



UNIVERSITÀ DEGLI STUDI DI TRENTO

DEPARTMENT OF PHYSICS
MASTER DEGREE IN PHYSICS

Optical Bistability As Neural Network
Nonlinear Activation Function

Supervisor:
Dr. Paolo Bettotti

Graduant:
Davide Bazzanella

Co-supervisor:
Dr. Enver Sangineto

20th March 2018

Contents

Preface	1
1 Artificial Neural Networks	7
1.1 History	7
1.2 Basis of Neural Networks	8
1.2.1 Applications of ANNs	10
1.3 Working Principles of ANNs	10
1.3.1 Learning Process	10
1.3.2 Validation Process	12
1.3.3 Testing Process	13
1.3.4 Datasets	13
1.4 Feed Forward ANN	13
1.4.1 Perceptron	13
1.4.2 Other Feed Forward ANNs	14
1.5 Other Types of ANNs	15
1.5.1 Recurrent ANN	15
1.5.2 Reservoir Computing	16
1.6 FFNN Simulation	16
1.6.1 PyTorch	17
1.6.2 Dataset	18
1.6.3 Simulated ANN operation	18
2 Integrated Photonics	23
2.1 Silicon Photonics	23
2.2 Guided-wave photonics	25
2.2.1 Waveguides	25
2.2.2 Microring optical cavity	29
2.3 Nonlinear Optics	35
2.3.1 Electronic nonlinearities	35
2.3.2 Charge carrier nonlinearities	36
2.3.3 Thermal nonlinearities	37
2.3.4 Nonlinear perturbation of microring response	37
2.4 Integrated photonics applied to ANNs	40
2.4.1 Weighted sum of inputs	40
2.4.2 Nonlinear Activation Function	40
2.4.3 Simulations	41
3 Samples, setup and experiments	45
3.1 The samples	45
3.1.1 Initial characterization	46
3.2 Characterization of the Activation Function	52

3.2.1	Bistability wavelength dependence	53
3.3	Optical Activation Function	55
3.3.1	Optical bistability as nonlinear activation function	55
3.3.2	Optical Nonlinear Response in ANN	57
Conclusions		59
Bibliography		61
Acknowledgements		65

Acronyms

AI	Artificial Intelligence	MAC	Multiply-and-Accumulate
ADF	All-Drop Filter	MLP	Multi-Layer Perceptron
ANN	Artificial Neural Network	MSE	Mean-Square Error
APF	All-Pass Filter	MZI	Mach-Zehnder Interferometer
AWG	Array Waveguide Grating	OSA	Optical Spectrum Analyzer
CMOS	Complementary Metal-Oxide-Semiconductor	PD	photo-detector
CNN	Convolutional Neural Network	PIC	Photonics Integrated Circuit
CPU	Central Processing Unit	ReLU	Rectified Linear Unit
CEL	Cross-Entropy Loss	RC	Reservoir Computing
DOF	Degree Of Freedom	RNN	Recurrent Neural Network
EM	Electro-Magnetic	SGD	Stochastic Gradiend Descent
EYDFA	Erbium-Ytterbiym-Doped Fiber Amplifier	SOI	Silicon-on-Insulator
FFNN	Feed Forward Neural Network	TE	Tranverse Electric
FCA	Free Carrier Absorption	TEM	Transverse Electro-Magnetic
FCD	Free Carrier Dispersion	TIR	Total Internal Reflection
FEM	Finite Element Method	TM	Tranverse Magnetic
FSR	Free Spectral Range	TOC	Thermo-Optic Coefficient
FWHM	Full-Width-Half-Maximum	TOE	Thermo-Optic Effect
GPU	Graphic Processing Unit	TPA	Two Photon Absorption
HBP	Human Brain Project	XOR	exclusive-or
ICT	Information and Communication Technology	VOA	Variable Optical Attenuator
		WG	waveguide

Preface

The research field

Research in the field of Artificial Neural Networks (ANNs) has became more and more popular in the last few decades as the computing power available to the average institution increased further. As a matter of fact, in the last two decades the calculation power which once belonged only to bulky, power hungry, and very expensive supercomputers, gradually became possible also for relatively more compact, efficient, and definitely cheaper servers and workstations.

