strength at the center, corresponding to the position of the non-diffracting order after the double-pass through the DOE, is the strongest. Figure 2.6(b) corresponds to the vertical profile (blue stars) and horizontal profile-cuts (red circle) through the central position of the panel (a). Despite a slight asymmetry, the experimental result matches well with the pyramid coupling profile of the model (black line). We notice that, for a coupling radius larger than two, the experimentally measured coupling strength is not strictly zero. This is attributed to the non-negligible contribution of the higher diffraction orders of the DOE, which can be clearly observed in Fig. 2.4(b).

2.1.4.3/ NETWORK SIZE LIMITATION

It is essential to estimate the maximum number of nonlinear nodes that can be coupled by diffraction and also to understand the limits of this method.

Coupling is based on two interdependent parts. The first is diffraction by the DOE, the second is imaging. There is a fundamental difference in trigonometric relationships between diffraction angle φ^{diff} in Eq. (2.5) which involves according to a sine function, and angle φ^{im} due to imaging, which depends on a tangent function. These two angles are similar within the limits of the paraxial approximation, *i.e.* for small angles. Due to its importance a full characterization of the deviation away from the coupling condition was performed [17]. Only a summary presenting the major results of this study will be presented in this thesis.

The experimental setup illustrated in Fig. 2.4(a) uses, a single mode optical fiber (Thorlabs TW670R5A2) to emulates different positions of the SLM pixels. The fiber is placed on a micrometric xy-stage (Thorlabs ST1XY-S/M) and is imaged on a camera (IDS USB 3µEye LE). A first microscope objective (Nikon Plan N, NA = 0.25 MAG = 10) collimates the fiber's output which is then imaged by a second microscope objective (Nikon N4X-PF, NA = 0.13, MAG = 4) onto the camera. The distance between the two objectives

is less than 50 mm and the DOE is placed approximately in the middle. For each position of the optical fiber along the horizontal or vertical axis, the image of the camera is recorded. The image is then fitted by nine Gaussian profiles.

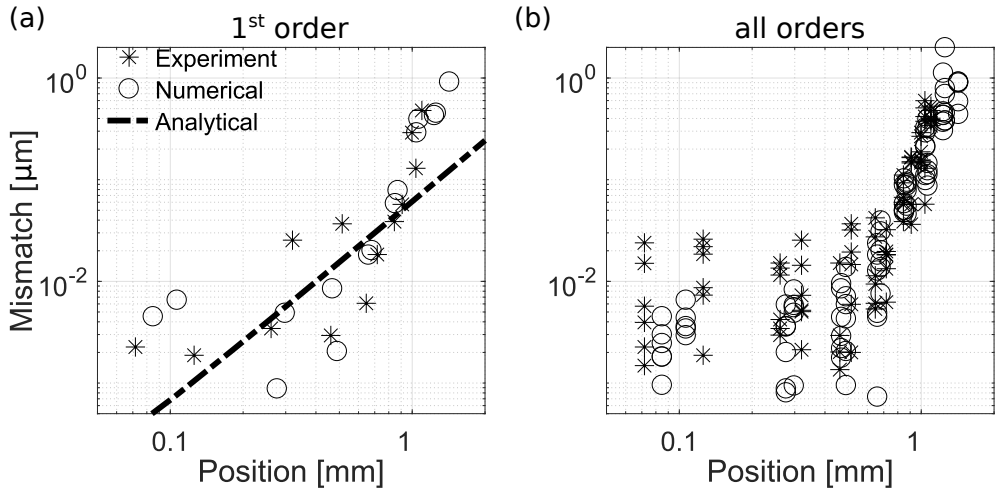

Figure 2.7: (a) Mismatch between the first diffractive order and nominal position according to the position of the emitter along the horizontal axis. Experimental results (stars) are excellently agreed with the numerical results (circle). The analytical solution (dashed line) is calculated according to grating equation. (b) As is (a) but for all diffractive orders, along horizontal and vertical axis. The sharp increase of the mismatch beyond 1 mm results from the vignetting of the beam.

Figure 2.7(a) illustrates on a double-logarithmic scale the experimental difference (stars) between the nominal position and the real position of diffractive order $m = -1$. Below a fiber displacement radius of 1 mm, the positioning error is below 0.1 µm. By taking a typical distance of 10 µm between photonic emitters in for discrete array [16, 121], coupling is therefore possible for more than 30,000 nodes. Nevertheless, the error increases sharply for coupling positions located by more than 1mm away from the system's center. In this outer region, the experimental data diverge strongly from the analytical solution calculated from Eq. (2.3). However, by calculating the numerical solution (circle) based on an optical propagation by angular spectrum method without paraxial approximations [17], the result is in excellent agreement with the experimental measurements. By examining the limiting factor beyond a radius of 1 mm, we realize that the limit is ultimately the entrance pupil of the second microscope objective, which creates a vignetting effect. This information is important because it shows that the main limitation of the coupling is not intrinsic to the concept of diffraction but to the imaging system. Figure 2.7(b) shows the coupling error for all diffraction orders. First, the error obtained confirms the limit obtained for the first order. Second, the low detection noise of the camera allows a very precise fit which brings an experimental resolution below 40 nm.

In addition to an excellent agreement between the positions of the diffraction orders, Figure 2.8 illustrates the width of the diffractive orders (stars) and the theoretical diffraction limit as a function of the position of the emitter. For a position less than one millimeter from the center, *i.e.* avoiding vignetting, the imaging system remains diffractive-limited. Results are slightly better than the diffraction limit, which is due to the uncertainty of the fiber's NA, which creates an uncertainty of the diameter of the collimated beam.

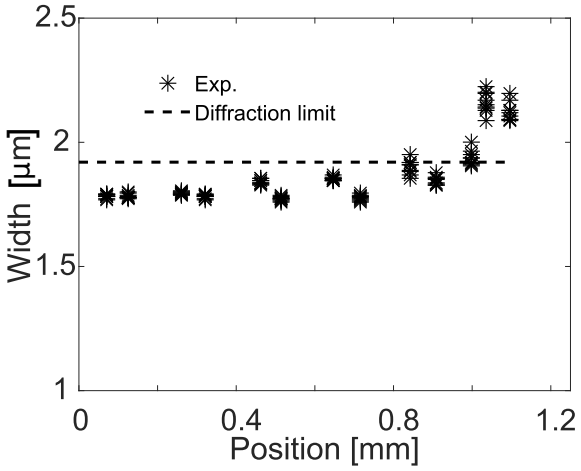

Figure 2.8: Width of the diffractive orders image according to the emitter position. Within an area of one millimeter in radius, the imaging system remains diffraction limited.

For the current electro-optic neural network, having a pixel pitch of the SLM of 12.5 µm, the diffraction coupling technique allows a connectivity between more than 20,000 nonlinear nodes. Importantly, the number of photonic nodes depends on the area, and therefore is proportional to the square of the radius. In addition, the DOE is a passive element, and an increase in the number of nodes has no influence on the cost of their connectivity, whether in terms of energy, size or time. Consequently this coupling technique has an excellent scalability if the vignetting challenge can successfully be addressed.

The optical path through the neural network could be reduced using a more compact mechanical layout in order to reduce the effect of vignetting. In addition, the microscope objectives of the imaging system can be changed, in particular M03 which could have a numerical aperture of 0.5. These modifications increase the vignetting-free imaging radius, but the SLM itself could also be changed to one with a smaller pixel pitch (Holoeye, GAEA-2, $\mathrm{p}^{\mathrm{SLM}} = 4$ µm). These technical improvements would then increase the number of connected nodes beyond several million.

2.1.5/ READOUT WEIGHTS

The weights assigned to each nonlinear node allowing the learning of the network are implemented without modifying the connections of the reservoir. For that purpose we use the PBS as the output-port of the reservoir. An additional BS allows SLM illumination and readout. The BS has a 50/50 splitting ratio in order to maximize the output power. A lens (Thorlabs AC254-400-B) images the pixels of the SLM on the surface of a DMD (DLi4120 XGA). A DMD is a micro-opto-electro-mechanical system composed in our case of a 1024x768 micro-mirrors array, having a mirror pitch of 13.68 µm. Each micro-mirror has two possible orientations, at ± 12° from the normal. Only the -12° degree orientation reflects the SLM light towards the detector and contributes to the output signal. Consequently, the physically implemented readout weights are strictly Boolean.

The imaging system between the SLM pixels and the DMD micro-mirrors has a magnification of 20. Considering the pitch of the SLM and the DMD, the image of a single pixel of the SLM on the DMD is a rectangular area with a size approximately equivalent to 18×18

micro-mirrors. The DMD is mounted on a rotating stage (Thorlabs CRM1L/M) allowing precise alignment between the SLM and DMD axis. The area of interest of the DMD is then carefully defined using an alignment camera (IDS USB 3µEye LE) to visualize the DMD reflection in the +12° degree direction. A first pixel pattern and its complementary is displayed on the SLM and the DMD, respectively. This allows to precisely identify the area of our reservoir's nodes on the DMD, to define its orientation and its size. Variation of the DMD pattern to the nearest pixel allows alignment accuracy of less than 2 mrad. Subsequently, 18×18 micro-mirrors modulate the detected power of a single pixel of the SLM, and in the following they are always addressed as one "super-pixel".

The micro-mirrors which compose the DMD make the device periodic and therefore the reflected light is diffracted. Using the grating equation in reflection with a mirror pitch of 13.68 µm, the difference of angle propagation between the specular order and the first diffractive order is approximately 50 mrad. Therefore, the optical beam after the DMD is intentionally not collimated. Thus, after a light propagation of 20 cm, the diffractive orders are spatially separated by a centimeter. Spatial filtering is then applied in order to physically reject the diffracted orders. In addition, this increases the signal to noise ratio by minimizing the detection of stray light. Finally, the detector (DET, Thorlabs PM100A, S150C) is placed behind a lens that collects the signal coming from the micro-mirrors turned towards its direction.

Figure 2.9 is a photography of the DMD's micro-mirrors. In the center, the white and black square is the area of interest which constitutes the output weights. Each white pixels in the image, corresponding to the value 1 in the weight vector $\mathbf{W}^{\mathrm{DMD}}$, are oriented -12° towards the output detector.

Figure 2.9: Image of a random 31×31 matrix $\mathbf{W}^{\mathrm{DMD}}$ on the digital micro-mirrors device. Each white or black pixel is an array of 18×18 DMD pixels. Only the white super-pixels are turned towards the output detector.

This spatial implementation of the output weights does not require temporal modulation as it is the case for example for reservoir implementations in delay systems [122]. Indeed, after learning to solve a specific task, the configuration of the DMD obtained is then passive. The readout weights are passive mirrors and therefore do not limit the bandwidth of information processing. In addition, the allocation of the weights is fully parallel, and

therefore the number of nodes in the network does not impact the speed of applying $\mathbf{W}^{\mathrm{DMD}}$. The Boolean character of the readout weights could be modified to a gray level if necessary. For this, each 18×18 super-pixel could be subdivided into several sub-areas, then giving a fraction of weight to each. The main constraint is to have a system sensitive enough to measure the variations on then finer scale.

2.2/ NEURAL NETWORKS: DYNAMICAL SYSTEMS

Now that we have described each part of the optical recurrent neural network and explained the physical implementation of the fundamental blocks, namely the non-linear nodes serving as neurons, their self-coupling and their connections to each other, as well as the assignment of the readout weights allowing learning, we will consider this experiment as a dynamical system. Indeed, the feedback from the camera to the SLM pixels implements an echo property. This can create dynamics that can take very different forms, ranging from damping to a chaotic regime depending essentially on the feedback strength.

2.2.1/ INPUT LAYER

The input layer is not yet implemented in hardware, but the external information is injected into the neural network via the SLM by a computer (see Fig. 2.1).

External information is a discrete sequence $u(n)$ normalized between 0 and 1. This vector is multiplied by an injection matrix $\mathbf{W}^{\mathrm{inj}}$ consisting of elements randomly distributed between 0 and 1. This injection matrix remains fixed for all experiments in this thesis. At each discrete time step n, this matrix distributes the scalar input information $u(n)$ to each active pixel of the SLM. The size of the injection matrix depends therefore on the number of nodes in the network, typically $45 \times 45 = 2025$ or $31 \times 31 = 961$. The external information injection is then

$$\mathrm{input}(n+1) = \gamma \mathbf{W}^{\mathrm{inj}} u(n+1) \,, \tag{2.7}$$

where γ is the injection strength which is optimized for each task.

2.2.2/ RESERVOIR

In order to study the dynamical aspect, we have to add a temporal context to the description of the experiment. Illustrated in Fig. 2.1, the camera-state $\tilde{\mathbf{x}}^C$ is recorded by a control computer which also controls the SLM-state $\mathbf{x}^{\mathrm{SLM}}$ via a Matlab program. Therefore, the reservoir forms a single unit where the camera and the SLM are linked, the first serving of input for the second. To close the loop, the state of the camera is multiplied by the feedback gain β, to which external information and a phase offset matrix $\boldsymbol{\theta}$ are added. The result is sent to the SLM as

$$\mathbf{x}^{\mathrm{SLM}}(n+1) = \beta\, \tilde{\mathbf{x}}^C(n) + \gamma \mathbf{W}^{\mathrm{inj}} u(n+1) + \boldsymbol{\theta} \,. \tag{2.8}$$

We can then define the reservoir-state $\mathbf{x}$ which is the result of the nonlinearity $\mathbf{f}(\cdot)$ of the nodes

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}^{\mathrm{SLM}}(n+1)) \,. \tag{2.9}$$

We notice that the reservoir-state is physically located just after the PBS which gives the nonlinear function to the system. In other words, the reservoir-state is never measured experimentally. Equations (2.8) and (2.9) have the same structure as in the classical concept of reservoir computing [15]. The only difference comes from the electro-optic nature of the neural network. The reservoir state is defined according to the optical field while the update is according to the optical intensity. Indeed, as shown in Eq. (2.6) which defines the camera-state, the camera records the square of the optical field.

The evolution of the reservoir dynamics is therefore described by a coupled Ikeda map [123] according to

$$x_i(n+1) = \alpha|\mathrm{E}_i^0|^2 \cos^2\left[\beta\cdot\alpha\left|\sum_j^N W_{i,j}^{\mathrm{DOE}}\mathrm{E}_j(n+1)\right|^2 + \gamma W_i^{\mathrm{inj}}u(n+1) + \theta_i\right]. \tag{2.10}$$

The overall update rate of the entire system is only 3 Hz. This low rate is currently limited by the Matlab script controlling the SLM, a programming language closer to the machine code would greatly accelerate the system. Indeed, the SLM currently used has a maximum frame rate of 50 Hz, which corresponds to the true limiting physical factor.

Despite the apparent complexity of Eq. (2.10), the only matrix multiplication provided by the control computer is the injection matrix. All other operations are performed optically, fully in parallel. The whole point of the hardware implementation of a neural network is that most of the computational operations, if not all, are done through physical phenomena. Simulating Eq. (2.10) by a computer may seem trivial but by looking more closely we realize that the reservoir-state of node i at time $n+1$ depends on the reservoir-state at time n of all the nodes in the network. Coupling between nodes is equivalent to matrix multiplication, for a computer with serial operation the problem scales with N^2. Coupling between nonlinear nodes, which is the intrinsic heart of a neural network, is therefore the major source of computational load in serial emulations of neural networks. While interconnectivity is easily implemented with the parallel DOE's, its computer simulation is complex in terms of time and storage space for large numbers of nonlinear nodes [119, 124].

2.2.3/ READOUT LAYER

As explained in sections 2.1.1 and 2.1.5, the output of the system contains as much information as the reservoir but is modulated by the weight vector $\mathbf{W}^{\mathrm{DMD}}$. The output of the system is therefore

$$\begin{aligned} y^{\mathrm{out}}(k) &\propto \left|\sum_i^N W_i^{\mathrm{DMD}}(k)\cdot[\mathrm{E}_i^0 - \mathrm{E}_i(n+1)]\right|^2 \\ &\propto \left|\sum_i^N W_i^{\mathrm{DMD}}(k)\cdot x_i(n+1)\right|^2, \end{aligned} \tag{2.11}$$

where k is the current learning iteration. Expressed as a vector in Eq. (2.11), $\mathbf{W}^{\mathrm{DMD}}$ corresponds in the experiment to a binary square matrix as illustrated in Fig. 2.9.

2.2.4/ BIFURCATION DIAGRAMS

With the evolution of the reservoir dynamics being established, we now turn to dynamics and its variation as a function of the feedback gain. It has been observed many times in dynamical systems and in particular in optics [123, 125–127] that a small gain variation can drastically modify dynamics behavior, and one then typically speaks of a "bifurcation". A tool for studying dynamic systems is the bifurcation diagram. It highlights potentially stable states of a dynamical system as a function of a parameter variation and indicates qualitative changes in the dynamical state of a system.

Figure 2.10 shows four bifurcation diagram examples for reservoir node (23,12) recorded for a network of $N = 45 \times 45 = 2025$ nodes. No information was injected ($\gamma = 0$) and the phase offset are fixed. The starting reservoir state is initialized in such a way that the result of the nonlinear function is minimum ($\mathbf{x}^{\mathrm{SLM}}(1) = 66$). In order to statistically measure the state of the system, we let the dynamics evolve for two thousand iterations. However, reservoir state $\mathbf{x}$ is not directly measured. Thanks to a careful nonlinear function characterization for each pixel as illustrated in Fig. 2.2(b), it is however possible to transform the state $\mathbf{x}^{\mathrm{SLM}}(n)$ into reservoir state $\mathbf{x}(n)$ with Eq. (2.9). The two thousand values of $\mathbf{x}$ are compiled into a histogram which translates the state probability in gray scale (colorbar). We scan the feedback gain ranging from 0.1 to 5 with a step of 0.1 in the graphs presented Fig. 2.10.

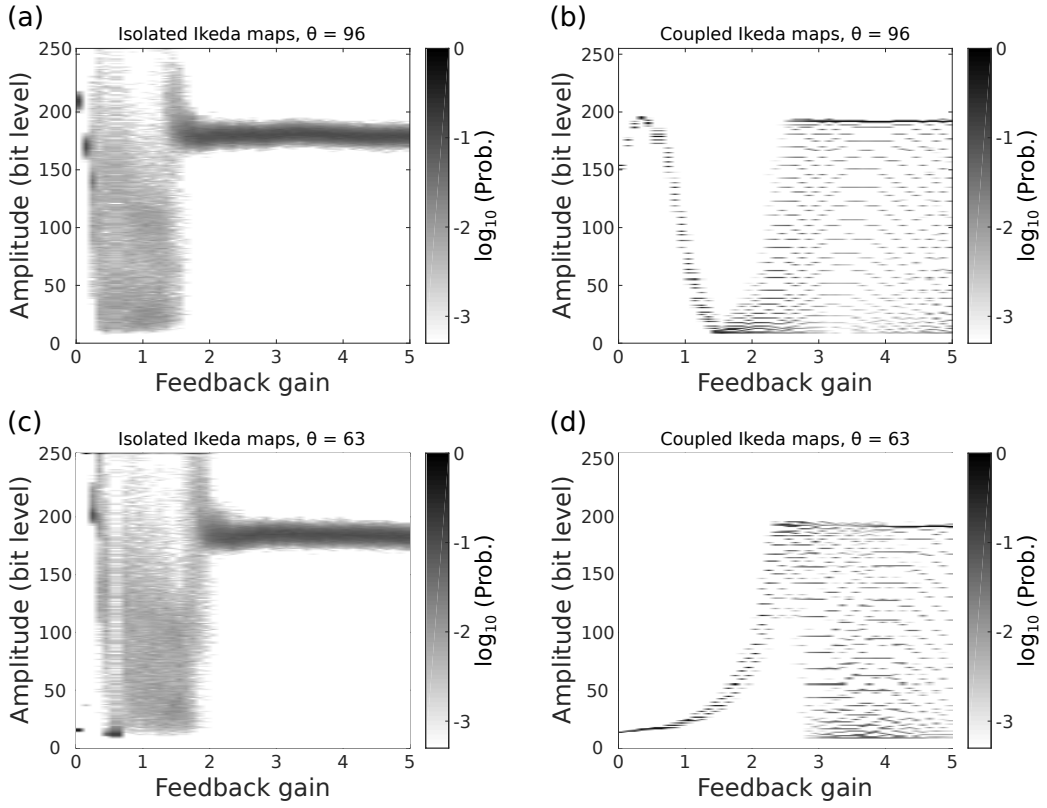

Figure 2.10: Four exemplary bifurcation diagrams for node (23,12). For panels (a) and (c), data was obtained without DOE in the network. For panels (b) and (d), the presence of the DOE excludes direct measurement of the reservoir state. This one is derived thanks to the nonlinear function previously experimentally characterized for all nodes. Panels (a) and (b) are measured with a phase offset $\boldsymbol{\theta} = 96$, for panels (c) and (d) $\boldsymbol{\theta} = 63$.

The data of panels (a) and (c) are obtained without the DOE, therefore for a self coupled Ikeda map. The data of panels (b) and (d) are obtained with the DOE included in the optical path, and we can observe the strong impact of $\mathbf{W}^{\mathrm{DOE}}$ on the node's dynamics. Panels (a) and (c) were measured for $\boldsymbol{\theta} = 63$ and panels (b) and (d) for $\boldsymbol{\theta} = 96$, where $\boldsymbol{\theta} = 63$ is located around the first minimum of the nonlinear function, $\boldsymbol{\theta} = 96$ is located in the middle of the linear part of the cosine function (see Fig. 2.2(b)).

We see that this change in $\boldsymbol{\theta}$ has little impact on the general dynamics of the uncoupled reservoir. On the other hand, when there is coupling between the nodes, the change is significant (panel (b) and (d)).

For uncoupled Ikeda maps (panel (a) and (c)), there is a stable solution for low value of the feedback gain. As soon as $\beta > 0.2$, the system enters chaotic dynamics. Note in panel (c) that a small stability window is observable for $\beta \approx 0.8$. For $\beta > 2$, the system again stabilized due to the saturation of the nonlinear function. For this particular node, its argument $\mathbf{x}^{\mathrm{SLM}}(n+1)$ exceeds 255 for $\beta > 2$, which results in an artificial restriction. This limitation at 255 causes a stabilization of the node dynamics at $f(\beta x) = 255$.

An important feature of coupling matrix $\mathbf{W}^{\mathrm{DOE}}$ is its ability to increase the steady state of the system. Indeed, by adding the DOE, the slight coupling asymmetry observed in Fig. 2.6 and the higher diffraction orders observed in Fig. 2.4(b) induce inhomogeneity in the coupling topology which is found in Fig. 2.5. As asymmetric coupling can accumulate stronger for some nodes than for others, some nodes have a coupling strength larger than the majority of the network. To avoid saturating the camera, the attenuation of the optical beam by ND filters was therefore increased by 60%. Consequently, the saturation of the SLM is pushed back for beta values greater than 2.5.

In the chaotic regime of the coupled system (panel (b) and (d)), the probability distribution is no longer continuous but certain amplitudes of the available range of 8-bit gray scale range have a greater probability. This is attributed to the inevitable noise when we characterized the nonlinear function (see Fig. 2.2(b)). This function used to derive the reservoir state by approximating the combined action of SLM and PBS, is therefore not smooth. The averaging of multiple measurements of the nonlinear function reduces this effect, while fitting the nonlinear function would remove it entirely

2.3/ PHOTONIC LEARNING

Learning strategies are an important component of neural networks. Most studied and used algorithms such as "gradient descent" or "error back propagation" are methods requiring the knowledge of all nodes states [15], of the connection weights or potentially even their gradients [5]. With a hardware network, knowledge and control of these values are far from trivial, in particular if the system is based on analogue phenomena. When the number of nodes is large the instrumentation allowing the required measurements can quickly become predominant compared to the network itself. Such learning methods are inconsistent with the idea of efficient hardware-based neural networks.

Returning to the concept of reservoir computing presented in section 1.3, learning optimizes readout weights matrix $\mathbf{W}^{\mathrm{out}}$ which, once multiplied onto the reservoir state, will ideally give the desired output result. This class of learning is called "supervised" due to the knowledge of examples of desired results, also being called training target y^T. The

most efficient method to obtain a reservoir's readout weight matrix is matrix inversion. Without taking into account that the matrix inversion computing time increases at least in $O(n^2 \log(n))$ [128] with $n \times n$ the matrix size, this method can only be performed on a computer. Furthermore, we do not wish to (i) measure the network's state, and (ii) Eq. (2.11) is nonlinear, making matrix-inversion impossible. Targeting an all-in-one neural network, we would like to implement an online learning.

Learning must therefore be able to run with little computing power, ideally on a simple electronic circuit. To overcome slow parameter drifts, the learning process must be in real time, allowing a continuously maintain good performance over the long term. In addition, the noise level of our analogue system also has to be taken into account. Therefore we implement local learning based only to the evolution of the system performances.

2.3.1/ EVOLUTIONARY LEARNING

The operating principle of evolutionary learning [129] is to optimize the DMD's configuration by measuring the impact of DMD mirrors' modifications onto computing error $\epsilon(k)$. The objective is then to modify $\mathbf{W}^{\mathrm{DMD}}(k)$ during $k = 1, 2, ..., K$ learning epochs in such a way that output $y^{\mathrm{out}}(K, n+1)$ best approximates training target $y^T(n+1)$ [16].

In the first learning iteration ($k = 1$), the N readout weights $\mathbf{W}^{\mathrm{DMD}}(1) \in \mathbb{Z}\{0, 1\}$ are randomly initialized, the first output signal $\mathbf{y}^{\mathrm{out}}(1)$ is measured and error $\epsilon(1)$ is determined. For the next ($k = 2, ..., K$) learning iteration, the Boolean learning algorithm can be divided into three conceptual sections:

MARKOVIAN MUTATION

$$\mathbf{W}^{\mathrm{select}}(k) = \mathrm{rand}(N), \tag{2.12}$$

$$l(k) = \max(\mathbf{W}^{\mathrm{select}}(k)), \tag{2.13}$$

$$W^{\mathrm{DMD}}_{l(k)}(k+1) = \neg W^{\mathrm{DMD}}_{l(k)}(k). \tag{2.14}$$

The vector $\mathrm{rand}(N)$ is composed of N random elements, independently and identically distributed between 0 and 1. The function $\max(\cdot)$ returns the largest entry's position $l(k)$ of $\mathbf{W}^{\mathrm{select}}(k)$. This selected position determines the Boolean readout weight $\mathbf{W}^{\mathrm{DMD}}_{l(k)}(k)$ to be mutated through the logical inversion operator $\neg(\cdot)$, *i.e.* the NOT operator.

Equations (2.13) and (2.14) indicate that only one readout weight is mutated at each learning epoch. According to Eq. (2.12), the readout weight matrix evolution follows a "Markovian process", in other words mutations depend only on the present and knowledge of the past has no impact. By definition vector l is then a discrete-time Markov chain.

ERROR AND REWARDS SIGNAL

$$\epsilon(k) = \frac{1}{T} \sum_{n=1}^{T} \left[y^T(n+1) - \tilde{y}^{\mathrm{out}}(k, n+1) \right]^2, \tag{2.15}$$

$$r(k) = \begin{cases} 1 & \text{if } \Delta\epsilon(k) < 0 \\ 0 & \text{if } \Delta\epsilon(k) \geq 0 \end{cases}, \tag{2.16}$$

$$\epsilon^{\min}(k) = [1 - r(k)]\epsilon(k-1) + r(k)\epsilon(k), \tag{2.17}$$

$$k^{\min} = [1 - r(k)]k^{\min} + r(k)k. \tag{2.18}$$

In order to compare the training target y^T having values between 0 and 1, the output y^{out} is normalized. We subtracted its mean and divided it by its standard deviation, resulting in signal $\tilde{y}^{\text{out}}$. For each learning epoch, mean square error $\epsilon(k)$ is then obtained from a sequence of T data points according to Eq. (2.15). In Eq. (2.16), $\Delta\epsilon(k) = \epsilon(k) - \epsilon(k-1)$ is the error variation induce by the weight modification. Reward $r(k)$ determines then if the system's performance has improved compared to the previous learning epoch. If $r(k) = 1$, the minimum error $\epsilon^{\min}(k)$ and the best learning epoch $k^{\min}$ are updated following Eqs. (2.17) and (2.18).

DESCENT ACTION

$$W^{\text{DMD}}_{l(k)}(k) = r(k) \cdot W^{\text{DMD}}_{l(k)}(k) + [1 - r(k)]W^{\text{DMD}}_{l(k)}(k-1). \tag{2.19}$$

Based on reward $r(k)$, the current configuration of the DMD is modified according to Eq. (2.19). The new DMD configuration is kept if its performance has been improved by the current modification. Otherwise the DMD returns to its previous configuration.

Technically, the exploration strategy is a stochastic gradient descent where the Eqs. (2.16) and (2.19) reinforce modifications which were found beneficial. The learning algorithm is currently executed from the computer control via Matlab for convenience. However, to avoid using a computer, the algorithm could be embedded in a microcontroller with the following instructions which summarize the learning routine: choose a DMD pixel to modify, measure the neural network output, compare its value with the previous one in function of the target, and according to the result, keep modification or return to the previous DMD configuration.

2.3.2/ GREEDY LEARNING

In order to increase the algorithm's efficiency, the exploration of the error landscape is favoured, generating therefore a "greedy learning". Compared with the Markovian learning algorithm, only the section concerning the mutation is slightly modified. Equation (2.12) is transformed and a fourth step is added to the routine. The new version of the mutation is the following:

$$\mathbf{W}^{\text{select}}(k) = \text{rand}(N) \cdot \mathbf{W}^{\text{bias}}, \tag{2.20}$$

$$l(k) = \max(\mathbf{W}^{\text{select}}(k)), \tag{2.21}$$

$$W^{\text{DMD}}_{l(k)}(k+1) = \neg W^{\text{DMD}}_{l(k)}(k), \tag{2.22}$$

$$\mathbf{W}^{\text{bias}} = \frac{1}{N} + \mathbf{W}^{\text{bias}}, \quad W^{\text{bias}}_{l(k)} = 0. \tag{2.23}$$

The matrix $\mathbf{W}^{\text{bias}} \in [0, 1]$ in Eq. (2.20) is randomly initialized at $k = 2$ with entries uniformly distributed between 0 and 1. To favour the exploration of untested weights, see Eq. (2.23), the entries of $\mathbf{W}^{bias}$ are increased by $\frac{1}{N}$ at each epoch, and, crucially, its entry at position $l(k)$ is set to zero. Once the selection-bias for a modified weight is reset to 0, it increases linearly until approaching unity N learning iterations after. A recently optimized weight is therefore statistically less likely to be quickly reselected. In simulation and in experiment, the greedy mutation shows a faster learning convergence, which we attribute to a better efficiency to explore the relevant dimensions of $\mathbf{W}^{\text{DMD}}$.

2.4/ TESTING

2.4.1/ BENCHMARK TASK

Our learning routine described in section 2.3 can be applied to optimize a neural network to solve a variety of computational classes. Indeed, according to Eq. (2.15), learning is ultimately a least square optimization. Due to the relevance of recurrent neural networks for dynamical signal processing, we explore one-step-ahead prediction of a chaotic sequence. All the results presented are obtained from prediction of the chaotic Mackey-Glass sequence [130], which is a common benchmark test for the evaluation of such systems [15, 131, 132].

In general terms, the system settings are as follows:
Parameters of Mackey-Glass sequence are identical to [16], using an integration step size of 0.1. Two hundred points of the chaotic sequence are used as injected training signal $u(n + 1)$. Since the task is a forecast, the target signal $y^T(n + 1)$ is therefore $u(n + 2)$. The first thirty values of the output are removed due to their transient nature, the mean of the remainder is subtracted, then normalized by its standard deviation, resulting in signal $\tilde{\mathbf{y}}^{\text{out}}$ as mentioned section 2.3.1.

2.4.2/ MITIGATION OF UNIPOLAR SYSTEMS LIMITATION

In general, the state of artificial neurons uses real numbers, node states or connection weights can therefore be positive or negative. However, our system faces a challenge which was already presented during the first realization of an optical neural network [110]. Signals in the system is always positive and the readout or internal connections between nodes are only constructive. The illumination of the SLM has a positive intensity, the polarization filtering realizing the nonlinear function has only a positive effect, as well as the internal coupling by diffraction and the binary readout weights. Finally, the optical detectors are sensitive only to the modulus square of the electric field.

The only possibility of obtaining a difference operation in an optical system is to create destructive interferences. An interferometer requires an optical path precision of the wavelength's order, under these conditions the temperature and the mechanical vibrations must be controlled with precision. Consequently, to avoid having a complex system which is hypersensitive to all external parameters, this choice was not retained.

The system functionality is then significantly limited by this unipolarity. Indeed, the multiplication of a negative weight allows not only to perform a subtraction between nodes,

but also to modify the symmetry of the node's nonlinear transformation. A first assessment based on the prediction of a chaotic Mackey-Glass sequence shows that the neural network has limited performance.

We then implemented a strategy to mitigate the unipolarity of the system by taking advantage of the periodic $\cos^2(\cdot)$ nonlinearity. We distribute phase offset matrix $\boldsymbol{\theta}$ randomly around two values. After testing different combinations, the optimal performance is obtained for the first value $\theta_0 = 42 \triangleq 0.17\pi$ slightly before the first minimum of the nonlinear function and the second value $\theta_0 + \Delta\theta = 106 \triangleq 0.43\pi$ close to the next maximum (see Fig. 2.11(a)). Finally, in order to diversify the phase offset of each SLM pixel we assign to θ_0 not directly a constant value but a randomly Gaussian distribution $G(\cdot)$ in the vicinity of θ_0 optimum value. As illustrated in Fig. 2.11(b), μ nodes have a phase offset $\theta_i = G(\theta_0)$ and then have a response featuring a positive slope, $1 - \mu$ nodes a negative slope with a phase offset $\theta_i = G(\theta_0) + \Delta\theta$. Using the positive and negative slopes of the nonlinear function breaks the reservoir state symmetry to a certain extent.

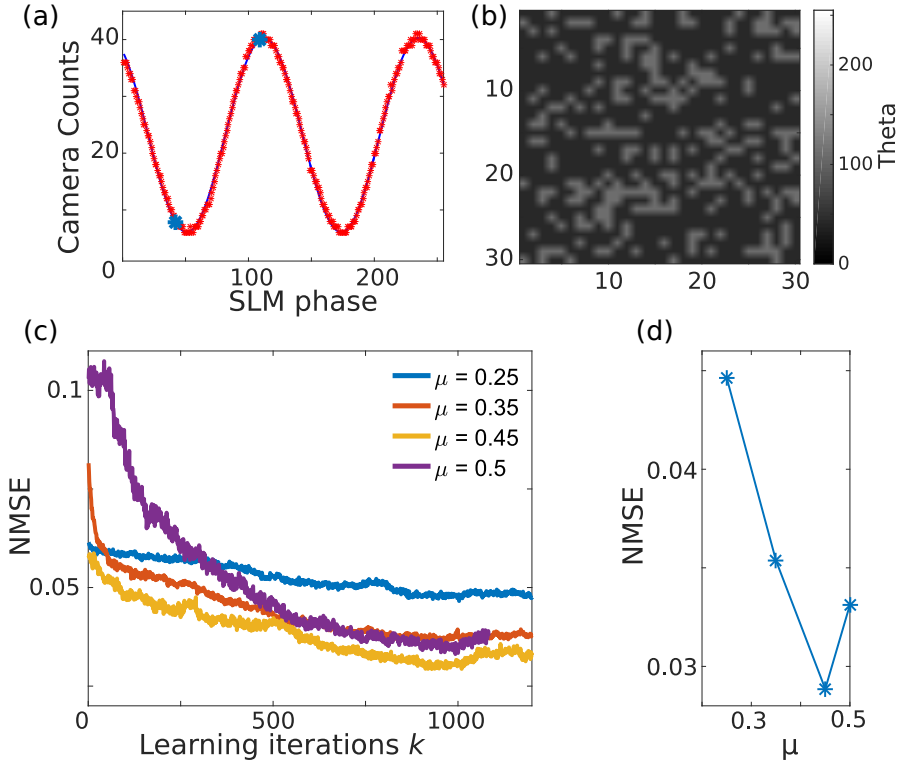

Figure 2.11: (a) Nonlinear function example of a node (red stars). According to a probability μ, the phase offset of each node is distributed around the first minimum of the nonlinearity and the following maximum (blue stars). (b) Spatial distribution example of the phase offset within the network nodes. (c) Learning curves for different probabilities μ, using $\gamma = 0.25$ and $\beta = 0.2$. (d) Results of the best performance for each probability μ of data panel (c). The phase offset distribution of the nodes around the two operating points breaks the symmetry of the nodes response and has a strong impact on the system performance.

In Figure 2.11(c) we see the influence of the probability μ in the learning of a chaotic Mackey-Glass sequence prediction. The blue, red, yellow and purple learning curves

were obtained for $\beta = 0.2$, $\gamma = 0.25$ and probability-ratios $\mu = [0.25, 0.35, 0.45, 0.5]$, respectively. The best performance of each learning is reported in Fig. 2.11(d). We observe that the symmetry breaking of the node response has a strong impact on the performance of the system. The best ratio between the two operating points is obtained for $\mu = 0.45$. It is important to note that changing μ from 0.25 to 0.45 reduces the system's prediction error by almost 50%. This strong improvement demonstrates to the effectiveness of this method. Consequently, the absence of negative weights in the readout layer and in the internal connectivity can be partially compensated for by using of a nonlinear function having a positive as well as negative slope. Since many optical neural networks are subject to unipolarity, this result is highly significant for neural networks hardware implementation.

2.4.3/ PERFORMANCE

The results of the neural network which will be presented in this section and have already been published in [16], were obtained with the optimized parameters $\beta = 0.4$, $\gamma = 0.25$, $\theta_0 = 42 \triangleq 0.34\pi$, $\theta_0 + \Delta\theta_0 = 106 \triangleq 0.86\pi$, $\mu = 0.45$. The number of nodes is $N = 961$, the training sample has a size of 200 steps.

Figure 2.12(a) shows the convergence of a learning curve where the prediction error (blue line) is efficiently reduced until convergence at $\epsilon = 0.013$. After learning, the prediction performance is tested on a sequence of 4500 consecutive data points which are independent of the training dataset. The result is illustrated by the red line on the same panel. With a slightly higher error than training, the prediction test passed successfully. This result demonstrates that the neural network is capable of extracting and synthesizing the general properties of a chaotic signal coming from a restricted sample during its learning phase. More details regarding this performance can be seen in Fig. 2.12(b) which shows 700 data points of the chaotic Mackey-Glass sequence used to test the neural network prediction. The reservoir output power is represented along the left vertical axis (blue line), while the normalized target signal is on the right vertical axis (red dots). The difference between the two signals is difficult to see, the prediction error ϵ (yellow dashed line) remains small throughout the test.

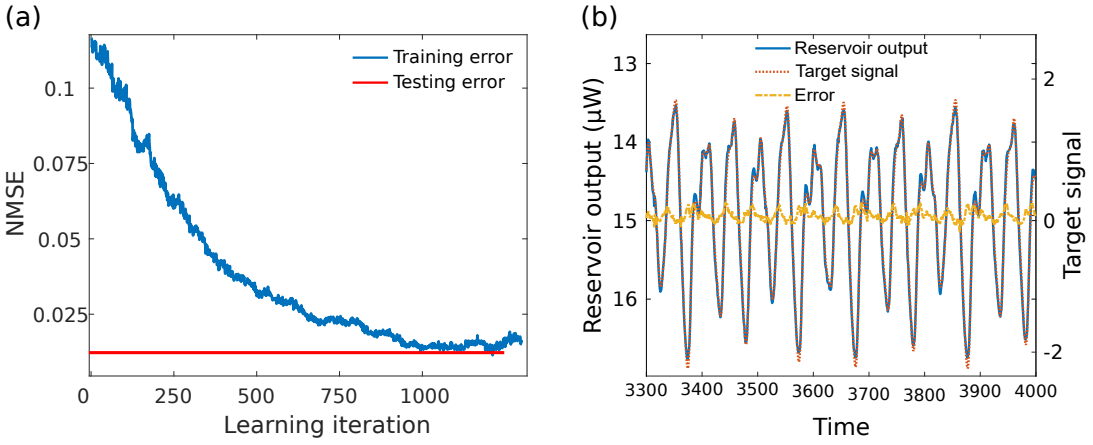

Figure 2.12: (a) Learning curve (blue line) and testing error (red line) at optimal parameters ($\beta = 0.8$, $\gamma = 0.4$ and $\mu = 0.45$). (b) Reservoir output power (blue line) predicting a chaotic Mackey-Glass sequence (red dots). The prediction error is given by the yellow dashed data.

In order to compare the performances obtained with other systems [131, 132], the injected signal is down-sampled by a factor 3 to be in similar conditions. In comparison with the reservoir computing system based on a semiconductor laser [131] and with the system based on a Mach-Zehnder modulator [132], the prediction error of our system under these conditions ($\epsilon = 0.042$) was larger by factors 2.2 and 6.5, respectively. Nevertheless, this result must be interpreted taking into account the difference of hardware implementation level of these three neural networks. Indeed, the readout weights of the two other systems were applied digitally in an off-line procedure using weights with double precision. In [132] authors identify that computational performance is strongly impacted by the digitalization resolution. This result then suggests that the error of our system can be significantly reduced by increasing the resolution of the readout weights.

2.5/ SUMMARY

In this chapter, we have presented a first demonstrator of an optical reservoir computer in which many vector matrix products were optically realized fully in parallel. All stages presenting a strategic interest or a technological challenge have been successfully implemented in hardware. The physical output layer and the idea of a learning algorithm that can be implemented in hardware are the first steps towards a true autonomous neural network. The entire system has been characterized and each element has been discussed in detail throughout this chapter.

In summary, the nodes are the SLM pixels, thus electro-optical. Their cosine analogue nonlinearity comes from polarization filtering performed by a PBS. Interconnectivity between nonlinear nodes is achieved by optical diffraction. This analogue and parallel process currently creates a coupling between a thousand nodes and has demonstrated its scalability, allowing a coupling of hundreds of thousands of nodes. The evolution of the reservoir dynamics is described by a coupled Ikeda map where all the matrix products except the information injection are realized in parallel by physical processes. The network readout is also fully implemented in hardware via a digital micro-mirrors array which creates Boolean weights. In order to demonstrate the autonomy of such a neural network, online learning has been implemented. This scheme acts in real time acting in real time, can fundamentally be implemented by low level on-board electronics. The algorithm concept is simple and its ability to efficiently explore an error landscape according to a scheme called "greedy learning" makes learning faster. The problem of unipolarity which affects many optical neural networks has been partially mitigated by modifying the symmetry of the nonlinear response of the nodes. Splitting the phase offset matrix around two operating points makes it possible to use the two slopes of the nonlinear function and to significantly increase the performance of the system. Once this neural network and its learning algorithm have been completed, the ability of the system to predict a step ahead of a chaotic Mackey-Glass sequence has been tested. Learning is really effective since the system is able to generalize its learning to an unknown sequence, the prediction error then being around 1%. This value, similar to the best performance during learning, attests to the non overfitting.

This experiment highlights all the interest of an analogue system, taking advantage of a massive parallelization enabled by optics. The nonlinear nodes are arranged on a 2D surface but the coupling takes place according to the third dimension. The classic problem related to the interconnectivity infrastructure size growing faster than the increase in the

nodes number, thus limiting it, disappears with this paradigm shift. Indeed, by the intrinsic nature of light, all the networks information can be contained in the system's optical signals along the third dimension without mixing. These characteristics allow a great energy and time efficiency, and a transition to neural networks of much larger scale.