This trend has been dictated by the fact that artificial neural networks are usually simulated on conventional computers, which are on the other hand based on the von Neumann, or Princeton, architecture. Implementations of this general purpose architecture are able to do any logical computation with a series of instructions. However the downside of this suitability to generic computations is that it is often difficult to parallelize single operations and therefore certain tasks are inherently inefficient, both in energy and in time. Moreover another problem is that it requires explicit programming for each singular task.

The formalization of the von Neumann (or Princeton) architecture dates back to 1945 [1] and it is as old as the first attempt in artificial neural networks [2]. Shortly after, research and industry efforts focused only on the development of the Princeton architecture, choosing it de facto as the primary design for computing devices.

In the recent years, help came from acceleration units known as Graphic Processing Units (GPUs), which were developed for a completely different objective. These GPUs, which are nowadays being designed specifically to accelerate calculations of artificial neural network simulations, are able to carry out a restricted set of instruction compared to CPUs, but are much more efficient both in power and in time, because of their parallel execution [3], [4].

Todays software implementations

Between the most famous examples of simulated neural network there certainly are the older Watson supercomputer from IBM and the more recent AlphaGo artificial intelligence. Both of them arose to popularity because they defeated us humans to our own games. Specifically the effort of Watson was aimed at prevailing the most strongest human contestants at *Jeopardy!*, an American television game, and succeeded in 2011. Watson was powered by a IBM supercomputer and used 0.22 MW of power. AlphaGO, nevertheless, is a deep neural network that in 2016, through reinforced learning algorithms, defeated Lee Sedol, considered the most formidable master of Go, a complicated board game [5].

Another interesting case is given by SpiNNaker, of the University of Manchester. It is composed by 10^4 neurons connected by more than 10^7 synapses, and is simulated on over 65 thousand 18-core ARM processors connected together. Unlike the other two, its main goals are to simulate several neural mechanisms, such as the operation of visual cortex. The SpiNNaker system is part of the Human Brain Project (HBP), funded by the European Union [5], [6].

Similarly to SpiNNaker, another effort in gaining better understanding of the human brain is *The Blue Brain Project*, which is leaded by École Polytechnique Fédérale de Lausanne in collaboration with over one hundred other international research institutions and it is funded by the European Union. This project is simulated on IBM Blue Gene supercomputers, each

fitted with more than 100 thousand processors and consuming several MW of power [5].

Present objectives and future trends

We can distinguish two principal objectives in this research field. The first one is a technological objective and is that of developing a more powerful computational instrument. The second objective is instead a topic of more fundamental research: to use the tools of artificial neural networks to better understand biological neural networks such as our brain [6], [7].

If the comprehension of the complex mechanism at the base of our brain has always been a very arduous assignment, the tasks that artificial neural network are facing become increasingly difficult while time passes. Whereas in the past years public and private research groups trained themselves with “easier” problems, such as television or board games, today they are concentrating their efforts on more challenging goals. Probably the most clear example of this tendency is given by the research on autonomous driving, where artificial neural networks are used primarily for real-time processing of data from several sensors [8].

A great number of businesses formed in recent years around these problems and many other will probably do the same for similar areas. It is very likely that to achieve their objectives, even more powerful and complex artificial networks will be required. Therefore more processors will be developed and put together to achieve impressive computing power. However, one cannot expect to increase indefinitely the power, without significant improvements in efficiency. Hence alternatives are sought [9].

Overview of current alternatives to simulations

Apart from the vast majority of the research on neural networks, which is being carried out with simulations on powerful processors, a smaller portion of research is focused on building neural network physically. Great efforts are made nowadays by many research institutions and companies to develop original computing architectures that allow artificial neural networks to run physically, instead of being simulated on conventional computers [9]. Figure 1 shows the increase of interest in the scientific community for the development of new architectures based on neural networks.

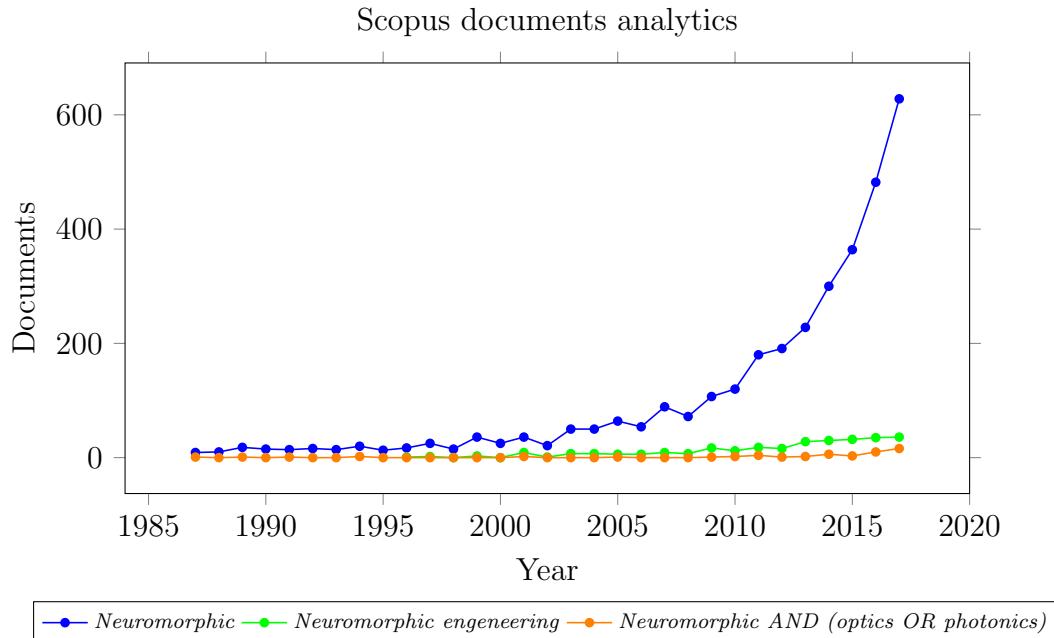


Figure 1: Scopus data on ‘neuromorphic’ vs ‘neuromorphic photonics’

The reason why considerable improvements can be expected is that the path of designing

hardware devices ad hoc for neural network computation has not yet been covered comprehensively. Majors performance or efficiency enhancements are coming from the fact that those new architectures are conceived with parallel execution of specific operations in mind. This, however, makes them inevitably less adaptable to changes, unlike software simulations on conventional computers, which can be modified just by changing few lines of code. A good example is given by size of the network, which in simulation is easily altered, even of orders of magnitude, while in hardware implementations it depends strictly by the architecture features and scalability.

Many of these use the electronics framework, mainly because of the many benefits of the mature silicon technology and its CMOS-compatible production chain. Specifically footprint and efficiency of CMOS devices has not yet stopped to grow since its discovery, in the past century. However, a relatively small portion of these works also inquired simpler network fabrication with organic electronic materials, achieving remarkable results [5].

Among the most interesting projects which uses electronics there are TrueNorth of IBM, Neurogrid of Stanford University, and BrainScaleS designed in Hidelberg and Dresden. The former is part of a long-term research project funded by DARPA and has the aim of building an artificial neural network with the capabilities of brains of small mammals, such as cats or mice. In 2014 IBM showed a chip, named TrueNorth, containing 1 million artificial neurons and 256 millions synapses. This network, relying on more than 5 billion transistors organized on an area of 4.3 cm^2 , consumed only 60 mW. The researcher also proved that connecting 48 of this chips together they obtained the equivalent of the brain of a mouse [5].

Similarly, the intention behind Neurogrid chip is to explore numerous hypothesis concerning mammalian brains, specifically about the inner mechanisms of operation of the cerebral cortex. Stanford's implementation reaches the line of 1 million neurons, while only requiring 5 W of power consumption. Moreover the communication lines between single neurons, the synapses, are provided by FPGAs and banks of SRAM [5].

Finally the BrainScaleS is built by 20 silicon wafers, each containing 200,000 biologically realistic neurons and 50×10^6 plastic synapses. Likewise SpiNNaker is funded by the European Union and it is part of the Human Brain Project. A particularity of this system is that does not execute pre-programmed code, but evolves according to the physical properties of the devices themselves [6].

Neuromorphic photonics

Besides electronics, other kind of physical implementations have been attempted from research teams. Probably the most intriguing example is given by integrated photonics. Photonics Integrated Circuits (PICs) have never been competitive in comparison to electronics in the development of von Neumann like processors, which are based on boolean logic gates. No photonic integrated logic of comparable performance to electronics has been proposed yet. However, by implementing a computing architecture that adheres to the artificial neural network rules integrated photonics might emerge as a successful platform [10]. Even if research in neuromorphic photonics is just in its embryonic state, many progresses have been accomplished lately. The aim of neuromorphic photonics is to match electronics performances by overcoming it in power consumption, computational speed, and intrinsic suitability to parallel computing (see Table 1).