3

FEATURES OF A HARDWARE EMBODIED OPTICAL RECURRENT NEURAL NETWORKS

Unlike a digital system, where discretization by thresholding practically eliminates noise and its propagation, the electro-optical hardware neural network presented in the previous chapter is inherently noisy. Indeed, all analogue processes are subject to noise. The main elements of the network like the nonlinear nodes, the coupling and the readout weights are all analogue, they therefore do not stop the propagation of noise and on the contrary modulate it, even could amplify it. Noise is therefore an intrinsic part of ANNs. Since we are fundamentally unable to remove it, it is essential to study and to understand the impact of noise on the characteristics and performances of the system.

On one hand, noise has advantages, it is for example artificially added in reservoir computing simulations in order to avoid overfitting [133]. Noise adds a universality characteristic linked to the generalization property of reservoir computing. This characteristic could even be essential to adapt to the "real world" which itself is subjected to constant perturbations. The example of the biological brain shows us its exceptional efficiency in information processing while neuronal activity is noisy. [134]. Therefore, such a neural network seems very resilient to noise.

However, we will see in this chapter that perturbations have a real impact on the neural network itself. In the first part, we will give a quick overview of noise sources inside our electro-optical neural network and their characteristics. Then we will focus on the noise impact on the system's dynamics and study the resilience of the reservoir's response. Finally, a detailed investigation on learning under noise [135] will be presented, in which many characteristics emerge but each new answer appears to lead to more questions.

3.1/ NOISE INVESTIGATION

In this section we will first present experimental observations, then a first model will be proposed. We will discuss specific characteristics of the system that will allow us to better understand the impact of noise on the neural network's dynamics and its interaction with learning, which will be discussed in the following sections.

3.1.1/ NOISE DISTRIBUTION

We will focus in this part on noise sources inside the reservoir. For this characterization, no input information is injected into the network, so injection strength γ in Eq. (2.7) is set to zero. In addition, the evaluation is performed with and without coupling between nonlinear nodes. As a passive element, the DOE does not add noise to the system, but it can modify its distribution.

Since the impact of noise is intrinsically linked to the level of the signal, we will use the signal to noise ratio (SNR) defined as

$$\mathrm{SNR}(\tilde{x}_i^C) = \frac{\mathrm{mean}(\tilde{x}_i^C)}{\mathrm{std}(\tilde{x}_i^C)}, \tag{3.1}$$

where $\mathrm{mean}(\cdot)$ and $\mathrm{std}(\cdot)$ are the average and the standard deviation of the signal $\tilde{x}_i^C$ for pixel i of the camera. Remember that in the absence of DOE the camera directly records the reservoir state $\mathbf{x}$, whereas we no longer have access to it in the presence of DOE in the experiment.

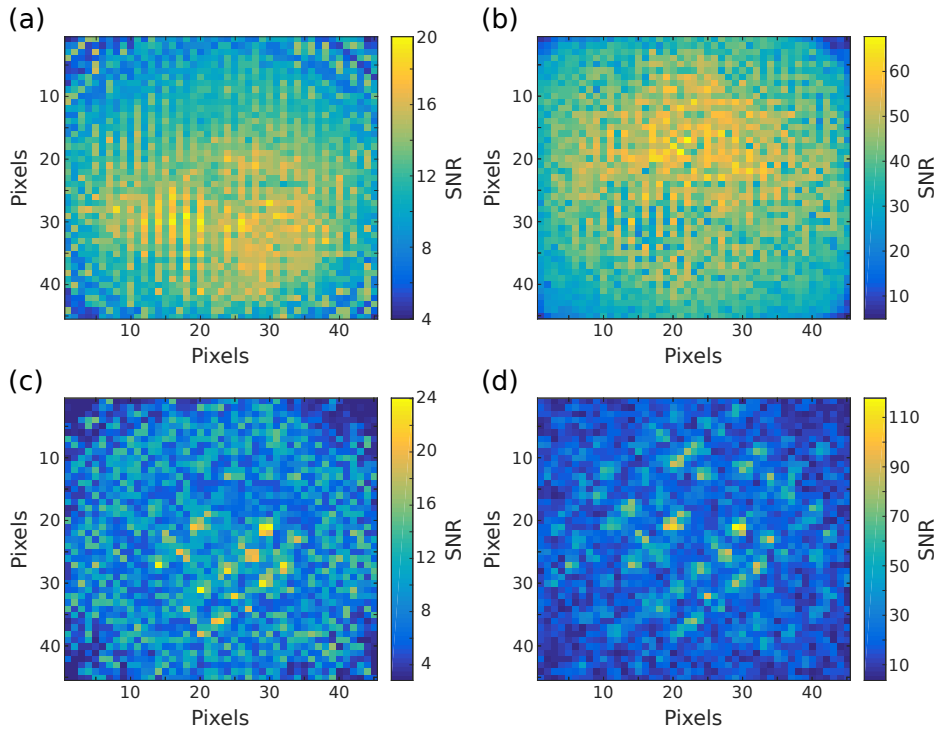

Figure 3.1: Spatial distribution of the signal to noise ratio (SNR). (a) and (c) are measured for a phase offset $\theta = 63$, while $\theta = 126$ for (b) and (d). We observe that in the absence of coupling between nodes (panels (a) and (b)) the SNR distribution is close to the Gaussian intensity illumination of the SLM, while in the presence of the DOE in the experiment (panels (c) and (d)), the SNR becomes heterogeneous, forming clusters.

Figure 3.1 shows the spatial SNR distribution in several configurations, in particular for values of phase offset θ discussed in 2.4.2. The characterization was performed over approximately sixty hours with a reservoir time scale of a few Hertz, resulting in 60,000 data points per pixel. On panels (a) and (c) the phase offset is $\theta = 63$, on panels (b) and (d) $\theta = 126$. These two values represent the parts of the nonlinear function which are

used in order to mitigate the unipolarity of our system as explained in section 2.4.2. For this characterization all the SLM pixels have the same phase offset, *i.e.* $\mu = 1$. Panels (a) and (b) are the result of the experiment without DOE, panels (c) and (d) with it.

As the illumination intensity of the SLM is Gaussian, the maximum intensity is in the center of the network. We therefore find this spatial distribution in configurations without the DOE. In panel (a), the maximum of SNR is 20 but decreases rapidly when approaching the edges of the network. The phase offset of 126 for panel (b) corresponds to a much higher light intensity in the reservoir, the SNR then increases by more than a factor of 3. In this configuration the signal of nearly all nodes has a high SNR. The introduction of the DOE into the reservoir changes the spatial distribution of noise. The coupling between the nodes distributes the light non-uniformly, and the signal amplitude for some nodes exceeds significantly the one of the remaining nodes. Consequently, we have to add additional neutral density filters in the experiment to avoid saturating the camera, further reducing the SNR of the nodes already having a low light intensity. We therefore find for the corresponding data shown in panel (c) a much greater heterogeneity, with clusters of nodes having a large SNR surrounded by much more noisy signals. In panel (d), the spatial distribution is even less uniform: a large part of the nodes have a relatively low SNR while clusters, with a wider spatial distribution, reach very high values. Finally, even though the DOE reorganizes the spatial distribution of the SNR, it still remains a function of a pixel's signal amplitude.

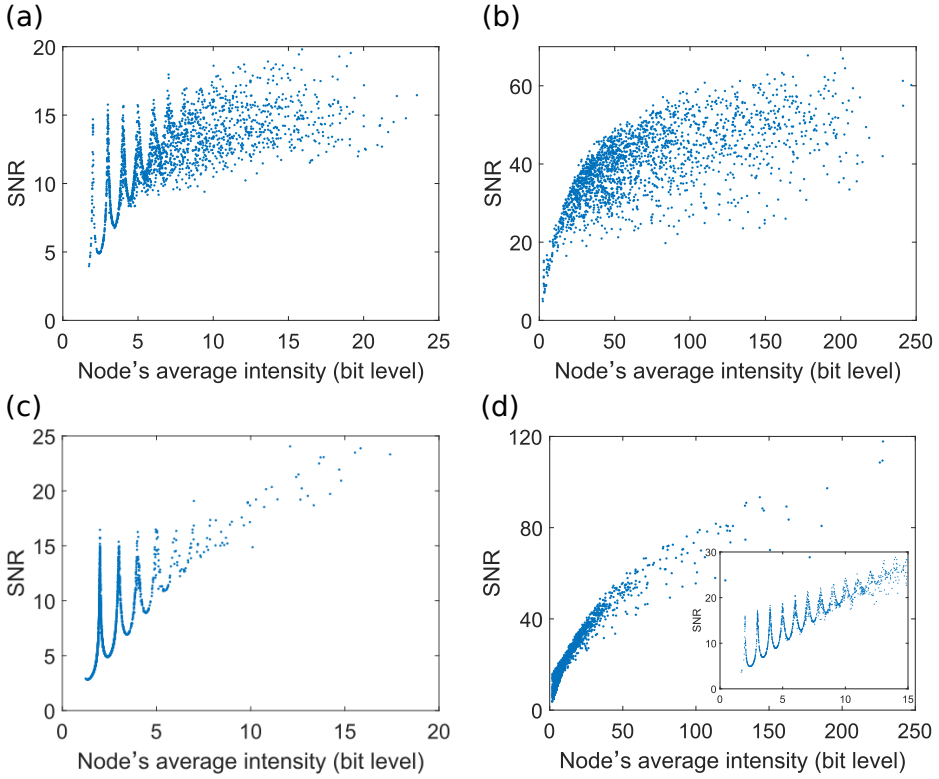

Figure 3.2: SNR as a function of the average intensity of each node. For (a) and (c) the phase offset is $\theta = 63$, for (b) and (d) $\theta = 126$. Without the DOE, panels (a) and (b), the SNR is distributed along the x-axis while in (c) and (d) coupling pushes the SNR of the majority of the nodes towards the bottom of the distribution. In (a), (c) and the zoom in the inset of (d) we observe SNR oscillations for low average intensities, consequence of the digitization of the analogue signal by the camera.

Data in Fig. 3.2 were obtained from the same measurements previously discussed, but the SNR is represented as a function of the average intensity of each node. The different panels correspond to the same conditions as in Fig. 3.1. We observe that the SNR non-linearly increases with the average node intensity $\mathrm{mean}(\mathbf{x})$. Without the DOE (panels (a) and (b)) the distribution is relatively well balanced, a large number of nodes having a SNR at the top of the distribution. The presence of the DOE (panels (c) and (d)) reduces the SNR of the majority of nodes at the start of the distribution, however, for some nodes the SNR is significantly increased.

Another interesting feature is the oscillations at the start of the distribution which can be observed in panels (a), (c) and in the zoom inset of (d). We observe that the maximum of the peaks corresponds to signal amplitudes where the average value is an integer. When the average intensity is between two integers, the digital camera rounds the analogue value to the nearest integer. Between two integers a small perturbation by noise can therefore lead to a significantly larger signal modification and an SNR oscillation with unity as its period. This variation causes a very significant variation in the SNR mostly for small average signal amplitudes, and these are stronger for low signal average intensities, panel (b) therefore does not present this feature. Importantly, this behaviour is not an artefact but an intrinsic feature of digitization if present in ANNs and must therefore be taken into account.

3.1.2/ CHARACTERIZATION AND SNR MODELLING

Several types of noise have to be considered, and their implications on the signal may be different. Noise sources are intrinsic to the system, such as the quantization noise or intensity noise of the laser. A second category is external noise, such as stray-light, temperature variations, mechanical vibrations. These perturbations can cause noise but mostly lead to long-term drifts. Finally, as we have seen in Fig. 3.2, there is digitization noise during analogue to digital conversion.

All these noises manifest themselves in different ways. The contribution of noise can be additive or multiplicative to the signal. Additive noise amplitude D_A is added to the noiseless signal $\tilde{s}$, hence the noisy signal s becomes $s = \tilde{s} + D_A$. In the case of multiplicative noise amplitude, the noiseless signal is multiplied by the multiplicative noise D_M, the noisy signal then becomes $s = \tilde{s} \cdot D_M$.

Figure 3.3 schematically illustrates the evolution of the SNR according to the influence of different types of noise. For additive noise in panel (a), the SNR evolves linearly with the mean of the signal, whereas for a multiplicative noise in panel (b), the SNR is constant. Indeed, according to Eq. 3.1 the SNR of the additive case increases linearly with the average growth since the standard deviation remains constant. In the multiplicative case, the average and standard deviation of the noise both increase linearly, therefore the SNR remains constant. Panel (c) is the characteristic of a *mixed noise*, the additive character dominates for a small signal amplitude signal, then the multiplicative aspect takes over for a larger signal amplitude. The SNR measurements illustrated in Fig. 3.2 present a curved trajectory, proof of the presence of mixed noise in the neural network. In order to identify the origin of this property, we characterized all components of the network individually, and identified the camera as the source of additive and multiplicative noise.

The nodes of the network are spatially distributed, and it is important to identify whether noise is different for each node or, on the contrary, if a single source of noise is common to

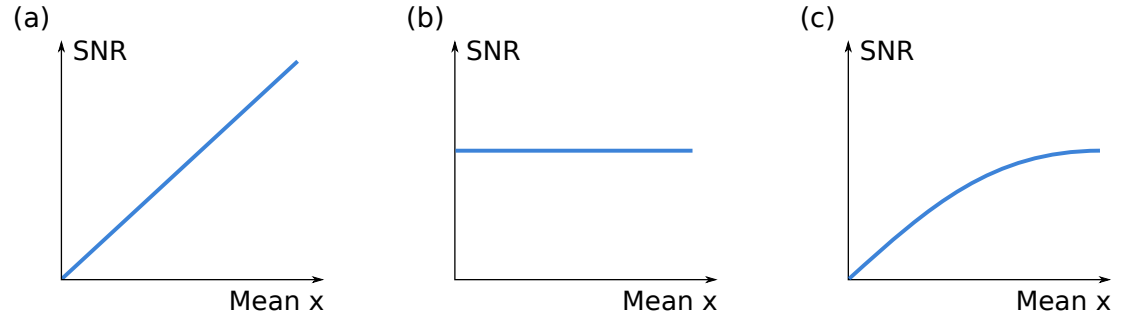

Figure 3.3: Signal to noise ratio characteristic according to the mean of the signal for an additive noise (panel (a)), a multiplicative noise (panel (b)) and a mixed noise (panel (c)).

all nodes. The former is uncorrelated and the latter is correlated noise. For that purpose we calculated the normalized cross-correlation $\Phi_{i,j}$ between the SLM pixels according to

$$\Phi_{i,j} = \frac{\text{mean}(\hat{x}_i(t) \cdot \hat{x}_j(t))}{\sqrt{\text{mean}(\hat{x}_i^2(t)) \cdot \text{mean}(\hat{x}_j^2(t))}}, \tag{3.2}$$

where $\hat{x} = x_i(t) - \text{mean}(x_i(t))$ is the deviation from the mean value.

Figure 3.4 illustrates the cross-correlation between noise of one node (red pixel) and the rest of the network. We observe a weak correlation when the pixel considered is close to the network edge (panels (a) and (c)) while correlation is strong for a pixel in the center (panel (b)). This indicates correlated multiplicative noise. We initially suspected the laser to be the source of this type of noise as amplitude fluctuations would be correlated across all SLM pixels and would result in multiplicative noise. However, the characterization of the laser falsified this hypothesis, as its SNR is larger than one thousand, which therefore is negligible compared to the measurements reported in Fig 3.1.

We additionally found that camera noise is uncorrelated, which therefore leaves the SLM as source of correlated noise. Unlike uncorrelated noise which can be suppressed by averaging a large number of node states, for example via the network's connectivity [136], correlated noise cannot easily be mitigated this way and hence potentially is an important factor for the performance of hardware neural networks.

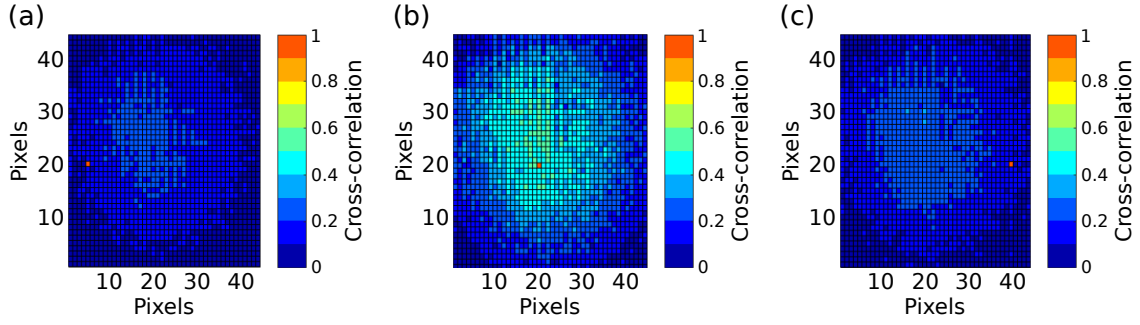

Figure 3.4: Correlation between SLM induced noise of each pixel and that of red pixel (20,5), (20,25) and (20,40) for panels (a) (b) and (c), respectively. The strong variation of the correlation indicates the presence of a multiplicative correlated noise.

Without the DOE and without input information, the reservoir state $\mathbf{x}$ measured directly by the camera is

$$\mathbf{x} = \tilde{\mathbf{x}} + \hat{\mathrm{N}}(\tilde{\mathbf{x}}), \tag{3.3}$$

where the noiseless reservoir state $\tilde{\mathbf{x}} = \alpha|\mathbf{E}^0|^2 \cos^2[\beta\alpha|\mathbf{E}|^2 + \boldsymbol{\theta}]$ is perturbed by the noise operator $\hat{\mathrm{N}}(\cdot)$. We can then decompose the noisy signal into the noiseless signal and the different sources of Gaussian white noise experimentally observed [136]. The camera is the source of an additive and multiplicative noise of amplitude D_A^C and D_M^C, respectively. The SLM is the source of a correlated multiplicative noise of amplitude $D_{M,\mathrm{C}}^{\mathrm{SLM}}$. The reservoir state can therefore be written according to

$$\mathbf{x} = \tilde{\mathbf{x}} \cdot (1 + D_M^C)(1 + D_{M,\mathrm{C}}^{\mathrm{SLM}}) + D_A^C. \tag{3.4}$$

Since the amplitude of additive noise is attributed only to the camera, obtaining D_A^C is straightforward simply by recording a series of camera images. However, the amplitude of the multiplicative noise comes from both, the camera and the SLM. By calculating the SNR illustrated in Fig. 3.5, Eq. (3.4) excellently matches (black line) the experimental data (red dots) measured for $\boldsymbol{\theta} = 126$. From this fit, we obtained $D_M^C = 5 \cdot 10^{-5}$ and $D_{M,\mathrm{C}}^{\mathrm{SLM}} = 1.1 \cdot 10^{-4}$, and $D_A^C = 0.11$.

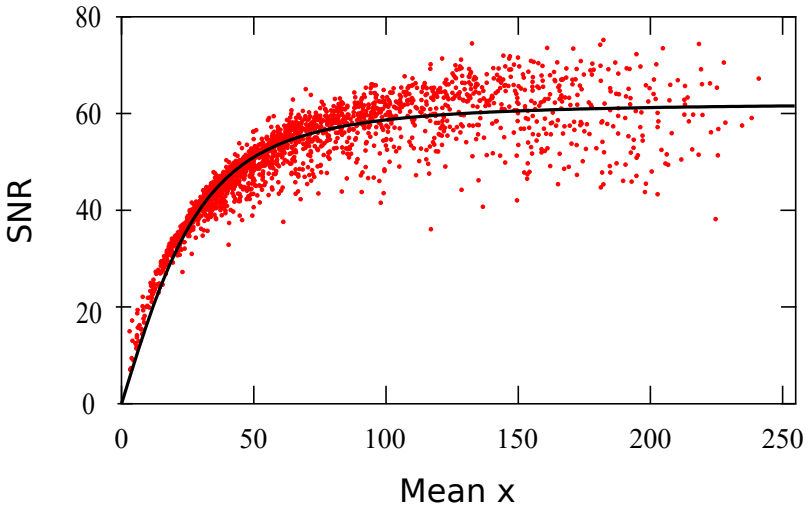

Figure 3.5: SNR against the node's average intensity for a reservoir without the DOE, without input information and for a phase offset $\boldsymbol{\theta} = 126$. Red points are experimental results, the black line is fit via Eq. (3.4).

This experimental characterization allowed us to identify the different noise sources and in particular their different characteristics. The derived Eq. (3.4) excellently describes noise's impact for different amplitudes of the nodes' states. In addition, these experimental results serve as targets and validations for theoretical research [136].

3.2/ CONSISTENCY OF THE DYNAMICAL RESPONSE

In the previous part, we gave an overview of the noise present in the reservoir and studied the resulting steady state characteristics. In this section we will focus on its impact on system dynamics by measuring the consistency of the reservoir's response.

Let us recall the reservoir properties defined in section 1.3.3. Two closely related pieces of input information should lead to two similar reservoir responses, and two sets of input information belonging to a different class should result in different responses. However, in an analogue hardware neural network, these two properties are put to the test. Indeed, the presence of noise could result in similar reservoir dynamics for two different classes of input information, or, on the contrary move two similar examples of input information apart. An essential issue for information processing is that a system computing the same information twice gives a sufficiently similar response answer and we speak of a consistent response. This notion was introduced by Uchida [137] in 2004 in the context of nonlinear dynamical systems as: "consistency is defined as the reproducibility of response waveforms in a nonlinear dynamical system driven repeatedly by a signal, starting from different initial conditions of the system". Mathematically, the consistency C between two states x_1 and x_2 is defined as

$$C = \frac{\mathrm{corr}(x_1 - \mathrm{mean}(x_1)\ ,\ x_2 - \mathrm{mean}(x_2))}{\mathrm{std}(x_1) \cdot \mathrm{std}(x_2)}, \tag{3.5}$$

where $\mathrm{corr}(\cdot,\cdot)$ is the cross-correlation function.

For our characterization we use a network of $\mathrm{N} = 45 \times 45 = 2025$ nodes, feedback strength $\gamma = 0.8$ and phase offset is $\theta = 63$ with $\mu = 1$. We inject into the reservoir the value $u(n)$ at each iteration n, with the input information vector $\mathbf{u}$ made up of 500 random data points, identically and independently distributed between 0 and 1. The evolution of the dynamics of each node is measured by the camera at each iteration. The entire process is repeated a hundred times, always using the same input vector $\mathbf{u}$. Camera state $\tilde{\mathbf{x}}^C(i,n,r)$ is therefore a three-dimensional matrix, with the node number i going from 1 to N, the iteration n ranging from 1 to 500 and the repetition r ranging from 1 to 100. As we want to measure the consistency of the reservoir state $\mathbf{x}$, we calculate it in post-processing from $\tilde{\mathbf{x}}^C$ using the nonlinear function characterized experimentally for each node.

In order to study the impact of short-term noise but also the impact of the system's long-term drifts, we studied the consistency on two time scales. Short term consistency C_{ST} is calculated with two consecutive repetitions with r and $r+1$, while the long term consistency C_{LT} is always calculated between the first $r = 1$ and the current repetition r:

$$C_{\mathrm{ST}}(i,r) = \frac{\mathrm{corr}\Big(x(i,r) - \mathrm{mean}(x(i,r))\ ,\ x(i,r+1) - \mathrm{mean}(x(i,r+1))\Big)}{\mathrm{std}(x(i,r)) \cdot \mathrm{std}(x(i,r+1))}, \tag{3.6}$$

$$C_{\mathrm{LT}}(i,r) = \frac{\mathrm{corr}\Big(x(i,1) - \mathrm{mean}(x(i,1))\ ,\ x(i,r) - \mathrm{mean}(x(i,r))\Big)}{\mathrm{std}(x(i,1)) \cdot \mathrm{std}(x(i,r))}. \tag{3.7}$$

The spatial distributions of short and long term consistency averaged over all repetitions are shown in Fig. 3.6. The color scale encoding consistency is set from 0.8 to 1 in order to better highlight variations. We can observe that the high values of consistency are aggregated in the center, implying the distribution of illumination intensity as the cause. However, the distribution is not uniform, some nodes at the edge of the network have a high consistency while very close to the center the consistency can locally be very low. Nevertheless, a very large number of nodes have a high consistency, proof that the hardware neural network maintains the consistency property required for reservoirs

despite the presence of noise. Unlike the short term consistency shown in panel (a), the long term consistency in panel (b) is on average slightly lower. However, the distribution remains similar and consistency does not significantly degrade over the more than forty hours during our measurements.

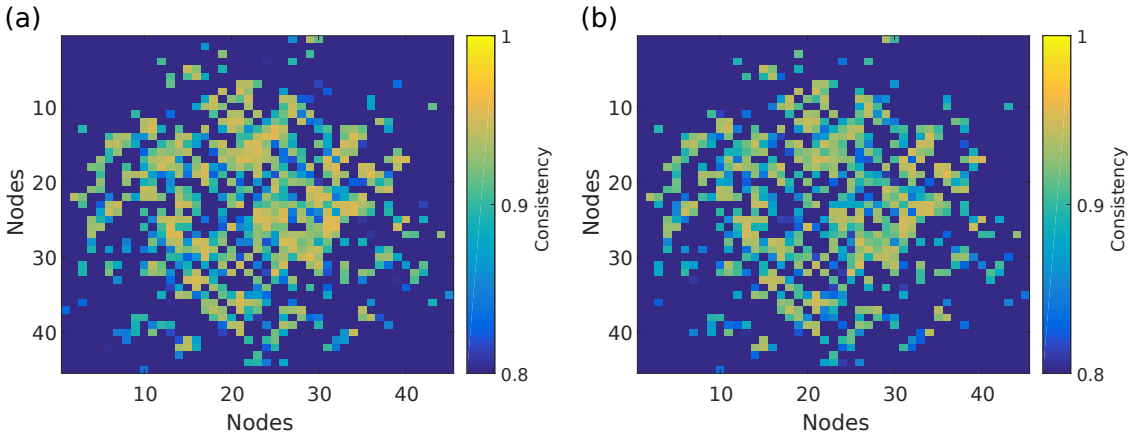

Figure 3.6: Spatial distribution of the short and long term consistency averaged according to 100 repetitions in panel (a) and (b), respectively. Although consistency is generally higher in the reservoir's center, we can observe some clusters outside and very low consistency values in the middle.

Although on average the consistency, both short and long term, is high among the nodes in the center of the network, it is interesting to examine its variation during repetitions. In Fig. 3.7 four different qualitative behaviors are presented for neighboring nodes, with the nodes' row and column coordinates (23,27), (22,27), (22,26), (21,27) represented on the panels (a), (b), (c) and (d), respectively. This shows that these characteristics do not necessarily belong to distinct spatial regions.

We notice that for panel (a) the short-term (blue line) and long-term (red line) consistency are very high, practically identical, with low fluctuations and remain constant throughout the experiments duration. This evolution is typical for nodes with high consistency. For panel (b), the average consistency value is slightly lower, however there is a considerable drop in consistency during one repetition. This type of instability is observed for other nodes and can occur at any time. Such singular behavior therefore seems to be the result of a disturbance very localized in space and time, which suggests noise as the underlying cause. The behavior seen in panel (c) is uncommon for the investigated conditions. While short term consistency remains high and constant, it decreases in the long term. However, this decrease is not constant, we observe a plateau followed by an increase. This variation could raise problems for the processing at different times of two identical tasks. Panel (d) is the more common behavior of nodes with relatively low consistency. Finally, we also find nodes with erratic alternations of very large and very low values over variable periods. As explained in section 2.4, the learning algorithm is based on a comparison of several hundred output measurements. If the consistency of a node evaluated in a particular learning epoch is erratic, its selection by the readout does not change the performance of the system.

The spatial distribution presented in Fig. 3.6 shows that there is on average a higher consistency where illumination is strongest. However, it is important to investigate a potential direct link between consistency and the input weight of each node. Indeed,the

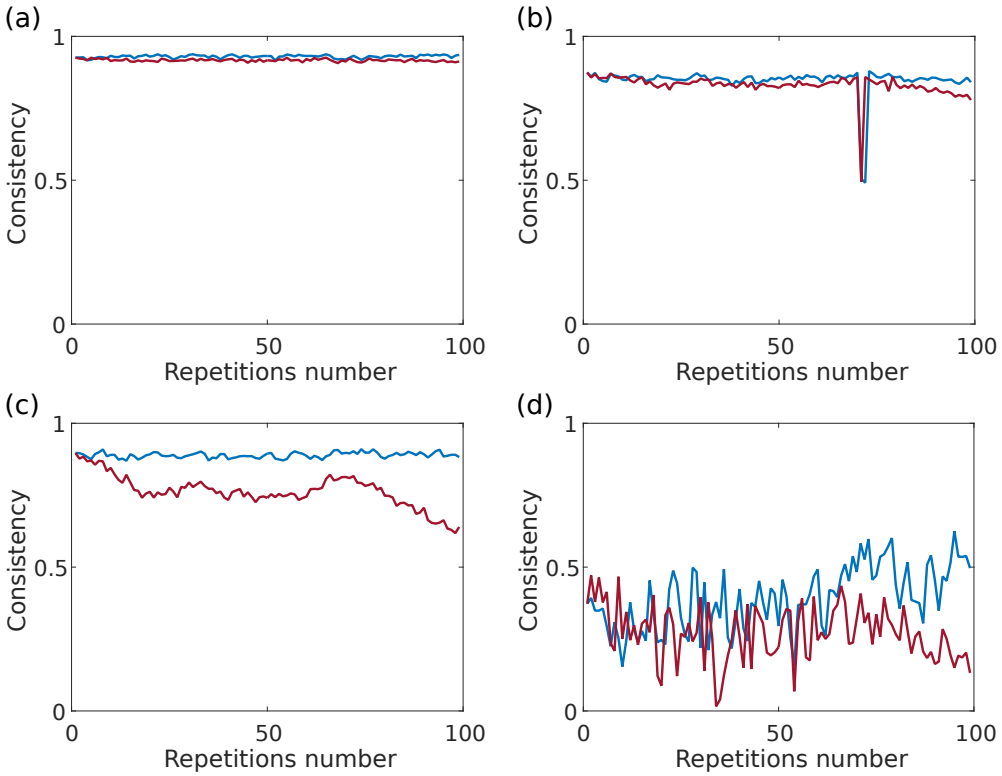

Figure 3.7: Four typical evolutions of short (blue line) and long (red line) term consistency shown with four representative nodes. In panel (a) both lines are constant and have small fluctuations, while in panel (b) a short perturbation happens. In panel (c) exhibits a slow drift, the short term consistency stay constant while the long term consistency decreases. The erratic noisy evolution shown in panel (d) is typical for nodes which have low consistency.

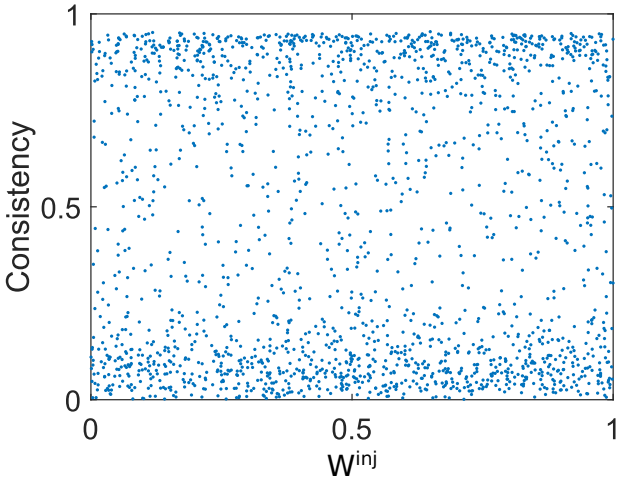

Figure 3.8: Long term consistency of all the nodes in function of their injection strength W^{inj}. There is no clear relationship between both quantities, which means that we can use the entire injection weight range without restriction.

constant injection matrix W^{inj} directly influences the reservoir state of each node. Thanks to Fig. 3.8 which represents the consistency according to the injection weights, we can reject a potential link between these two quantities. It is interesting to see that on average consistency tends to be either high or low, regardless of W^{inj}. Furthermore, short and long term consistency are both limited to below 0.96, and it would be interesting to vary the noise level to potentially observe a quantitative change in the limit.

Our experiments show that short or long term consistency can be high, in particular at the center of the network where illumination is strongest. This indicates that the neural network can process information with limited penalization by ambient noise. Finally, the input weight has no influence on the consistency, which is important since it allows to use the full range of the physically possible injection matrix entries.

3.3/ LEARNING UNDER NOISE-PERTURBATIONS

Important criteria of a learning algorithm are convergence speed and the final error. However, in an analogue hardware system, reproducibility and robustness to noise also play an important role. In this section we will first describe the overall specificities of learning, then study the impact of noise on the network's output reproducibility, investigate the system topology and finally by measuring the scalability of the learning algorithm.

The learning task remains one-step prediction of the chaotic Mackey-Glass sequence. Two hundred points of the chaotic sequence are used as injected training signal $u(n+1)$, and parameters have been optimized to $\beta = 0.8$, $\gamma = 0.25$, $\theta_0 = 55 \mathrel{\hat{=}} 0.44\pi$, $\theta_0 + \Delta\theta_0 = 119 \mathrel{\hat{=}} 0.95\pi$, $\mu = 0.45$. Using the results of noise and consistency studies, we keep the network at $\mathrm{N} = 31 \times 31 = 961$ nodes to have a balance between the size of the system and a high SNR as well as consistency. Finally, the testing error is determined with an independent set of 9000 data-points unused in the training sequence.

We previously discussed learning in a space of $N+1$ dimensions, called the *error landscape*. The position within this topology is given by the N-entries of the system's output weights $\mathbf{W}^{\text{DMD}}(k)$, while the landscape's height along dimension $N+1$ is given by the system's error $\epsilon(k)$. The reward $r(k)$ drives the configuration from $\mathbf{W}^{\text{DMD}}(1)$ to a local minima at $\mathbf{W}^{\text{DMD}}(k^{\min})$ of error value $\epsilon^{\min}$. According to our algorithm, learning varies the position in the error landscape by steps of one. We will refer to one complete learning process for $k : 1 \rightarrow k^{\min}$ as a minimizer.

3.3.1/ GENERAL AND LOCAL FEATURES OF CONVERGENCE

As noise is intrinsic to the system, we measure a statistical series of learning in order to determine average characteristics. We wish to study the algorithm's exploration of the error landscape, and to achieve this, the initial random position $\mathbf{W}^{\text{DMD}}(1)$ is identical for each minimizer. The results of different minimizers (gray curves) are shown in Fig. 3.9, their average is represented by the red crosses. Panel (a) and (b) illustrate 20 greedy minimizers and 14 Markovian minimizers, respectively.

All minimizers differ since the exploration sequence is random for each repetition, yet they all converge towards a low error; none gets stuck in a local minima with bad performance. This first result is very significant because, unlike traditional reservoir computing which

ensures the convexity of the linear readout weight regression [62], the nonlinearity of the optical detector at the network's output does not preserve this property. The error landscape can therefore not be assumed convex, yet measurements in Fig. 3.9 show that on a macroscopic scale there still appears a smooth topology. The final performances are highly reproducible, both in terms of value and speed of convergence. For the greedy exploration in Fig. 3.9(a), we obtained error $\epsilon^{\min} = (14.2 \pm 2.9) \cdot 10^{-3}$ after $k^{\min} = 973.6 \pm 63.7$ learning epochs. Markovian exploration shown in Fig. 3.9(b) arrives in average at a slightly smaller error $\epsilon^{\min} = (13.4 \pm 1.9) \cdot 10^{-3}$ but after a longer learning duration $k^{\min} = 1856.5 \pm 175.1$ [135]. No correlation between the error at the beginning $\epsilon(1)$ and the final performance $\epsilon(k^{\min})$ was found. The testing error, shown by the green line in Fig. 3.9, excellently matches its training error, *i.e.* $15.2 \cdot 10^{-3}$ and $16.6 \cdot 10^{-3}$ for greedy and Markovian exploration, respectively. This result once again excludes over fitting.

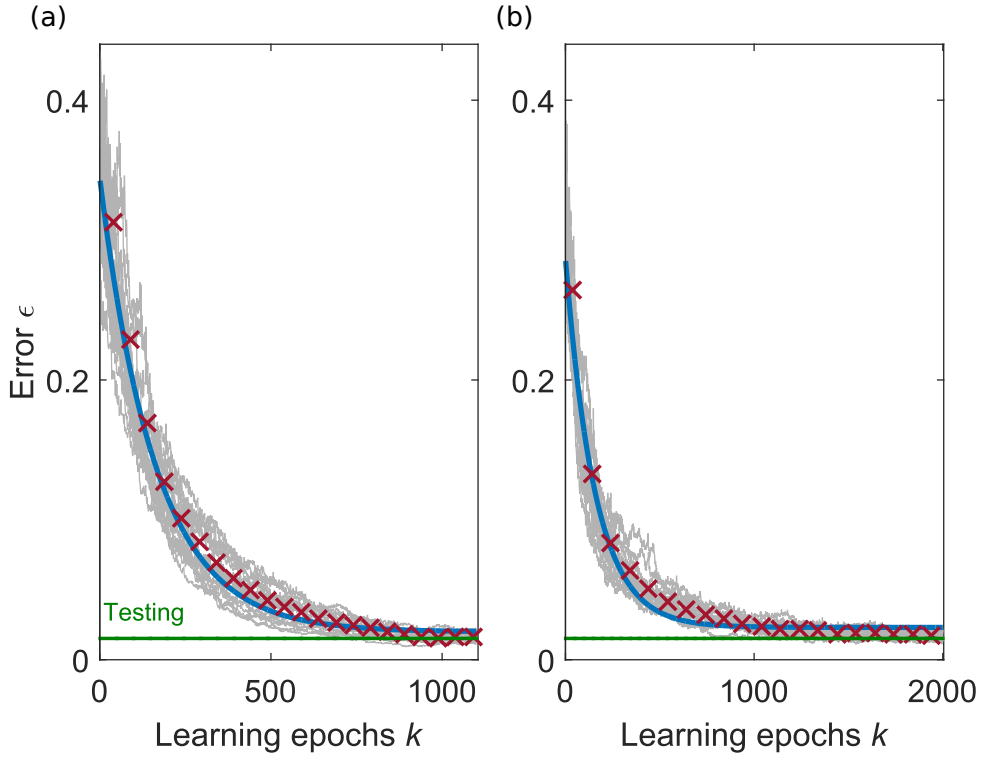

Figure 3.9: Learning performance for greedy and Markovian explorations in (a) and (b), respectively. Red crosses are the average of individual minimizers (gray data). The testing error illustrated by the green line is lower than the average of optimized training error, thus excluding overfitting. The excellent agreement between the fit (blue data) and the average dependency suggests an exponential convergence for greedy learning.

Let us now focus on the learning convergence. Although each minimizer is different, they statistically follow a relatively systematic trajectory. The blue curve in Fig. 3.9 is an exponential fit of the minimizers' average (red crosses). For greedy exploration (panel (a)), the fit ($y = 0.315e^{-0.006} + 0.02$) matches excellently with our data, therefore convergence is on average an exponential decay. This characteristic seems to be less the case for Markovian exploration (panel (b)), where we find a certain discrepancy between the data and exponential decay.

In order to better illustrate the local topology of the error landscape, five minimizers of greedy exploration are shown in Fig. 3.10(a). Zooming in on the first 300 learning epochs highlights the different topological exploration of each minimizer. The blue curve seems to follow a quite smooth descent with a slight slope, while the yellow is much steeper. The other minimizers descend rather in steps, and we even observe sudden jumps.

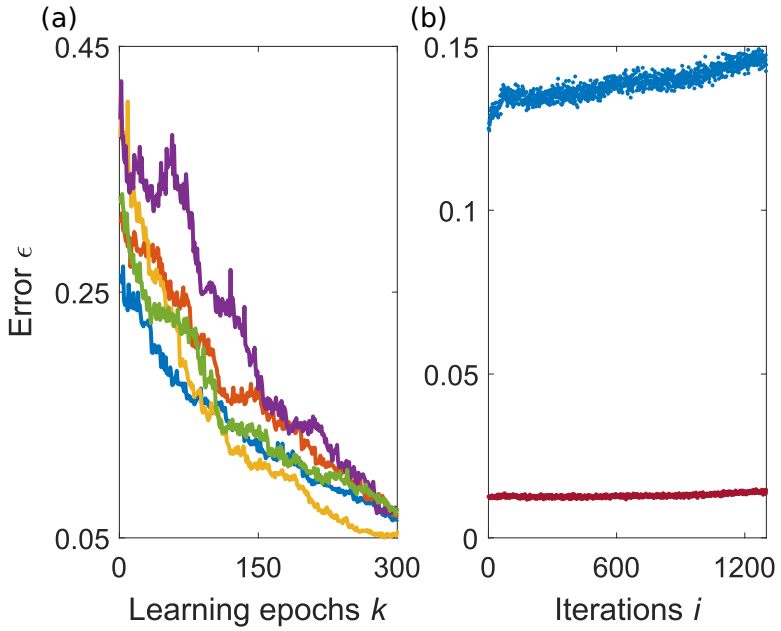

Figure 3.10: (a) Detailed zoom showing a selection of five exemplary learning curves displayed in Fig. 3.9. These curves reveal different learning-trajectories though the complex error landscape. (b) Measures of the prediction error's temporal evolution for two fixed yet different DMD configurations. Blue data is obtained for the DMD configuration after 150 learning epochs, red after 961. Averages are respectively $(13.85 \pm 0.22) \cdot 10^{-2}$ and $(12.9 \pm 0.3) \cdot 10^{-3}$.

Figure 3.10 (b) depicts the temporal stability of the network's output for constant readout matrices obtained after 150 (blue curve) and 961 (red curve) learning epochs. The readout weights alternates between $\mathbf{W}^{\mathrm{DMD}}(150)$ and $\mathbf{W}^{\mathrm{DMD}}(961)$ to simultaneously measure network output's stability for both weight configurations under similar conditions. We observe that both curves undergo a small drift over long time scales. Nevertheless, for optimized readout weights $\mathbf{W}^{\mathrm{DMD}}(961)$, the prediction drifts only by 1.5% relative to the minimum error after 12 hours of characterization and by 10% after 24 hours. The error variation ultimately remains within the training error standard deviation range, the system is therefore quite robust to long term drifts.

Most importantly, we find that the impact of noise changes during the different stages of the learning process. Indeed, the standard deviation of the performance after 150 and 961 learning epochs decreases sharply from $4.4 \cdot 10^{-3}$ to $0.6 \cdot 10^{-3}$ respectively for an error average ranging from $138.5 \cdot 10^{-3}$ to $12.9 \cdot 10^{-3}$. The topology of the error landscape explains this result: the closer we get to a minimum the more shallow becomes the landscape's slope. The same level of system noise therefore disturbs the prediction less and less. The variations of errors during the different minimizers depicted in panel (a) are therefore mostly not related to noise but are the topological consequence of the various error landscape explorations.

Remaining with the study the learning convergence, we calculate the effective error gradient. Thus, at each learning epoch we determine the change of local error relative to the previous lowest error according to $\dot{\epsilon}(k) = \epsilon(k) - \epsilon(k^{\min})$. We define $\dot{\epsilon}^+$ and $\dot{\epsilon}^-$ which contains all $\dot{\epsilon}$ which are positive and negative, respectively. Then we calculate the average of all $\dot{\epsilon}^+$ and $\dot{\epsilon}^-$ obtained at each epoch k for each minimizer, see red and blue data respectively in Fig. 3.11. Therefore, blue data corresponds to a error reduction and red to an increase.

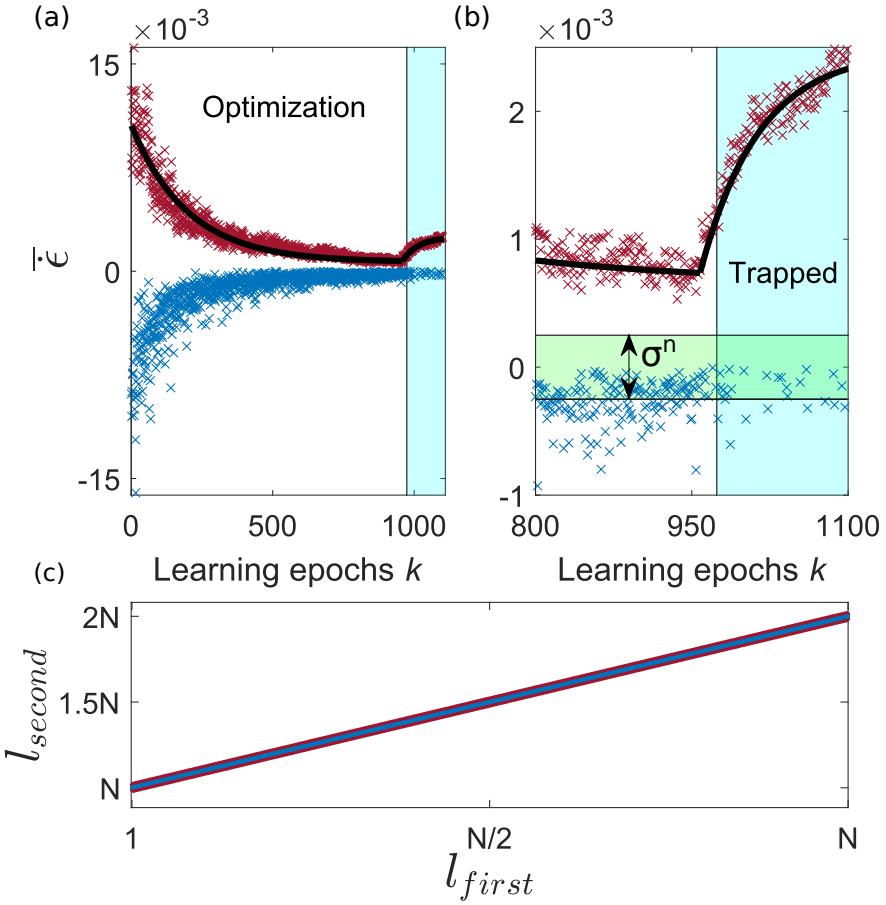

Figure 3.11: (a) Average of positive (red) and negative (blue) error gradients of the minimizers displayed in Fig. 3.9(a). Black curves are two different exponential fits (b) Zooms of the two curves around the epoch $k = 950$. The beginning of the blue zone corresponds to the average epoch where the minimizers are trapped in a minimum. Regarding the blue curve, we notice that on average most of the values in this region are in the noise level (green zone). On the other hand, due to the mutation algorithm, the red curve is structured according to a growing exponential. (c) Statistical repetition of greedy mutation sequence. The epoch at which a $l(k)$ was selected for a first time is stored in set l_{first}, the epoch of its second selection in l_{second}. The initial mutation sequence is repeated after $k > N$, as illustrated by the clear linear correlation between l_{first} and l_{second}.

Figure 3.11(a) depicts on a global scale the average distribution of positive and negative gradients for greedy exploration. As the convergence of the greedy algorithm is on average exponential, its derivative follows this law as well, which matches our finding for the positive gradient curve (red curve). However, we observe that this is not the case for the negative gradient (blue curve). Finally, the red curve in addition indicates that positive gradients show a qualitative change when the system is close to optimal performance. As

clearly visible from data shown in panel (b) which is a zoom of panel (a), the increase of the red curve starts just before the 974th learning epoch (blue region). This iteration corresponds statistically to the moment when the minimizers reach their minimum. Learning is composed of two parts: first an optimization during which the error and its gradient decrease, and after reaching an error minimum the system remains trapped and the negative gradients are in the noise level (green zone). Note that does not provide a complete prove of our system being trapped in a local minima, since this would require to retest the 960 nearest neighboring position. Nevertheless, by measuring multiple minimizers of up to 1500 learning epochs, we never observed the error decrease below the first minimum reached at around $k \approx 974$. The learning algorithm is therefore not noise limited since the red curve does not drop below the noise level shown in green, it is simply limited by the topology of the error landscape.

Another interesting feature emerges from Fig. 3.11(b) when on average minimizers are trapped in minima: the red curve exhibits a highly systematic increase, well described by an exponential law (black line). This means that, on average, gradients according to the error landscape's dimensions tested, follow a structure. Indeed, due to the learning rule Eq. (2.23), the probability of one readout weight to be modified again increase linearly in $1/N$. Therefore, after N epochs the selection of the position $l(k)$ statistically repeats the selection sequence carried out during the optimization.

Figure 3.11(c) excellently illustrates this repetition. By statistically repeating the mutation sequence, *i.e.* the selection of the dimension to be explored, the epoch at which a node l is selected a second time is displayed in function of the epoch of its first selection. The linearity of the average distribution plotted in blue demonstrates the statistical reiteration of a readout weight every N epochs. Its standard deviation of 27.3 plotted in red, shows the consistent pattern in this behaviour. Therefore, once trapped in a minimum, minimizer on average explorers the landscape according to the same sequence which lead to this particular minima. Unlike during $k \rightarrow k^{\min}$, for $k > k^{\min}$ we find that positive gradients increase according to $\bar{\dot{\epsilon}} = 1.8 \cdot 10^{-3}(1 - e^{-1.8 \cdot 10^{-3} k}) + 7 \cdot 10^{-4}$. Consequently, after optimization the dimensions which have contributed to greatly reduce the error at the beginning of the learning process, degrade the system performance the least after arriving at the minima.

3.3.2/ ERROR LANDSCAPE TOPOLOGY AROUND THE MINIMA

We have seen that the error landscape appears macroscopically convex but is in fact composed of many local minima of comparable error. As full characterization of the error landscape is physically impossible due to its size ($2^{961} \simeq 2 \cdot 10^{289}$), we focus on the topology where the error is lowest. We first turn to probing the error evolution around an error landscape minimum. The result is displayed in the Fig. 3.12 where the blue line is the average of twenty random walks (gray curves), which start from the same minima and walk away with a step size of changing 10 $\mathbf{W}^{\mathrm{DMD}}$ weights at a time. These measurements are parallelized in order to minimize long-term drifts inducing artefacts between different random walks, and we find that the error monotonously increases when departing from a minima along a random trajectory.

Let us now explorer the error's evolution between two minima M_a and M_b. We optimized readout weights via two minimizers starting at different random positions $\mathbf{W}^{\mathrm{DMD,a}}(1)$ and $\mathbf{W}^{\mathrm{DMD,b}}(1)$, which arrived at two distinct local minima $M_a = \mathbf{W}^{\mathrm{DMD,a}}(k^{\mathrm{min,a}})$ and

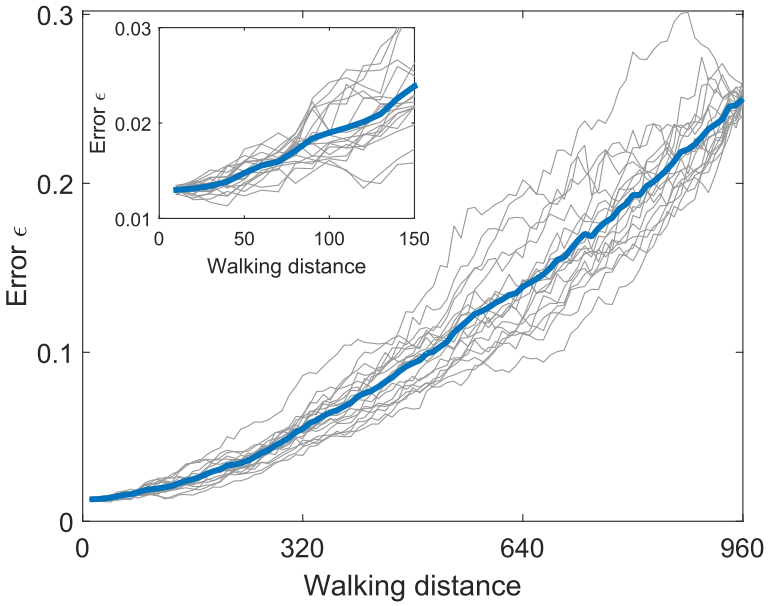

Figure 3.12: Error topology around a minimum for random walks. In gray, twenty random walks starting from the same minima, in blue their average. The inset is a zoom close to the minimum.

$M_b = \mathbf{W}^{\mathrm{DMD,b}}(k^{\mathrm{min,b}})$. Once there, we determined the list of m weights where M_b differs from M_a. The list is randomly arranged in sequence $\mathbf{l} \in [l(1), l(2), \ldots, l(m-1), l(m)]$ according to which we invert the corresponding weights $W^{\mathrm{DMD,a}}_{l(k)}(k)$ and $W^{\mathrm{DMD,b}}_{l(k)}(k)$ for $k \in [1, m]$. Importantly, this mutation is always kept and no optimization based on reward $r(k)$ is taking place. Starting from M_a (M_b), this results in a random path $P_a : M_a \longrightarrow M_b$ ($P_b : M_b \longrightarrow M_a$) of length m. We denote by $\mathbf{W}^{\mathrm{DMD},P_a}(m)$ ($\mathbf{W}^{\mathrm{DMD},P_b}(m)$) the DMD configuration along the path P_a (P_b). It should be noted that the two walks are not identical, as shown in Fig. 3.13(a), but symmetrically follow in inverted trajectory.

The results of this experiment are depicted in Fig. 3.13(b). Red and blue are the errors for P_a and P_b, respectively Minima M_a and M_b are depicted by the green and purple crosses, respectively. The standard deviation of noise σ^{noise} is shown in gray. By calculating the error gradients $\dot{\epsilon}^a(m) = \epsilon^a(m) - \epsilon^a(m-1)$ and $\dot{\epsilon}^b(m) = \epsilon^b(m) - \epsilon^b(m-1)$ between two steps m, we find that on average 41% of the variations are below the noise level. Importantly, it is clear that the error variation during the walk exceeds the level of noise. The error landscape is not flat between both minima and along these two particular paths, yet in comparison to the inset in Fig. 3.12, the error constantly remains low. These results are a first experimental indication that minima are linked by saddles [138, 139]. Following an error gradient would then be an optimization strategy to find a lower minima, as for example M_c (orange cross) which is 23% lower than M_a and 12% lower than M_b.

We repeat the same experiment between seven different minima found in parallel. One random path and its symmetry is probed in parallel between each minima. This leads to 42 walks between minima. Figure 3.14 shows in gray the moving average of the paths with a window of ten steps. The red line is the paths average in one direction, the blue line the average on the other, and the red and blue areas are their standard deviations, respectively. Error values during this experiment are higher than all previous ones. We

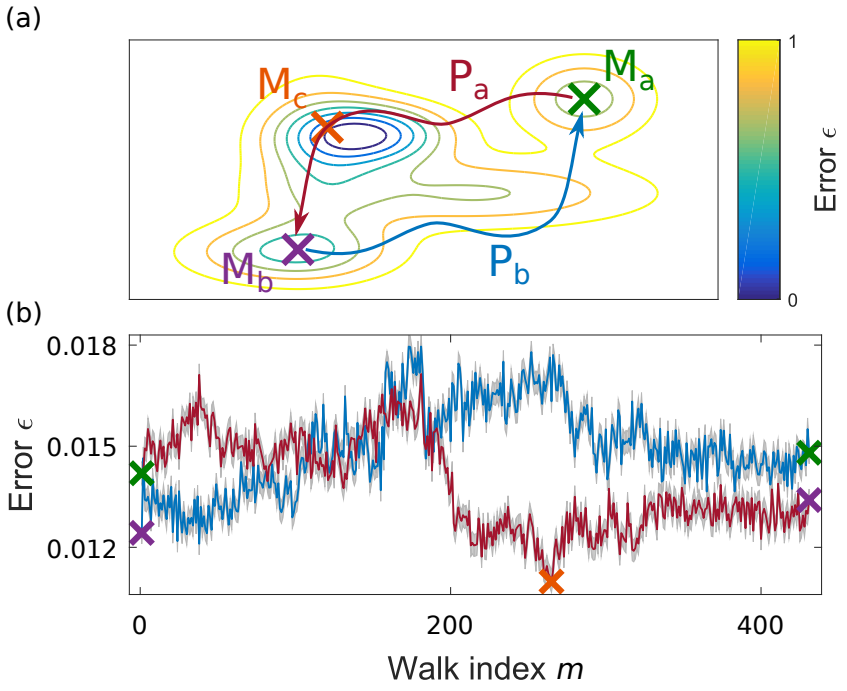

Figure 3.13: (a) Diagram of the error landscape, the curves represent the iso-error lines. This graph illustrates the characterization of the error landscape by going from one minimum to another. The red and blue arrows explicit the symmetry of the two paths, thus making it possible to explore the independence of the readout weights of the system. The passing through M3 shows that there are lower local minimums but not attainable by the learning algorithm when the DMD configuration is trapped in another local minimum. (b) Green and purple crosses are the error of two DMD optimized configurations, respectively in M_a and M_b. The two configurations have m different read weights which are flipped one by one until their position is completely reversed.. The red and the blue curves are the error measured during this transformation. In gray is the standard deviation of the noise during this process.

attribute this constant bias to the fact that this was the first measurement after a complete realignment of the experiment. However this does not modify the following results since we are interested in the relative, not absolute behavior of error ϵ.

On average, the error landscape between the different minima appears rather flat. While we probe only 42 of the $42 \cdot 2^{425}$ possible connections, their randomization provides statistical information. Whereas the characterization of an undirected random walk starting from a minima results in a significantly and monotonously increasing error, the error between minima always remains low. No significant error maximum separating two minima along any of the random connections was observed, and minima seem therefore to be connected.

The evolution of the error along a path connecting two minima exceeds the system's noise. We also found that many readout weights did not imply a variation in error greater than noise. This suggests a closer look into the role and relevance of individual weights: how many weights induce a systematic contribution to convergence, and whether their gradients depend on the sequence of previous mutations. In order to avoid the drifts potentially present within the long time characterization presented in Fig. 3.14, we will use

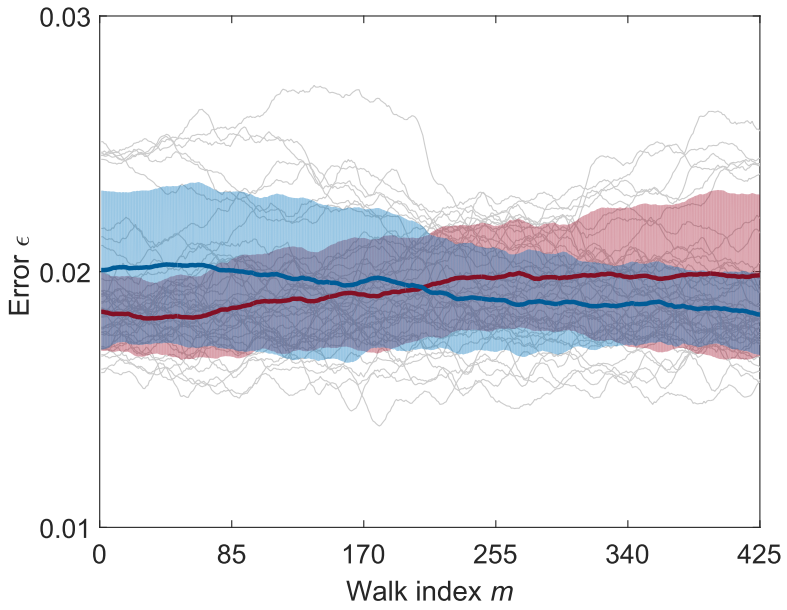

Figure 3.14: Error landscape topology between minima. The 42 gray curves are the error measurements of 21 random paths and their symmetrical between 7 different minima. The red line (area) is the average (standard deviation) of the error of the paths in one direction, in blue for their symmetrical. The 42 different walks, of more than 400 steps, do not see their error increase significantly.

the data of Fig. 3.13(b) which were obtained during a shorter time window. By studying the error gradients $\dot{\epsilon}^a(m)$ and $\dot{\epsilon}^b(m)$ respectively along P_a and P_b, we probe a fundamental characteristic of the error landscape, which is the dependence or independence of the readout weights, which is a measure related to orthogonality between the different dimension.

At each iteration m the positions $\mathbf{W}^{\mathrm{DMD},P_a}(m)$ and $\mathbf{W}^{\mathrm{DMD},P_b}(m)$ change along the same dimension, however with inverted direction. Therefore, the difference between $\dot{\epsilon}^a(m)$ and $\dot{\epsilon}^b(m)$ depends on the configuration of the other W^{DMD} dimensions To estimate this independence, we calculate

$$S(m) = |\dot{\epsilon}^a(m) - \dot{\epsilon}^b(m)| - \frac{1}{2}\sigma^n. \tag{3.8}$$

For a negative $S(m)$ the difference between modifying the same weight at m, but in the context of the different W^{DMD} configurations, is below the noise level. Hence, this dimension of the error landscape is potentially linearly independent from the others. This necessary criterion is not sufficient since a test of all linear combinations is required to prove linear independence. However, for a positive $S(m)$ the same modification imperatively depends on the other weight configurations. It is therefore a sufficient criteria to show that a particular weight is linearly dependent on the rest of the readout weights.

Figure 3.15 shows the impact of modifying entries of W^{DMD} in the ($\dot{\epsilon}^b(m)$ - $\dot{\epsilon}^a(m)$) plane. Data inside the gray circle with diameter σ^n do not contain exploitable information, and these correspond to 11% of the total measurements. The anti-diagonal in red symbolically represents linear independence. Regardless of the position in the error landscape, changing a linearly independent Boolean readout weight will cause an error gradient op-

posed to the opposite modification. The green zone of width σ^n corresponds to $S < 0$, and all weights contained inside this region are potentially independent, which are 30% of the 430 readout weights tested. The remaining 59% are outside of the green or gray zone, therefore they are linearly dependent [135]. We observe that the distribution in Fig. 3.15 does not show a distinctive structure, in particular along the red line. This suggests that the majority of dimensions could be linearly dependent.

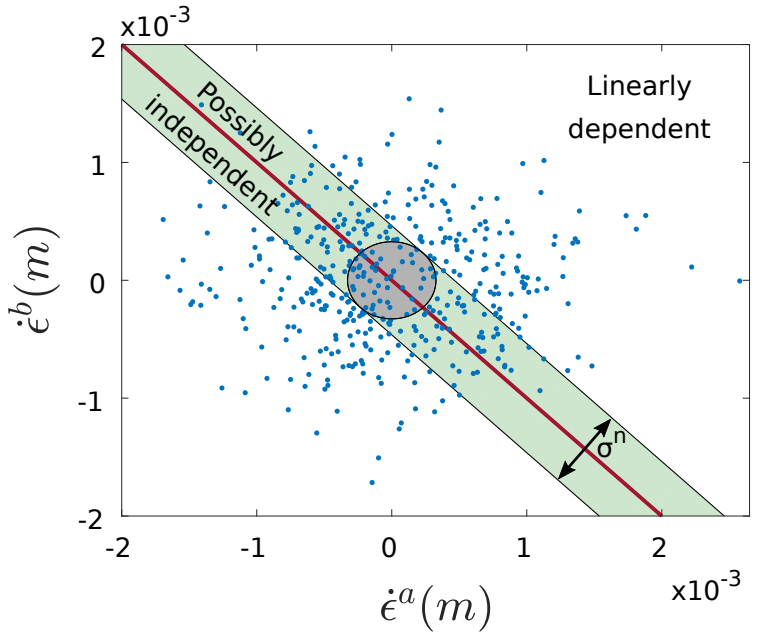

Figure 3.15: Error gradient (blue dots) for all readout weights encountered along paths P_a and P_b between minima M_a and M_b. The red anti-diagonal represents the linearly independent dimensions. The green zone depict the noise level around this line. Within this green zone, the dimensions tested are possibly independent, while outside they are linearly dependent. Inside the circle of diameter σ^n no conclusion can be drawn due to the system's noise.

These results are in agreement with previous results which show that learning seems to take place in the context of previously optimized weights, for example see Fig. 3.11(a,b). Moreover, the error landscape does not have 2^{961} orthogonal dimensions and accurate characterization of the system's dimensionality is unrealistic at the current stage.

3.3.3/ NOISE SENSITIVITY

The overall similarity of the minimizers indicates that the topology of the error landscape seems to have no local minimums with a large error. All the optimization paths arrive at different local minima with comparable errors. Despite similar performances, the correlation between the optimal DMD configurations of different minimizers are negligible. The separation between two Boolean readout configurations $\mathbf{W}^{\mathrm{DMD},a}(k^a)$ and $\mathbf{W}^{\mathrm{DMD},b}(k^b)$ can be determined by their *Hamming distance*:

$$H(k^a, k^b) = \sum_i^N \left| W_i^{\mathrm{DMD},a}(k^a) - W_i^{\mathrm{DMD},b}(k^b) \right|. \tag{3.9}$$

For the 20 different minimizers measured in Fig. 3.9(a), we obtain $20(20-1)/2 = 190$ distances between their respective minima. The normalized histogram of inter-minima distances and a fit by a Gaussian is shown in Fig. 3.16. The distribution is centered around a distance of 419 with a half width at $1/e$ of 14 [135]. This result shows a very specific and unusual error landscape: local minima appear not to be irregularly distributed, nor located in a particular region. Indeed, the minima are on average all separated by a Hamming distance of 419. Not presented here, we also find that Markovian exploration results in a similar distribution.

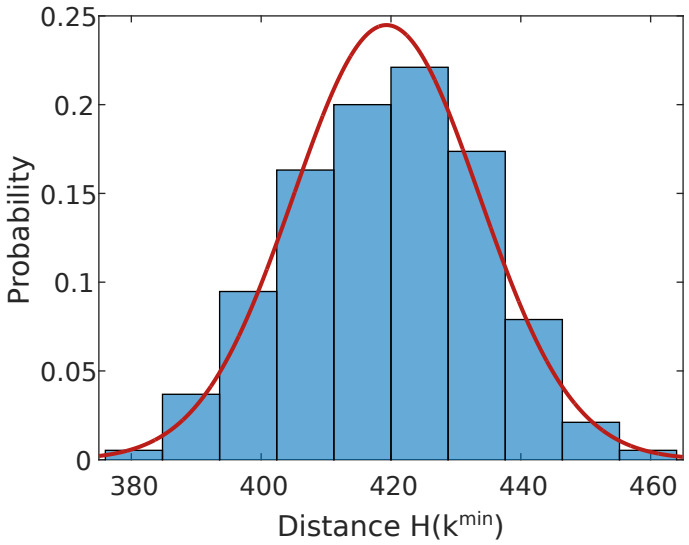

Figure 3.16: Probability distribution of the Hamming distances between the twenty minima of the minimizers displayed in Fig. 3.9(a). Red curve is a Gaussian fit with a mean of 419 and a standard deviation of 29.

While these results appear to indicate an almost arbitrary configuration of W^{DMD}, we are able to demonstrate the opposite. We repeatedly inverted 419 randomly selected weights of several $W^{\mathrm{DMD}}(k^{\mathrm{min}})$, and determined their error ϵ. We found that the error significantly increases from $8.3 \cdot 10^{-3}$ to between 0.17 and 0.44. This clearly shows that minima are not arranged on a regular grid with a spacing of 419.

To further investigate this phenomena, we reduce the number of uncertainties during learning. We therefore measure three minimizer paths starting at identical position $\mathbf{W}^{\mathrm{DMD}}(1)$ in parallel. One of the three minimizers acts as a "master" and defines mutation sequence $l(k)$, $k \in [1 \cdots K]$, which the other two minimizers follow as "slaves". The sequence in which the error landscape's dimensions are tested is therefore identical for all three, but their own rewards $r(k)$ are independently computed via the error of each minimizer. Keeping in mind the potentially systematic error of a slow experimental parameter drift, we measure the three minimizers in parallel. The result of this experiment for greedy exploration are plotted in Fig. 3.17(a). The three solid lines are the minimizer's errors on a semi-logarithmic scale, the blue line corresponds to the master, the red and the green lines to the slaves. Panel (b) shows the result between one master (blue line) and a single slave (red line) for a Markovian exploration. The error of the different minimizers have an exceptionally high average correlation of 99.4% and 99.72% for greedy and Markovian exploration, respectively.

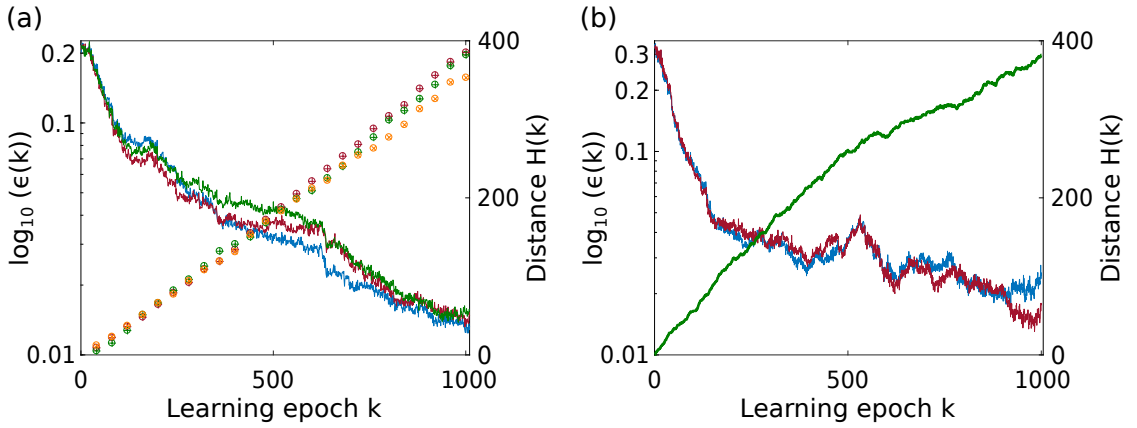

Figure 3.17: (a) In solid lines, three minimizers having a greedy exploration and starting from identical position are measured in parallel. At each learning epoch, the two slaves (red and green lines) test the mutation l_k determined by the master (blue line). Crucially, reward $r(k)$ of each minimizer is independent. The crossed-circles depict the Hamming distance's evolution between the minimizers. (b) Identical experiment for a Markovian exploration between one master (blue line) and a single slave (blue line). The green curve is the evolution of the Hamming distance.

Despite this high degree of agreement, the individual W^{DMD} configurations diverge during the optimization. The evolution of the Hamming distance $H(k)$ between minimizers is plotted on the right y-axis in Fig. 3.17. On panel (a), the red and green crossed-circles are the Hamming distance between the slaves and the master, the yellow is the distance between both slaves, the separation is linear throughout the learning process. For Markovian exploration on panel (b), the Hamming distance evolution between master and slave (green line) is not linear. Without noise the rewards $r(k)$ of the minimizers would be identical and they should therefore follow the same trajectory from $\mathbf{W}^{\mathrm{DMD}}(1)$ to the same minimum $\mathbf{W}^{\mathrm{DMD}}(k^{\mathrm{min}})$. The Hamming distance between the master and the slave would therefore be zero.

In order to understand the non-zero evolution of the Hamming distance measured, we therefore have to consider the impact of noise and learning upon the system's error $\epsilon(k)$ [135]. The response of error $\epsilon(k)$ to a modification in the system's output $\Delta y^{\mathrm{out}}(k)$ is

$$\Delta\epsilon(k) = \dot{\epsilon}(k) \cdot \Delta y^{\mathrm{out}}(k), \tag{3.10}$$

where $\dot{\epsilon} = d\epsilon/dy^{\mathrm{out}}$ is the error gradient and $\Delta y^{\mathrm{out}}(k)$ is the mean modification of output $y^{\mathrm{out}}(k, n+1)$ within a certain window, which during training contains 170 sample points.

Several general considerations are relevant at this stage. Due to the SLM's illumination by a collimated Gaussian beam, the amplitudes of all network nodes $\tilde{x}_i$ are Gaussian distributed. Randomly changing one readout weight therefore results in $\Delta y^{\mathrm{out}}(k)$ according to a normalized Gaussian distribution with a width of $\Delta y^{\mathrm{out,learn}}(k) = \sigma^l(k)$. The noise of all elements in our opto-electronic neural network, and their accumulated impact upon $y^{\mathrm{out}}(k)$ is excellently approximated by Gaussian white noise with a width of $\Delta y^{\mathrm{out,noise}}(k) = \sigma^n(k)$ [136]. The large number of minimizer measured shows that readout weights $\mathbf{W}^{\mathrm{DMD}}$ remain evenly distributed between zeros and ones for all k. Following the reservoir computing concept, the learning does only modify readout connections. It neither modifies $\tilde{x}_i$ nor the system's noise, which are then both independent of learning. We can therefore

assume that modifications to y^{out} induced by learning and noise remain constant for all k, hence $\sigma^l(k) = \sigma^l$ and $\sigma^n(k) = \sigma^n$.

Noise (σ^n) and learning (σ^l) modify both the system's error according to Eq. (3.10). This identical relationship for the two phenomena is of general importance. Convergence during learning is characterized by $\epsilon(k)$ and $\dot{\epsilon}(k)$, and both depend on the particular computational tasks. Due to the same constant $\dot{\epsilon}(k)$, the ratio between the error's susceptibility towards σ^n and σ^l remains constant. This in turn imposes the same constant relative susceptibility of reward $r(k)$ towards σ^n and σ^l. Thereby, neither $\dot{\epsilon}(k)$ nor $\epsilon(k)$ modify the interaction between noise and learning upon the system's weights. The following discussion, results and observations are therefore fully general and independent of task and learning algorithm - as long as these do not change amplitudes σ^n and σ^l. Noise and weight modifications are therefore independent actors, whose action upon learning is somehow competitive.

The objective of modifying a readout weight is to probe the error landscape's gradient. However, this action is contaminated with noise which can potentially be greater than the systematic gradient in the opposite direction. In this case, the sign of $\Delta\epsilon(k)$ change, which results in an inversion of the reward $r(k)$. We call C, resulting on the relative amplitudes of σ^n and σ^l, the constant probability of such a modification occurring.

The increasing separation between two identical minimizers is generated by this probability C. Two scenarios are then possible. The first situation is a reward $r(k)$ different between the two minimizers. This happens if one reward is the result of its systematic value while the second is inverted by noise. The probability that this situation occurs is $C(1-C) + (1-C)C = 2C(1-C)$. The second situation is two identical rewards, which can either be the consequence of the systematic results of the two minimizers, or rewards are both inverted by noise. The combined probability of this second case is $(1-C)^2 + C^2 = 1 - 2C(1-C)$. The first situation leads to $H(k+1) \neq H(k)$, the second to $H(k+1) = H(k)$, and the Hamming distance's rate equation is

$$\Delta H(k+1) = \rho^{\text{id}}(k) \cdot 2C(1-C) - \rho^{\text{op}}(k) \cdot 2C(1-C), \tag{3.11}$$

where $\rho^{\text{id}}(k)$ and $\rho^{\text{op}}(k)$ are the probability of finding both minimizers' weights $l(k)$ to be identical or opposite, respectively. Using $\rho^{\text{id}}(k) = 1 - \rho^{\text{op}}(k)$, the evolution of the Hamming distance is

$$H(k+1) = H(k) + \tilde{C}(1 - 2\rho^{\text{op}}(k)), \tag{3.12}$$

and then the Hamming distance's gradient is

$$\Delta H(k+1) = \tilde{C}(1 - 2\rho^{\text{op}}(k)), \tag{3.13}$$

where $\tilde{C} = 2C(1-C)$.

For Markovian exploration, the mutation is fully random (see Eq. (2.12)), the probability of a weight to be selected is identical at every k. Thus, the Hamming distance at the previous epoch k determines the probability of two weights being opposite in their configuration: $\rho^{\text{op}}(k) = H(k)/N$. For greedy learning, the bias term in Eq. (2.20) makes it improbable to repeat the same weight for an interval specified by $k = k' + aN, k' \in [1, N]$, with non-negative integer a. The probability of both minimizers to be configured opposite for all k' and a specific a is therefore their Hamming distance at the end of the previous interval: $\rho^{\text{op}}(k) = H(aN)/N$, which results in constant slopes $\Delta H(a)$ for each k'.

Figure 3.18(a) shows the evolution of Hamming distance $H(k/N)$ for optimizations starting at the same position $\mathbf{W}^{\text{DMD}}(1)$, hence we have $H(1) = 0$. The red and blue lines are the

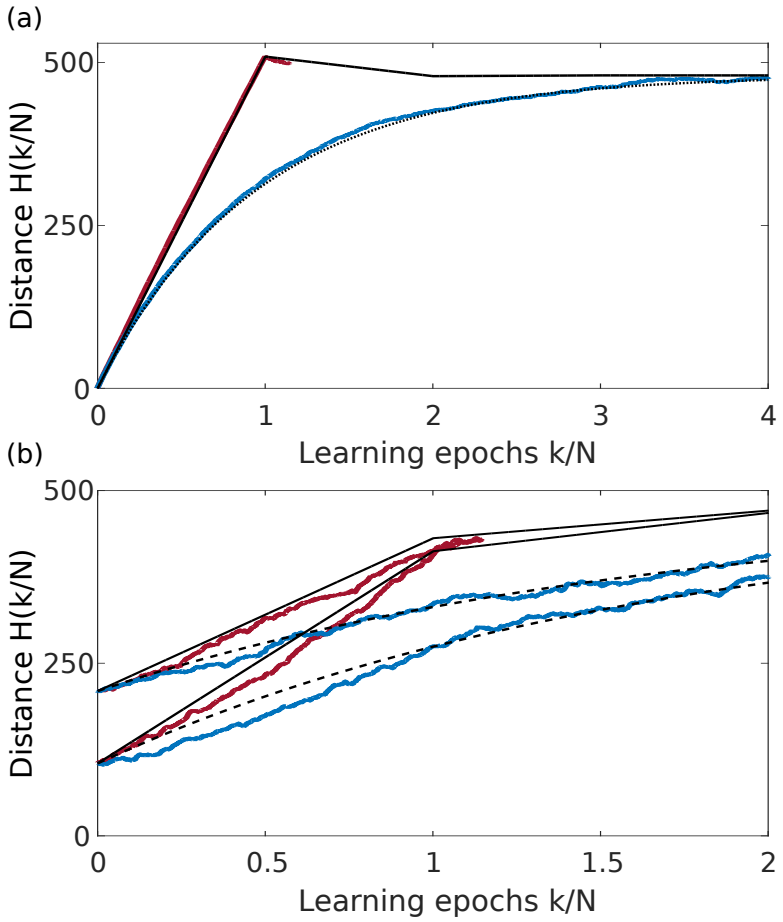

Figure 3.18: (a) Evolution of average Hamming distance with learning epoch k normalized by the network's Size N. Random mutation (blue line) leads to a smooth saturation function behaviour, while greedy mutation (red line) results in linear intervals of length N. Black dashed line and black solid line are respectively their analytical solutions. (b) The same characteristics are found for two minimizers starting in different positions separated by a distance $H(1) > 0$.

average Hamming distance computed from 20 minimizers, respectively from a greedy and Markovian exploration. The analytical solutions are drawn in solid black line and in black dashed line for greedy and random mutations, respectively. An excellent agreement between the experimental data and the analysis is obtained for a parameter $\tilde{C} = 0.53$ identical to the two situations. It is noteworthy that we always arrive at $H(k)|_{k\to\infty} = N/2$, regardless of the algorithm [135]. Curiously, the value of the parameter $\tilde{C}$ leads to a complex constant C, resulting in an overshoot for the greedy mutation.

Figure 3.18(b) shows in red and blue line experimental results with a starting separation of $H(1) = 105$ and $H(1) = 210$, respectively for a greedy and a Markovian exploration. The analytical solutions are again the solid black line and the solid dotted line. Now the two minimizers start from different positions, the Hamming distance's evolution starts with a separation $H(1) > 0$ and its slope are less steep. Despite the differences between the analytical solutions and the experimental data, one can note that even without averaging, the overall evolution is well described.

In conclusion, the Hamming distance's evolution, described in Eq. (3.12), is therefore governed by constant $\tilde{C}$ which is due to noise, and by how the learning algorithm picks weight $W^{\mathrm{DMD}}_{l(k)}(k)$ from a population with a certain $\rho^{\mathrm{op}}(k)$. It is important to remember that neither error nor gradient play a role in this evolution.

3.4/ LEARNING ALGORITHM SCALABILITY

As motivated in the previous chapter, one of the main arguments supporting optical neural network is related to the massive parallelization of network connections. In such a network, one could greatly increase the nodes number without reducing the speed of the system. Nevertheless, a network undergoes a learning phase where each connection is trained, and for the scalability for neural network hardware the learning time is an essential characteristic. The reservoir computing concept has the advantage of only training the output weights, which already reduces the required epochs for achieving convergence. We therefore study the scaling behaviour when optimizing our Boolean readout layer.

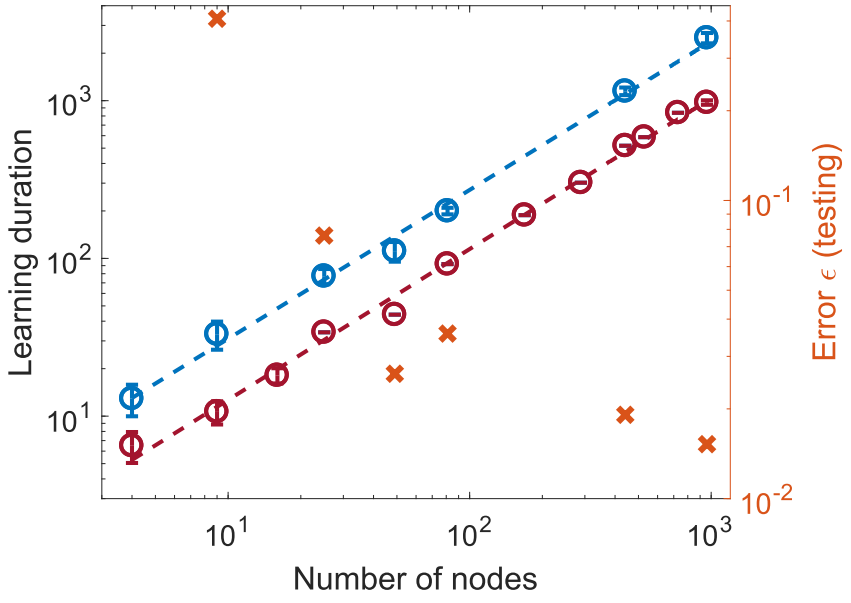

Figure 3.19: Scaling of the required time to convergence for networks of different sizes. Red (blue) data were obtained for greedy (Markovian) learning. We find that learning almost perfectly scales linearly with a polynomial coefficient of 0.96 (0.94). Orange data are testing errors, which continue to decrease.

We measure the learning duration $k^{\min}$ across networks of sizes spanning nearly three orders of magnitude, ranging from 4 to 961 nodes. Results [140] of the greedy and Markovian algorithm are respectively the red and blue circles in the double log-plot of Fig. 3.19. The red and blue dashed lines are linear fit, and we find that $k^{\min}$ for greedy and Markovian learning scales with $k^{\min} = 1.41 \cdot N^{0.96}$ and $k^{\min} = 3.55 \cdot N^{0.94}$, respectively. In both cases learning therefore scales almost perfectly linearly with the nodes number, however the greedy mutation improves the training by a factor of 2.5. The orange data are the testing error on 9000 previously unseen data points. The testing error continues

to reduce for each increase in network size, showing that even for the largest network we have not yet reached the limit of over fitting.

On a final note, let us mention that recent results in theoretical computer science also have an interesting say in our study. Seminal work of Wegener and Witt [141], [142] has paved the way around complexity analysis for simple randomized evolutionary algorithms based on local search. Despite the simplicity of the models, mostly restricting the analysis to Markovian dynamics in the configuration space, a host of theoretical lower and upper bounds are available in the recent literature [143], [144], [145] and these bounds provide interesting benchmarks with which to compare our empirical findings. An interesting result in this sense is theorem 9.4 from [143], which can be summarized as follows: the optimisation time of the Markovian evolutionary algorithm where each bit has uniform probability of being selected (i.e. memoryless dynamics) for maximizing any function on $\{0,1\}^N$ with unique global maximizer has as lower bound $N\log(N)$. Note that the asymptote when $N \gg 1$ of this lower bound is N. The main conclusion that can be therefore drawn from this theoretical bound is that our two evolutionary learning algorithms, when the nodes number is large, accurately match the best possible completion time for any binary optimisation problem using a simple Markovian evolutionary hill-climbing type method.

3.5/ SUMMARY

In this chapter we have studied the characteristics of our hardware neural network.

We first presented a study of the noise present in the reservoir. An important result is the presence of a relatively high SNR following a spatial distribution clustered by the action of the DOE. We have also characterized the different sources of noise in the system which were found to be additive, multiplicative, correlated or uncorrelated noise. This study began with an experimental characterization serving us as a target, with a bottom-up approach, by modelling the interactions and characteristics of noises. One of the objectives was to improve the SNR distribution, typically by shifting the operating point along the nonlinear function. We presented a first analytical model reproducing excellently our experimental observations. Numerical models now allow to understand the characteristics of noise, how it propagates and thus test hypotheses. Recent work [136] shows that uncorrelated noise tends to be strongly suppressed when the number of nodes is large enough. These fundamental results can be generalized for other types of neural networks.

Moreover, the system is realized with off-the-shelf components. Although optical experiments are known to be sensitive, our neural network has demonstrated great stability and to be resilient to noise. For example, a single adjustment of the hyper parameters was sufficient for nine months of night and day measurements. This stability is mainly due to the optical cage system used but also to the control and stability of the laser temperature and the meticulous experimental adjustments.

We then studied the noise impact on the dynamic response of the system. We measured the nodes consistency, *i.e.* their ability respond similarly to the same input. Whether short or long term, consistency is high for many nodes, in particular in the network's center. We have observed the presence of rare cases of long term drifts in consistency, and nodes with very low consistency, mainly at the network periphery. Crucially, we have shown that consistency is not correlated with the injection matrix, there is therefore no restriction to

use the full range of injection values. The upper bound of consistency is 0.96, but we were unable to identify the cause of this limitation below the noise and digitization level.

A very broad characterization of learning under noise has been also presented. We have found that error convergence due to learning is exponentially decreasing. An important point is the macroscopic convexity of the error landscape, learning never got stuck at low performance. However, the error landscape has multiple local minimums, but all minima we have found had an comparably low error. On first sight their spatial configuration seems to follow a very particular periodic arrangement, but we have been able to demonstrate that this is a consequence of the noise during optimization. Our analytical model matches excellently with the experimental results for both, greedy and Morkovian learning.

The Hamming distance between the readout weights of two systems always converges to N/2, and no particular preference for any groups of nodes was found. Therefore, even for entirely identical networks, the configurations of the readout weights will never be the same [146]. This result can probably be extended to non-Boolean case and to weights between layers of analogue deep neural network. The field of learning implemented in physical and therefore noisy substrates is still at the beginning [16, 108, 122], and the confirmation of our findings in other material systems would prove the generality of our result. Finally, the human brain is a very noisy network [134], which suggests that our findings may also have interesting implications for the field of theoretical neuroscience.

We have seen that a large part of the readout weights are linearly dependent. What this means in practical terms is that each modification of a weight has to be interpreted in the context of all previous modifications. Optimization of an output weight therefore encodes the history of the changes in the reward during the previous learning periods, including those which were wrongly assigned due to noise. From an application's point of view, a direct consequence is the potential difficulty to transfer or exchange the optimized configurations between two ANNs.

We characterized the topology between several minima, the error landscape always seems to remain at low error values. However, this region is not flat and configurations with better performance can be found. No barrier appears to separate local minima, which suggests that the minima are probably related to each other. An optimization strategy based on a gradient would then become interesting to find lower minima in this particular region.

Finally, we measured the scalability of our two optimization algorithms. Across network sizes of three orders of magnitude, the optimization time scales linear with the number of nodes. This result is of great importance because it indicates the scalability of our learning algorithm to very large networks optimizations.