Since PIC manufacturing is not as developed as fabrication of electronic devices, integrated photonics research on artificial neural networks concentrates on integrating single optical devices rather than building a full-grown all-optical network. The great majority of researches integrates optical structures and electronic ones on the same network, building hybrid optoelectronic artificial neural networks. Moreover, the networks implemented mostly are of recursive type or trained with the reservoir computing method. Both of them allow to build functioning ANNs with a low number of (controllable) computing nodes.

Chip	MAC rate	Energy per MAC (pJ)	Area per MAC (μm^2)
Photonic hybrid III-V/Si	20 GHz	1.3	205
Sub- λ photonics (future trend)	200 GHz	0.0007	20
HICANN	22.4 MHz	198.4	780
TrueNorth	2.5 kHz	0.27	4.9
Neurogrid	40.1 kHz	119	7.1
SpiNNaker	3.2 kHz	6×10^5	217

Table 1: Comparative table on current and future trends in electronic and photonic micro-architectures for neuromorphic computing. MAC stands for *Multiply and ACCU-mulate* operation, a standard computational step. Taken from [10].

My work

My work in this thesis has two related objectives. The primary aim is to implement a proof of concept for a fundamental component of artificial neural networks: the neuron’s activation function. I plan to do so by employing the framework made available by integrated photonics. The second intent is to bring together two until now distant fields of research: photonics and machine learning. In fact generating knowledge and tools to be used in common by the two research fields is at least as much important as the primary objective.

The idea at the base of our physical implementation is to employ integrated optical devices, which have been manufactured for other projects, to create a proof of concept for the activation function. Therefore in this work I will study a particular device, a microring resonator, and characterize its response to different inputs, for several working conditions. This device can process information because it has, under certain circumstances, a nonlinear response function to its inputs. Moreover, the same device might be used in another important part of artificial neural network nodes: the weighted sum. With this two parts together one could claim to have built a rudimentary artificial neuromorphic network, the *perceptron* (see Subsection 1.4.1). However, since integrated structures able to carry out a weighted sum of optical signals have already been discussed in literature and it requires less effort to produce acceptable results, it will not be at the center of this study.

During the thesis I achieved two main results. I have demonstrated the use of the nonlinear response of a microring optical cavity to produce an activation function by exploiting the optical bistability regime. In addition, with the aim of testing such response function, I simulated an artificial network with custom activation function by using standard software libraries used by Information and Communication Technology (ICT) community to model neural networks.

Structure of the thesis

The first chapter of this thesis is intended as an introduction to the world of artificial neural networks. After a brief historical summary, I describe the main aspects of neural networks and their working principles. Then I focus on the description of the simplest type of artificial networks, the feedforward network, and finally an overview on other kind of networks. Later, the chapter is concluded with an introductory description of the Python language and the PyTorch library, used in this work for artificial network simulations.

The second chapter introduces the physics which I will study. Again, after a short explanation on the basic devices used in the field, I will describe in depth the device studied for the task. Having clarified the physics underneath, I ought to explain how I am going to use this physics to implement (part of) an artificial neural network. I will address these topics both from the theoretical point of view and from the numerical one, with simulation of the specific (simplified) system.

The third chapter is dedicated to the description of the experimental work. At the beginning

the problem of the selection of the sample is discussed, along with the definition of the device chosen. After, I characterize the different behaviors with detailed data. Finally, an explanation on how the device will be used in the neural network viewpoint and the corresponding tests of operation.

Chapter 1

Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational systems which elaborate information in a way that is loosely inspired by the operation of biological neural networks (brains). Biological networks are superior in performance to computers and extremely more efficient in difficult tasks such as classification (e.g. image recognition) and prediction (e.g. pattern recognition). The underlying idea is to copy some of their mechanisms and exploit them for computational applications. These systems are: intrinsically parallel in operation, able to modify their behavior through a learning process, and sharing similar dynamics to solve different tasks.

Topologically, ANNs are made of collection of interconnected nodes, each one of them elaborating information. ANNs can be either simulated on computers or physically built on hardware designed ad hoc. At this time, the improvement of software implemented ANN is withheld by limitation in computing power and efficiency. Training a complex artificial neural network with computers at the state of art might take even weeks. With the aim of improving operating performance from simulated networks, research on hardware architectures is surging: digital, analog, electrical, and optical devices are being developed.