4

MULTI-LASER RECURRENT NEURAL NETWORKS: OVERCOMING THE ENGINEERING CHALLENGE

We described in chapter 2 an electro-optical neural network, and we studied its architecture and its different characteristics. The spatial light modulator, realizing the electro-optical nodes of the system, is a technology which limits the speed of the system to a few tens or hundreds of Hertz.

Designing an all-optical neural network is a paradigm shift that could fundamentally change the field of machine learning. Optics can potentially allow massive parallelization and the computational bandwidth limiting factor will then be the speed of light or the response time of the photonic artificial neuron. The characteristic time of such a neural network can readily reach several Gigahertz [147]. A large-scale, parallelized system operating at this timescale will then have low energy per operation and a high power efficiency [99].

In this chapter, we will see the consequences of the transition from electro-optical nonlinear nodes to optical nonlinear nodes and will study the key points of realizing such a system. This technological change will therefore be first treated as an engineering challenge including an in-depth characterization of its optical nodes.

The overall architecture of the system will first be presented, then we will describe the structure of the quantum dots micropillar lasers (QDMLs) [18] which will be the nonlinear nodes of the network. The characteristics necessary to achieve an array of coupled lasers acting as nonlinear nodes of a recurrent neural network will be discussed. We will show why the wavelength homogeneity of the QDMLs is essential. Engineering solutions regarding mechanical mounting and cryogenic solutions will also be discussed.

Still wishing that the neural network could evolve into a much larger system, QDML's population inversion is performed by optical pumping to avoid the scalability problem of electrical connections that cannot intersect. The pump laser characteristics will be analysed and the design and implementation of a laser for pumping the QDMLs' wetting layer, which was constructed from scratch, will be presented in appendix B. We will then investigate a quantitative study of the QDMLs pumping enabling to select between two different optical pumping approaches and to optimize the experimental parameters. In addition,

this characterization will give us access to the efficiency limitation of current QDMLs and will therefore allow us to suggest improvements to the next QDMLs design. Finally, a first simultaneous pumping of 9 QDMLs will be performed.

4.1/ OPTICAL NEURAL NETWORKS: OVERVIEW

Our goal is to implement a partially scalable all-optical neural network. To achieve this, we will use a similar concept as the neural network presented in chapter 2 and modify the nonlinear nodes to be fully optical.

Our choice of QDMLs was made for several reasons. QDMLs are semiconductor lasers, their response time is dominated by the carrier's relaxation oscillation frequency which can be of the order of 10 Gigahertz [147]. As an artificial neuron, a semiconductor laser has numerous sources of non-linearity such as the laser threshold, gain saturation or amplitude-phase coupling due to Henry's alpha factor [148]. Moreover, QDMLs pumping can be done optically, which allows a high scalability by using a single diffracted pump beam to operate a large number of QDMLs. Using the same principle, input information can be injected all optically [149]. However, QDMLs do not operate at room temperature, and the sample should be cooled to below 150 K, which complicates the neural network design from an engineering point of view.

Shown in purple in the experimental scheme of Fig. 4.1, 20×20 QDMLs are arranged in an array and form the set of nonlinear nodes of the system. The network's recurrent connections can be optically implemented via feedback of the laser emission back into the QDMLs using a diffractively multiplexed external cavity [17, 113]. The reservoir, shown in gray, is then composed of an QDMLs array and an external cavity produced by a 4f imaging system of magnification 1 with two identical microscope objectives MO^{S} and MO^{F} (Olympus, LMPLN Plan Achro 10x) having $f_{\mathrm{MO}^{\mathrm{S}}} = f_{\mathrm{MO}^{\mathrm{F}}} = 18$ mm as focal length. The network nodes connectivity, visible in red, is provided by a 3×3 DOE. As explained in detail in section 2.1.4, the diffraction angle of the DOE and the pitch of the QDMLs array are adjusted in such a way that the feedback of a QDML emission, after passing the 3×3 DOE twice, is injected into itself and its twenty-four near neighbors [16, 113]. The input/output of the reservoir can be realized through a beam splitter. The QDMLs power source is optical, the population inversion is achieved by a pump laser. In order to pump multiple QDMLs simultaneously, the laser beam is separated into N diffraction orders, with N the number of lasers pumped. By adjusting the diffraction angle according to the QDMLs array pitch, each order is focused on a different micropillar. The information laser, displayed in green, can use the same concept to inject information into the neural network. The information laser having the same wavelength as the QDMLs is separated from the pump laser by a dichroic mirror. The readout of the network can be implemented identically to that of the electro-optical network presented in section 2.1.5. The readout layer shown in blue in Fig. 4.1 is therefore based on a DMD which will select the nodes which will be detected. A spectral filter can be added in order to remove any residue from the pump laser.

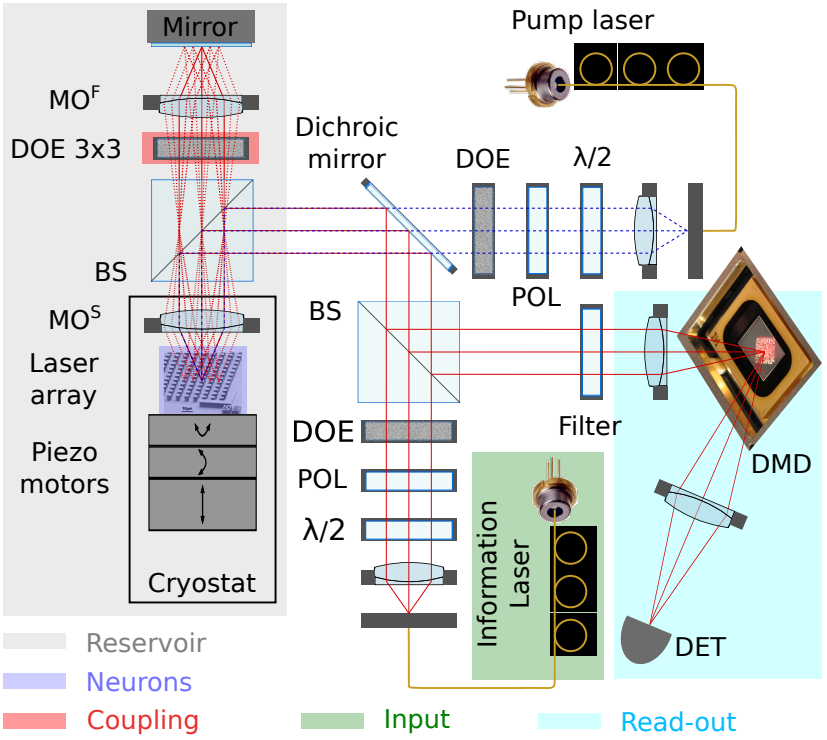

Figure 4.1: Photonic spatio-temporal neural network. A laser array attached to piezomotors inside a cryostat, forms the nonlinear nodes of the network. They are powered by a pump laser which, once diffracted, is focused on numerous QDML. A identical technique can be used to inject information into the nodes. The feedback of the QDMLs emission can be implemented by an external cavity. The nodes interconnectivity is achieved by diffraction, each QDML is coupled to its neighbors in addition to itself. The readout layer can optically be implemented through a DMD which selects the nodes to be detected.

4.2/ LASER NEURONS

The nonlinear nodes that we use are semiconductor QDMLs. The samples used during this work were produced by the research group at the *Technical University of Berlin (TUB)* led by Prof. Stephan Reitzenstein in the framework of the Volkswagen Foundation NeuroQNet project [121]. A detailed overview of this technology can be found here [150]. In this section we will detail the structure of the QDMLs and in particular the characteristics required to build our optical neural network.

4.2.1/ QUANTUM DOT MICROPILLAR LASERS

The QDML is a vertical-cavity surface-emitting laser (VCSEL). The fabrication is performed by successive growth and etching steps.

The first stage is the epitaxial growth of the sample on a GaAs substrate using metal-organic chemical vapor deposition (MOCVD). Schematically illustrated in Fig. 4.2(a), the sample consists of a central one-λ thick GaAs resonant cavity ($L_{\mathrm{LC}} = 279$ nm) sandwiched between a lower and an upper distributed Bragg reflector (DBR) acting as mirrors. In order to ensure sufficient gain for lasing operation, the intra-cavity gain medium is composed of

three separate layers of self-assembled InAs Quantum Dots (QDs). The three QD layers are fabricated via the growth controlled monoatomic layers of InAs on top of GaAs. This thin film is called wetting layer (WL). Beyond a critical thickness, growth continues through nucleation and coalescence of islands, which form the QDs [151].

Our layers of self-assembled QDs are formed with a density of about $10^{10}\ \mathrm{cm}^{-2}$. The inset of Fig. 4.2(b) shows scanning electron microscopy (SEM) micro-graph of the QDs where we can observe their random spatial arrangement. Considering an average diameter of the micropillars of 4 µm, the gain medium is composed of several thousand QDs. QD s-states provide optical gain by light-matter interaction with the fundamental cavity mode. However, due to the QD's inhomogeneous broadening, only a small part, typically 10%, aligns with the cavity's resonance and can hence contribute to laser emission [152, 153].

The lower and upper DBR are respectively composed of 27 and 23 $\lambda/4$-thick $Al_{90}Ga_{10}As$ / GaAs mirror pairs. Once sample growth is complete, the micropillars are nanoprocessed by high-resolution electron beam lithography in combination with reactive-ion etching [18]. Targeting lasing at 980 nm, the thickness of the laser cavity measures approximately 280 nm and that of the lower and upper DBRs 3.71 µm and 3.43 µm, respectively. A SEM micrograph of our QDMLs is shown in Fig. 4.2(c). The mirror pairs of the DBRs and the resonant cavity are clearly visible, and the lower DBR remains partially buried. More information about this particular sample can be found in [18].

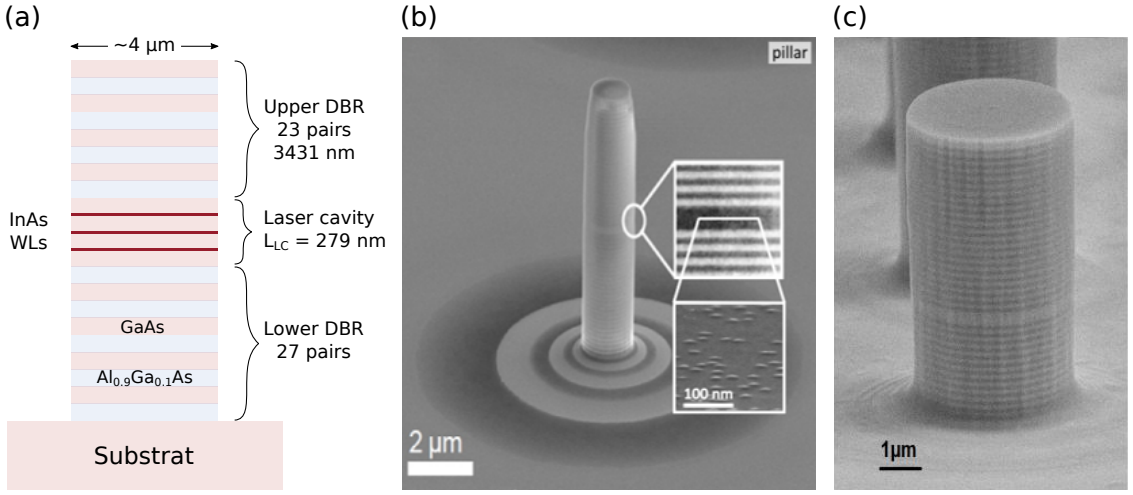

Figure 4.2: (a) Schematic illustration of QDML where the three WLs are represented in red. (b,c) QDML images obtained by scanning electron microscope (SEM). (b) Micropillar laser from [152] where we can observe the resonant cavity sandwiched between a lower and an upper distributed Bragg reflectors in the inset and the QDs in its zoom. (c) Micropillar laser example that we use for the optical neural network [18]

4.2.2/ QUANTUM DOTS MICROPILLAR LASERS ARRAYS

The objective is to fabricated QDML arrays (QDMLA) to obtain a large number of nonlinear nodes. The QDMLA must have two fundamental characteristics: a regular spacing of the micropillars [17, 113] and homogeneity of the laser emission [131].

The QDMLA's pitch must be adjusted to use a 3×3 DOE to couple the lasers. We

made the choice to buy a commercial DOE and adjust the distance between the QDML accordingly. As shown by the grating formula in Eq. (2.3), the diffraction angle varies with the wavelength, therefore we have characterized the DOE according to this criterion. We have used three lasers with 840.9 nm, 916 nm and 975.5 nm wavelength, which was measured with a Yokogawa AQ6370D optical spectrum analyser (OSA). The laser beam diffracted by the DOE was focused on a camera with a pixel pitch of 2.2 µm with a lens of $f_{\text{charac}} = 100$ mm focal length. In combination with the pixel pitch, this focal length allows determining the diffraction angle with 22 µrad resolution. Each diffraction spot is fitted by a 2D Gaussian function, enabling determining the position of diffractive orders with precision larger than the pixel pitch of the camera. The distances measured between the diffracted orders are multiplied by the ratio $f_{\text{MOS}}/f_{\text{charac}}$ to obtain the physical distances ΔX and ΔY between diffractive orders for the QDMLA setup.

Results are shown in Fig. 4.3 where ΔX (red crosses) and ΔY (blue circles) are linearly fitted by the red and blue curves, respectively. The QDMLs emit at 980 nm, ΔX(980 nm) = 8.41 µm and ΔY(980 nm) = 8.33 µm, the micropillars pitch will be averaged at 8.37 µm. In addition, this experimental characterization using a laser injected into a singlemode fiber (Thorlabs SM800-5.6-125) collimated with a lens of 35 mm focal length allowed us to determine the width at half maximum $\text{FWHM}(980\text{ nm}) = 4.2 \pm 0.1$ µm of the diffraction orders imaged the QDMLA. This width is compatible with the mode field area of the QDMLs emission.

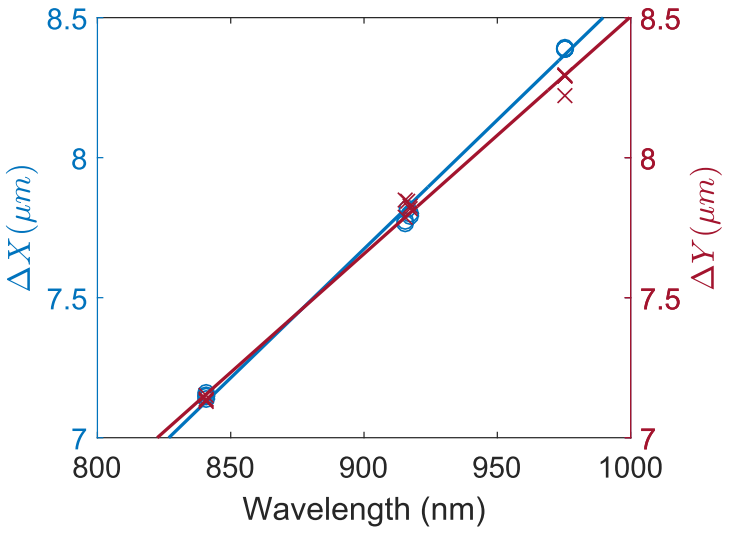

Figure 4.3: Calibration of the DOE 3x3 in function of the wavelength. ΔX (red crosses) and ΔX (blue circles) are the vertical and horizontal distance between the diffractive orders on the QDMLA. Red and blue curves are the linear fit of ΔX and ΔY which result respectively in $\Delta X_{fit} = 9.21 \cdot 10^{-3}\lambda - 0.614$ µm and $\Delta Y_{fit} = 8.47 \cdot 10^{-3}\lambda + 0.036$ µm with λ in nanometer.

We would like to create a reservoir, and hence the emission of QDMLs will be reinjected into the neighboring micropillars to establish the required recurrent network connections. In addition, in the future information will need to be injected into the network via an external injection laser. However, to obtain injection locking, the power of the injected laser increases with the frequency difference squared of the two oscillators [154]. It is therefore essential to obtain a very homogeneous QDMLA if one wishes to have sufficient power to drive a large number of micropillars. The locking range is the frequency detuning range

between an injection laser and a slave laser over which phase, and hence frequency of the slave lasers is identical to the injection laser. Achieving such injection locking is essential for neural network and reservoir computing with semiconductor lasers [131, 149]. The locking range $\Delta\nu_{\text{locking}}$ is defined as

$$\Delta\nu_{\text{locking}} \propto k_{\text{eff}} \sqrt{1+\alpha^2} \sqrt{\frac{P_{\text{inj}}}{P_{\text{pillar}}}}\,, \tag{4.1}$$

where k_{eff} is the effective injection ratio, α the linewidth enhancement factor [148], P_{pillar} the output power of the QDML and P_{inj} the injection laser power. In the case of edge emitters where the reflectivity of the surface is quite low, k_{eff} is the inverse of the laser cavity round trip time [155–157]. Regarding the QDMLs, typical values are $\alpha \approx 3$ [158] and $k_{\text{eff}} \approx 10$ Ghz is related to the photon lifetime [154, 159].

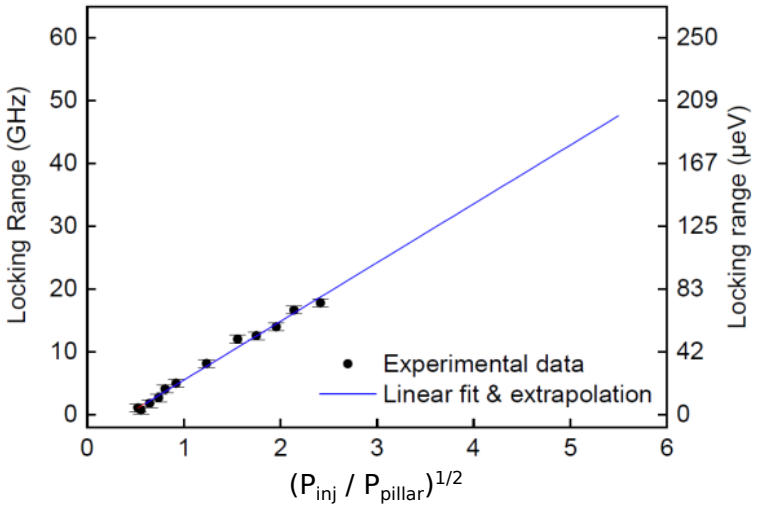

Figure 4.4: Locking range according to the injection strength ratio. Black dots are the experimental characterization, the blue line is a theoretical fit and extrapolation until 50 GHz which corresponds to the range required for locking the entire QDMLA [18].

An experimental characterization of injection locking lasers in our sample was performed by our collaborators at TUB [18]. The results are presented in Fig. 4.4, where the locking range is represented as a function of the square root of the ratio of the injection power and the pillar power. The experimental data (black dots) have been fitted and extrapolated up to a locking range of 50 GHz, which is comparable to the one of our QDMLA [18]. To obtain this range, the injection power has to be approximately thirty times higher than the output power of the QDML. Considering an output power $P_{\text{pillar}} = 1$ µW per QDML and an array comprising 900 lasers, the information injection laser should deliver 27 mW on the sample. This power can realistically to be attainable in spite of elements along the optical path such as beamsplitters or the diffractive element causing losses.

A frequency variation of 50 GHz at 980 nm corresponds to a variation of 0.16 nm. The thickness of the QDML's cavity is $(\lambda \pm 0.16)/n_{\text{GaAs}} = 280 \pm 0.05$ nm, with $n_{\text{GaAs}} \simeq 3.5$ the GaAs refractive index. Considering the GaAs lattice constant is 0.565 nm [160], the precision of the resonant cavity length should be higher than a mono-atomic layer, therefore not physically achievable. However, there is a relationship between the diameter of the mi-

cropillar and the wavelength [161, 162]. The nanostructuring of micropillars can therefore potentially mitigate for inhomogeneities during the materials growth [121, 163]. Figure 4.5 shows the result of this technique applied to a sample of 900 micropillars which have a central wavelength of 1002.8 nm. 850 micropillars (94.44%) have an optical frequency variation of less than 50 Ghz with a standard deviation of 0.132 nm. A SEM micrograph of a QDMLA at 980 nm center wavelength is shown in Fig. 4.6. The spacing between the micropillars is 8.37 µm. Each QDML has its own particular diameter in order to homogenize the array's wavelength.

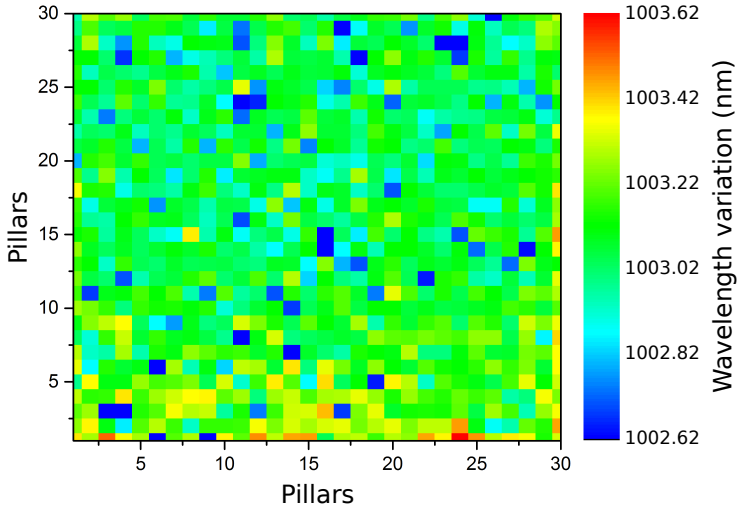

Figure 4.5: Wavelength characterization by TUB of a QDMLA optimized for a emission wavelength around 1 µm. The average wavelength of 850 out of the 900 QDML is 1002.8 ± 0.066 nm, *i.e.* almost 95% have a laser frequency detuning lower than 50 GHz.

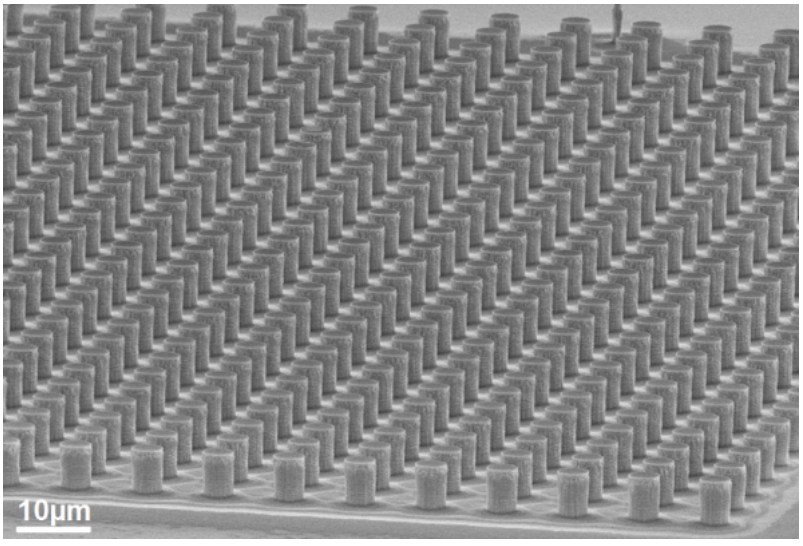

Figure 4.6: SEM image of 900 QDMLs lasing at 980 nm [121]. The array's pitch is 8.37 µm and each micropillar diameter is optimized for wavelength homogeneity.

4.2.3/ CRYOGENIC COOLING AND MECHANICAL STABILITY

The QDMLA were optimized to operate at 77 K, the temperature of liquid nitrogen. Such cooling is necessary to reduce non-radiative losses by thermal escape [18], and the ensemble of QDs' gain profile was adjusted to spectrally match with the optical mode at this temperature. The sample placed in the chamber of a liquid nitrogen cryostat (Oxford Instrument, OptistatDN-V). To obtain thermal insulation to the room temperature of the exterior, and to avoid condensation of water and other gases present in the atmosphere, a secondary vacuum ($\simeq 10^{-5}$ mbar) is created in the chamber.

In order to ensure the precision required for optical pumping of the micropillar lasers, the sample is mounted on three Attocube piezomotors stages, two to compensate for a potential sample inclination (ANGt101 and ANGp101) and a third (ANPz102) for translation of 5 mm range to adjust the sample's position relative to MO^S's focal plane. The copper sample holder includes a photodiode which measures the pump laser power in transmission of the sample and a thermistor to measure the local temperature.

The cryostat's nitrogen tank is located at the top of the cryostat chamber, and initially the piezomotors stack and the sample holder were directly suspended vertically from the cold finger. The latter is a 300 mm long copper rod submerged in the liquid nitrogen tank, whose extremity ends in the cryostat vacuum chamber. This standard configuration resulted in two major problems. The first is low thermal conduction of the piezomotors, the second is a strong mechanical instability of the sample. By imaging the QDMLA, we observe erratic movements of a micrometer scale every few seconds. We attribute this instability of the cold finger. The latter and the stack of piezomotors act comparable to a cantilever of 300 mm in length. A small angular cold finger instability inside the nitrogen tank is therefore amplified at the sample location. In this configuration, controlled optical experiments were rendered impossible due to these mechanical instabilities.

The strategy to solve this problem was to mechanically decouple the cold finger from the sample, see Fig. 4.7. To achieve this we build a non conductive mechanical support, a plastic (PVC) tube 150 mm high. Its lower part is supported by the bottom of the vacuum chamber and only this lower section is in contact with the vacuum chamber walls which are nearly at room temperature. At the top of the tube, an in-house fabricated manual XY translation stage was added. The piezomotors stack and the sample holder were suspended from this XY stage and inside the plastic tube. The thermal conduction is performed by an assembly of flexible copper foils which connect the cold finger to the copper translation stage, and then to the sample holder.

However, this layout is subject to strong constraints which are notably linked to the cryostat chamber diameter. Apart from its base, the plastic tube must not touch the chamber wall to avoid additional thermal links to the warm cryostat walls. To ensure enough space for the piezomotors and for lateral displacements of one millimeter of the sample while maintaining sufficient mechanical rigidity, the wall of the plastic tube is 2.15 mm thick and features four openings to let the corners of the piezomotors pass. Only a 0.55 mm clearance is left between the vacuum chamber walls and the outside plastic tube. This vacuum distance is sufficient to be thermally low-conductive and the plastic tube length and its low thermal conductivity theoretically provide thermal insulation of the sample from the outside.

The internal diameter of the tube measures 31.75 mm and the width of the piezomotors is 24 mm. By deducting the maximum lateral movements of the sample, there remain

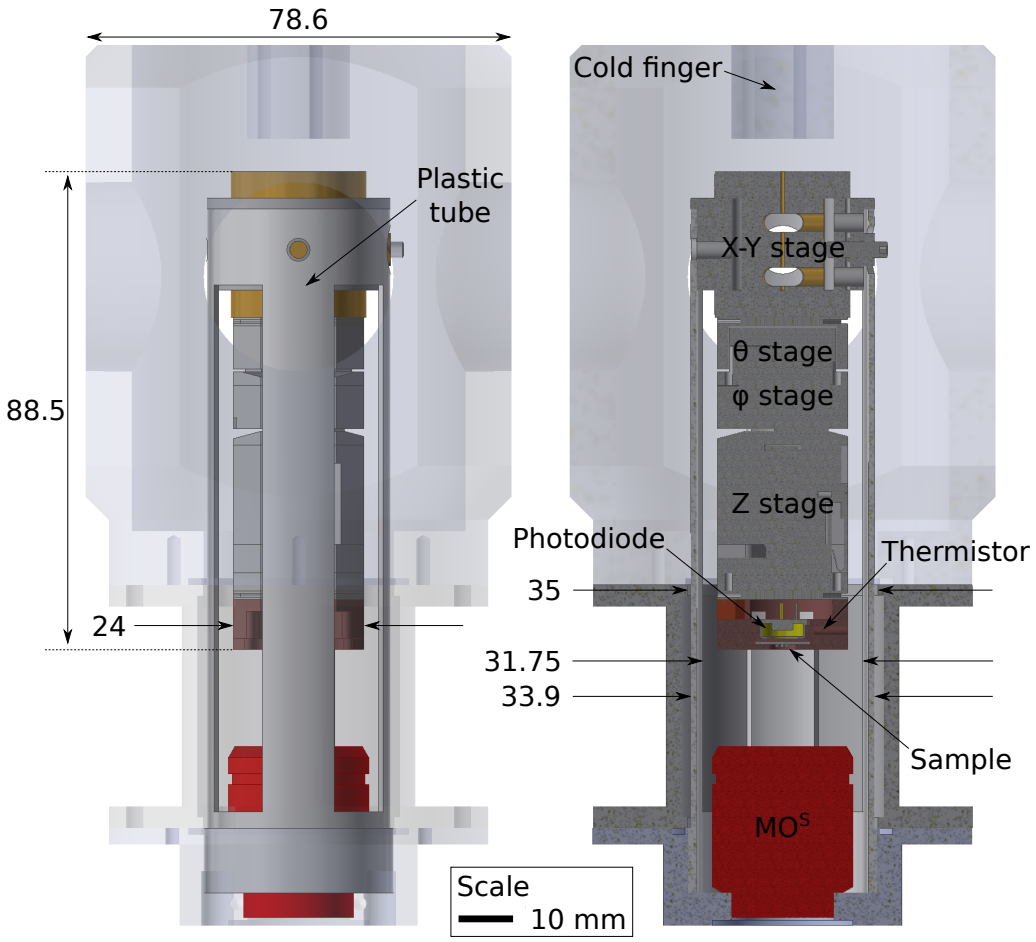

Figure 4.7: Mechanical decoupling of the sample from the cold finger. The objective microscope MO^S is located inside the vacuum chamber to reduce optical aberrations. The thermal bridges performed by flexible copper foils connecting the cold finger to the XY stage, then from this to the sample holder, are not shown for clarity. The drawing on the right is a sectional view where is shown the sample, a photodiode and a thermistor.

2.9 mm to allow ten electrical connections to pass for the photodiode, the thermistor and the three piezomotors and a thermal bridge made of flexible copper foils to cool the sample. An additional difficulty is that the sample must be able to move vertically along the optical axis in 5 mm range. The thermal and electrical connections must then be designed for the Z piezomotor in the extended position. However, in the contracted position, the connections have no other choice than to take up space radially, and the maximum lateral distance between a stretched and loose wire increases with its length (more details in appendix A). Faced with these constraints, the thermal bridge in form of a double helix of ten copper foils, each foils 0.09 mm thick, was realized. In this configuration, the maximal lateral space decreases with the number of turns around the piezomotors stack. However, the length of the thermal bridge also increases with the number of turns, thus reducing the heat flow. In our experimental situation, we choose a single turn as a compromise.

Illustrated in Fig. 4.8, we then performed experimental measurements of the temperature during the cooling of the cryostat. Several configurations were tested, from the empty chamber to the complete sample holder mount. Each element has been introduced step

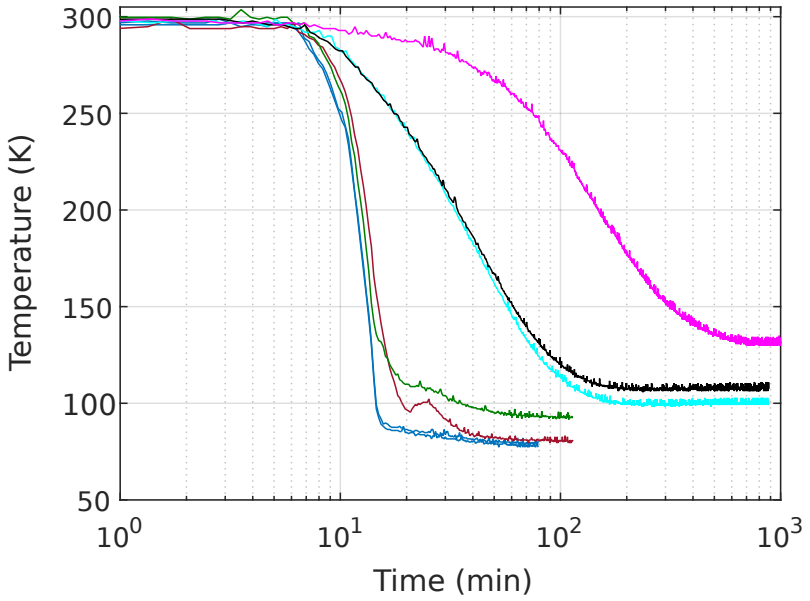

Figure 4.8: Temperature characterizations during the cryostat cooling. The blue and red lines are the temperature of the cold finger. The green line represents the extremity of the first copper foils. When connected to the XY stage, the temperature above and below this part is cyan and black lines, respectively. Finally, the magenta line is the sample holder temperature when all the elements are present in the chamber.

by step in order to characterize their own impact. Two thermistors were first calibrated in contact with the cold finger at 77 K (blue lines). Then we added the first copper foils. One of the thermistors remained in place (red line) and the other is placed at the extremity of foils (green line) where the temperature reaches ~ 90 K. We then connected the foils to the XY stage mounted on the plastic tube. The characterization is repeated with one thermistor above the XY stage (cyan line) and the second below (black line). Noteworthy, below the XY stage it takes ten times longer for the reaching thermal equilibrium. This delay is the result of the stages relatively large mass, and we find the bottom of the stage to be at ~ 100 K. We then see an eight degree temperature difference between the upper and lower sides of the XY stage. Finally, we placed the piezomotors and the second foils, we measured the temperature directly in contact with the sample holder (magenta line). The final temperature stabilizes at 130 K after twelve hours. We note that this temperature obtained is significantly higher than our theoretical calculations. A first explanation is the highly confined space, which makes it difficult to install the thermal bridge between the cold finger and the XY stage and to ensure maximum thermal conductivity an improvement in the contact surfaces is one possible solution. The second explanation is linked to potential heat loss between the XY stage and the sample holder. Given the limited spaces, it is likely that contact between the plastic tube and the cryostat vacuum chamber has occurred, thus creating a thermal bridge to the outside.

Figure 4.9 is a characterization from the group at TU Berlin of the QDML's lasing threshold at different temperatures. The experimental data (black squares) well matches a linear dependency (red line). For the sample operated at 130 K instead of 77 K, the efficiency of QDML is reduced by 38%. The temperature which is obtained experimentally is therefore

not optimal but is sufficient for the operation of the QDML and the first optical tests.

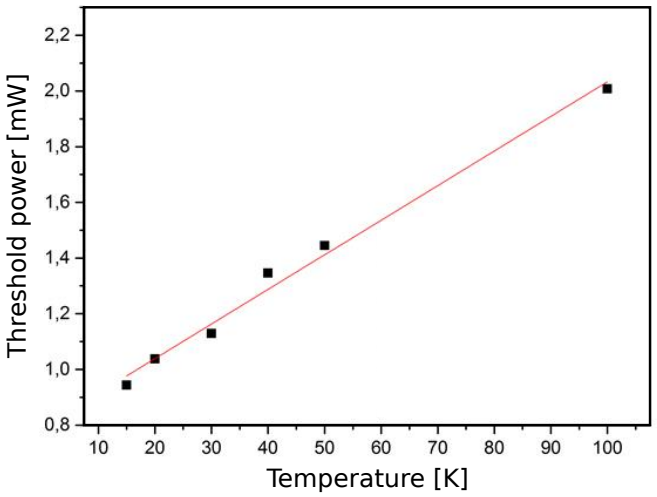

Figure 4.9: Characterization from the group at TU Berlin of the QDML laser threshold as a function of their temperature. The experimental data (black squares) follow a linear dependence. We estimate the laser threshold at 130 K and deduce a loss of QDMLs efficiency around 38% in comparison with an optimal operation at 77 K.

4.3/ OPTICAL PUMPING

Pumping of a QDML can be performed electrically or optically. In the first case, two electronic contacts at the top and the bottom of the QDML are deposited during fabrication. Furthermore, DBRs have to be doped with donors and acceptors to allow electrons and holes to reach the gain material, where upon relaxation into the QDs they form excitons that enable optical amplification via stimulated emission. This technique is well controlled, efficient and practical due to the control of the laser by off-the-shelf electronic power supplies. However, the application of this technique to a QDMLA comprising a large number of QDMLs becomes complicated. In the case of a neural network, each laser has to be individually addressed, one wire per laser therefore has to be realized without any of the wires crossing in the same plane [115]. This quickly becomes a circuitry problem and the space dedicated to connections becomes predominant. As our motivation is creating a demonstrator of a large-scale optical neural network, optical pumping is considered. In this configuration, no electrodes are needed. A pump laser of an energy higher (wavelength shorter) than the QDML lasing transition is focused onto the QDML, creating excitons which provide the required population inversion.

Pumping depends on many parameters which are challenging to independently determine experimentally. In this first part, we will introduce the most important contributions to the overall processes. The diagram in Fig. 4.10 summarizes the different interactions present during the optical pumping of QDMLs. In order to be consistent with the laser rate equations [164] the arrows in Fig. 4.10 represent rate, *i.e.* a number per time unit.

Only a fraction κ of the pump photons is absorbed by the gain medium. The rest is either reflected or absorbed by the DBR, or passes through the QDML without interaction. Absorbed pump photons create electron-hole pairs called excitons n_e. These excitons can

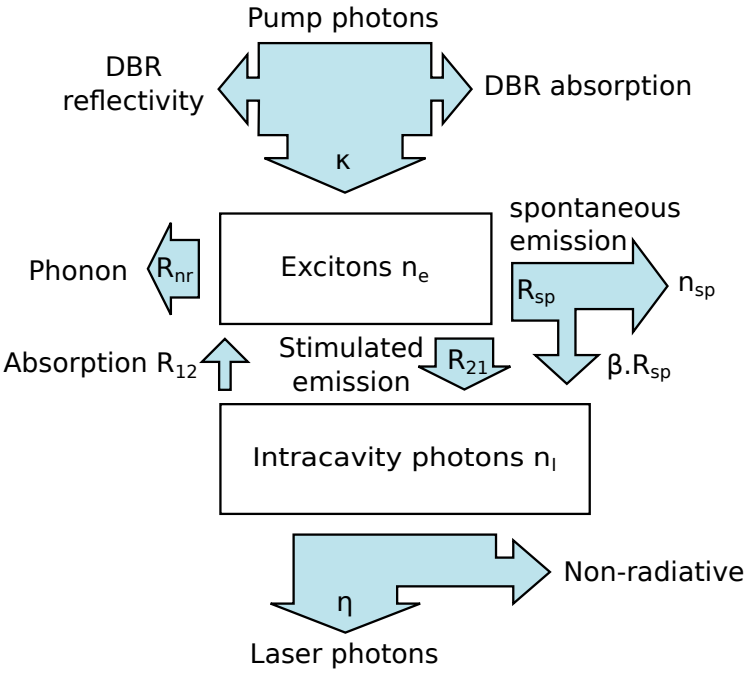

Figure 4.10: Flowcharts of QDMLs optical pumping representing the different processes and their effects. Arrows represent a rate, *i.e.* a number per time unit.

recombine in several ways. Not radiatively with phonon radiation at rate R_{nr} resulting in heating. Another recombination is radiative through spontaneous emission with a rate R_{sp}. A certain fraction β of the spontaneous emission photons is captured in the QDML's lasing mode. Factor β corresponds to the spontaneous emission coupling ratio [165]. Here, we limit ourselves to the single mode case. The captured photons n_l resonate in the laser cavity and induce stimulated emission and absorption with rate R_{21} and R_{12}, respectively. Stimulated absorption regenerates the excitons at the lasing level, which is depleted by stimulated emission. A fraction of lasing mode photons stay in the cavity, a fraction η leaves the QDML due to outcoupling and another part is absorbed by the DBR.

4.3.1/ PUMP ABSORPTION

Figure 4.11 describes the processes inside the semiconductor sample relevant for optical pumping of QDMLs. Figure 4.11(a) illustrates the bandgap energies quantum states at different positions inside the QDML as well as photon energies of the different pump concepts. When pumping around 912 nm, the pump photon energy creates excitons only inside the three wetting layers. The situation is different for pumping with wavelengths shorter than 830 nm, where GaAs of laser cavity and in the DBR is absorbing pump photons.

The pump first passes the top DBR where transmission $T^{\mathrm{DBR}}(\lambda)$ depends of the DBR reflectivity $R^{\mathrm{DBR}}(\lambda)$ and the DBR absorption $A^{\mathrm{DBR}}(\lambda)$. Since we do not have direct access to the transmission, we can measure the DBR reflection as a function of the wavelength as shown in Fig. 4.12. The DBR absorption $A^{\mathrm{DBR}}(\lambda) = 1 - e^{-\alpha^{\mathrm{DBR}}(\lambda,T)L^{\mathrm{eff}}}$ can be calculated by absorption coefficient $\alpha^{\mathrm{DBR}}(\lambda, T)$ and the effective absorption length L^{eff}. By neglecting

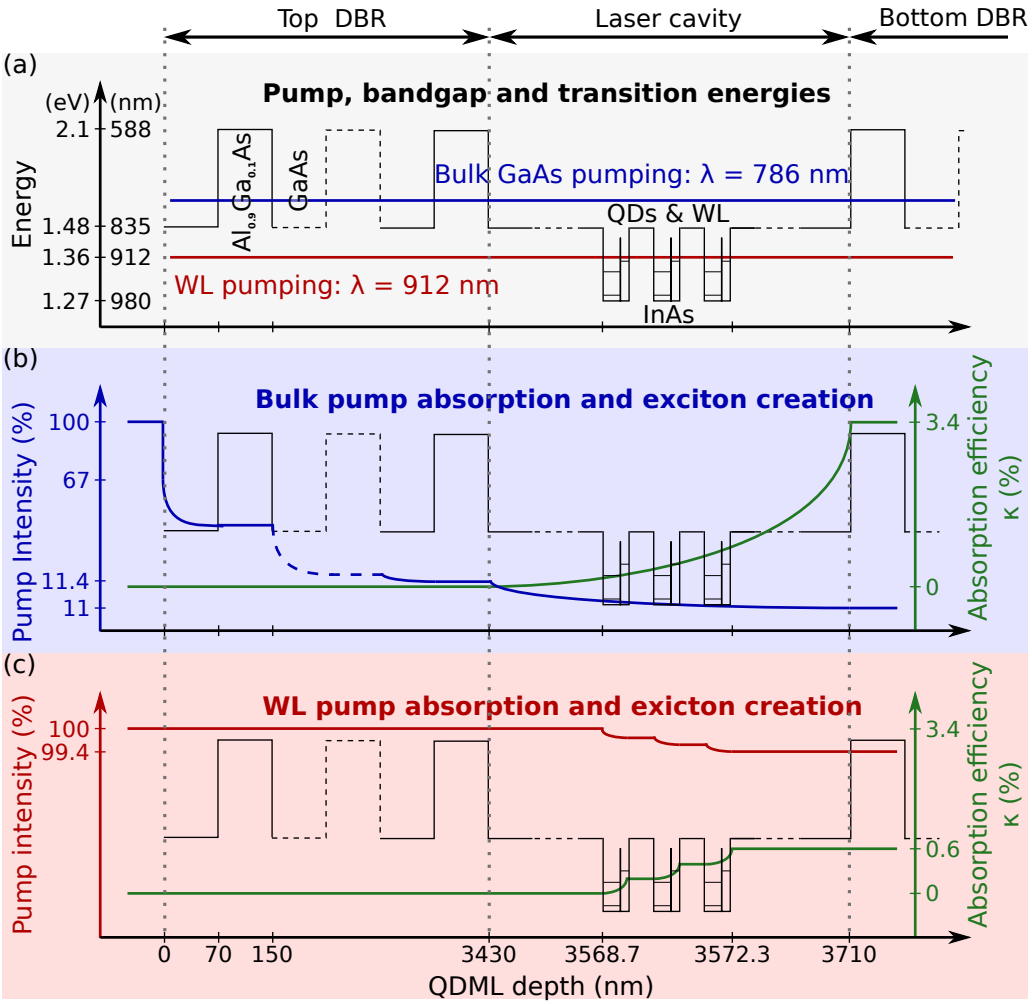

Figure 4.11: Schematic illustration of the relevant mechanisms for optically pumping QDMLs. Panel (a) gives the energies of wetting layer and bulk GaAs pump lasers, bulk material bandgaps and the WL and QD transitions. Panel (b) and (c) illustrate the pump intensity as pump photons traverse the sample, and the resulting useful exciton creation efficiency for bulk GaAs and WL pumping, respectively. The left (right) y-axis in (b) and (c) give the local pump intensities (pump-photon to pump exciton conversion ratio).

other possible losses, the transmission of the top DBR is

$$T^{\mathrm{DBR}}(\lambda) = [1 - R^{\mathrm{DBR}}(\lambda)]e^{-\alpha^{\mathrm{DBR}}(\lambda,T)L^{\mathrm{eff}}} . \tag{4.2}$$

The transmitted photons arrive in the laser cavity and are converted into an exciton through absorption of the gain medium according to

$$A^{\mathrm{g}}(\lambda) = 1 - e^{-\alpha^{\mathrm{g}}(\lambda,T)L^{\mathrm{g}}(\lambda)} , \tag{4.3}$$

where $\alpha^{\mathrm{g}}(\lambda, T)$ and $L^{\mathrm{g}}(\lambda)$ are the absorption coefficient and length of the optically pumped medium creating gain, respectively. The result is an absorbed pump power

$$P^{\mathrm{abs}}_{\mathrm{pump}} = \kappa(\lambda)P_{\mathrm{pump}} , \tag{4.4}$$

with $\kappa(\lambda) = T^{\mathrm{DBR}}(\lambda)A^{\mathrm{g}}(\lambda)$ as the absorption efficiency.

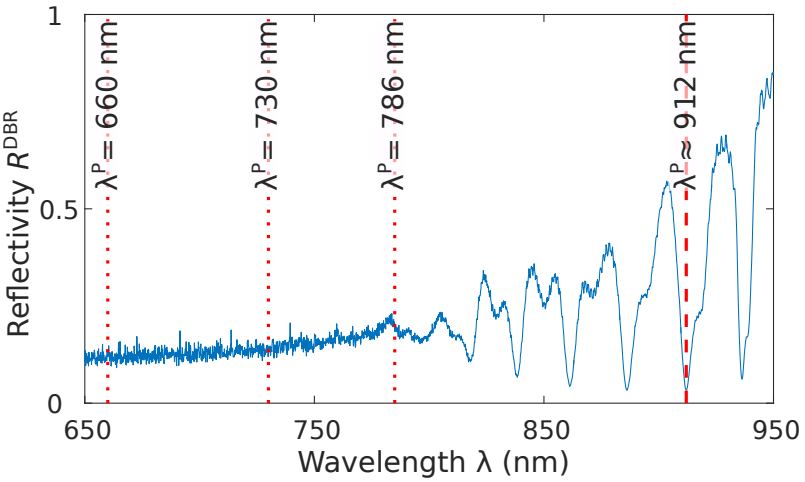

Figure 4.12: DBR reflection spectrum measured with a supercontinuum. Pumping wavelengths for the wetting layers (WL) and GaAs are indicated by the red dashed or the dotted lines, respectively.

Pumping the optical transition of the bulk in GaAs (see Fig. 4.11(b)), which corresponds to a continuum of states with transition energies higher than its bandgap at 1.494 eV (831 nm) at 130 K [166, 167]. We investigate the GaAs bulk pumping with three wavelengths, $\lambda = 786$ nm (Thorlabs FPL785S-250), $\lambda = 730$ nm (Roithner LaserTechnik HL7302MG), $\lambda = 660$ nm (Thorlabs LP660-SF20). Optically pumping in this range has an undesired side effect: the GaAs layers of the upper DBR have a strong absorption by creating excitons which consequently results in losses. Indeed, the quantum tunneling of these excitons towards the QDs is strongly suppressed by the energy-barrier of the $Al_{0.9}Ga_{0.1}As$ layers, and therefore not contributing to the lasing process. Pump energies used are below the $Al_{0.9}Ga_{0.1}As$ bandgap at 2.06 eV (600 nm), hence this material is transparent for all optical pumps in explored in this thesis. The effective absorption length $L^{\text{eff}} = 1603$ nm corresponds to the 23 GaAs layers thickness. Using the data of reference [168], which we extrapolated to 130 K according to reference [169], we calculated the absorption coefficients of GaAs $\alpha^{\text{DBR}}(786\text{ nm}, 130\text{ K}) = 1.248$ µm^{-1}, $\alpha^{\text{DBR}}(730\text{ nm}, 130\text{ K}) = 1.532$ µm^{-1} and $\alpha^{\text{DBR}}(660\text{ nm}, 130\text{ K}) = 2.615$ µm^{-1}. Furthermore, the DBR reflection at these three wavelength plotted in dotted lines in Fig. 4.12 cannot be neglected. Combining DBR reflection and absorption, we obtained $T^{\text{DBR}}(786\text{ nm}) = 0.114$, $T^{\text{DBR}}(730\text{ nm}) = 0.074$ and $T^{\text{DBR}}(660\text{ nm}) = 0.014$ [170]. Regarding exciton creation, the entire laser cavity absorbs pump photons, which in the end can relax into the QDs due to carrier diffusion ranges on the order of 10 µm [171]. The length and the absorption coefficient of the gain medium are then $L^{\text{g}}(\lambda) = L_{\text{LC}} = 279$ nm and $\alpha^{\text{g}}(\lambda) = \alpha^{\text{DBR}}(\lambda)$, respectively. The absorption efficiency κ was calculated for different pumping wavelengths for the eight QDMLs characterized (details will be provided in Sec. 4.3.5). The large length of the gain medium compensates for the low DBR transmission, and we get as absorption efficiency $\kappa^{\text{GaAs}}(786\text{ nm}) = 3.43 \cdot 10^{-2}$, $\kappa^{\text{GaAs}}(730\text{ nm}) = 2.64 \cdot 10^{-2}$ and $\kappa^{\text{GaAs}}(660\text{ nm}) = 0.71 \cdot 10^{-2}$ [170]. The average of the eight absorption efficiency κ^{GaAs}, represented by the black line and the uncertainty by the red area in Fig. 4.13(a), shows a strong dependence on the pump wavelength. The highest absorption efficiency κ^{GaAs} is then for pump laser with photons energy a little lower than the GaAs band gap [169]. The pump absorption efficiency varies by approximately a factor of 30 between a wavelength of 600 nm and 800 nm.

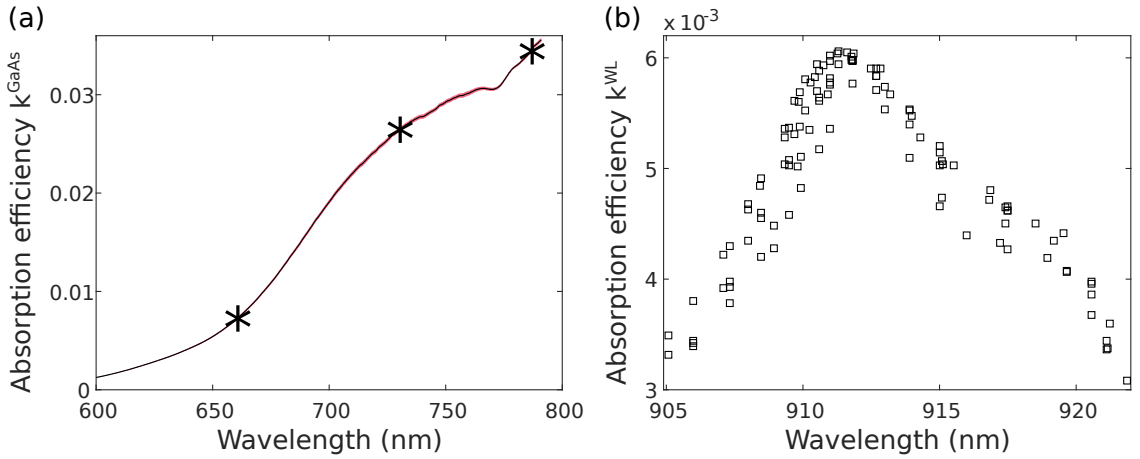

Figure 4.13: Wavelength dependence of the absorption efficiency κ of the gain medium for eight different QDMLs. The average of κ^{GaAs} is represented in panel (a) by the black line and its uncertainty by the red area. The stars represent the three pumping wavelength used. The result κ^{WL} for the WL pumping is shown in panel (b).

In Fig. 4.11(c) we consider the attenuation and absorption of pump photons as well as exciton creation for pumping at the energy of the wetting layer. Directly pumping the WL transition is attractive in principle for several reasons. Firstly, the high-energy excitons created by the pump are in direct vicinity to the QDs, suggesting a high probability that these excitons are captured by the QDs and will hence efficiently populate their s-states. Secondly, the WLs transition wavelength is approximately 912 nm, which is below the bandgap energy of the DBR materials and no pump photons are lost due to absorption by the top DBR. The DBR transmission can therefore be simplified in $T^{\mathrm{DBR}}(\lambda) = 1 - R^{\mathrm{DBR}}(\lambda)$. And, finally, this transition is located inside a DBR transmission window, hence minimizing back reflection of the optical pump by the top DBR.

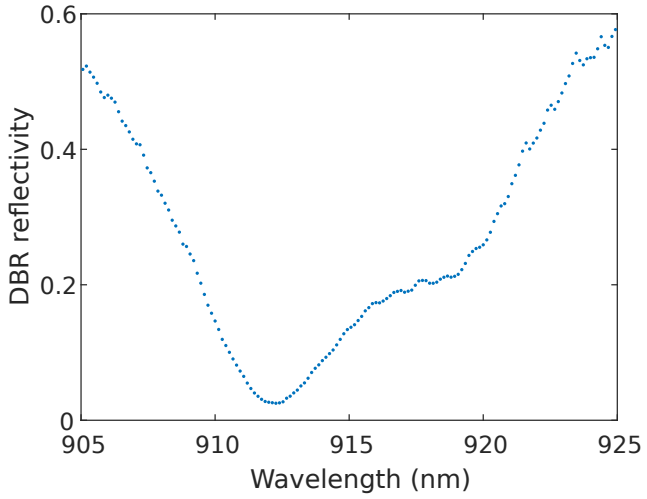

Figure 4.14: DBR transmission window around 912 nm measured with a supercontinuum.

Figure 4.14 shows an exemplary QDML reflection spectrum within the range of the DBR's reflection minimum at the WL's transition, where we a minimum $R^{\mathrm{DBR}}(912.3) = 0.025$,

hence a transmission of $T^{\mathrm{DBR}}(912.3) = 0.975$. However, these positive effects have to compete with the small absorption of pump photons by the wetting layers. Combined, the three WLs are only $L^{\mathrm{g}} \approx 1.8$ nm thin [160], which, with an absorption coefficient of $\alpha^{g}(912, 130\ \mathrm{K}) = 3.4\ \mu\mathrm{m}^{-1}$ [172]. The consequence of this very thin absorbing layers for gain is a low absorption efficiency with a maximum $\kappa^{\mathrm{WL}}(912) \approx 6 \cdot 10^{-3}$ [170]. The wavelength dependence of κ^{WL} is shown in Fig. 4.13(b).

4.3.2/ Optical pumping theory

The parameters of general pumping described in Fig. 4.10, in particular the rate linking pump excitons to lasing level photons is difficult to be experimentally determined independently or directly. For this reason we use a simplified model, illustrated by the flowchart Fig. 4.15, and introduce parameter δ. Exciton conversion efficiency δ is the fraction of excitons which ultimately contribute to the lasing transition.

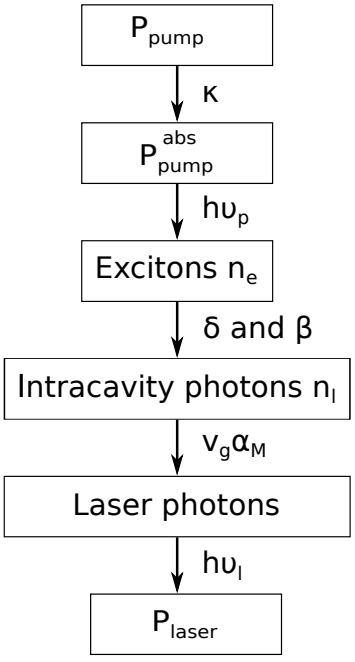

Figure 4.15: Flowcharts of a QDMLs optical pumping simplified model by introducing δ, which is the exciton conversion efficiency into a photon. Arrows represent a flow, *i.e.* a number per time unit.

Starting from a laser-rate equation description [164], the connection between pump and lasing-mode photons in the cavity n_l is given by

$$P^{\mathrm{abs}}_{\mathrm{pump}} = \frac{h\nu_l\gamma}{\beta\delta} \cdot \left[\frac{n_l}{1+n_l}(1+\xi)(1+\beta n_l) - \xi\beta n_l \right], \tag{4.5}$$

with h the Planck constant, ν_l the laser frequency and $\xi = n_0\beta/\gamma\tau_{sp}$ which depends on the cavity decay rate γ, the transparency carrier number n_0 and the spontaneous emission lifetime τ_{sp} [152, 165, 173]. Parameter ξ can be interpreted as the photon number in the lasing mode at transparency n_0. Equations (4.4) and (4.5) link the number of lasing

photons n_l to the pump power P_{pump} via material characteristic κ and fundamental laser parameters γ, β and δ. Finally, the QDML's output power P_{laser} is given by

$$P_{\text{laser}} = v_g \alpha_M h\nu_l \cdot n_l \,, \tag{4.6}$$

with the group velocity v_g and the DBR's outcoupling losses α_M calculated via

$$\alpha_M = \frac{1}{L_{\text{LC}}} \log\left(\frac{1}{\sqrt{R^{\text{DBR}}(\lambda_l)}}\right). \tag{4.7}$$

The laser cavity length is $L_{\text{LC}} = 279$ nm and the top DBR reflectivity at the laser wavelength $R^{\text{DBR}}(\lambda_l) = 0.9957$ calculated based on the number of DBR mirror pairs and their refractive index.

4.3.3/ FREE PARAMETER DETERMINATION

The QDML input-output dependency given by Eq. (4.5) describes the characteristic s-shaped curve on a double-log scale. Below threshold, *i.e.* $n_l \ll 1$, spontaneous emission dominates and Eq. (4.5) becomes linear in n_l:

$$P^{\text{abs}}_{\text{pump}}(\lambda) \approx \frac{h\nu_l \gamma}{\beta\delta}\Big[1 + (1-\beta)\xi\Big] n_l. \tag{4.8}$$

For the case of $\beta \ll 1$, this leads to $P^{\text{abs}}_{\text{pump}}(\lambda) \propto (n_0/\delta) n_l$. Above threshold, *i.e.* $1 \ll n^{\text{L}}$, similar arguments with no assumption regarding β result in $P^{\text{abs}}_{\text{pump}}(\lambda) \propto (1/\delta) n_l$. The input-output curve's slope below and above threshold therefore depend on δ while n_0 modifies the ratio between both, and β is sensitive on the nonlinearity at the lasing transition. The different operating regimes of Eq. (4.5), *i.e.* below, around and above threshold therefore allow determining n_0, β and δ, given that $\kappa(\lambda)$, γ and τ_{sp} are known.

Here, we use $\tau_{\text{sp}} = 1$ ns and the QDML's quality factor $Q = 1.5 \cdot 10^4$ reported in reference [18] for identical QDMLs to determine $\gamma = (2\pi\nu_l)/Q$. As explain in detail in Sec. 4.3.1, pump absorption efficiency $\kappa(\lambda)$ was calculated using material absorption, $R^{\text{DBR}}(\lambda)$ and QDML dimensions. β depends on the QDML cavity-geometry and the spectral overlap between QD-emission and the lasing mode [174], δ on the exciton relaxation channels. Our QDMLA is spectrally highly uniform [18], hence with similar QD-emission and lasing wavelength conditions. A similar local temperature, assuming local heating induced by P_{pump} is comparable to the same pumping conditions, should therefore result in comparable β for all QDMLs. However, taking into account the high difference of pump absorption between the WL and bulk GaAs pumping, resulting in a local heating difference, β should be different for these two pumping mechanisms. Exciton relaxation is a property of gain material and pump process, and hence δ should mostly depend on the quantum transition driven by the pump laser.

When fitting Eq. (4.5) to our output curves measured at varying conditions we use the dispersion of fit parameters as validation of our hypotheses. We therefore study the impact of having n_0, β and δ as free fit parameter of each laser, wavelength or only pumping mechanisms. We pay particular attention to unexpected wavelength dependencies of fit parameters, as this would indicate a shortcoming of our closed-form fitting routine.

4.3.4/ PUMP LASERS

The role of the pump laser is to supply energy to the QDMLs. If it is desired to simultaneously pump a large number of QDMLs, the overall pump efficiency must be optimized. This issue is crucial if we want to limit the required pump laser power to a realistic amount (typically < 1 W in continuous wave).

In order to optimize the pumping efficiency, a parameter to be controlled is the spatial profile of the pump laser's focal spot on top of the QDMLs. For a maximum efficiency the overlap integral between the spatial mode of pump laser and QDML has to be unity - ideally. The diameter of the QDML used ($\sim$4 µm) makes the lasers to emit in a single mode spatial profile for the confined mode [175]. The pump laser must be singlemode and the imaging system must be adjusted such that the pump spot mode diameter matches the QDML's mode.

The optical pumping efficiency is also determined by the absorption efficiency κ of the QDML gain medium and the exciton conversion efficiency δ. We have introduced two different optical pumping mechanisms. Regarding the bulk GaAs pumping, since the continuum of states absorbs all wavelengths below its bandgap. The only specific criteria for the pump laser is its spatial singlemode character. We can therefore use a commercial continuous wave semiconductor laser. On the contrary, WL pumping is around 912 nm and the DBR transmission window illustrated in Fig. 4.14 is narrow. Furthermore, the optimum pumping wavelength depends on details of the fabrication process and the thermal environment of the QDMLS, and hence, a pump laser tunable over several nanometers is required for this experiment. The realization of this pump laser particularly constructed for this task is detailed in appendix B.

4.3.5/ EXPERIMENTAL SETUP

We characterize the QDMLs' efficiency by measuring input-output curves with continuous wave pump laser. As illustrated in Fig. 4.16, we use a half wave-plate and a polarizer to accurately control the intensity of the pump laser on top of the QDMLs. The pump laser is detected by a bulk powermeter placed after a short-pass filter (SP, Thorlabs FESH0900) which suppresses the QDMLs laser intensity by more than 50 dB. The QDML emission is measured in a similar way after three consecutive long-pass filters (LP, Thorlabs FELH0950), which remove crosstalk from the pump-laser by suppressing its photons by 15 orders of magnitude. The recordings of both detectors are carefully post-processed in order to account for the transmission coefficients of each component inside the respective beam paths into account in order to obtain the pump and emission intensity at the top of each QDML.

We characterize in great detail a total of eight QDMLs located in different regions of the array. We target the extraction of quantitative values, which is significantly more challenging than obtaining simply qualitative dependencies. We therefore have to pay great attention to the calibration of all involved components and to the creation of reference systems. A particular challenge is that we vary the pump wavelength by as much as 250 nm, and any chromatic aberrations and modifications of pump spot diameters have to be compensated for. For this reason we have implemented strict and reproducible alignment criteria which results in equivalent experimental conditions for the different pump wavelengths.

First, we use a reference laser at 980 nm (Thorlabs BL976-SAG300), the same wave-

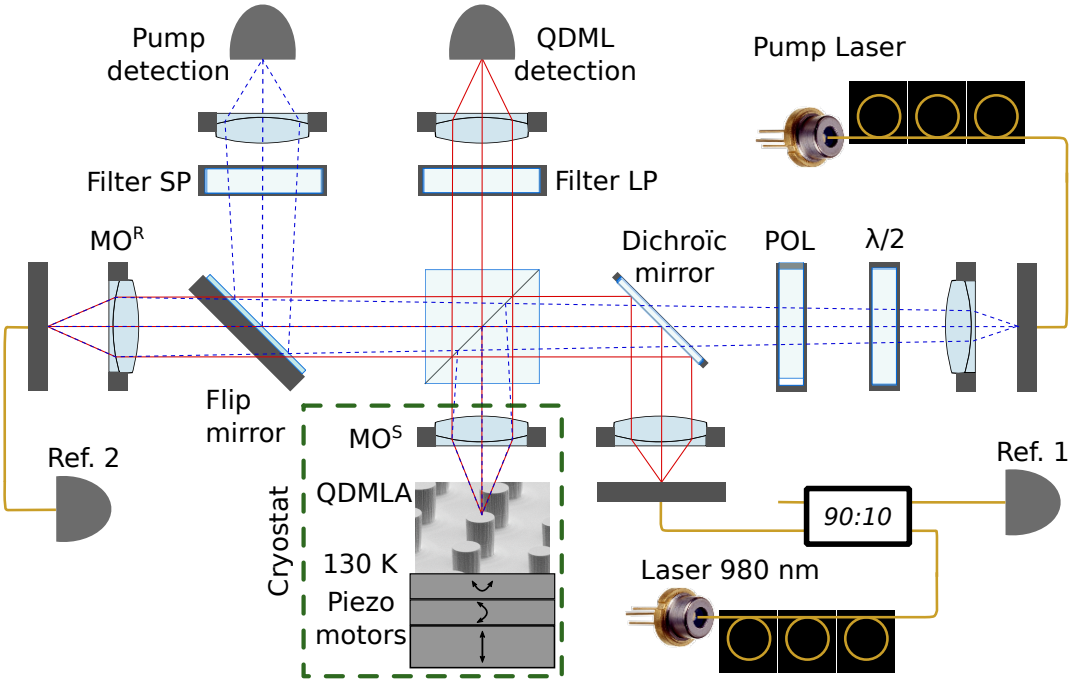

Figure 4.16: Experimental QDMLs scheme. The laser at 980 nm is used to adjust the MO^S focal distance relative to the QDMLs, for which the 90:10 output port acts as confocal reference (Ref. 1) The pump laser with a shorter wavelength is decolimated in order to maximize coupling into the reference fiber (Ref. 2) and therefore focusing on the top of the QDML.

length as the QDML emission. Before being injected into the experiment's free-space path, this laser passes through a 90:10 four-port fiber-splitter (Thorlabs TW1064R2A2A). The 90% output port is connected to the experiment via collimation optics, who we carefully align using a mirror placed inside the optical path directly following the collimation section. The mirror angle, the distance of the collimation lens (Throlabs AC254-030-B-ML) relative to the singlemode fiber output and the fiber's (x, y)-position are optimized in a such way that the feedback collected by the fiber is maximized, for which we use the 10% output-port of the fiber-splitter in backwards direction. This collimates the 980 nm beam based on a highly sensitive confocal measurement, and we created reference 1 (Ref. 1). Second, we proceed from there and reproduce the QDML injection optics outside the cryostat using the identical microscope objective (MO^R) and emulate optical injection into QDMLs with a singlemode optical fiber (Thorlabs 780HP) connected to a power-meter. Alignment of MO^R and the singlemode fiber is optimized to maximize the collection efficiency of Ref. 1 into this single-mode fiber, hence to conditions precisely matching the QDMLs' emission wavelength. This creates reference 2 (Ref. 2). The third step is adjusting the sample's position inside the cryostat. For that, we again use Ref. 1 and ensure that its focal spot is formed on top of a QDML, which was achieved by optimizing the sample's z-position such that the back reflection into the 10% output port of Ref. 1 is optimized. This creates alignment condition which again leverages confocal detection for high sensitivity, and two piezo-motors (θ, φ) control the parallel alignment, while the third motor controls the distance of the sample relative to the focal plane of the MO^S at 980 nm, respectively. And as the fourth and final step, we align each of the different pump lasers by maximizing their power injected into Ref. 2. Compensate for the small wavelength in-

Wavelength (nm)	912	786	730	660
MO transmission	0.9	0.86	0.85	0.77
Injection coefficients	-	0.59	0.57	0.64

Table 4.1: Summarize of the MO transmission and the injection coefficients through the reference single mode fiber depending on the wavelength.

duced shifts in focal distance of MO^S by slightly decolimating the pump. This procedure ensures that each pump's focal spot is precisely located on the top of the QDMLs. Only after these four steps we can be certain that the pump laser will always be focused in the same plane, regardless of the particular pump wavelength.

For calibration purposes we determine the injection coefficient of each pump laser into Ref. 2. The injection coefficient is determined by the convolution between the mode field areas of the fiber and focused incident pump laser. MO transmission values as well as the collection efficiencies at our pump wavelengths are given in table 4.1. Noteworthy, we obtain highly reproducible collection efficiencies which have a deviation of only 7%, which is excellent considering that we measured across 250 nm. This reproducibility ensures that each pump laser's spot on the QDMLA will be aligned to close to identical conditions.

The structure of the DBR was spectrally characterized using a supercontinuum source from 600 nm to 1000 nm (see Fig. 4.12). In order to probe the micropillars structure in the identical configuration as used for the pumping, the broadband laser is coupled into a singlemode fiber connected to a 50:50 four port fiber-splitter. The supercontinnum is collimated in free space and focused onto the QDMLs. The backreflection was collected by the same fiber and was measured by the OSA. The spectra were normalized by the feedback of the supercontinuum reflected off a mirror instead of the QDMLs, hence providing a calibration of the inherent losses of the setup.

4.3.6/ RESULTS

We first study a QDML emission for a constant pump power. The optical spectrum, see Fig 4.17(a), is measured by the OSA through collection by a confocal system. Main mode emission is at 979.105 nm. When the pump power is high enough, higher order modes start to appear with significantly lower powers (-16 dB and - 26dB for the second and third modes). Figure 4.17(b) shows the sample illuminated by LED (Thorlabs M1050L2) and the laser emission of a QDML, the pump laser being suppressed for image acquisition by a camera (IDS uEye UI-3482LE). We find that the emission of QDML is spatially singlemode.

BULK GAAS PUMPING

By selecting the wide GaAs pumping range from 660 nm to 786 nm, we test the validity of our approach for a variety of absorption coefficients. Starting with the pump at 786 nm, 40 input-output curves have been measured on the same eight QDMLs. An example of an input-output curve is shown in Fig. 4.18(a) where the photon number in the cavity n_l is plotted as a function of the pump power (red squares), the fit corresponding to Eq. 4.5 and its uncertainty are plotted as red and yellow lines, respectively. We obtain an average

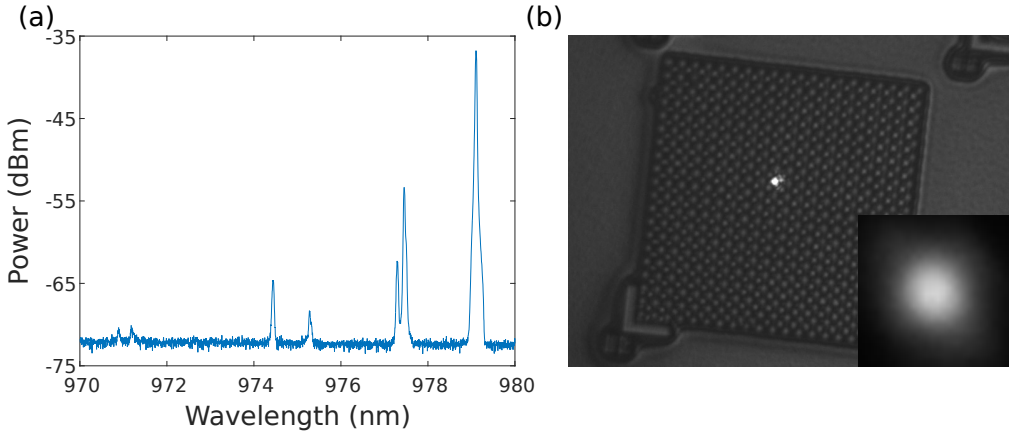

Figure 4.17: (a) Optical spectrum of a QDML. (b) Image of the QDMLA illuminated by LED where one micropillar is optically pumped beyond its lasing threshold. The pump laser was suppressed for the image acquisition. The inset is the QDML emission without illumination.

lasing threshold of $I_{\text{pump}} \approx 2$ mW. However, for shorter pump wavelengths the GaAs absorption increase and consequently the absorption by the top DBR increases. This causes an increase of non-radiative decay which locally heats the QDMLs. That induces a relative shift between wavelength and gain profile which causes a reduction in the optical gain [164, 176] in an effect known as thermal roll-over. For 730 nm and 660 nm pumping, this effect is so strong that it prohibits the QDMLs from crossing the lasing threshold. A comparison between the three different GaAs pump wavelengths is given in Fig. 4.18(b). The gray dots are the collective data of 40 input-output curves obtained for $\lambda^{\text{P}} = 786$ nm. The blue, red and green data correspond to below threshold emission for pumping at 786 nm, 730 nm and 660 nm, respectively. Here, we used a more sensitive photodetector in order to obtain a higher signal to noise ratio at the nW power-levels below threshold.

We fit experimental data with Eq. (4.5) for 786 nm pumping using $\kappa^{\text{GaAs}}(786\text{ nm}) = 3.34 \cdot 10^{-2}$. As explained in Sec.4.3.3, n_0 depends on the pumping and the QDML structure, we therefore used an independent parameter $n_0^{\text{GaAs}}(786\text{ nm})$ per QDML. However, we set one global $\delta^{\text{GaAs}}(786\text{ nm})$ and $\beta^{\text{GaAs}}(786\text{ nm})$ to fitting all QDMLs at this particular wavelength. The parameters obtained from the 40 fits are $\delta^{\text{GaAs}}(786\text{ nm}) = 1.91 \cdot 10^{-2} \pm 0.01 \cdot 10^{-2}$, $\beta^{\text{GaAs}}(786\text{ nm}) = 2.24 \cdot 10^{-2} \pm 0.03 \cdot 10^{-2}$ and $n_0^{\text{GaAs}}(786\text{ nm}) = 2913 \pm 70$ [170]. β and δ are determined with an uncertainty in the % range, and n_0 has a comparable standard deviation across the eight lasers. The β factor is within the range commonly determined in other experiments, and we find a very low exciton conversion efficiency δ. Only a couple of percent of excitons created by the pump end up in a QD transition contributing to lasing.

For $\lambda \leq 730$ nm our QDMLs do not lase, yet data recorded below threshold allows us to confirm our findings. Based on the calculated $\kappa^{\text{GaAs}}(\lambda)$ we rescale the input-output curves for all GaAs pump wavelengths and show the resulting data in Fig. 4.18(b). By following the arguments presented in Sec.4.3.3, we now define n_0 as a laser parameter independent of the GaAs pump wavelength. Regarding $\beta^{\text{GaAs}}(\lambda)$ we assume a negligible heating effect between the three GaAs pumping wavelengths, in particular below the lasing threshold, which leads us to keep the parameter constant. Finally, although we would expect essentially identical exciton conversion efficiencies since the three pumping levels are linked through efficient intra-band relaxation, we set δ^{GaAs} to be independent

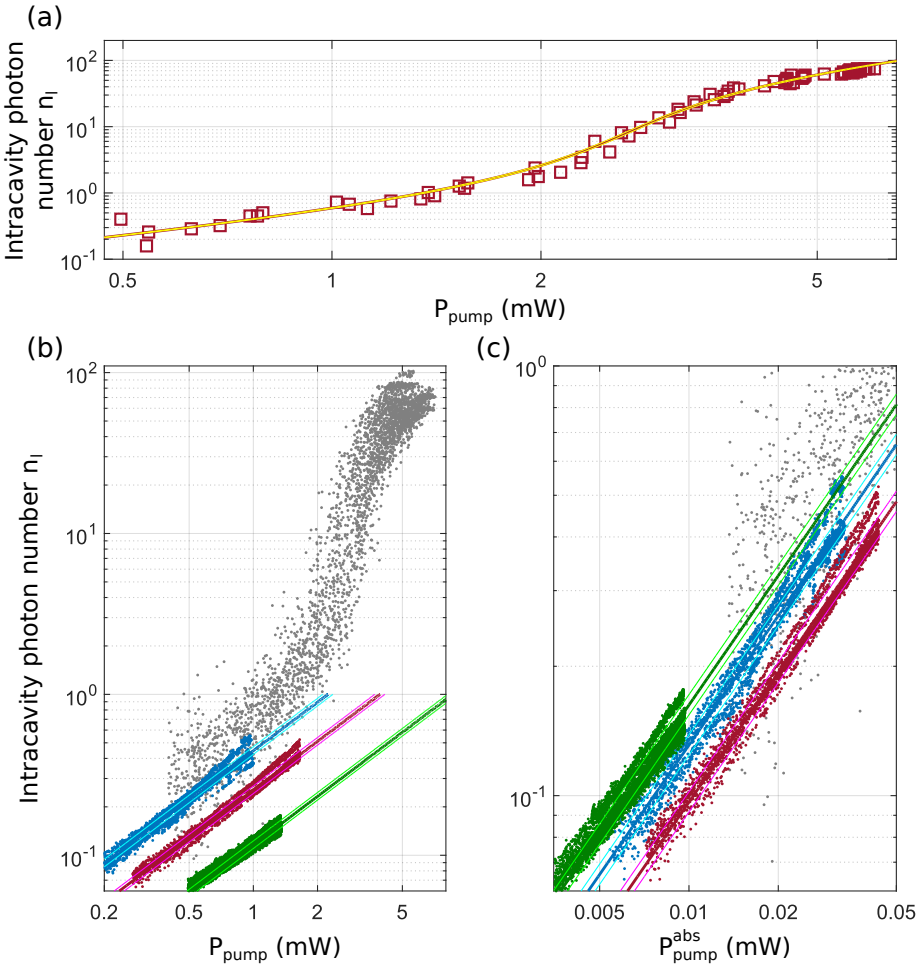

Figure 4.18: GaAs pumping results. (a) Input-output curve from one QDML pumped at 786 nm (red squares) is fitted (red line and its uncertainty in yellow) by the Eq. (4.5). All the input-output curves at 786 nm (gray dots) are plotted in panels (b) and (c) where the measurements below threshold for 786 nm, 730 nm and 660 nm pumping are plotted in blue, red and green dots. (c) Rescaling to the absorbed pump power. The superimposition of the curves shows that the main impact of the GaAs pumping wavelength is related to the absorption efficiency $\kappa^{\mathrm{GaAs}}(\lambda)$.

of the QDMLs, yet a free parameter for all three pump wavelengths. From the linear fitting of the experimental curves depicted in Fig. 4.18(b) via Eq. (4.8), we obtain a averaged $n_0^{\mathrm{GaAs}} = 3496 \pm 254$ and $\delta^{\mathrm{GaAs}}(786\ \mathrm{nm}) = 2.4 \cdot 10^{-2}$, $\delta^{\mathrm{GaAs}}(730\ \mathrm{nm}) = 1.8 \cdot 10^{-2}$ and $\delta^{\mathrm{GaAs}}(660\ \mathrm{nm}) = 3 \cdot 10^{-2}$ [170]. The fits are plotted in Fig. 4.18(b) with the same color as their respective data points, the uncertainty is represented by the lighter lines. Crucially, the very small variations in δ^{GaAs} mean that the calculated $\kappa(\lambda)$ accounts for the QDMLA wavelength dependence when pumping GaAs. On a loglog-scale efficiency variations for pumping at different wavelengths result in an offset between the curves, therefore pumping at 730 nm and 660 nm are respectively 43% and 74% less efficient than pumping at 786 nm. Plotting the data as a function of the pump power absorbed by the gain medium in Fig. 4.18(c), we find that the three curves are practically superimposed. It can therefore be deduced that the difference in pumping efficiency resulting from the wavelength

is mainly due to the change in the GaAs absorption coefficient; internal exciton relaxation and photon capture processes remain essential independent of the pump wavelength.

In order to increase the global efficiency, we tried a GaAs pumping from the side. Instead of focusing the pump laser on the top of the DBR, we readjusted the focus 3.5 µm deeper, at the beginning of the laser cavity. The idea is to minimize the absorption of DBR thanks to the part of the pump beam which does not suffer losses by avoiding DBR from the side. Ultimately the results for the three wavelengths showed no significant improvement (+0.61% for pumping at 786 nm), if not a degradation (-1.08% and -5.0% for 730 nm and 660 nm, respectively).

WETTING LAYER PUMPING

We characterized pumping in the range from 905 nm to 922 nm, which spans the DBR reflection minimum indicated by the dashed line in Fig. 4.12 and detailed in Fig. 4.14. Crucially, for a sample temperature of ~ 130 K we expect the WL's transition within this range. A typical input-output curve obtained at the DBR's transmission maximum at 912.3 nm is given in Fig. 4.19(a). The photon number in the cavity n_l is plotted as a function of the pump power (blue squares) with its fit (blue curve) and uncertainty (cyan curve). We obtain a lasing threshold of around 10 mW.

We recorded a total of 114 input-output curves for the eight identical QDMLs. According to our previous line of argument explain in Sec. 4.3.3, we defined a single β^{WL} and δ^{WL} for fitting all data. Parameter n_0^{WL} was initially free for each input-output series since it depends on the QDML's characteristics and on the pumping configuration itself. We found that the parameter n_0^{WL} strongly depends on the pump wavelength. This dependence is shown in Fig. 4.20(a) where normalized n_0^{WL} are averaged by a moving filter of width 12 (black squares). Our approach already takes DBR reflection into account, and the resulting shape and width (full width at half maximum of 14.8 meV compared to ∼10 meV in [177]) suggest that the variation is in fact related the WL's absorption spectrum. We have therefore fit (blue line) Fig. 4.20(a) by a polynomial and have incorporated it's normalized version into the absorption efficiency κ^{WL}. Results of κ^{WL} before (blue dots) and after (black squares) this post-treatment are shown in Fig. 4.20(b). Since the WL's absorption spectrum is now taking into account by κ^{WL}, we can assign a unique n_0^{WL} per QDML, and no longer for each input-output curves. After filtering, we fit the remaining 102 input-output curves and obtained: $\delta^{\mathrm{WL}} = 7,35 \cdot 10^{-3} \pm 0.07 \cdot 10^{-3}$, $\beta^{\mathrm{WL}} = 9.7 \cdot 10^{-2} \pm 0.2 \cdot 10^{-2}$ and $n_0^{\mathrm{WL}} = 627 \pm 151$ [170].

Figure 4.19(b) plots all input-output curves (gray dots). The blue, green and red curves are three examples of the same QDML for three different pumping wavelengths ($\lambda_b = 921.3$ nm, $\lambda_g = 915.5$ nm and $\lambda_r = 911.8$ nm, respectively), which results in three different absorption efficiency $\kappa^{\mathrm{WL}}(\lambda_b) = 3.1 \cdot 10^{-3}$, $\kappa^{\mathrm{WL}}(\lambda_g) = 5 \cdot 10^{-3}$ and $\kappa^{\mathrm{WL}}(\lambda_r) = 5.5 \cdot 10^{-3}$, respectively [170]. We see in Fig. 4.19(b) an increase in pumping efficiency as a function of absorption efficiency, and in general the three curves are spaced far apart. In Fig. 4.19(c) we plot the same data according to the number of pump photons absorbed by the WL, and in this case the blue, green and red curves are superimposed. The contrast between panels (b) and (c) shows that when including the derived rescaling according to the DBR's transmission and the WL's absorption spectra the input-output curves are correctly renormalized. This confirms the validity of our approach and the wavelength independence of δ^{WL} and β^{WL}.

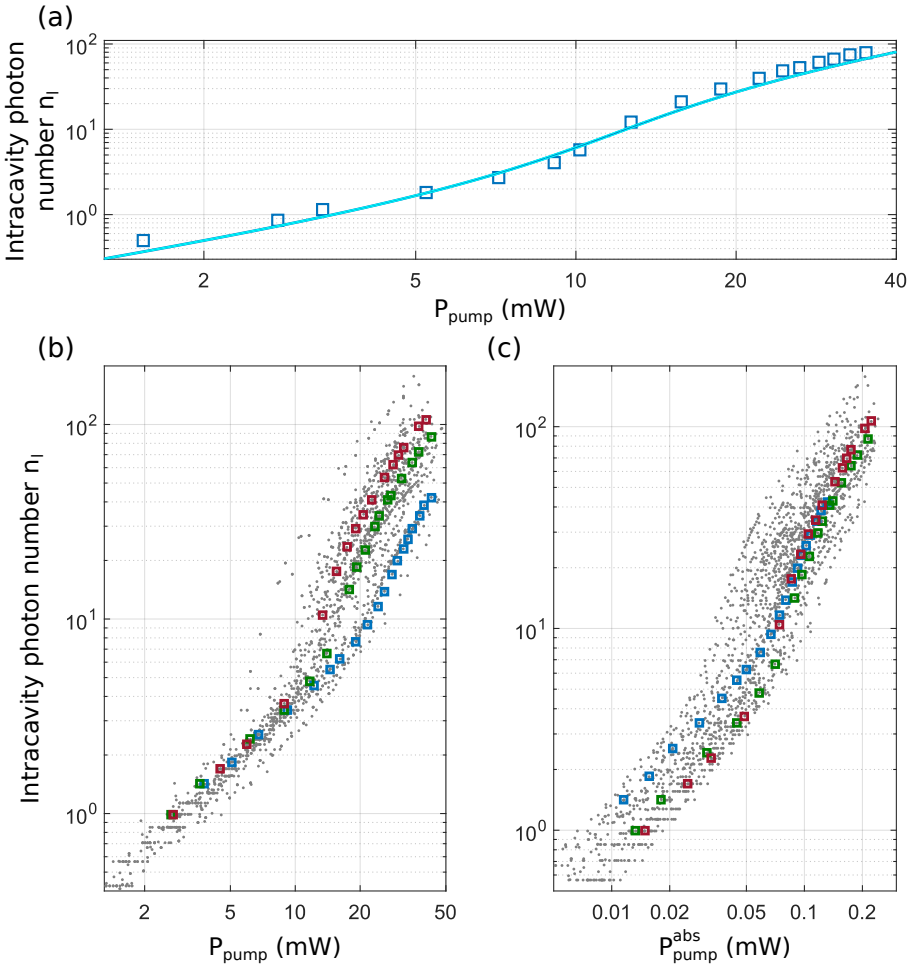

Figure 4.19: WL pumping results. (a) The input-output curve from one micropillar (blue squares) is fitted (blue line and its uncertainty in cyan) by the Eq. (4.5) where the WL spectral profile takes into account in the absorption efficiency. All the input-output curves (gray dots) are plotted in panel (b) and (c). The fact that the three curves are superimposed only in panel (c) indicates that absorption efficiency κ is related to the pump wavelength while β and δ are independent from it.

PUMPING COMPARISON AND DISCUSSION

We have characterized with great detail the input-output curves of eight QDMLs for two different pumping mechanisms. We find that the exciton conversion efficiency is over 2.6 times higher for pumping bulk GaAs than for WLs. Due to the direct contact between WLs and QDs, initially we intuitively expected an opposite ratio, and this result surprised us. However, excitons in the GaAs matrix can relax into the WL and the QDs, and this combined recombination channels could be the reason for $\delta^{\mathrm{GaAs}} > \delta^{\mathrm{WL}}$. We also find that despite its lower efficiency, WL pumping presents a significant advantage with respect to the laser's spontaneous emission confinement factor ($\beta^{\mathrm{WL}} \approx 0.1$) compared to pumping into the bulk GaAs ($\beta^{\mathrm{GaAs}} \approx 0.02$).