1.1 History

It is widely acknowledged that the opening work of this research field was made in 1943 by Warren McCulloch and Walter Pitts, a neurophysiologist and a mathematician respectively. In their research work, they described the operating mechanism of biological neurons by modeling a simple electronic circuit [11].

The following important work was made by Donald O. Hebb, who hypothesized the neural plasticity in 1949 [12]. He pointed out the fact that neural pathways are strengthened each time they are used, a concept known as *Hebbian learning*.

During the late 1950s and the early 1960s, as computers became more powerful, many promising works were published. Some simulated operation of artificial networks on calculators, e.g. Farley and Clark [13] and Rochester [14]. Other produced circuitry that implemented on hardware such networks, e.g. Rosenblatt [2], [15]. However, despite the early successes of neural networks, the traditional computing architecture (von Neumann architecture) was chosen as the preferred computing architecture.

The reason why this happened, probably, was due to several concurring facts. First of all, it was believed that more complex networks such as multiple- or deep-layered networks were impossible to train due to the vanishing gradient problem. Moreover, in the same time period (1969) a research paper by Minksi and Papert [16] identified an important problem: the basic perceptron was not able to execute the exclusive-or (XOR) operation, unlike the logical circuits at the base of von Neumann architecture.

In addition to this research, the early successes of some works on neural networks pointed to

an overestimation of the artificial neural networks potential, also held back by the technological capacity of the time. Finally, important questions of philosophical nature came to light, such as the fear fueled debate on the impact on our society of a class of computers able to think. This very controversy, i.e. the Artificial Intelligence (AI) problem, is discussed still today [17].

During the 1980s the interest for this computing method was reinvigorated by a large number of works suggesting methods to train multi-layered networks by distributing pattern recognition errors through all the layers in the network. This method is now called *backpropagation*.

In today's research and technology, artificial neural networks are used in numerous applications: from the most serious ones, such as medical image classification [18], to the lighter ones, like photo enhancement [19] and generation of polyphonic music [20], [21].

1.2 Basis of Neural Networks

A neural network is a collection of processing elements, or nodes, interconnected with arbitrary topology. From its input nodes, the network accepts information, which distribute across inner nodes through the interconnections and will get elaborated at each node. A layer of output nodes reads to outcome of the inner (hidden) nodes computation and transfer it to the outside. Inner nodes are also called hidden, because they are not meant to be accessible to the external world. A generic scheme of such network is shown in Figure 1.1.

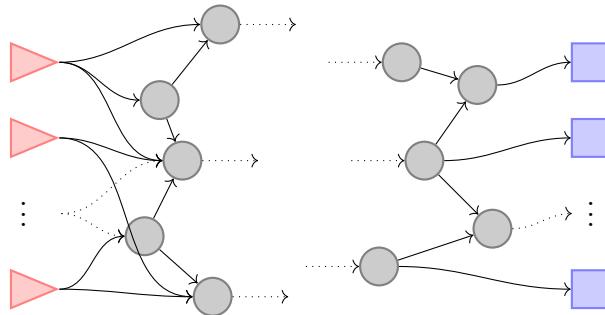


Figure 1.1: Generic scheme of a neural network. Triangles (red) are input nodes, circles (grey) are inner nodes, and squares (blue) are output nodes. Interconnections among nodes are represented by arrows.

Nodes can all implement the same function or behave differently, depending on the type. The operation of nodes resembles that of neurons: various input gets collected and elaborated together to obtain an output, which will become one of the many inputs for subsequent neurons/nodes. Specifically, the most used model for neurons is the McCulloch–Pitts (MCP) neuron. It is divided into two parts, as shown in Figure 1.2: the first part is a weighted sum of the inputs, while the second part is given by the so called activation function.

The node is described mathematically by equation (1.1)

$$y = f_a \left(w_0 + \sum_{i=1}^n w_i x_i \right), \quad (1.1)$$

where f_a is the activations function, evaluated on the sum of the input x_i weighted with w_i , plus a bias w_0 .

Each node accepts values at its inputs and produces an output accordingly. However, in addition to the input, the output depends also on the node's parameters: the weights and the bias (see Section 1.3). Generally these parameters are modified during the training of the network, but not while it is operating.