Regarding the global pumping efficiency of QDMLs, we obtain a lasing threshold of

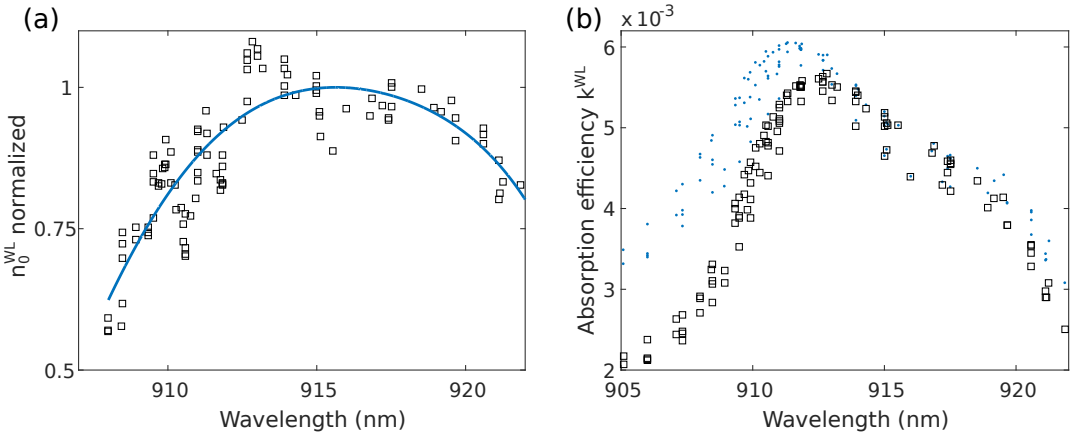

Figure 4.20: (a) Wavelength distribution of n_0^{WL} (black squares) for the WL pumping and its polynomial fit (blue line) This is the consequence of our model not incorporating the WL's absorption spectrum. (b) Pump absorption κ^{WL} for WL pumping before to take into account the WL's extracted absorption spectra (blue dots) and after (black squares).

~10 mW for WL pumping at the maximum DBR transmission. This value is too high for optical pumping of a large number of QDMLs. However, we find that the changing material's absorption coefficient has a major influence on the QDMLs' performance: when pumping at 786 nm, we obtain a lasing threshold of ~2 mW, *i.e.* a performance 5 times higher than WL pumping. Thereby, it should be possible to pump a hundred QDMLs with a commercial semiconductor laser, but further optimization of efficiency would be desirable.

At some stage we attempted to increase the absorption efficiency κ by focusing the pump laser waist at the top of the GaAs cavity in order to minimize the impact of DBR absorption by partially pumping the QDMLs through their side walls. Regarding GaAs pumping, the associated results showed no significant improvement or even a slight degradation. However, the margin of these modifications were too small to arrive at a conclusion. Ultimately, the highest WL pumping efficiency occurs when the pump beam waist is focused on the top of the QDMLs.

Our results identify two major challenges for increasing the optical pumping efficiency. When pumping the monolayer thin absorption medium of the WL unavoidably results in a very low pump absorption efficiency. We therefore find that WL pumping will always be limited to below 1% power efficiency. Similar, the bulk GaAs pumping efficiency is currently limited to a couple percent due to absorption by the top DBR. However, this effect can be highly mitigated by designing a top DBR comprising of materials with a bandgap higher than the one of the QDML's cavity. An ambitious strategy is to slightly modify the DBR composition, in particular the GaAs layers. By changing these to $Al_{0.1}Ga_{0.9}As$, the bandgap increases by 125 meV, to approximately 751.6 nm (760.4 nm) for 77 K (130 K). The GaAs refractive index would decrease by $5.4 \cdot 10^{-2}$, and as a consequence DBR reflectivity at 980 nm would only decrease from 99.57% to 99.12%. Following such a strategy the absorption efficiency will increase by a factor of 7.4, thus exceeding 20%, which leads to a lasing threshold of 270 µW. We also find that the exciton conversion efficiency κ stays low. The non-radiative decay of the excitons could then be mitigated. We propose to passivate QDMLs [178, 179] in order to reduce the defects located at the QDML's etched sidewalls which contribute to losses. For example, GaAs passivation by nitridation

allows reducing the surface state density by a factor of 6 and the effect is stable over time [180]. Such improvements using established techniques therefore promise a significant improvement of QDML's conversion efficiency by one to two orders of magnitude.

4.4/ QDMLA PUMPING

Optically pumping a large number of QDMLs simultaneously is not trivial. Uniformly illuminating a large area with the pump laser is not an option. All energy deposited between the micropillars, corresponding to the majority of the sample's surface ($\simeq$ 82% in our case), would not contribute to optical pumping. Even worse, it will contribute and hence increase parasitic sample heating. We will therefore diffract the pump laser such that each diffraction order is precisely focused on a different QDML [113].

In order to increase the flexibility of the system, the diffraction pattern for creating multiple pump spots in parallel is not created with a passive element (DOE) but using a SLM (Santec SLM-200) in reflection. The pattern can then be adjusted to the particular QDMLA. Several constraints must be taken into account, which can lead to difficulties in building the architecture of the optical system presented in Fig. 4.16. Since we want to pump a large number of QDMLs at the same time, the fewest optical elements should be used, in order to maintain the highest pump power at the sample. The incidence angle of the laser on the SLM must be close to normal (93 degrees in our experiment) to maximize the SLM's efficiency and use a fiber polarization controller removes reflection losses of waveplates when adjusting to the required polarization (p-polar for phase modulation by the SLM). In addition, the diffraction orders having an angle in comparison to the zero order, the distance between the SLM and the microscope objective is restricted to 208 mm in order to avoid beam clipping by the MO^S clear aperture of 11.32 mm [17].

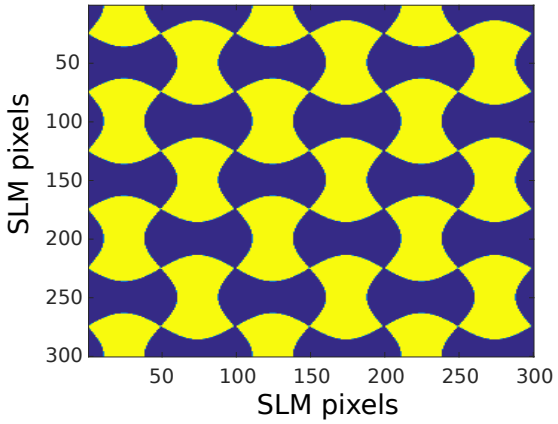

Figure 4.21: Example of the diffraction grating used to create a 3x3 optical pattern. Three parameters must be adjusted such that the diffracted orders are superimposed on the QDMLs. Once the phase modulation amplitude, grating orientation, and spatial period are properly adjusted, the pattern is then extended to all pixels in the SLM.

We start with a phase mask with a periodic structure creating a 3×3 diffraction pattern as in [16, 113], and to move to masks creating $N \times N$ in the future. An example of the phase mask for a 3×3 diffraction pattern is shown in Fig. 4.21, which is identical to the DOE in section 2.1.4.

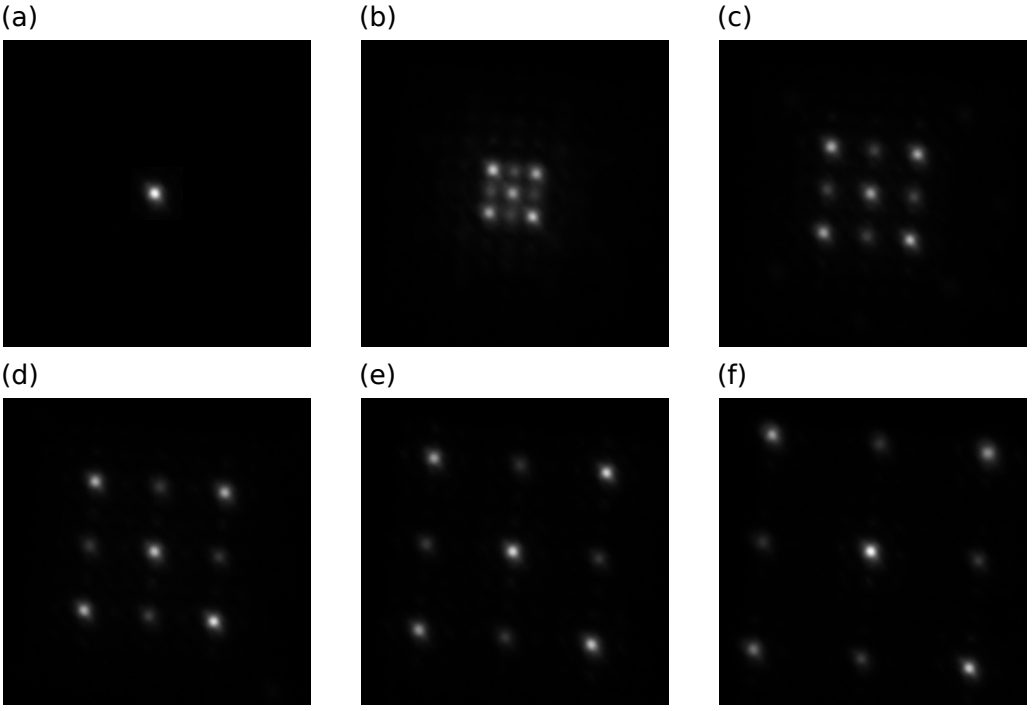

Figure 4.22: Laser emission images of micropillars optically pumped. The pump laser is spectrally suppressed. (a) The pump laser is not diffracted and pumps a single micropillar. Then the diffraction grating with optimized parameters is displayed on the SLM, a 3x3 optical pattern simultaneously pumps nine QDMLs. (b) The spatial period of the grating is adjusted in such a way that the diffraction orders are superimposed on the neighboring micropillars of the specular order. (c), (d), (e) and (f), these spatial periods are divided by two, three, four and five, respectively.

The phase modulation amplitude of the binary phase mask must be adjusted to obtain an uniform optical power between the different diffractive orders. In addition, the phase mask's orientation and the spatial period must be adjusted such that the diffractive orders are superimposed on top of the individual QDMLs. The emissions of a 3×3 array of optically pumped QMPLs are shown in Fig. 4.22. In panel (a) the pump laser is not diffracted, only one QMPL is pumped. The optimized phase mask is then displayed by the SLM, and in panel (b) the emission of the optically pumped 3×3 QDMLs is shown. By modifying the spatial period of the phase mask, we can readily adjust the distance between the different diffractively created pump spots. As a consequence, we can selectively pump 3×3 QDMLs with different separation between the lasers, see Fig. 4.22(c-f). The laser emission presented in panels (c), (d), (e) and (f) are obtained for a pitch between the diffracted pump spots equivalent to 2, 3, 4 and 5 times the QDMLA pitch, respectively. We find that even with an optimized phase modulation, the intensity of the laser emission is not uniform. QDMLs in the diffracted pattern corners receive more pumping intensity than those located on the horizontal or vertical diffraction axis. Nevertheless, the intensity of the non-diffracted order is not predominant, and the emission of the central QDML in Fig. 4.22(b) represents only 12% of the 9 QDML emission.

The orientation and the diffraction grating period is studied in more detail in Fig. 4.23. The laser emission of all nine pumped QMPLs is integrated on one detector for different values of the two parameters.

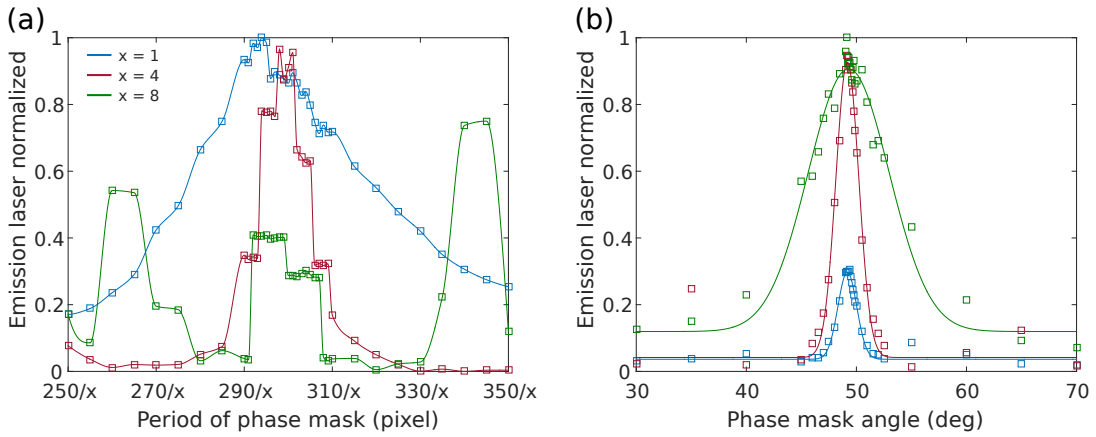

Figure 4.23: Characterization of phase mask parameters for creating a 3×3 pump spot array. Nine QDMLs are pumped simultaneously and their combined power is detected. (a) The spatial period is characterized for optimal pattern orientation. The blue curve corresponds to a pumping nine neighboring QDMLs ($x = 1$). The spatial period of the phase mask is divided by $x = 4$ and $x = 8$ for the red and blue curves, respectively. (b) The spatial period corresponding to the blue curve is now set to 295 pixels, then divided by 4 and 8 for the red and green curves, respectively. The optimal orientation for all three measurements excellently matches.

When we study the phase mask's period such as in Fig. 4.23(a), the phase mask's rotation was previously optimized. A first measurement is performed with a spatial period ranging from 250 to 350 SLM pixels (blue curves). We observe that the laser emission of the nine QDMLs increases to a maximum which corresponds to the configuration of Fig. 4.22(b). In order to increase the sensitivity of the characterization, two further measurements were performed. The spatial period of the network is divided by $x = 4$ (red curves) to increase the impact of its relative variation. This is reflected in the measurements by a thinner emission peak of the red curve. A space period of the phase mask results in a larger separation between the individual pump spots of the 3×3 pattern, and the red curve corresponds to Fig. 4.22(e) where the QDMLs distant by 4 times the QDMLA pitch from the central micropillar are pumped. The green curve is obtained by dividing the grating spatial period by $x = 8$, further increasing the relative variation of the period. The consequence is the appearance of two peaks on each side of the central peak at a distance of ≈ 5 pixels, these correspond to the laser emission of radial neighboring QMPLs. In other words, this phase mask has sufficient resolution to distinguish the diffraction angle necessary to pump the 7th (right peak) from the 8th (central peak) and 9th (left peak) QDML starting from the central micropillar. This technique applied to this SLM would therefore have the physical possibility of pumping at least 361 QDMLs. The optimal spatial period of the phase mask is close to 295 pixels, *i.e.* 2.36 mm with an SLM pixel pitch of 8 µm, around four periods can therefore be displayed on the SLM (1920x1200 pixels). In order to satisfy the diffraction principle and obtain the desired interferences, the diameter of the pump laser beam should be at least equal to one period but ideally larger.

The orientation characterization of the phase mask is presented in Fig. 4.23(b), the phase mask's period was set to the optimum 295 pixels. The different colors represent the same configurations as in panel (a), *i.e.* the spatial period of the blue data is 295 pixels, those of the red and green data are divided by four and eight, respectively. In order to superimpose

the diffractive orders of the pump beam on the 9 QDMLs, proper angular orientation of the diffracted pattern is crucial. We then measured the 9 collective QDMLs emission for a phase mask orientation between 30 and 70 degrees. The three results (blue, red and green squares) and their Gaussian fit (blue, red and green lines, respectively) match excellently. As expected, the best sensitivity to rotation is obtained by the blue curve. We obtain an optimal angle of 49.18 ± 0.08 degrees with 95% confidence bounds.

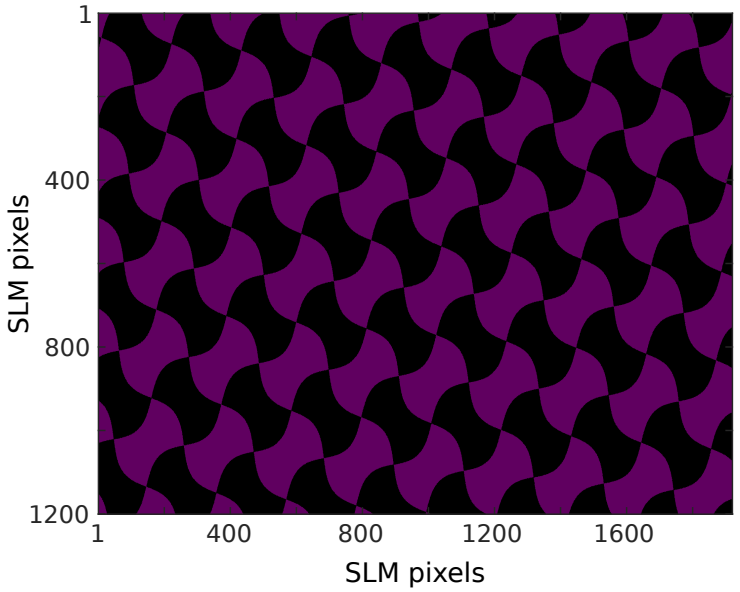

Figure 4.24: Phase mask optimized displayed across the entire SLM screen

The resulting, optimized phase mask is therefore, oriented 49.2 degrees clockwise and with a period of 295 pixels as illustrated in Fig.4.24. Based on this configuration, we will now characterize the laser emission of nine QDMLs optically pumped simultaneously. We use pumping the central micropillar without using the phase mask as a reference. The input-output curve is shown in Fig. 4.25(a) where the laser emission power of the central QDML is represented as a function of the pump power. We then load the phase mask on the SLM and measure the input-output relationship of the nine QDMLs on the same powermeter. The blue curves in Fig. 4.25(b) illustrate the result obtained. The same experiment was performed again by dividing the spatial period of the diffraction grating by seven, thus increasing the spacing between the diffraction orders by the same amount. The result obtained is plotted in red in the same figure. We observe in Fig. 4.25(b) that for a pump power nine times higher than panel (a), the emission of the nine QDMLs is only 4 to 5 times higher. Therefore, we only lose about 50% of pumping efficiency during the 3×3 optical pumping. One reason might be a reduced homogeneity between the diffraction orders of the pump laser.

The homogeneity between the individual diffractive orders is further reduced by creating from the same principle a 5×5 diffraction pattern (see Fig. 4.26). The concept of optically pumping the QDMLs should therefore be further developed to diffractively create $N \times N$ pump spots. Such diffractive beam splitters have already been obtained experimentally with excellent homogeneity [181–184], which shows that such a strategy is feasible.

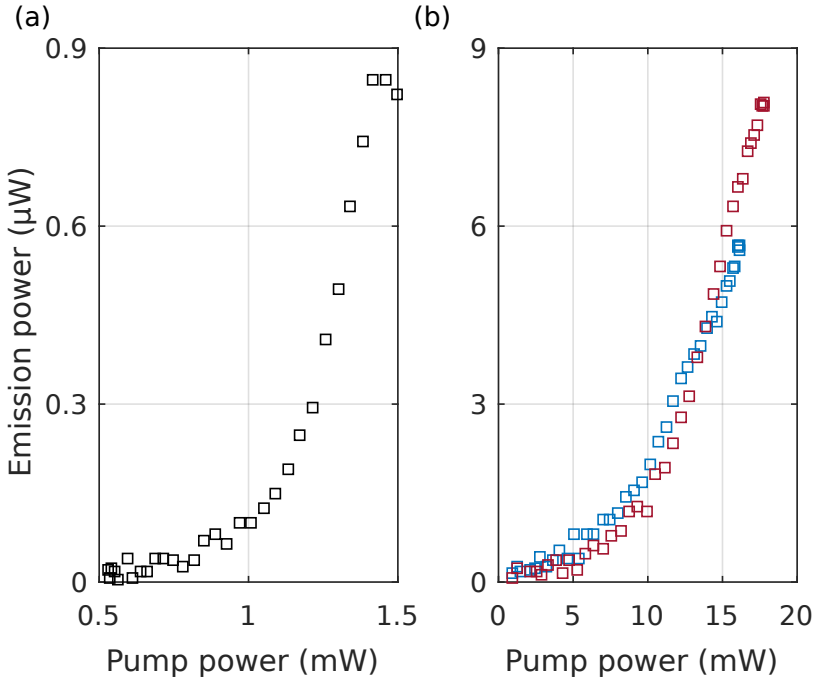

Figure 4.25: Efficiency characterization of the simultaneous optical pumping of nine QDMLs. The global emission power of all micropillar pumped is plotted according to the pump power. Panel (a), reference for a single pumped QDML. Panel (b), the pump laser is diffracted. The blue curve corresponds to a spatial period of the phase mask of 295 pixels, the red to a period divided by seven, hence 42 pixels.

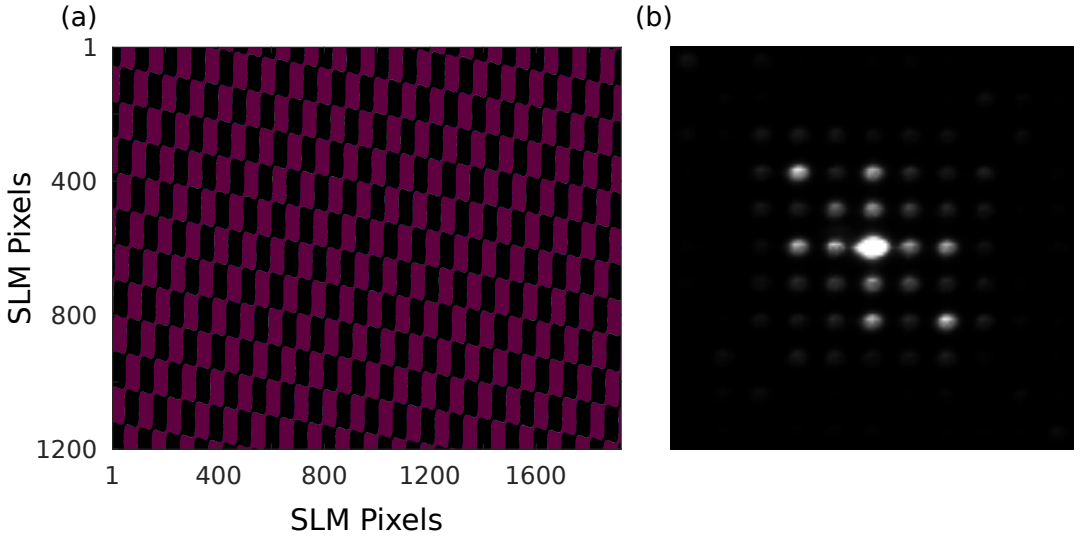

Figure 4.26: (a) Diffraction grating based on the phase profile shown in Fig. 4.21. The added spatial frequencies create the 5x5 diffraction pattern shown in panel (b). We can then observe the inhomogeneity between the diffraction orders produced by this method.

4.5/ SUMMARY

In this chapter we have discussed the basic building blocks that will be used to build a spatial optical recurrent neural network. The objective is to design a demonstrator that can be generalized for a large-scale neural network. To achieve this we use the neural network architecture described in chapter 2 by replacing the electro-optic nodes with QDMLs. The diffractive nodes coupling method, the readout layer and the learning will be similar, hence we will be able to profit from the experience obtained with these methods based on the other setups.

Nevertheless, the realization of an all-optical network remains a significant engineering challenge. In this chapter we have studied several technological issues, the current limiting factors, and described the different engineering strategies deployed to face the difficulties linked to the experimental reality. First of all, we use a QDMLA with a very high quality factor, and the wavelength homogeneity required for optical injection locking results in a strong technological challenge during fabrication. Additionally, we were confronted with unexpected mechanical instabilities when cryogenically cooling the sample. The solution put in place required engineering work, from a mechanical, cryogenic and vacuum point of view. The result obtained certainly still provides significant amount for improvement, as, the sample is stabilized but its temperature is 130 K instead of 77 K. A change of cryostat will ultimately be desired to gain ease of use and in particular to potentially reduce the lasing threshold of QDMLs by 38% when additionally lowering its temperature. Nevertheless, this compromise allowed us to continue our initial investigations and to develop the procedures for preparing the sample, handling the now complex and very confined vacuum chamber, cooling and aligning the QDMLA.

We were able to perform a detailed pumping characterization of eight QDMLs from an array with homogenized emission wavelengths. We have studied pumping either WL or the bulk GaAs. The DBR effects and the gain medium absorption were taken into account and careful calibration enabled us to characterize internal transfer mechanisms of the QDMLs. This allowed us to compare pumping mechanisms over several hundred nanometers in wavelength and to extract a quantitative description. Considering the free parameters one can determine using a simple rate-equation description of QDMLs, we measured the QDML's spontaneous emission confinement factor β, the exciton conversion efficiency δ and the number of excitons at transparency n_0. We lend significant robustness to our results by fitting a large number of QDMLs at a variety of pumping conditions. The number of carriers at transparency n_0 was individually optimized for each data or laser, while β and δ were kept uniform for a particular laser or pumping condition. For the correct global description of optical pumping, n_0 showed no dependency of the particular pumping wavelength, and we achieved this for describing GaAs and WL pumping alike. Using the bulk GaAs pumping, the global efficiency is currently high enough to optically pump simultaneously approximately one hundred QDMLs. Our characterization allowed us to understand two limiting factors and we proposed two improvements which would increase the pump absorption efficiency κ and the exciton conversion efficiency δ. Under these conditions, the global pumping efficiency would be increased by one or two orders of magnitude.

Simultaneous optical pumping up to 9 QDMLs was performed. The strategy adopted was to use a diffraction pattern already known from the electro-optical experiment. A first characterization of a 3×3 pattern has been carried out, which constitutes a first part of the

demonstrator. By adding spatial frequencies to the diffraction pattern, we can therefore increase the number of diffraction orders. However, this method turned out not to be applicable for a large number of diffractive orders due to insufficient power homogeneity. This work could be continued using a phase retrieval algorithm that will seek the optimal phase mask to generate the desired diffraction pattern [185–187].

CONCLUSION AND PERSPECTIVES

CONCLUSION

We saw in Chapter 1 the important role of neural networks in performing tasks that are simple for a human but extremely challenging to translate into traditional computer language. Computers are extremely efficient at executing a pre-established sequence of commands, *i.e.* an algorithm, which must be entirely and without ambiguity designed by the programmer. For example, pattern recognition was not long ago still extremely difficult for a computer due to the need of interpreting its general characteristics, a task difficult to formalize. On the contrary, the learning of neural networks is generally inductive. They are composed of a large number of neurons which are highly interconnected. By adjusting the connections during their optimization referred to as learning, the neural network is able to generalize characteristics from examples. The computational power of a neural network comes partly from the number of neurons but also from the number of interconnections. The human brain, capable of generalizing an unknown rule only on the basis of few examples, has approximately 80 billion neurons [66] which each have between five and sixty thousand connections [67]. One of the fundamental issues in the design of artificial neural networks is therefore their number of nonlinear nodes and their massive and parallel connections.

Artificial neural networks have recently emerged as a viable approach to tackle major challenges [77, 188] based on an increase in the computational power of computers. However von Neumann architecture, which is the concept on which computing is still largely based, is very different of the characteristics observable in biological neural networks, most importantly their massive parallel connectivity. This contradiction could potentially be ignored if the computing power continued to increase strongly. However, due to the current sizes of the components of state of the art integrated circuits the margin for further reduction is limited [11]. The heat dissipation further constraints an excessive density of integrated circuits and limits the clock frequencies of CPUs [82]. In addition, an increase in the components number by increasing the surface area of the processors is a difficult challenge. Indeed, the connectivity is constrained both by the intrinsic characteristics of electronics and its manufacturing processes. Electronic architectures still remain mostly planar and electrical connections cannot cross each other. In addition, the information flow of an electrical connection depends on the ratio between its diameter and its length [81]. In view of these constraints, the design of massive and parallel connectivity between artificial electronic neurons is an important challenge. Because of these difficulties, computing may not continue to accelerate sufficiently to meet the needs of neural networks [84]. Training the best performing systems requires a large amount of energy and already necessitates data centers[80].

A paradigm shift currently emerges to address these challenges by implementing neural networks in optics [85, 92, 107]. Light is well known for its high capacity for information transport [14], more specifically because optical communication channels can cross or

even overlap in space, thus allowing strong parallelization. Light is therefore suitable for creating the high connectivity [113] which is essential for a neural network's functioning. In addition, to allow scalability of systems, a 3D architecture where optical nonlinear nodes are arranged on a surface and their interconnections are realized by optical diffraction along the light propagation is a promising integration strategy [189].

We studied in chapter 2 an electro-optical neural network [16] based on the reservoir computing concept [15]. Each part of the system has been studied in detail, in particular the realization of artificial neurons by the combination of the SLM pixels and a polarizing filter, and the network's recurrent internal connections made by diffraction. Such optical coupling between the nodes of the spatially distributed network gives high scalability to the system since this technique could physically achieve a million connections [17]. The readout layer is implemented using a DMD, allowing programming of fully parallelized Boolean readout weights. The neural network being electro-optical, therefore analogue, the learning algorithm is designed to operate without knowledge of all node states in order to avoid an auxiliary information processing system. The algorithm is therefore based on evolutionary learning [16], comparing the current error with the previous one, which makes it simple and realistically implementable on a low complexity support hardware. Importantly, the idea was to create an autonomous demonstrator, learning is performed therefore in real time.

Despite the drawbacks of an electro-optical implementation, whether the necessity to use a computer to control the electro-optical nodes, or the low computational speed of the system, this approach gives a scientific advantage. Indeed, the parameters are easily adjustable and reservoir states measurable, which turns this experience into a white box. Using this access we studied in chapter 3 the different noise sources and in particular the impact of additive or multiplicative noise on the SNR [136]. The system is resilient to noise and has high consistency. A long study of learning under noise has also been performed [135]. Although the error landscape has multiple local minima, we find that these have a similar error. Learning never got stuck in a local minima with low performance. We have shown that the configuration of the readout weights to solve a task is the result of an optimization in the context of noise. The training result is therefore specific to each situation. As the algorithm operates in real time, it is possible to continue optimization in order to strengthen the neural network's resilience to noise. In addition to the parallelization of the couplings and the readout layer, we have shown that the learning time of our algorithm scales linear with the number of network nodes.

While keeping the same architecture and therefore its strengths, in Chapter 4 we laid the foundations to make an all-optical neural network. Nonlinear nodes are now implemented by a QDMLA [18]. This node substrate change has multiple consequences, both on its future performance and on its experimental implementation. The bandwidth of the system will be of the GHz order, with limitations either due to the optical cavity roundtrip time, or the QDMLs's characteristic time scale. In addition, optical pumping of the QDMLs array, although still under study, allows scalability by avoiding the connection problems that electric pumping faces. Homogenization of the QDMLA's emission wavelength was carried out, maximizing the potential coupling strength of the lasers [18]. We have performed a quantitative study of optical pumping by exploring pumping either the samples WL or the bulk GaAs of the laser cavity. Despite the absorption of the top DBR, bulk GaAs pumping is more efficient than WL pumping. The quantitative results have helped to understand the internal mechanisms of QDMLs and we propose two structural improvements to increase the pumping efficiency by one to two orders of magnitude. Counting in the positive

effect of further decreasing the sample's temperature, the QDML threshold could be lowered to between 10 and 100 µW. This low pump power is essential for the design of a large optical neural network.

PERSPECTIVES

Regarding the electro-optical neural network, this version is only at its first stage, new modifications are already ready to be implemented. The Gaussian illumination of the SLM will be modified to give way to flat illumination thanks to a Köhler integrator implementation. This new illumination will notably allow to increase the SNR by best exploiting the dynamic range of the camera. The implementation of the input layer is also planned, the architecture and its realization have already been considered. A laser will be injected into a multimode fiber whose output facet will be imaged on the screen of the camera. Some commercial fibers offer rectangular cores, the superposition of its image with the image of the SLM on the camera will allow to all-optical injection information into the system. The random but constant injection matrix will be directly produced by the coupling between the modes of the optical fiber [118, 190]. Making the system more compact, in particular the distances between the microscopes objectives, would significantly reduce the problem of beam vignetting and therefore strongly increase the number of nodes in the network. Abandoning the use of commercial mounts and manufacturing a tailor-made global mount, would both optimize the network geometry and would add additional optical stability and precision.

To take advantage of the noise source investigation that discovered the presence of correlated noise, it would be interesting to work on noise reduction. Indeed, unlike uncorrelated noises which have a zero overall average, the correlation leads to a non-zero mean. One approach could be to subtract the mean of the noise that is continuously measured at an additional neural network output.

A new PhD student is starting to work on modification of system use in order to obtain spike neural network. This work will bring the system even closer to biological networks, and will open new doors for research.

From a machine learning point of view, there is still a need to diversify the learning tasks. With regard to pattern recognition, a strategy recently implemented is the creation of spatial clusters within the phase offset matrix. This break in spatial symmetry of the state of the reservoir should increase performance.

This last point was not a priority, in particular due to the low speed of the system, but now a change of SLM is underway. The new fast SLM allows acceleration of the system by a factor 100. This new paradigm will revolutionize the study of this neural network. A current learning curve takes a full day, with the new system few minutes will only be required. This may seem trivial but for the study of the system, the exploration of learning, the understanding of all the factors involved, such as the role of noise, the study fundamentally requires robust statistics. While current exploration require weeks or months of measurements, a few hours or days will then be necessary. This improvement will even allow to explore certain characteristics of the system which until now were not possible. This simple characteristic time change can therefore result in a leap in understanding neural networks.

About the exploration of the error landscape, a further study of the topology would be relevant. Thanks to the frequency increase of the system it would now be possible to map the error landscape according to 18 dimensions. Even if it may seem small mapping in comparison to this gigantic dimension space, this will allow a better understanding of the topology and thereby refine the learning algorithm when the system error is close to local minimums.

Regarding to the all-optical neural network, the first need is to continue implementing a diffraction grating on SLM to obtain the desired diffraction patterns. The implementation of an optical feedback will allow to study QDML dynamics, then the addition of a DOE will allow coupling between the lasers. The information input will be carried out by injection locking via a tunable laser and the homogenizing work of the QDMLs frequency will simplify this step. The readout layer and the learning algorithm can be directly imported from the electro-optical experiment.

I would particularly like to conclude by bouncing back on the vision of Herbert Jaeger [191] who emphasizes that wanting to calculate like a brain means not to simulate a neural network but, on the contrary, to physically embody "computing" inspired by the brain. If we therefore want to take neuromorphic hardware seriously, this means going beyond the comparison with traditional computing and taking into account a certain number of real phenomena. The first is that the calculation then becomes physical. The system is therefore analogue and subject to noise, to stochastic phenomena, then the question of consistency arises. Drifts or system aging must also be taken into account. It would even be interesting to study the system adaptation in response to the failure of a component. The neural network performance is intrinsically linked in its connections number, it is therefore important to be concerned about their parallelization in order to obtain a scalable system. A large analogue neural network has as a corollary a certain unobservability. To maximize the efficiency of the system, all states are no longer measurable. Besides, probing a low energy state could disturb it. More importantly, unlike electrical computing where the clock requires high power consumption [83], the dynamics within biological neural networks are not clocked globally or locally. To further mimic biology and save energy, systems should therefore be based on continuous time. Even if these perspectives are the fruit of an in-depth vision of hardware neural networks, the research presented in this manuscript falls, less ambitiously, in this line and already take up several challenges.

PAPERS AND CONFERENCES RELATED TO THIS PHD THESIS

ARTICLES IN INTERNATIONAL REVIEWED PUBLICATION

Andreoli, L., Porte, X., Heuser, T., Große , J., Moeglen-Paget, B., Furfaro, L., Reitzenstein, S., and Brunner, D., "Optical pumping of quantum dot micropillar lasers", *Optics Express* **29**, 6 (2021), 9084–9097.

Andreoli, L., Porte, X., Chrétien, S., Jacquot, M., Larger, L., and Brunner, D., "Boolean learning under noise-perturbations in hardware neural networks", *Nanophotonics* **9**, 13 (2020), 4139–4147.

Semenova, N., Porte, X., Andreoli, L., Jacquot, M., Larger, L., and Brunner, D., "Fundamental aspects of noise in analog-hardware neural networks", *Chaos* **29**, 10 (2019), 103128.

Maktoobi, S., Froehly, L., Andreoli, L., Porte, X., Jacquot, M., Larger, L., and Brunner, D., "Diffractive Coupling For Photonic Networks: How Big Can We Go?", *IEEE Journal of Selected Topics in Quantum Electronics* **26**, 1 (2019), 1–8.

CONFERENCE PROCEEDING

Porte, X., Andreoli, L., Jacquot, M., Larger, L., and Brunner, D., "Reservoir-size dependent learning in analogue neural networks", *In Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions* (2019), pp. 184–192.

INVITED INTERNATIONAL CONFERENCE

Andreoli, L., Porte, X., Heuser, T., Große, J., Jacquot, M., Chrétien, S., Reitzenstein, S., Larger, L., and Brunner, D., "Greedy Boolean Learning in Photonic Recurrent Neural Networks", *European Materials Research Society*, Warsaw - Poland, 2019.

INTERNATIONAL CONFERENCES

Andreoli, L., Porte, X., Jacquot, M., Larger, L., and Brunner, D., "Scaling laws and topology-properties of Boolean reinforcement learning in photonic neural networks", *European Quantum Electronics Conference (EQEC). Optical Society of America (CLEO)*, Munich - Germany 2019. p. jsi_2_3.

Andreoli, L., Semenova, N., Semenov, V., Maktoobi, S., Jacquot, M., Fischer, I., Larger, L., and Brunner, D., "Impact and mitigation of noise in analogue spatio-temporal neural network", *Cognitive Computing: Merging Concepts With Hardware*, Hannover - Germany, 2018.

Andreoli, L., Jacquot, M., Larger, L., and Brunner, D., "Photonics spatio-temporal networks of 2025 Ikeda Maps", *International Symposium on Nonlinear Theory and its Applications (NOLTA)*, Tarragona - Spain, 2018.

POSTER PRESENTATION

Porte, X., Andreoli, L., Semenova, N., Semenov, V., Jacquot, M., Larger, L., and Brunner, D., "Noise and consistency of analogue spatio-temporal photonic neural networks", *Workshop on Dynamical Systems and Brain-Inspired Information Processing*, Konstanz - Germany, 2019.

BOOK CHAPTER

Brunner, D., Soriano, M. C., and Van Der Sande, G. (Eds.), "Photonic Reservoir Computing: Optical Recurrent Neural Networks", Walter de Gruyter GmbH & Co KG, 2019.

ARTICLES IN PREPARATION

Chrétien, S., Andreoli, L., Porte, X., and Brunner, D., "Convergence and scaling of Boolean-weight optimization for hardware reservoirs", in preparation.

Andreoli, L., Porte, X., Chrétien, S., and Brunner, D., "Boolean error landscape topology around the minima in optical neural networks", in preparation.

Andreoli, L., Porte, X., and Brunner, D., "Consistency of analogue spatio-temporal photonic neural networks", in preparation.

A

CRYOSTAT VACUUM CHAMBER: COPPER FOILS

In order to achieve the thermal transfer between the cold finger of the cryostat and the sample, we have made two thermal bridges in serial using copper foils. Our thermal bridges are an assembly of about fifteen thin (0.02 mm) superimposed copper foils, combining mechanical flexibility with thermal conductivity. The first thermal bridge going from the cold finger above the translation micrometer stage is fixed. The second connects the underside of this part of the sample, therefore the thermal bridge must have enough degrees of freedom for the sample to be mobile, in particular longitudinally with a maximum displacement $\Delta = 5$ mm.

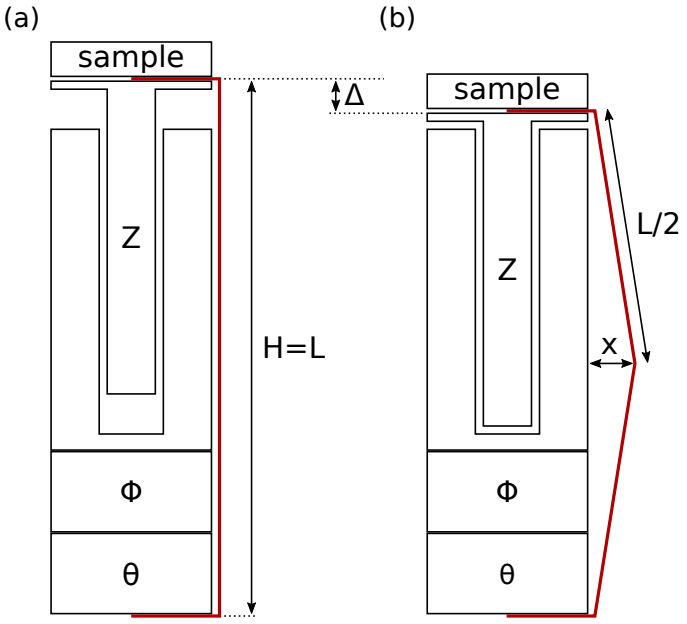

Figure A.1: Schematic illustration of the piezomotors and copper foils (red line) providing a thermal bridge ensuring the heat transfer between the sample and the cold source. Panel (a), piezomotor in unfolded position, the copper foil is then stretched to the maximum. Panel (b), piezomotor in folded position, the copper foils are then deflected laterally.

As illustrated in Fig. A.1(a), when the piezomotor is in the unfolded position, the stacking

of motors has a height $H = 55$ mm. The copper foil must then measure at least the length $L = H$. When the piezomotor is folded up, the stacking of motors has a height $H - Delta = 50$ mm. The copper foil always measures the same length and must therefore bend laterally. The constraint is to obtain a lateral displacement x less than 3 mm in order that the copper foil does not touch the plastic tube which surrounds it and create an unwanted thermal bridge.

Let us calculate the maximum deflection which takes place for the situation presented in Fig. A.1(b).

$$\begin{aligned} \mathrm{x}^2 &= \left(\frac{L}{2}\right)^2 - \left(\frac{L-\Delta}{2}\right)^2 \\ \mathrm{x} &= \sqrt{\left|\left(\frac{L}{2}\right)^2 - \left(\frac{L-\Delta}{2}\right)^2\right|} \\ \mathrm{x} &= \frac{1}{2}\sqrt{\left|\Delta^2 - 2\Delta L\right|} \\ &. \end{aligned} \tag{A.1}$$

We note that the deflection x increases with the foil length L. Considering our situation, we get $\mathrm{x} = 11.5$ mm, which is far too large.

An obvious solution to the problem is then to fold the copper foil into an accordion, but due to the small size the realization is difficult and folding a thin foil risks causing mechanical stress points which are conducive to damage. As illustrated in Fig. A.2(a), we then realized an alternative which consists in winding the foil around the piezomotors in the unfolded position. The propeller has a diameter $d_0 = 28$ mm. When the piezomotor contracts, shown in Fig. A.2(b), the diameter of the propeller becomes $d_1 > d_0$ but the propeller also compresses on itself. Figure A.2(c) is a diagram of the two situations when the propeller is unfolded, N is the number of turns of the propeller around the piezomotors. The length of the foil L remains the same in both configurations, we then obtain

$$\begin{cases} L^2 &= H^2 + (N\pi d_0)^2 \\ L^2 &= (H-\Delta)^2 + (N\pi d_1)^2 \end{cases} \tag{A.2}$$

By identifying the two equations of the Eq. A.2, we can derive the number of turns N according to the other parameters

$$\begin{aligned} (N\pi d_1)^2 - (N\pi d_0)^2 &= H^2 - (H-\Delta)^2 \\ N^2 &= \frac{H^2 - (H-\Delta)^2}{\pi^2(d_1^2 - d_0^2)} \\ N &= \sqrt{\frac{H^2 - (H-\Delta)^2}{\pi^2(d_1^2 - d_0^2)}}\,. \end{aligned} \tag{A.3}$$

By setting $d_1 = 31$ mm the maximum diameter, it is then necessary to make at least 0.55 turns. This solution is then realistic. The diameter d_1 according to the other parameters is

$$d_1 = \sqrt{\frac{H^2 - (H-\Delta)^2}{N^2\pi^2} + d_0^2}\,. \tag{A.4}$$

To ensure a safety margin, we then choose to perform an entire turn, *i.e.* $N = 1$. In the configuration of the folded piezo motor, the diameter of the propeller is then $d_1 = 29$ mm.

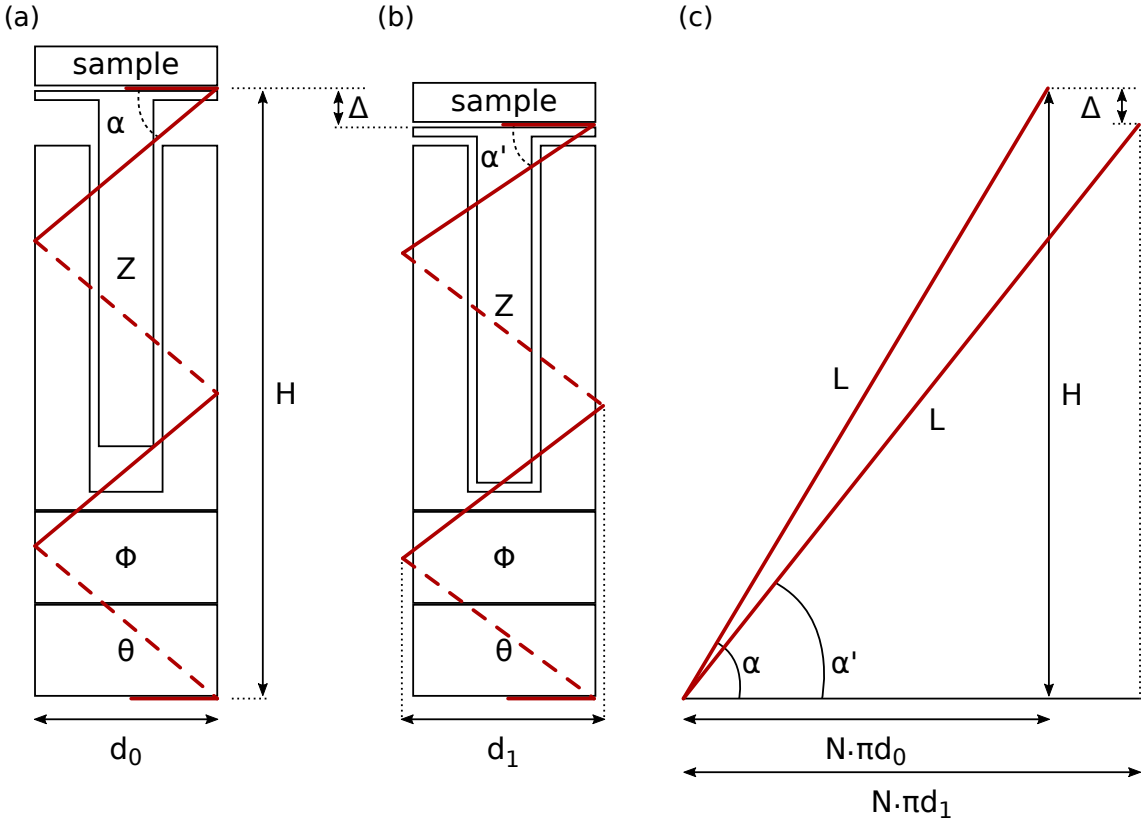

Figure A.2: Schematic illustration of the copper foil (red line) wound around the piezomotors. In unfolded position, panel (a), the diameter of the propeller is d_0. In the folded position, panel (b), the helix compresses and its diameter d_1 widens, *i.e.* $d_1 < d_0$. When the propeller formed by the copper foil is unfolded in panel (c), the diameter d_1 then clearly depends on the number of turns N of the propeller.

Figure A.3: Sample ready to be installed in the vacuum chamber of the cryostat. The piezomotors and the thermistor are electrically connected, a double propeller of copper foils is realized in order to establish a thermal transfer to the sample. The piezomotors are mounted on a translation micrometer stage which is itself fixed to an insulating plastic tube providing support to the assembly. To achieve good thermal transfer of the sample, the copper foils must under no circumstances be in contact with the plastic tube.

B

TUNABLE PUMP LASER

The total design of the device, including the external laser cavity, its amplification and its injection into a single-mode optical fiber is shown in Fig. B.9.

B.1/ EXTERNAL CAVITY LASER

The tunable pump laser was produced from a CW semiconductor laser at 915 nm and nominal power 300 mW (Thorlabs M9-915-0300). The objective is to create an external laser cavity in order to obtain a single-mode laser with a fine spectral width, stabilized and tunable in wavelength over several nanometers.

A laser is a light oscillator amplified by a gain medium. The wavelengths are defined by the length of the laser cavity where several longitudinal modes meet the criteria for resonance. An infinity of modes exists but only a part is amplified by the gain medium which depends on the wavelength. Adding an external cavity laser makes it possible to add an additional resonance criterion, the longitudinal modes must satisfy both the internal and external laser cavity. This results in spectral thinning, by finely adjusting the dimension of the external cavity, it is then possible to obtain a single-mode laser longitudinally. Such feedback within a laser can cause significant instabilities [192] but on the contrary if the feedback force is high enough (greater than 10 %) the feedback stabilizes the optical mode [193, 194].

To successfully tune the laser wavelength, many techniques exist [195], in our case we use a diffraction grating to create the external laser cavity [196]. More precisely, the grating used (Thorlabs GR13-1210) is a blaze grating in Littrow configuration, the angle of the diffraction order is confused with the incidence angle [197]. The laser is collimated with an aspherical lens and around 90% of the power is reinjected inside the laser via the first order of the diffraction grating. The diffraction angle being related to the wavelength by the grating formula (see Eq. (2.3), a variation of the grating angle amounts to changing the angle of incidence, therefore the wavelength. Thus, varying the angle of the grating is one way to tune the wavelength of the laser [198]. However, in this configuration the angle of the non-diffracted order also varies with the orientation of the grating [199], which makes it difficult to use the laser output. We have therefore installed a mechanism which modifies the orientation of the mirror to return simultaneously to the grating [200] in a such way that the angle variation turns into a very slight translation of the output beam. This mechanism is visible in Fig. B.1 where the optical network and the deflection mirror are fixed on the same support whose angle can be adjusted.

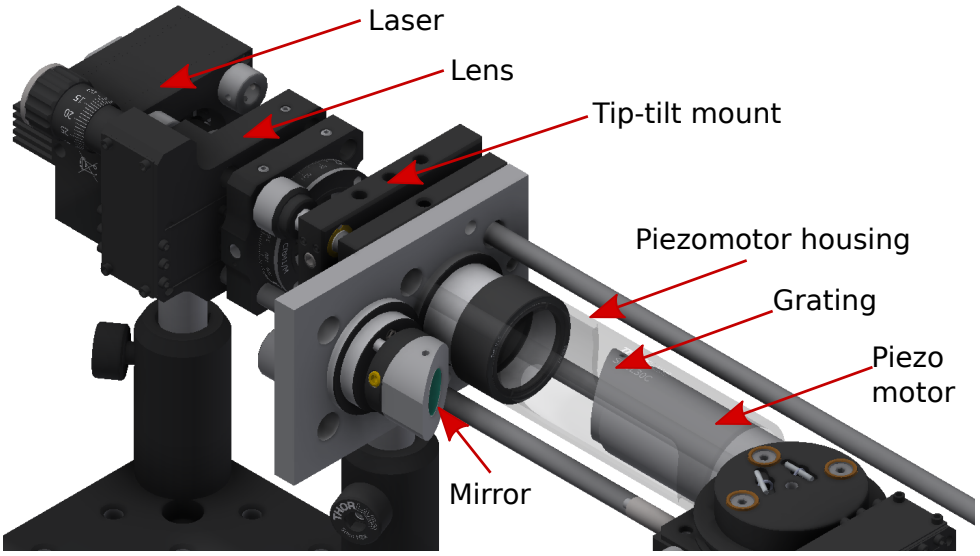

Figure B.1: External cavity laser. The grating in Littrow configuration is attached to a tip/tilt/Z piezomotor mounted inside a tube fixed to a tip-tilt mount. The output mirror reflecting the output of the external cavity is attached to the same mount too. Changing the orientation of the grating results in a slight translation of the beam reflected by the mirror.

One method to align the optical feedback inside the laser is calling laser threshold reduction. To achieve this, the intensity of the pump current is increased very gradually until the gain of the laser amplifies the main mode. The principle is to maximize the output of the laser for a fixed current by modifying the orientation of the grating and the position of the lens. Care must be taken to ensure that the laser always remains collimated. In order to obtain the highest sensitivity during this adjustment, the current must be reduced more and more in order that the laser always remains close to the stimulated emission threshold. This method was performed and as indicated by the difference between the L-I curves with (red squares) and without (blue circles) feedback in Fig. B.2, almost 90% of the laser power is reinjected. The inset is a zoom near the laser threshold which shows its reduction of approximately 10% in the feedback case.

The laser spectrum is shown in Fig. B.3. The resolution of the spectrum analyzer (Yokogawa AQ6370D) is 20 pm. The side mode suppression ratio is 50 dB and the FWHM, limited by the spectrometer resolution, is 37 pm. The central wavelength can be varied from 905 nm to 923 nm, however the gain being centered at 915 nm, it is in this region that the highest optical power is obtained.

B.2/ AMPLIFICATION

The pump laser is then amplified through a semiconductor optical amplifier (SOA). The device that we use (Dilas, TA-090-2000) can amplify a laser seed between 10 and 30 mW around 915 nm to 2000 mW. The SOA input is a ridge of 4 µm wide and 1.4 µm thick mounted on a $7 \times 6.4 \times 4.3$ mm frame drilled with a 2.2 mm diameter hole to hold it in position. The aspect ratio of the input ridge is very close to that of the semiconductor

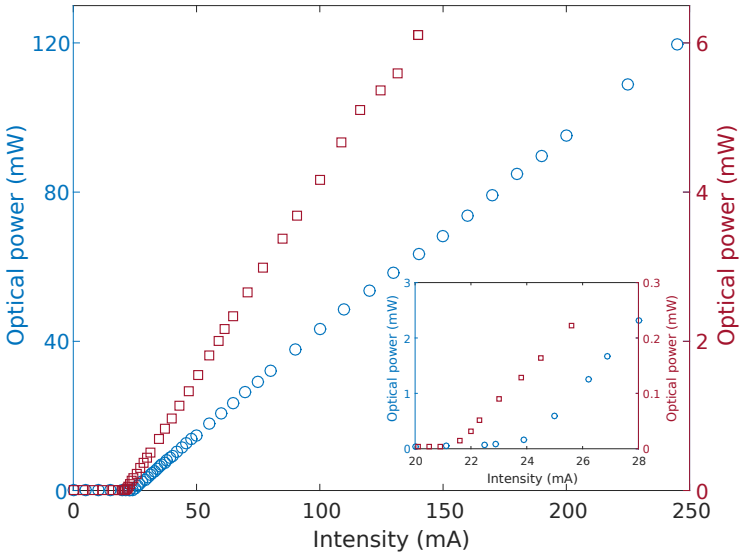

Figure B.2: L-I curves of the laser with (blue circles) and without (red squares) feedback. The inset is a zoom where we can observe the laser threshold reduction when the feedback is optimized.

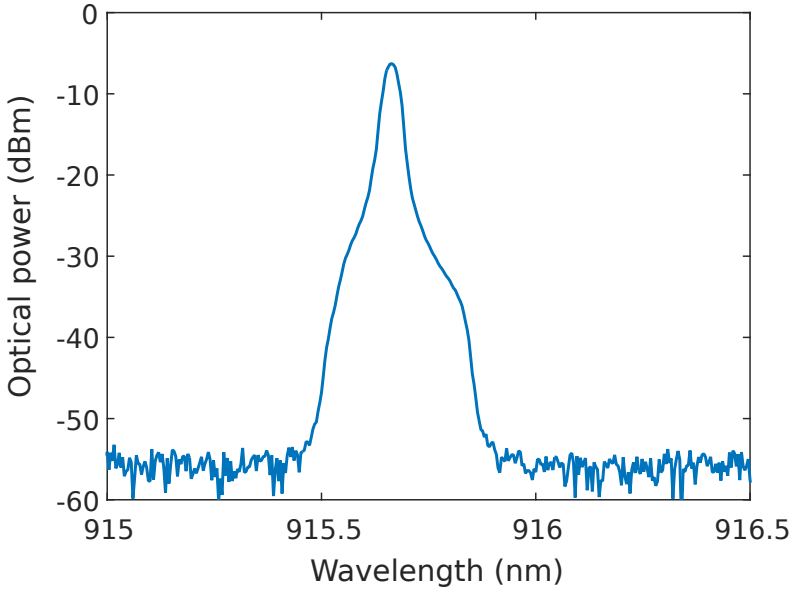

Figure B.3: Spectrum of the external cavity laser output. We can observe a side mode suppression ration of 50 dB and the full width at half maximum is 37 pm.

junction of the laser, their spatial position are therefore adjusted to optimize the injection of the laser into the amplifier.
As shown in Fig. B.4, a completely homemade mount was built in order that the SOA can be integrated into the optical cage system. Several characteristics are required for the mount. It must be thin enough to be able to get close the collimating lenses on either side of the device. The amplifier must be controlled in current but also in temperature. A thermistor is added in a copper block which ensures the heat transfer of the SOA.

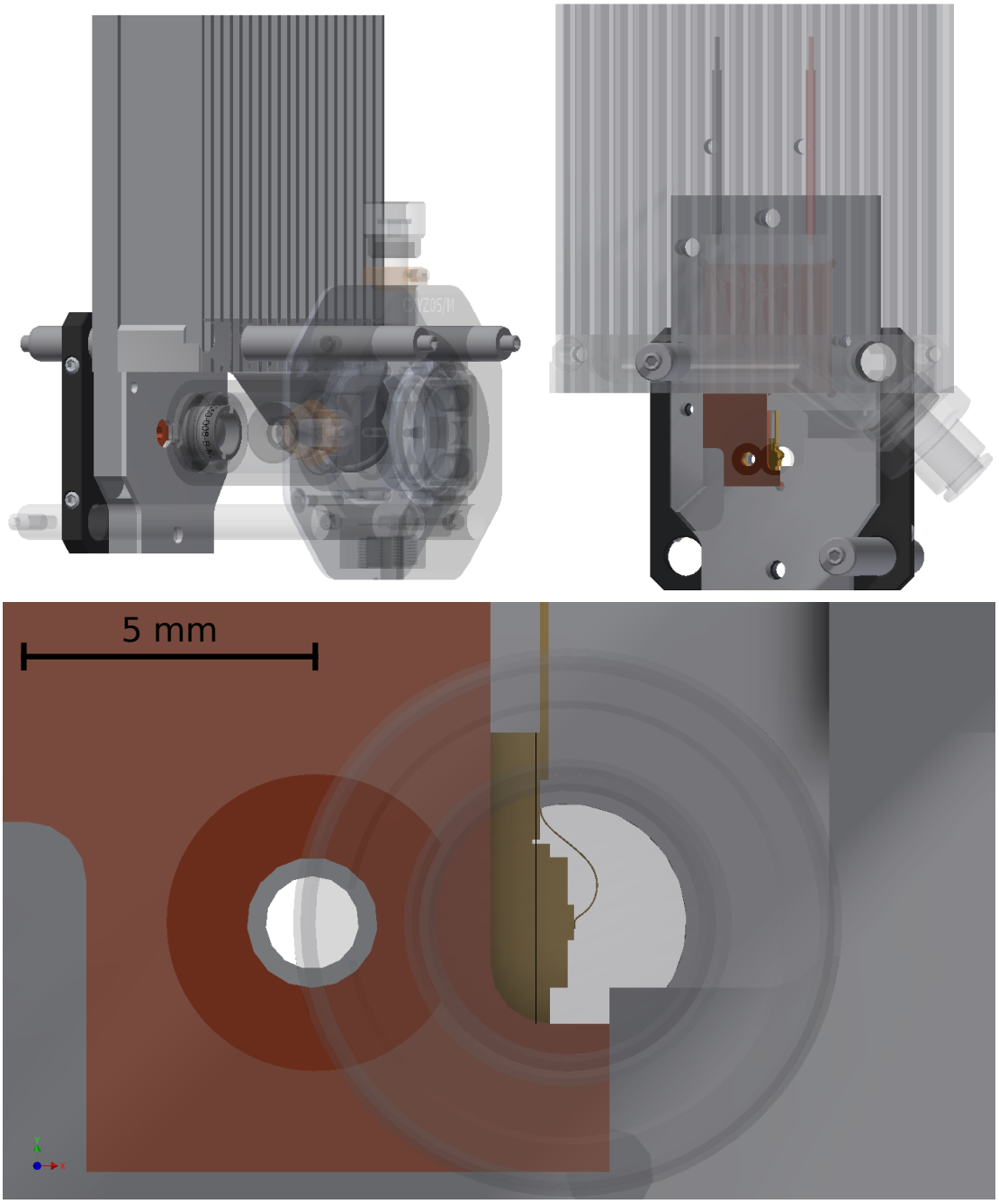

Figure B.4: Conception of the amplifier mount compatible with the optical cage system. The upper left panel shows the global view where the whole device is packed in a teflon box. The lens must be close enough and be movable to inject light inside the amplifier. In upper right panel, we can distinguish the amplifier in yellow. Through the radiator in transparency, we can see the Peltier module which performs the thermal exchange. To improve the efficiency and reach high current control (4 A), the radiator is cool down thank to air flow which comes to the pneumatic connection visible to the right. In the bottom panel, the view au the amplifier mount through the lens. The amplifier input with a width and thickness of 4 and 1.4 µm respectively, is not visible.

The copper part is connected to a Peltier module (Thorlabs TEC3-6) connected to a radiator. However, the heat dissipation is not sufficient, an airflow is then injected inside the radiator. In addition, the SOA must be electrically isolated from the cage system to avoid a short circuit, protect the semiconductor from an electrical contact, and avoid a loss of current (4 A). The SOA and its copper heat exchanger are therefore mounted inside a teflon box. The current and the temperature are regulated by PID controller.

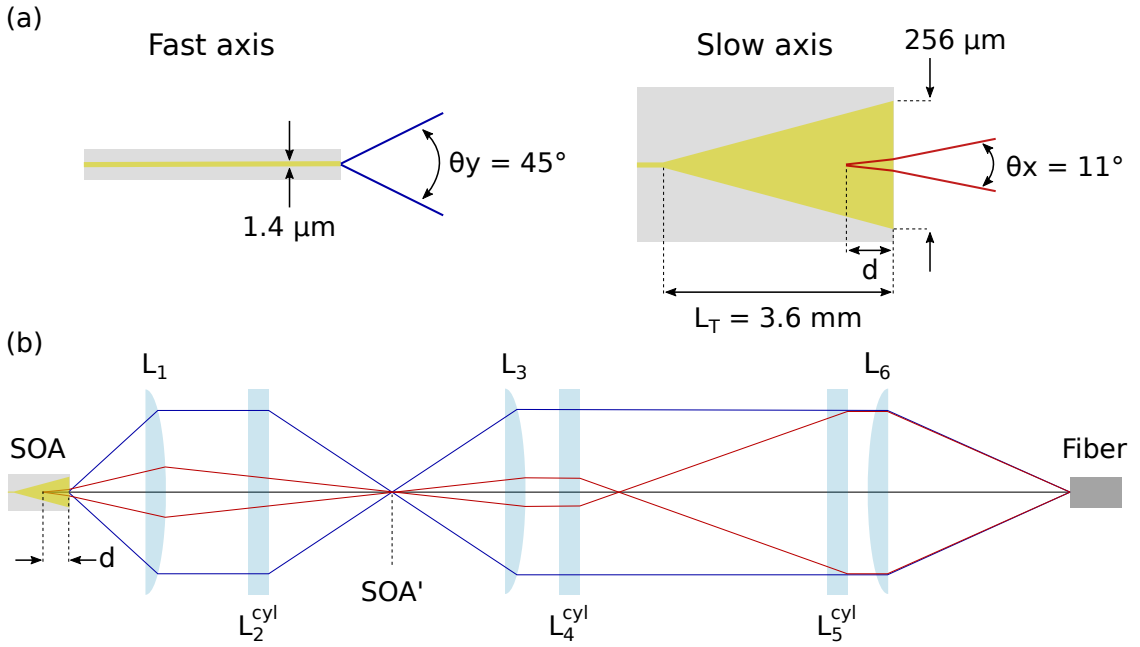

Figure B.5: Panel (a) is a schematic representation of the SOA. While on the fast axis, the divergence of the output is 45 degrees, this one of the fast axis is only 11 degrees. In addition, the output is very astigmatic with a position deviation of more than a millimeter between the source points along the two axes. Panel (b) is the optical scheme of the output collimation. The principle is to use one lens and one cylindrical lens to image the SOA output in the same plane SOA' which can be collimated for both axis by the lens L_3. The slow axis beam can be extended by a cylindrical telescope and finally the beam can be injected into the optical fiber.

Injecting the laser into the amplifier is difficult because there is no intermediate stage. The protocol followed consists of a rigorous scanning of the surface after focusing the laser on it. When the injection is performed, the device is very robust to potential mechanical perturbations. The collimation of the SOA output is delicate too. The amplifier is a 1.4 µm thick semiconductor junction. The operating principle is identical to that of a semiconductor laser but without a resonant cavity. To increase the relaxation effect of the carriers by stimulated emission, the interaction length is 3.6 mm. But to avoid decreasing the power density while keeping an adiabatic transition to maintain a singlemode profile, the guide gradually widens in the form of tap, going from 4 to 256 µm wide. In order to avoid an excessively high power density, the guide gradually widens in the shape of a cone while maintaining the single-mode profile thanks to an adiabatic transition, ranging from 4 to 256 µm in width. The taper shape eliminates the phenomenon of filament propagation, beam quality degradation, in particular due to the phenomenon of spectral or spatial hole burning. The output of the amplifier shown schematically in Fig. B.5(a) is a ridge 1.4 µm thick and 256 µm wide. The divergence of the output beam is then very different along the

two axes, called the fast axis and the slow axis, respectively 45 and 11 degrees. In addition, the device is highly astigmatic. The source point for the slow axis is inside the taper at a distance d from its output. This astigmatism distance d can be calculated according to

$$d \approx \frac{L_T}{n_T} \approx 1.2 \text{ mm} , \tag{B.1}$$

where L_T is the taper length and n_T its effective refractive index.

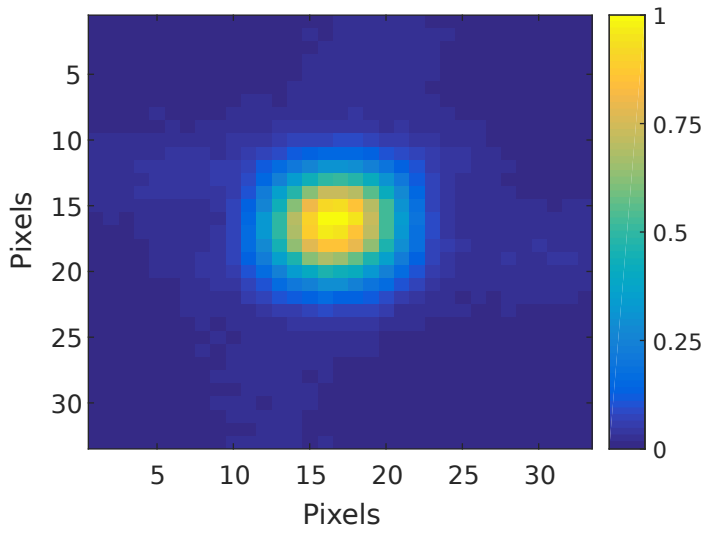

Figure B.6: Spatial mode profile of the pump laser beam after the SOA collimation.

The collimation of the amplifier output and its injection into an optical fiber uses six lenses [201]. Figure B.5(b) schematically illustrates the method used, where the blue and red rays respectively illustrate the fast and slow axis. The principle is to image the amplifier output in the SOA' plane in order to cancel the astigmatism effect. To achieve this, a first lens L_1 is adjusted to collimate the fast axis and focus the slow axis in the SOA' plane. An L_2^{cyl} cylindrical lens is used to focus only the fast axis in the SOA' plane. The SOA' plane is then the amplifier output image and can be collimated by the lens L_3. The difference in spatial width along the axes can be compensated by a telescope formed from L_4^{cyl} and L_5^{cyl} to widen only the slow axis. A sixth lens L_6 makes it possible to image the SOA' plane and therefore the amplifier output in the optical fiber. Figure B.6 is an image of the spatial profile of the amplified and collimated laser beam.

B.3/ SPATIAL FILTERING

The pump laser must be spatially filtered in order to be singlemode and thus increase the optical pumping efficiency of QDMLs. To achieve this, the pump laser is injected into a singlemode fiber. However the power density is too high for a standard fiber, therfore we are using a large mode area fiber (Thorlabs LMA-25). This fiber is a hexagonal photonic crystal which creates an effective core having a diameter at $1/e^2$ of 20.7 ± 2 µm. The fiber is cleaved with a ruby blade and polished at an 8 degree angle to avoid reflections. We use quick connectors (Thorlabs BFT1 and B30270C) to maintain the fiber because the glue used in standard fiber connections does not support the power density.

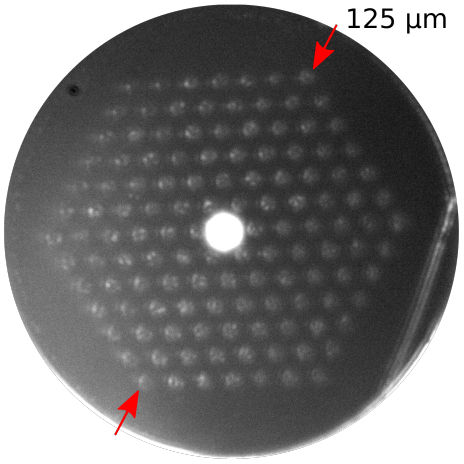

Figure B.7: Image of the large mode area fiber. We can see the periodic structure creating a photonic crystal which confines the mode in the center.

The great width of the mode confined in the fiber goes hand in hand with a small numerical aperture, NA(915 nm) = 0.03, *i.e.* a divergence of 3.4 degrees. The incidence angle of the pump laser beam is critical to maximize its injection. Thus, the fiber is fixed on a five axis mount, constituted by three translations and two rotations. Due to the low numerical aperture of the fiber, the difficulty in carefully collimating the amplified pump beam and the difference in spatial profile between the pump laser and the optical mode supported by the fiber, the efficiency of injection is only 20%.

Figure B.7 shows an image of the output of the fiber. One can observe the photonic crystal with a width of 125 µm and the confined mode at the center. The fiber mode is shown in Fig. B.8(a) where we can see its hexagonal profile. The full width at half maximum of the measured mode is 28.3 µm. Fig. B.8(b) shows the 2D-Gaussian fit of the mode fiber, the overlap integral between the two functions is 47%.

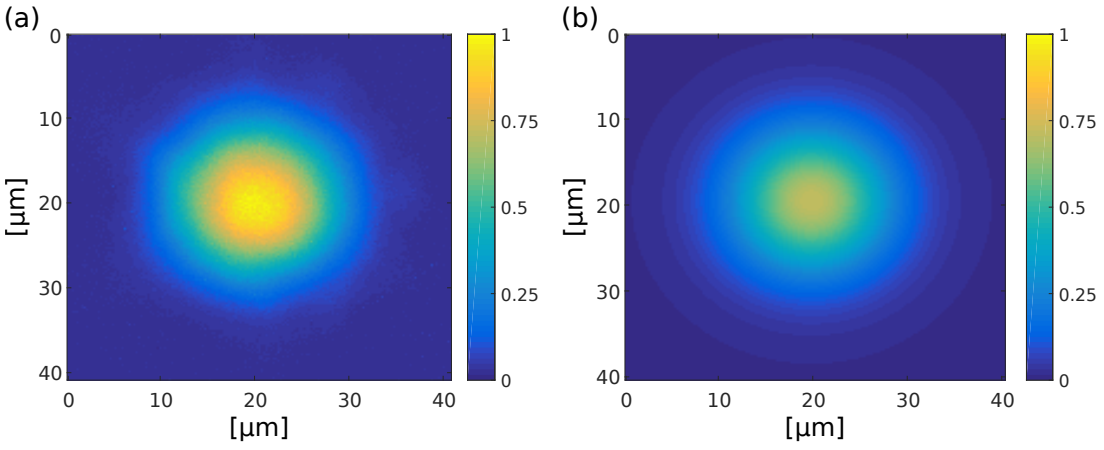

Figure B.8: Panel (a) is the spatial mode of the large mode area fiber. We can see its hexagonal shape, its full width at half maximum is 28.3 µm. Panel (b) is a 2D-Gaussian fit of the fiber mode, the overlap integral between the two functions is 47%.

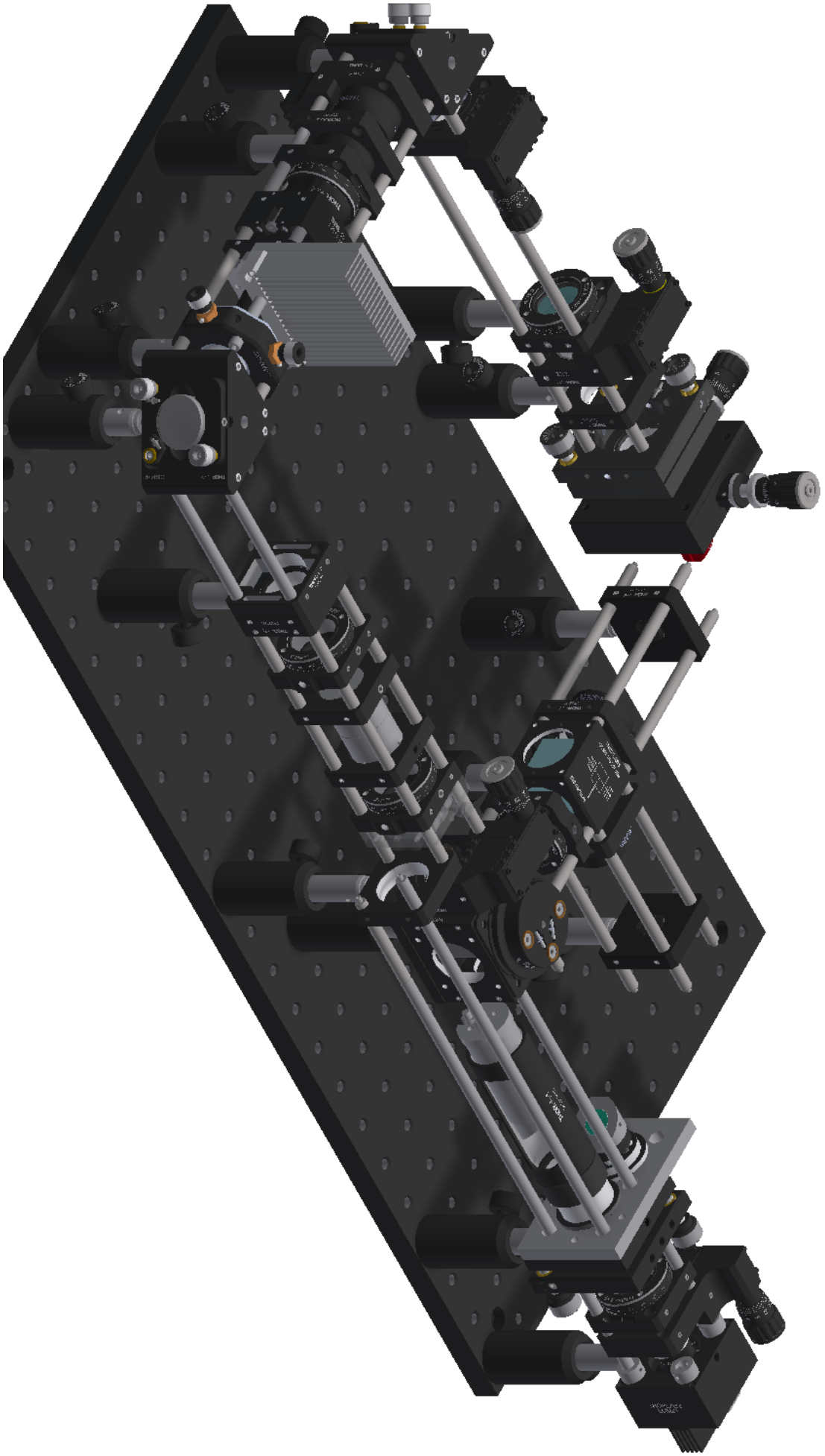

Figure B.9: Design of the external laser cavity including its amplification and its injection into a single-mode optical fiber.

BIBLIOGRAPHY

[1] ABIODUN, O. I., JANTAN, A., OMOLARA, A. E., DADA, K. V., MOHAMED, N. A. E., AND ARSHAD, H. **State-of-the-art in artificial neural network applications: A survey**. *Heliyon 4*, 11 (2018), 1–41.

[2] GRAVES, A., MOHAMED, A.-R., AND HINTON, G. **Speech recognition with deep recurrent neural networks**. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 6645–6649.

[3] GRAVES, A., LIWICKI, M., FERNANDEZ, S., BERTOLAMI, R., BUNKE, H., AND RGEN SCHMIDHUBER, J. **A Novel Connectionist System for Unconstrained Handwriting Recognition**. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 5 (2009), 855–868.

[4] MCCORDUCK, P. **Machines Who Think**, 2 ed. A K Peters/CRC Press, 2004.

[5] LECUN, Y., BENGIO, Y., AND HINTON, G. **Deep learning**. *Nature 521*, 7553 (2015), 436–444.

[6] ALTENRIED, M. **The platform as factory: Crowdwork and the hidden labour behind artificial intelligence**. *Capital and Class 44*, 2 (2020), 145–158.

[7] HAZELWOOD, K., BIRD, S., BROOKS, D., CHINTALA, S., DIRIL, U., DZHULGAKOV, D., FAWZY, M., JIA, B., JIA, Y., KALRO, A., LAW, J., LEE, K., LU, J., NOORDHUIS, P., SMELYANSKIY, M., XIONG, L., AND WANG, X. **Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective**. In *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)* (2018), pp. 620–629.

[8] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V., LANCTOT, M., DIELEMAN, S., GREWE, D., NHAM, J., KALCHBRENNER, N., SUTSKEVER, I., LILLICRAP, T., LEACH, M., KAVUKCUOGLU, K., GRAEPEL, T., AND HASSABIS, D. **Mastering the game of Go with deep neural networks and tree search**. *Nature 529*, 7587 (2016), 484–489.

[9] BOROWIEC, S. **AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol**. *The Guardian 15* (2016).

[10] TROMP, J., AND FARNEBÄCK, G. **Combinatorics of Go**. In *International Conference on Computers and Games* (2006), pp. 84–99.

[11] WALDROP, M. M. **The chips are down for Moore's law**. *Nature News 530*, 7589 (2016), 144.

[12] HARROD, W. **A journey to exascale computing**. In *2012 SC Companion: High Performance Computing, Networking Storage and Analysis* (2012), IEEE, pp. 1702–1730.

[13] JOUPPI, N., YOUNG, C., PATIL, N., AND PATTERSON, D. **Motivation for and Evaluation of the First Tensor Processing Unit**. *IEEE Micro 38*, 3 (2018), 10–19.

[14] GALDINO, L., EDWARDS, A., YI, W., SILLEKENS, E., WAKAYAMA, Y., GERARD, T., PELOUCH, W. S., BARNES, S., TSURITANI, T., KILLEY, R. I., LAVERY, D., AND BAYVEL, P. **Optical Fibre Capacity Optimisation via Continuous Bandwidth Amplification and Geometric Shaping**. *IEEE Photonics Technology Letters 32*, 17 (2020), 1021–1024.

[15] JAEGER, H., AND HAAS, H. **Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication**. *Science 304*, 5667 (2004), 78–80.

[16] BUENO, J., MAKTOOBI, S., FROEHLY, L., FISCHER, I., JACQUOT, M., LARGER, L., AND BRUNNER, D. **Reinforcement learning in a large-scale photonic recurrent neural network**. *Optica 5*, 6 (jun 2018), 756–760.

[17] MAKTOOBI, S., FROEHLY, L., ANDREOLI, L., PORTE, X., JACQUOT, M., LARGER, L., AND BRUNNER, D. **Diffractive Coupling For Photonic Networks: How Big Can We Go?** *IEEE Journal of Selected Topics in Quantum Electronics 26*, 1 (jul 2019), 1–8.

[18] HEUSER, T., GROSSE, J., HOLZINGER, S., SOMMER, M. M., AND REITZENSTEIN, S. **Development of Highly Homogenous Quantum Dot Micropillar Arrays for Optical Reservoir Computing**. *IEEE Journal of Selected Topics in Quantum Electronics 26*, 1 (2020), 1–9.

[19] TURING, A. M. **Computability and λ-Definability**. *The Journal of Symbolic Logic 2*, 4 (1937), 153–163.

[20] CHURCH, A. **An Unsolvable Problem of Elementary Number Theory**. *American Journal of Mathematics 58*, 2 (1936), 345–363.

[21] SIPSER, M. **Introduction to the theory of computation**, 2nd ed ed. Thomson Course Technology, 2006.

[22] COPELAND, J. **Colossus: The Secrets of Bletchley Park's Codebreaking Computers**. Oxford University Press, 2006.

[23] JACKSON, P. **Introduction to expert systems**, 3 ed. ed. Addison-Wesley Longman Publishing Co., Inc., 1998.

[24] GROSAN, C., AND ABRAHAM, A. **Intelligent systems: A Modern Approach**. Springer, Berlin, 2011.

[25] TVERSKY, A., AND KAHNEMAN, D. **The framing of decisions and the psychology of choice.** *Science 211*, 4481 (1981), 453–458.

[26] MELBY, P., WEBER, N., AND HÜBLER, A. **Dynamics of self-adjusting systems with noise**. *Chaos 15*, 3 (2005), 033902.

[27] GINTAUTAS, V., FOSTER, G., AND HÜBLER, A. W. **Resonant forcing of chaotic dynamics**. *Journal of Statistical Physics 130*, 3 (2008), 617–629.

[28] TRACHETTE JACKSON, A. R. E. **Applications of Dynamical Systems in Biology and Medicine**, 1 ed. Springer-Verlag New York, 2015.

[29] GANDOLFO, G. **Economic Dynamics**, 4 ed. Springer-Verlag Berlin Heidelberg, 2009.

[30] KREYSZIG, E. **Advanced Engineering Mathematics**, 10th editi ed. John Wiley & Sons, 2010.

[31] LEHALLE, C.-A. **Le contrôle non linéaire par réseaux de neurones formels : les perceptrons affines par morceaux**. PhD thesis, Université Paris 6, 2005.

[32] MARQUEZ, B. A. **Reservoir Computing Photonique et Méthodes Non-Linéaires de Representation de Signaux Complexes: Application à la Prédiction de Séries Temporelles**. PhD thesis, Université de Franche-Comté, 2018.

[33] STROGATZ, S. H. **Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering**. Westview Press, 2001.

[34] GLEICK, J. **Chaos: Making a New Science**, anniversar ed. Penguin Books, 2008.

[35] MCCULLOCH, W. S., AND PITTS, W. **A logical calculus of the ideas immanent in nervous activity**. *The Bulletin of Mathematical Biophysics 5* (1943), 115–133.

[36] DEBAO, C. **Degree of approximation by superpositions of a sigmoidal function**. *Approximation Theory and its Applications 9*, 3 (1993), 17–28.

[37] ŽILINSKAS, A. **Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms**. *Interfaces 36*, 6 (2006), 613–615.

[38] WU, H. **Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions**. *Information Sciences 179*, 19 (2009), 3432–3441.

[39] SUSSILLO, D., AND ABBOTT, L. F. **Random Walk Initialization for Training Very Deep Feedforward Networks**. In *ICLR 2015* (2015), pp. 1–10.

[40] ROSENBLATT, F. **The perceptron: A probabilistic model for information storage and organization in the brain**. *Psychological Review 65*, 6 (1958), 386–408.

[41] LECUN, Y. **Generalization and network design strategies**. Elsevier Publishers, 1989.

[42] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. **ImageNet Classification with Deep Convolutional Neural Networks**. In *Advances in Neural Information Processing Systems 25* (2012), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., pp. 1097–1105.

[43] ELMAN, J. **Finding Structure in Time**. *Cognitive Science 14* (1990), 179–211.

[44] JORDAN, M. I. **Serial Order: A Parallel Distributed Processing Approach**. In *Advances in Psychology*, vol. 121. Elsevier, 1997, pp. 471–495.

[45] SUN, G. Z., CHEN, H. H., LEE, Y. C., AND GILES, C. L. **Turing equivalence of neural networks with second order connection weights**. In *Proceedings. IJCNN - International Joint Conference on Neural Networks*, no. 2. 1992, pp. 357–362.

[46] PAKDAMAN, K., VIBERT, J. F., BOUSSARD, E., AND AZMY, N. **Single neuron with recurrent excitation: Effect of the transmission delay**. *Neural Networks 9*, 5 (1996), 797–818.

[47] SAK, H., SENIOR, A., AND BEAUFAYS, F. **Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition**. *arXiv preprint arXiv:1402.1128* (2014).

[48] LI, X., AND WU, X. **Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition**. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (2015), 4520–4524.

[49] CONNOR, J. T., MARTIN, R. D., AND ATLAS, L. E. **Recurrent Neural Networks and Robust Time Series Prediction**. *IEEE Transactions on Neural Networks 5*, 2 (1994), 240–254.

[50] RUMELHART, D. E., HINTONT, G. E., AND WILLIAMS, R. J. **Learning representations by back-propagating errors David**. *Nature 323*, 9 (1986), 533–536.

[51] GOH, G. B., HODAS, N. O., AND VISHNU, A. **Deep learning for computational chemistry**. *Journal of Computational Chemistry 38*, 16 (2017), 1291–1307.

[52] PASCANU, R., MIKOLOV, T., AND BENGIO, Y. **On the difficulty of training recurrent neural networks**. In *International conference on machine learning* (2013), pp. 1310–1318.

[53] HOCHREITER, S. **Untersuchungen zu dynamischen neuronalen Netzen**. PhD thesis, Institut f. Informatik, Technische Univ. Munich, 1991.

[54] HOCHREITER, S., AND SCHMIDHUBER, J. **Long Short-Term Memory**. *Neural Computation 9*, 8 (1997), 1735–1780.

[55] CHO, K., VAN MERRIËNBOER, B., GULCEHRE, C., BAHDANAU, D., BOUGARES, F., SCHWENK, H., AND BENGIO, Y. **Learning phrase representations using RNN encoder-decoder for statistical machine translation**. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* (2014), 1724–1734.

[56] MORRIS, R., AND HEBB, D. O. **The organization of behavior wiley: New York; 1949**. *Brain research bulletin 50*, 5-6 (1999), 437.

[57] OJA, E. **Simplified neuron model as a principal component analyzer**. *Journal of Mathematical Biology 15*, 3 (1982), 267–273.

[58] HAWKINS, D. M. **The Problem of Overfitting**. *Journal of Chemical Information and Computer Sciences 44*, 1 (2004), 1–12.

[59] JAEGER, H. **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), 13.

[60] MAASS, W., NATSCHLÄGER, T., AND MARKRAM, H. **Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations**. *Neural Computation 14*, 11 (2002), 2531–2560.

[61] JAEGER, H. **Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach**. GMD-Forschungszentrum Informationstechnik Bonn, 2002.

[62] GRIGORYEVA, L., AND ORTEGA, J. P. **Echo state networks are universal**. *Neural Networks 108*, August (2018), 495–508.

[63] JAEGER, H. **Short term memory in echo state networks. gmd-report 152**. In *GMD-German National Research Institute for Computer Science* (2002), German National Research Institute for Computer Science.

[64] VAN ESSEN, D. C., UGURBIL, K., AUERBACH, E., BARCH, D., BEHRENS, T. E., BUCHOLZ, R., CHANG, A., CHEN, L., CORBETTA, M., CURTISS, S. W., DELLA PENNA, S., FEINBERG, D., GLASSER, M. F., HAREL, N., HEATH, A. C., LARSON-PRIOR, L., MARCUS, D., MICHALAREAS, G., MOELLER, S., OOSTENVELD, R., PETERSEN, S. E., PRIOR, F., SCHLAGGAR, B. L., SMITH, S. M., SNYDER, A. Z., XU, J., AND YACOUB, E. **The Human Connectome Project: A data acquisition perspective**. *NeuroImage 62*, 4 (2012), 2222–2231.

[65] COSGROVE, K. P., MAZURE, C. M., AND STALEY, J. K. **Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry**. *Biological Psychiatry 62*, 8 (2007), 847–855.

[66] AZEVEDO, F. A., CARVALHO, L. R., GRINBERG, L. T., FARFEL, J. M., FERRETTI, R. E., LEITE, R. E., FILHO, W. J., LENT, R., AND HERCULANO-HOUZEL, S. **Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain**. *Journal of Comparative Neurology 513*, 5 (2009), 532–541.

[67] SABLONNIÈRE, B. **Le cerveau. Les clés de son développement et de sa longévité**. Odile Jacob, 2013.

[68] FISCHL, B., AND DALE, A. M. **Measuring the thickness of the human cerebral cortex from magnetic resonance images**. *Proceedings of the National Academy of Sciences of the United States of America 97*, 20 (2000), 11050–11055.

[69] MILOŠEVIĆ, N. T., AND RISTANOVIĆ, D. **Fractality of dendritic arborization of spinal cord neurons**. *Neuroscience Letters 396*, 3 (2006), 172–176.

[70] SAMPAIO-BAPTISTA, C., AND JOHANSEN-BERG, H. **White Matter Plasticity in the Adult Brain**. *Neuron 96*, 6 (2017), 1239–1251.

[71] BARTOL, T. M., BROMER, C., KINNEY, J., CHIRILLO, M. A., BOURNE, J. N., HARRIS, K. M., AND SEJNOWSKI, T. J. **Nanoconnectomic upper bound on the variability of synaptic plasticity**. *eLife 4*, 2015 (2015), e10778.

[72] RAICHLE, M. E., AND GUSNARD, D. A. **Appraising the brain's energy budget**. *Proceedings of the National Academy of Sciences of the United States of America 99*, 16 (2002), 10237–10239.

[73] ZHENG, Z., LAURITZEN, J. S., PERLMAN, E., ROBINSON, C. G., NICHOLS, M., MILKIE, D., TORRENS, O., PRICE, J., FISHER, C. B., SHARIFI, N., CALLE-SCHULER, S. A., KMECOVA, L., ALI, I. J., KARSH, B., TRAUTMAN, E. T., BOGOVIC, J. A., HANSLOVSKY, P., JEFFERIS, G. S., KAZHDAN, M., KHAIRY, K., SAALFELD, S., FETTER, R. D., AND BOCK, D. D. **A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster**. *Cell 174*, 3 (2018), 730–743.

[74] BERNARDI, R., CAKICI, R., ELLIOTT, D., ERDEM, A., ERDEM, E., IKIZLER-CINBIS, N., KELLER, F., MUSCAT, A., AND PLANK, B. **Automatic description generation from images: A survey of models, datasets, and evaluation measures**. *Journal of Artificial Intelligence Research 55* (2016), 409–442.

[75] JOHNSON, M., SCHUSTER, M., LE, Q. V., KRIKUN, M., WU, Y., CHEN, Z., THORAT, N., VIÉGAS, F., WATTENBERG, M., CORRADO, G., HUGHES, M., AND DEAN, J. **Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation**. *Transactions of the Association for Computational Linguistics 5* (2017), 339–351.

[76] SALLAB, A. E., ABDOU, M., PEROT, E., AND YOGAMANI, S. **Deep reinforcement learning framework for autonomous driving**. *Electronic Imaging 2017*, 19 (2017), 70–76.

[77] SILVER, D., HUBERT, T., SCHRITTWIESER, J., ANTONOGLOU, I., LAI, M., GUEZ, A., LANCTOT, M., SIFRE, L., KUMARAN, D., GRAEPEL, T., LILLICRAP, T., SIMONYAN, K., AND HASSABIS, D. **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play**. *Science 362*, 6419 (2018), 1140–1144.

[78] BROWN, N., AND SANDHOLM, T. **Libratus: The superhuman AI for no-limit poker**. *IJCAI International Joint Conference on Artificial Intelligence 0* (2017), 5226–5228.

[79] VINYALS, O., BABUSCHKIN, I., CZARNECKI, W. M., MATHIEU, M., DUDZIK, A., CHUNG, J., CHOI, D. H., POWELL, R., EWALDS, T., GEORGIEV, P., OH, J., HORGAN, D., KROISS, M., DANIHELKA, I., HUANG, A., SIFRE, L., CAI, T., AGAPIOU, J. P., JADERBERG, M., VEZHNEVETS, A. S., LEBLOND, R., POHLEN, T., DALIBARD, V., BUDDEN, D., SULSKY, Y., MOLLOY, J., PAINE, T. L., GULCEHRE, C., WANG, Z., PFAFF, T., WU, Y., RING, R., YOGATAMA, D., WÜNSCH, D., MCKINNEY, K., SMITH, O., SCHAUL, T., LILLICRAP, T., KAVUKCUOGLU, K., HASSABIS, D., APPS, C., AND SILVER, D. **Grandmaster level in StarCraft II using multi-agent reinforcement learning**. *Nature 575*, 7782 (2019), 350–354.

[80] STRUBELL, E., GANESH, A., AND MCCALLUM, A. **Energy and policy considerations for deep learning in NLP**. *arXiv preprint arXiv:1906.02243* (2019).

[81] MILLER, D. A., AND OZAKTAS, H. M. **Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture**. *Journal of Parallel and Distributed Computing 41*, 1 (1997), 42–52.

[82] BORKAR, S. **Design challenges of technology scaling**. *IEEE Micro 19*, 4 (1999), 23–29.

[83] MAGEN, N., KOLODNY, A., WEISER, U., AND SHAMIR, N. **Interconnect-power dissipation in a microprocessor**. *International Workshop on System Level Interconnect Prediction, SLIP*, 74 (2004), 7–13.

[84] REUTHER, A., MICHALEAS, P., JONES, M., GADEPALLY, V., SAMSI, S., AND KEPNER, J. **Survey of Machine Learning Accelerators**. 1–11.

[85] APPELTANT, L., SORIANO, M. C., VAN DER SANDE, G., DANCKAERT, J., MASSAR, S., DAMBRE, J., SCHRAUWEN, B., MIRASSO, C. R., AND FISCHER, I. **Information processing using a single dynamical node as complex system**. *Nature Communications 2*, 1 (2011), 1–6.

[86] SORIANO, M. C., ORTÍN, S., KEUNINCKX, L., APPELTANT, L., DANCKAERT, J., PESQUERA, L., AND VAN DER SANDE, G. **Delay-based reservoir computing: Noise effects in a combined analog and digital implementation**. *IEEE Transactions on Neural Networks and Learning Systems 26*, 2 (2015), 388–393.

[87] OMONDI, A. R., AND RAJAPAKSE, J. C. **FPGA implementations of neural networks**. Springer, Boston, 2006.

[88] WALTER, F., RÖHRBEIN, F., AND KNOLL, A. **Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks**. *Neural Networks 72* (2015), 152–167.

[89] TORREJON, J., RIOU, M., ARAUJO, F. A., TSUNEGI, S., KHALSA, G., QUERLIOZ, D., BORTOLOTTI, P., CROS, V., YAKUSHIJI, K., FUKUSHIMA, A., KUBOTA, H., YUASA, S., STILES, M. D., AND GROLLIER, J. **Neuromorphic computing with nanoscale spintronic oscillators**. *Nature 547*, 7664 (2017), 428–431.

[90] VANDOORNE, K., DAMBRE, J., VERSTRAETEN, D., SCHRAUWEN, B., AND BIENSTMAN, P. **Parallel Reservoir Computing Using Optical Amplifier**. *IEEE Transactions on Neural Networks 22*, 9 (2011), 1469–1481.

[91] MESARITAKIS, C., PAPATAXIARHIS, V., AND SYVRIDIS, D. **Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system**. *JOSA B 30*, 11 (2013), 3048–3055.

[92] LARGER, L., SORIANO, M. C., BRUNNER, D., APPELTANT, L., GUTIERREZ, J. M., PESQUERA, L., MIRASSO, C. R., AND FISCHER, I. **Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing**. *Optics Express 20*, 3 (2012), 3241–3249.

[93] FERNANDO, C., AND SOJAKKA, S. **Pattern Recognition in a Bucket**. In *European conference on artificial life* (2003), Springer, pp. 588–597.

[94] TANAKA, G., YAMANE, T., HÉROUX, J. B., NAKANE, R., KANAZAWA, N., TAKEDA, S., NUMATA, H., NAKANO, D., AND HIROSE, A. **Recent advances in physical reservoir computing: A review**. *Neural Networks 115* (2019), 100–123.

[95] KEYES, R. W. **Optical logic-in the light of computer technology**. *Optica Acta: International Journal of Optics 32*, 5 (1985), 525–535.

[96] MILLER, D. A. B. **Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators as efficient impedance converters**. *Optics Letters 14*, 2 (1989), 146–148.

[97] MILLER, D. A. **Rationale and challenges for optical interconnects to electronic chips**. *Proceedings of the IEEE 88*, 6 (2000), 728–749.

[98] FERREIRA DE LIMA, T., SHASTRI, B. J., TAIT, A. N., NAHMIAS, M. A., AND PRUCNAL, P. R. **Progress in neuromorphic photonics**. *Nanophotonics 6*, 3 (2017), 577–599.

[99] MILLER, D. A. **Attojoule Optoelectronics for Low-Energy Information Processing and Communications**. *Journal of Lightwave Technology 35*, 3 (2017), 346–396.

[100] BRUNNER, D., SORIANO, M. C., AND VAN DER SANDE, G. **Photonic Reservoir Computing: Optical Recurrent Neural Networks**. Walter de Gruyter GmbH & Co KG, 2019.

[101] WAN, K. H., DAHLSTEN, O., KRISTJÁNSSON, H., GARDNER, R., AND KIM, M. S. **Quantum generalisation of feedforward neural networks**. *npj Quantum Information 3*, 1 (2017), 1–10.

[102] GHOSH, S., OPALA, A., MATUSZEWSKI, M., PATEREK, T., AND LIEW, T. C. **Quantum reservoir processing**. *npj Quantum Information 5*, 1 (2019), 1–6.

[103] SELMI, F., BRAIVE, R., BEAUDOIN, G., SAGNES, I., KUSZELEWICZ, R., AND BARBAY, S. **Relative refractory period in an excitable semiconductor laser**. *Physical Review Letters 112*, 18 (2014), 1–5.

[104] HURTADO, A., AND JAVALOYES, J. **Controllable spiking patterns in long-wavelength vertical cavity surface emitting lasers for neuromorphic photonics systems**. *Applied Physics Letters 107*, 24 (2015), 241103.

[105] ROBERTSON, J., DENG, T., JAVALOYES, J., AND HURTADO, A. **Controlled inhibition of spiking dynamics in VCSELs for neuromorphic photonics: theory and experiments**. *Optics Letters 42*, 8 (2017), 1560–1563.

[106] LARGER, L., BAYLÓN-FUENTES, A., MARTINENGHI, R., UDALTSOV, V. S., CHEMBO, Y. K., AND JACQUOT, M. **High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification**. *Physical Review X 7*, 1 (2017), 1–14.

[107] VANDOORNE, K., DIERCKX, W., SCHRAUWEN, B., VERSTRAETEN, D., BAETS, R., BIENSTMAN, P., AND VAN CAMPENHOUT, J. **Toward optical signal processing using photonic reservoir computing**. *Optics express 16*, 15 (2008), 11182–11192.

[108] SHEN, Y., HARRIS, N. C., SKIRLO, S., PRABHU, M., BAEHR-JONES, T., HOCHBERG, M., SUN, X., ZHAO, S., LAROCHELLE, H., ENGLUND, D., AND SOLJACIC, M. **Deep learning with coherent nanophotonic circuits**. *Nature Photonics 11*, 7 (2017), 441–446.

[109] KATUMBA, A., HEYVAERT, J., SCHNEIDER, B., UVIN, S., DAMBRE, J., AND BIENSTMAN, P. **Low-Loss Photonic Reservoir Computing with Multimode Photonic Integrated Circuits**. *Scientific Reports 8*, 1 (2018), 1–10.

[110] FARHAT, N. H., PSALTIS, D., PRATA, A., AND PAEK, E. **Optical implementation of the Hopfield model**. *Applied Optics 24*, 10 (1985), 1469–1475.

[111] PSALTIS, D., AND FARHAT, N. **Optical information processing based on an associative-memory model of neural nets with thresholding and feedback**. *Optics Letters 10*, 2 (1985), 98–100.

[112] SMITH, S. D. **Lasers, nonlinear optics and optical computers**. *Nature 316*, 6026 (1985), 319–324.

[113] BRUNNER, D., AND FISCHER, I. **Reconfigurable semiconductor laser networks based on diffractive coupling**. *Optics Letters 40*, 16 (aug 2015), 3854–3857.

[114] MESARITAKIS, C., BOGRIS, A., KAPSALIS, A., AND SYVRIDIS, D. **High-speed all-optical pattern recognition of dispersive Fourier images through a photonic reservoir computing subsystem**. *Optics Letters 40*, 14 (2015), 3416–3419.

[115] HEUSER, T., PFLÜGER, M., FISCHER, I., LOTT, J. A., BRUNNER, D., AND REITZENSTEIN, S. **Developing a photonic hardware platform for brain-inspired computing based on 5×5 VCSEL arrays**. *Journal of Physics: Photonics 2*, 4 (2020), 044002.

[116] KATZ, O., HEIDMANN, P., FINK, M., AND GIGAN, S. **Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations**. *Nature Photonics 8*, 10 (2014), 784–790.

[117] ROTTER, S., AND GIGAN, S. **Light fields in complex media: Mesoscopic scattering meets wave control**. *Reviews of Modern Physics 89*, 1 (2017), 015005.

[118] SAADE, A., CALTAGIRONE, F., CARRON, I., DAUDET, L., DREMEAU, A., GIGAN, S., AND KRZAKALA, F. **Random projections through multiple optical scattering: Approximating kernels at the speed of light**. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2016), pp. 6215–6219.

[119] DONG, J., GIGAN, S., KRZAKALA, F., AND WAINRIB, G. **Scaling Up Echo-State Networks with Multiple Light Scattering**. In *2018 IEEE Statistical Signal Processing Workshop, SSP 2018* (2018), IEEE, pp. 448–452.

[120] MOUGHAMES, J., PORTE, X., THIEL, M., ULLIAC, G., LARGER, L., JACQUOT, M., KADIC, M., AND BRUNNER, D. **Three-dimensional waveguide interconnects for scalable integration of photonic neural networks**. *Optica 7*, 6 (2020), 640–646.

[121] HEUSER, T., GROSSE, J., KAGANSKIY, A., BRUNNER, D., AND REITZENSTEIN, S. **Fabrication of dense diameter-tuned quantum dot micropillar arrays for applications in photonic information processing**. *APL Photonics 3*, 11 (nov 2018), 116103.

[122] ANTONIK, P., HAELTERMAN, M., AND MASSAR, S. **Brain-Inspired Photonic Signal Processor for Generating Periodic Patterns and Emulating Chaotic Systems**. *Physical Review Applied 7*, 5 (may 2017), 054014.

[123] IKEDA, K., DAIDO, H., AND AKIMOTO, O. **Optical turbulence: Chaotic behavior of transmitted light from a ring cavity**. *Physical Review Letters 45*, 9 (1980), 709–712.

[124] PEURIFOY, J., SHEN, Y., JING, L., YANG, Y., CANO-RENTERIA, F., DELACY, B. G., JOANNOPOULOS, J. D., TEGMARK, M., AND SOLJAČIĆ, M. **Nanophotonic particle simulation and inverse design using artificial neural networks**. *Science Advances 4*, 6 (2018), 1–8.

[125] HOPF, F. A., KAPLAN, D.L., GIBBS, H.M. AND SHOEMAKER, R. **Bifurcations to chaos in optical bistability**. *Physical Review A 25*, 4 (1982), 2172–2183.

[126] GOEDGEBUER, J.-P., LARGER, L., PORTE, H., AND DELORME, F. **Chaos in wavelength with a feedback tunable laser diode**. *Physical review Eeview E 57*, 3 (1998), 2795–2798.

[127] SORIANO, M. C., GARCÍA-OJALVO, J., MIRASSO, C. R., AND FISCHER, I. **Complex photonics: Dynamics and applications of delay-coupled semiconductors lasers**. *Reviews of Modern Physics 85*, 1 (mar 2013), 421–470.

[128] RAZ, R. **On the Complexity of Matrix Product**. *SIAM Journal on Computing 32*, 5 (2003), 1356–1369.

[129] DOERR, B., AND JANSEN, T. **Theory of Evolutionary Computation: Recent Developments in Discrete Optimization**. Springer Nature, 2019.

[130] MACKEY, M. C., AND GLASS, L. **Oscillation and chaos in physiological control systems**. *Science 197*, 4300 (1977), 287–289.

[131] BUENO, J., BRUNNER, D., SORIANO, M. C., AND FISCHER, I. **Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback**. *Optics Express 25*, 3 (feb 2017), 2401–2412.

[132] ORTÍN, S., SORIANO, M. C., PESQUERA, L., BRUNNER, D., AND FISCHER, I. **A Unified Framework for Reservoir Computing and Extreme Learning Machines based on a Single Time-delayed Neuron**. *Nature Publishing Group*, October (2015), 1–11.

[133] LUKOŠEVIČIUS, M., AND JAEGER, H. **Reservoir computing approaches to recurrent neural network training**. *Computer Science Review 3*, 3 (2009), 127–149.

[134] SUAREZ-PEREZ, A., GABRIEL, G., REBOLLO, B., ILLA, X., GUIMERÀ-BRUNET, A., HERNÁNDEZ-FERRER, J., MARTÍNEZ, M. T., VILLA, R., AND SANCHEZ-VIVES, M. V. **Quantification of signal-to-noise ratio in cerebral cortex recordings using flexible MEAs with co-localized platinum black, carbon nanotubes, and gold electrodes**. *Frontiers in Neuroscience 12*, NOV (nov 2018).

[135] ANDREOLI, L., PORTE, X., CHRÉTIEN, S., JACQUOT, M., LARGER, L., AND BRUNNER, D. **Boolean learning under noise-perturbations in hardware neural networks**. *Nanophotonics 9*, 13 (2020), 4139–4147.

[136] SEMENOVA, N., PORTE, X., ANDREOLI, L., JACQUOT, M., LARGER, L., AND BRUNNER, D. **Fundamental aspects of noise in analog-hardware neural networks**. *Chaos 29*, 10 (2019), 103128.

[137] UCHIDA, A., MCALLISTER, R., AND ROY, R. **Consistency of nonlinear system response to complex drive signals**. *Physical Review Letters 93*, 24 (dec 2004), 1–4.

[138] BOCIORT, F., VAN DRIEL, E., AND SEREBRIAKOV, A. **Network structure of the set of local minima in optical system optimization**. In *Novel Optical Systems Design and Optimization VI* (2003), J. M. S. Manhart, R. J. Koshel, and P. K., Eds., vol. 5174, SPIE, pp. 26–34.

[139] BOCIORT, F., VAN DRIEL, E., AND SEREBRIAKOV, A. **Networks of local minima in optical system optimization**. *Optics Letters 29*, 2 (2004), 189–191.

[140] PORTE, X., ANDREOLI, L., JACQUOT, M., LARGER, L., AND BRUNNER, D. **Reservoir-size dependent learning in analogue neural networks**. In *Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions* (2019), pp. 184–192.

[141] WEGENER, I., AND WITT, C. **On the analysis of a simple evolutionary algorithm on quadratic pseudo-boolean functions**. *Journal of discrete algorithms 3*, 1 (2005), 61–78.

[142] WEGENER, I., AND WITT, C. **On the optimization of monotone polynomials by simple randomized search heuristics**. *Combinatorics, Probability and Computing 14*, 1-2 (2005), 225–2247.

[143] DOERR, B. **Probabilistic Tools for the Analysis of Randomized Optimization Heuristics**. Springer, 2020.

[144] DOERR, B., WITT, C., AND YANG, J. **Runtime Analysis for Self-adaptive Mutation Rates**. *Algorithmica 83*, 4 (2021), 1012–1053.

[145] YANG, J. **Designing Superior Evolutionary Algorithms via Insights From Black-Box Complexity Theory**. PhD thesis, Université Paris-Saclay, 2018.

[146] FREIBERGER, M., KATUMBA, A., BIENSTMAN, P., AND DAMBRE, J. **Training Passive Photonic Reservoirs with Integrated Optical Readout**. *IEEE Transactions on Neural Networks and Learning Systems 30*, 7 (2019), 1943–1953.

[147] LIU, K., YE, C. R., KHAN, S., AND SORGER, V. J. **Review and perspective on ultrafast wavelength-size electro-optic modulators**. *Laser and Photonics Reviews 9*, 2 (2015), 172–194.

[148] HENRY, C. H. **Theory of the Linewidth of Semiconductor Lasers**. *IEEE Journal of Quantum Electronics QE-18*, 2 (1982), 259–264.

[149] BRUNNER, D., SORIANO, M. C., MIRASSO, C. R., AND FISCHER, I. **Parallel photonic information processing at gigabyte per second data rates using transient states**. *Nature Communications 4*, January (2013), 1364–1367.

[150] REITZENSTEIN, S., AND FORCHEL, A. **Quantum dot micropillars**. *Journal of Physics D: Applied Physics 43*, 3 (2010), 1–25.

[151] EAGLESHAM, D., AND CERULLO, M. **Dislocation-Free Stranski-Krastanow Growth of Ge on Si(100)**. *Physical Review Letters 64*, 16 (1990), 1943–1946.

[152] REITZENSTEIN, S., BAZHENOV, A., GORBUNOV, A., HOFMANN, C., MÜNCH, S., LÖFFLER, A., KAMP, M., REITHMAIER, J. P., KULAKOVSKII, V. D., AND FORCHEL, A. **Lasing in high-Q quantum-dot micropillar cavities**. *Applied Physics Letters 89*, 5 (2006), 051107.

[153] KREINBERG, S., CHOW, W. W., WOLTERS, J., SCHNEIDER, C., GIES, C., JAHNKE, F., HÖFLING, S., KAMP, M., AND REITZENSTEIN, S. **Emission from quantum-dot high-ß microcavities: Transition from spontaneous emission to lasing and the effects of superradiant emitter coupling**. *Light: Science and Applications 6*, 8 (2017), 1–8.

[154] OHTSUBO, J. **Semiconductor Lasers: Stability, Instability and Chaos**, 4 ed. Springer International Publishing, 2012.

[155] PETITBON, I., GALLION, P., DEBARGE, G., AND CHABRAN, C. **Locking Bandwidth and Relaxation Oscillations of an Injection-Locked Semiconductor Laser**. *IEEE Journal of Quantum Electronics 24*, 2 (1988), 148–154.

[156] CHANG, C. H., CHROSTOWSKI, L., AND CHANG-HASNAIN, C. J. **Injection Locking of VCSELs**. *IEEE Journal on Selected Topics in Quantum Electronics 9*, 5 (2003), 1386–1393.

[157] LAU, E. K., SUNG, H.-K., AND WU, M. C. **Scaling of resonance frequency for strong injection-locked lasers**. *Optics Letters 32*, 23 (2007), 3373–3375.

[158] OSINSKI, M., AND BUUS, J. **Linewidth Broadening Factor in Semiconductor Lasers–An Overview**. *IEEE Journal of Quantum Electronics 23*, 1 (1987), 9–29.

[159] SCHLOTTMANN, E., HOLZINGER, S., LINGNAU, B., LÜDGE, K., SCHNEIDER, C., KAMP, M., HÖFLING, S., WOLTERS, J., AND REITZENSTEIN, S. **Injection Locking of Quantum-Dot Microlasers Operating in the Few-Photon Regime**. *Physical Review Applied 6*, 4 (2016), 1–7.

[160] REMEDIAKIS, I. N., AND KAXIRAS, E. **Band-structure calculations for semiconductors within generalized-density-functional theory**. *Physical Review B - Condensed Matter and Materials Physics 59*, 8 (1999), 5536–5543.

[161] PANZARINI, G., AND ANDREANI, L. C. **Quantum theory of exciton polaritons in cylindrical semiconductor microcavities**. *Physical Review B - Condensed Matter and Materials Physics 60*, 24 (1999), 16799–16806.

[162] GIPPIUS, N. A., TIKHODEEV, S. G., AND FORCHEL, A. **Angle dependence of the spontaneous emission from confined optical modes in photonic dots**. *Physical Review B - Condensed Matter and Materials Physics 59*, 3 (1999), 2223–2229.

[163] DOUSSE, A., LANCO, L., SUFFCZYŃSKI, J., SEMENOVA, E., MIARD, A., LEMAÎTRE, A., SAGNES, I., ROBLIN, C., BLOCH, J., AND SENELLART, P. **Controlled light-matter coupling for a single quantum dot embedded in a pillar microcavity using far-field optical lithography**. *Physical Review Letters 101*, 26 (2008), 30–33.

[164] AGRAWAL GOVIND P. AND DUTTA NILOY K. **Semiconductor Lasers**, 2 ed. Springer US, 1993.

[165] YAMAMOTO, Y. **Analysis of Semiconductor Microcavity Lasers Using**. *IEEE Journal of Quantum Electronics 27*, 11 (1991), 2386–2396.

[166] MONEMAR, B., SHIH, K. K., AND PETTIT, G. D. **Some optical properties of the AlxGa1-xAs alloys system**. *Journal of Applied Physics 47*, 6 (1976), 2604–2613.

[167] BLAKEMORE, J. S. **Semiconducting and other major properties of gallium arsenide**. *Journal of Applied Physics 53*, 10 (1982), R123–R181.

[168] ASPNES, D. E., KELSO, S. M., LOGAN, R. A., AND BHAT, R. **Optical properties of AlxGa1-x As**. *Journal of Applied Physics 60*, 2 (1986), 754–767.

[169] STURGE, M. D. **Optical absorption of gallium arsenide between 0.6 and 2.75 eV**. *Physical Review 127*, 3 (1962), 768–773.

[170] ANDREOLI, L., PORTE, X., HEUSER, T., GROSSE, J., MOEGLEN-PAGET, B., FURFARO, L., REITZENSTEIN, S., AND BRUNNER, D. **Optical pumping of quantum dot micropillar lasers**. *Optics Express 29*, 6 (2021), 9084–9097.

[171] LEVINSHTEIN, M. **Handbook series on semiconductor parameters**, 1 ed. World Scientific, 1997.

[172] ASPNES, D. E., AND STUDNA, A. A. **Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV**. *Physical review B 27*, 2 (1983), 985–1009.

[173] WANG, W. H., GHOSH, S., MENDOZA, F. M., LI, X., AWSCHALOM, D. D., AND SAMARTH, N. **Static and dynamic spectroscopy of (Al,Ga)AsGaAs microdisk lasers with interface fluctuation quantum dots**. *Physical Review B 71*, 15 (2005), 155306.

[174] GÉRARD, J. M., SERMAGE, B., GAYRAL, B., LEGRAND, B., COSTARD, E., AND THIERRY-MIEG, V. **Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity**. *Physical Review Letters 81*, 5 (1998), 1110–1113.

[175] ULRICH, S. M., GIES, C., ATES, S., WIERSIG, J., REITZENSTEIN, S., HOFMANN, C., LÖFFLER, A., FORCHEL, A., JAHNKE, F., AND MICHLER, P. **Photon statistics of semiconductor microcavity lasers**. *Physical Review Letters 98*, 4 (2007), 3–6.

[176] BAVEJA, P. P., KÖGEL, B., WESTBERGH, P., GUSTAVSSON, J. S., HAGLUND, Å., MAYWAR, D. N., AGRAWAL, G. P., AND LARSSON, A. **Assessment of VCSEL thermal rollover mechanisms from measurements and empirical modeling**. *Optics Express 19*, 16 (2011), 15490–15505.

[177] KSENDZOV, A., GRUNTHANER, F., LIU, J., RICH, D., TERHUNE, R., AND WILSON, B. **Absorption And Photoluminescence of Ultrathin Pseudomorphic InAs/GaAs Quantum Well**. *Physical Review B 43*, 18 (1991), 14574–14580.

[178] MATTILA, M., HAKKARAINEN, T., LIPSANEN, H., JIANG, H., AND KAUPPINEN, E. I. **Enhanced luminescence from catalyst-free grown InP nanowires**. *Applied Physics Letters 90*, 3 (2007), 033101.

[179] HIGUERA-RODRIGUEZ, A., ROMEIRA, B., BIRINDELLI, S., BLACK, L. E., SMALBRUGGE, E., VAN VELDHOVEN, P. J., KESSELS, W. M., SMIT, M. K., AND FIORE, A. **Ultralow Surface Recombination Velocity in Passivated InGaAs/InP Nanopillars**. *Nano Letters 17*, 4 (2017), 2627–2633.

[180] ALEKSEEV, P. A., DUNAEVSKIY, M. S., ULIN, V. P., LVOVA, T. V., FILATOV, D. O., NEZHDANOV, A. V., MASHIN, A. I., AND BERKOVITS, V. L. **Nitride surface passivation of GaAs nanowires: Impact on surface state density**. *Nano Letters 15*, 1 (2015), 63–68.

[181] WALKER, S. J., AND JAHNS, J. **Array generation with multilevel phase gratings**. *JOSA A 7*, 8 (1990), 1509–1513.

[182] MILLER, J. M., TAGHIZADEH, M. R., TURUNEN, J., AND ROSS, N. **Multilevel-grating array generators: fabrication error analysis and experiments**. *Applied Optics 32*, 14 (1993), 2519–2525.

[183] GALE, M. T., ROSSI, M., SCHÜTZ, H., EHBETS, P., HERZIG, H. P., AND PRONGUÉ, D. **Continuous-relief diffractive optical elements for two-dimensional array generation**. *Applied Optics 32*, 14 (1993), 2526–2533.

[184] KATZ, S., KAPLAN, N., AND GROSSINGER, I. **Using Diffractive Optical Elements: DOEs for beam shaping - fundamentals and applications**. *Optik & Photonik 13*, 4 (2018), 83–86.

[185] GERCHBERG, R. **A practical algorithm for the determination of the phase from image and diffraction plane pictures**. *Saxton, W.O. 35* (1972), 237–246.

[186] FIENUP, J. **Iterative method applied to image reconstruction and to computer -generated holograms**. *Optical Engineering 19*, 3 (1980), 297–305.

[187] ARRIZÓN, V., TESTORF, M., SINZINGER, S., AND JAHNS, J. **Iterative optimization of phase-only diffractive optical elements based on a lenslet array**. *Journal of the Optical Society of America A 17*, 12 (2000), 2157–2164.

[188] SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I., HUANG, A., GUEZ, A., HUBERT, T., BAKER, L., LAI, M., BOLTON, A., CHEN, Y., LILLICRAP, T., HUI, F., SIFRE, L., VAN DEN DRIESSCHE, G., GRAEPEL, T., AND HASSABIS, D. **Mastering the game of Go without human knowledge**. *Nature 550*, 7676 (2017), 354–359.

[189] DINC, N. U., PSALTIS, D., AND BRUNNER, D. **Optical Neural Networks: The 3D connection**. *Photoniques*, 104 (2020), 34–38.

[190] KOLENDERSKA, S. M., KATZ, O., FINK, M., AND GIGAN, S. **Scanning-free imaging through a single fiber by random spatio-spectral encoding**. *Optics Letters 40*, 4 (2015), 534–537.

[191] JAEGER, H. **Neuromorphic computing: A productive contradiction in terms**. In *Emerging Topics in Artificial Intelligence 2020* (2020), p. 114690T.

[192] MIN, S., DALJEVIC, T., LAFRENIERE-ROULA, M., MANLHIOT, C., NALLI, N., GRASEMANN, H., SCHWARTZ, S., KAMATH, B., NG, V., PAREKH, R., AND AL., E. **A pilot trial of genotype-guided dosing versus standard clinical dosing in**

pediatric solid organ transplants. *American journal of transplantation. Conference: 17th american transplant congress, ATC 2017. United states 17*, 3 (2017), 649–649.

[193] FLEMING, M. W., AND MOORADIAN, A. **Fundamental line broadening of single-mode (GaAl)As diode lasers**. *Applied Physics Letters 38*, 7 (1981), 511–513.

[194] FLEMING, M. W., AND MOORADIAN, A. **Spectral Characteristics of External-Cavity Controlled Semiconductor Lasers**. *IEEE Journal of Quantum Electronics 17*, 1 (1981), 44–59.

[195] WIEMAN, C. E., AND HOLLBERG, L. **Using diode lasers for atomic physics**. *Review of Scientific Instruments 62*, 1 (1991), 1–20.

[196] FAVRE, F., DE GUEN, D., SIMON, J. C., AND LANDOUSIES, B. **External-Cavity Semiconductor Laser with 15 nm Continuous Tuning Range**. *Electronics Letters 22*, 15 (1986), 795–796.

[197] LOEWEN, E., NEVIERE, M., AND MAYSTRE, D. **Grating efficiency theory as it applies to blazed and holographic gratings**. *Applied Optics 16*, 10 (1977), 2711–2721.

[198] SMITH, M. S. **Diffraction gratings utilizing total internal reflection facets in littrow configuration**. *IEEE Photonics Technology Letters 11*, 1 (1999), 84–86.

[199] ARNOLD, A. S., WILSON, J. S., AND BOSHIER, M. G. **A simple extended-cavity diode laser**. *Review of Scientific Instruments 69*, 3 (1998), 1236–1239.

[200] HAWTHORN, C. J., WEBER, K. P., AND SCHOLTEN, R. E. **Littrow configuration tunable external cavity diode laser with fixed direction output beam**. *Review of Scientific Instruments 72*, 12 (2001), 4477–4479.

[201] LIVAS, J. C., CHINN, S. R., KINTZER, E. S., WALPOLE, J. N., WANG, C. A., AND MISSAGGIA, L. J. **Single-Mode Optical Fiber Coupling of High-Power Tapered Gain Region Devices**. *IEEE Photonics Technology Letters 6*, 3 (1994), 422–424.

RÉSUMÉ EN FRANÇAIS

Ces dernières années, l'intelligence artificielle s'est largement diffusée dans nos sociétés numériques. Plus particulièrement, les réseaux de neurones artificiels sont utilisés dans de nombreux domaines tels que les sciences, les transports, la finance, les jeux, la médecine, l'agriculture, la sécurité, les médias, la gestion, l'assurance ou encore l'ingénierie [1]. Les réseaux de neurones ont même révolutionné l'interaction Homme-machine grâce à la reconnaissance vocale et manuscrite [2, 3]. Cependant, si les études des premiers réseaux de neurones ont commencé il y a plus de 70 ans [4], et que les concepts théoriques étaient déjà connus depuis la fin des années 1980, ce n'est que récemment qu'une prodigieuse accélération des applications a vu le jour avec notamment l'apparition du *deep learning* [5]. En plus des avancées de la recherche dans le domaine, le récent succès des réseaux de neurones peut aussi être expliqué par les progrès de l'informatique. La démocratisation du numérique au cours du XXI^{e} siècle, notamment avec l'avènement des réseaux sociaux [6, 7], a permis de stocker et labelliser de gigantesques bases de données utilisées pour entraîner les réseaux de neurones. De plus, la montée en puissance des ordinateurs a permis d'émuler numériquement de larges réseaux de neurones. La victoire au jeu de Go de l'algorithme AlphaGo [8] contre Lee Sedol en 2016 [9] a été un évènement majeur, rendant visible au monde entier les avancées exceptionnelles réalisées dans le domaine des apprentissages automatiques. Pour atteindre un tel niveau dans un jeu réputé pour son extrême complexité en raison du nombre extraordinairement élevé de combinaisons [10], l'algorithme a dû s'entraîner avec des dizaines de milliers de parties menées par des joueurs experts, puis jouer des millions de fois contre lui-même. Ainsi, l'entraînement de tels réseaux de neurones est réalisé en utilisant la puissance de calcul de *data-centers*.

Cependant, le fonctionnement d'un ordinateur est fondamentalement très éloigné de celui des réseaux de neurones. Alors que la puissance de ces derniers provient d'une connectivité neuronale massive et parallèle, les ordinateurs, basés sur l'architecture de von Neumann, fonctionnent principalement en série. La taille des transistors actuels étant déjà proche des limites physiques [11], leur nombre par processeur augmente difficilement [12]. Depuis quinze ans, la progression de la puissance de calcul numérique est principalement liée à l'augmentation du nombre de processeurs utilisés [12]. Cependant, l'évolution rapide des réseaux de neurones émulés par ordinateur, nécessitant un grand nombre de connexions, requiert une puissance de calcul de plus en plus importante. Afin d'augmenter l'efficacité des réseaux de neurones numériques, des innovations technologiques matérielles telles que la conception des *Tensor Product Units* (TPU) améliorent partiellement la parallélisation en microélectronique [13]. L'un des principaux défis des futurs réseaux de neurones artificiels est alors de parvenir à une connectivité hautement parallélisée.

Alors que la technologie électronique des ordinateurs commence à atteindre ses limites [11, 12], nous montrons dans le Chapitre 1 que la recherche en optique offre des pistes encourageantes pour les générations futures de calculateurs à base de réseaux

de neurones [106]. En effet, contrairement à l'électron, le photon n'a pas de charge électrique, ne souffre pas de dissipation d'énergie capacitive et inductive. Ces caractéristiques intrinsèques de la lumière sont déjà utilisées dans les télécommunications, où les informations sont codées et échangées optiquement dans un seul canal de communication, comme une fibre optique, avec des débits record [14]. La possibilité pour les informations optiquement modulées d'être échangées simultanément, avec un minimum de bruits, permet une parallélisation de la connectivité des réseaux de neurones.

Notre objectif de recherche est de créer des réseaux de neurones inspirés par l'architecture des réseaux de neurones biologiques. Le principe est d'utiliser des phénomènes physiques pour éviter d'émuler les réseaux neuronaux sur ordinateur, et ainsi de construire des systèmes physiques qui constituent intrinsèquement les réseaux de neurones. Nous quittons alors le monde du tout numérique et revenons à des systèmes analogiques pour nous rapprocher des caractéristiques physiques du cerveau. Pour poursuivre au mieux ces principes et réaliser des réseaux neuronaux avec une connectivité parallélisée, nous utilisons la lumière comme vecteur de propagation de l'information. Nos réseaux de neurones sont basés sur le concept de *reservoir computing* [15] qui est une sous-catégorie de réseaux de neurones récurrents. L'architecture du réseau est composée d'une couche de neurones d'entrée, d'une couche de neurones récurrents, appelée réservoir, et d'une couche de sortie. Les neurones du réservoir doivent avoir une caractéristique non linéaire. Les connexions neuronales de la couche d'entrée et du réservoir peuvent être aléatoires et doivent être invariantes. Cette simplicité d'architecture facilite fortement la réalisation d'un réseau de neurones optique. Cette classe de réseau est spécialement bien adaptée pour résoudre des tâches de classification et de traitement d'informations temporelles tel que la prédiction de série chaotique. D'un point de vue pratique, l'apprentissage du réseau à résoudre un problème spécifique consiste à ajuster uniquement les poids des connexions de sortie afin de minimiser l'erreur du système.

Nous présentons dans le Chapitre 2 un réseau neuronal récurrent complet [16]. La distribution spatiale des neurones permet d'augmenter la taille du réseau sans réduire la bande passante du système. Son support théorique, sa réalisation, son algorithme d'apprentissage et sa phase de tests y sont étudiés. Les neurones sont électro-optiques, mais l'ensemble du réseau neuronal a été conçu pour servir de prototype à un système tout optique, plus proche des objectifs précédemment définis, qui sera présenté au Chapitre 4. La connectivité du réseau neuronal est réalisée par diffraction optique qui permet une réelle parallélisation et offre une grande capacité d'évolution [17]. La couche de sortie est réalisée grâce à une matrice de micro-miroirs, permettant une attribution des poids de lecture entièrement parallélisée. L'algorithme d'apprentissage est conçu pour faire partie intégrante du système et fonctionne en temps réel.

Nous discutons dans le Chapitre 3 des caractéristiques du réseau de neurones récurrent. Alors que le *reservoir computing* est souvent souligné pour son caractère *boite noire*, la réalisation électro-optique du réseau neuronal nous permet de mesurer les états internes du système, de le caractériser et de tester de nombreuses configurations d'apprentissage. Ce système devient alors une expérience permettant d'étudier de nombreux champs scientifiques. En raison du caractère analogique de l'optique, le réseau de neurones est constamment soumis au bruit. Tout d'abord nous avons étudié les différentes sources, leurs spécificités et leurs impacts sur le système [136]. En particulier, nous avons mesuré la constance du réseau de neurones, c'est-à-dire sa capacité à traiter de manière similaire des informations identiques malgré la présence de bruit.

L'apprentissage a également été étudié en détail. Bien que la fonction d'erreur possède de multiples minima locaux, ceux-ci présentent une erreur similaire. L'apprentissage converge toujours et permet la prédiction d'un signal chaotique avec une erreur de l'ordre du pourcent. Nous avons démontré que la configuration des poids de lecture pour résoudre une tâche est le résultat d'une optimisation de la résilience au bruit au cours de l'apprentissage [135]. Ainsi, l'entraînement est spécifique à chaque situation. Ces résultats obtenus découlent de principes généraux qui doivent être pris en compte dans tous les réseaux de neurones analogiques, selon leur propre spécificité. Enfin, nous avons mesuré la durée d'apprentissage qui se révèle être proportionnelle au nombre de neurones, donnant au système une importante capacité d'évolution [140].

En remplaçant les neurones électro-optiques par des éléments optiques, le système deviendra alors un réseau neuronal entièrement optique. Les neurones artificiels seront une matrice de micropiliers laser à quantum dots [18]. Les concepts fondamentaux, l'architecture et le couplage par diffraction resteront cependant les mêmes. Nous avons étudié dans le Chapitre 4 les bases d'un tel réseau qui fonctionnera non plus à quelques Hertz, mais au Gigahertz. Le pompage optique simultané de plusieurs lasers semiconducteurs, bien que toujours à l'étude, permettra à terme de réaliser un réseau de grande dimension. Nous avons caractérisé le pompage optique des micropiliers laser selon deux mécanismes différents. En plus d'obtenir une efficacité de pompage suffisamment élevée pour créer un futur réseau d'une centaine de neurones optiques, notre étude quantitative nous a permis de proposer des modifications aux micropiliers laser qui améliorent leur efficacité énergétique d'un ou deux ordres de grandeur.

Enfin, nous concluons en présentant un résumé des recherches présentées tout au long de ce manuscrit. En poursuivant selon nos lignes directrices et en nous appuyant sur les résultats obtenus, nous proposons des perspectives visant notamment à améliorer le réseau de neurones électro-optiques, mais aussi les prochaines étapes nécessaires à la réalisation du réseau de neurones tout optique. À la fin de cette thèse, des annexes détaillées présentent les aspects plus techniques des travaux présentés dans le Chapitre 4. L'annexe A illustre la conception géométrique des feuilles de cuivre permettant à la fois le refroidissement des micro-piliers laser tout en assurant un degré de liberté suffisant pour déplacer l'échantillon au sein d'un cryostat. L'annexe B décrit le laser de pompe accordable construit à l'aide d'une cavité laser externe via un réseau optique, amplifié optiquement, puis filtré spatialement par une fibre optique monomode.

Document generated with LaTeX and:
the LaTeX style for PhD Thesis created by S. Galland — http://www.multiagent.fr/ThesisStyle
the tex-upmethodology package suite — http://www.arakhne.org/tex-upmethodology/