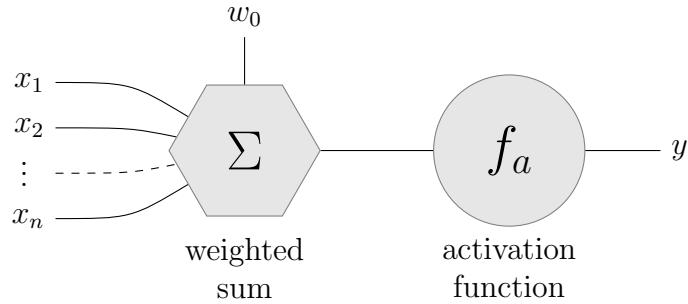


Figure 1.2: Generic node representation. x -values are inputs, y -values are outputs, w_0 is the bias.

Moreover it is mandatory for the activation function $f_a(\cdot)$ to be nonlinear, because otherwise a collection of nodes will result in just a weighted sum of its inputs. Nonlinear activation functions are distinguished for their properties, such as continuity, limitedness, and existence of the derivative. Two examples are shown below in Figure 1.3.

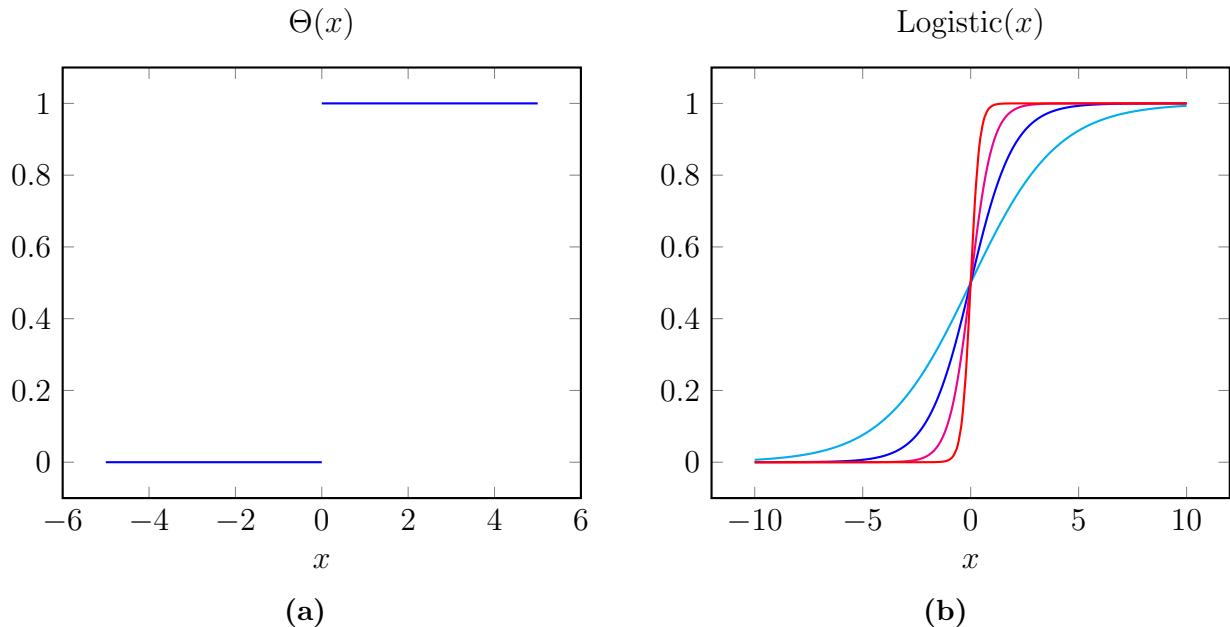


Figure 1.3: Examples of activation function: (1.3a) is the step function, or Heaviside Θ , (1.3b) depicts few Logistic functions.

Standard Representation

In the thesis I depict a generic neuromorphic network in a slightly different way compared to ICT literature. Usually weights are represented on the connection between the nodes, which are then designated to apply only the activation function. Moreover the input layer is linear as it feeds the inner nodes with the input data, while the output layer is actually given by the last nonlinear layer of the inner nodes. A generic network is shown in Figure 1.4a.

I consider the inner nodes as the only place where any kind of elaboration on the data happens. Inner nodes have a number of inputs, which are weighted and summed together to be entered as argument in the activation function. This leads to a natural separation between input/output nodes, which acquire the task of providing data from/to the outside, and inner nodes, which is where the weighted sum and/or the activation function are carried out.

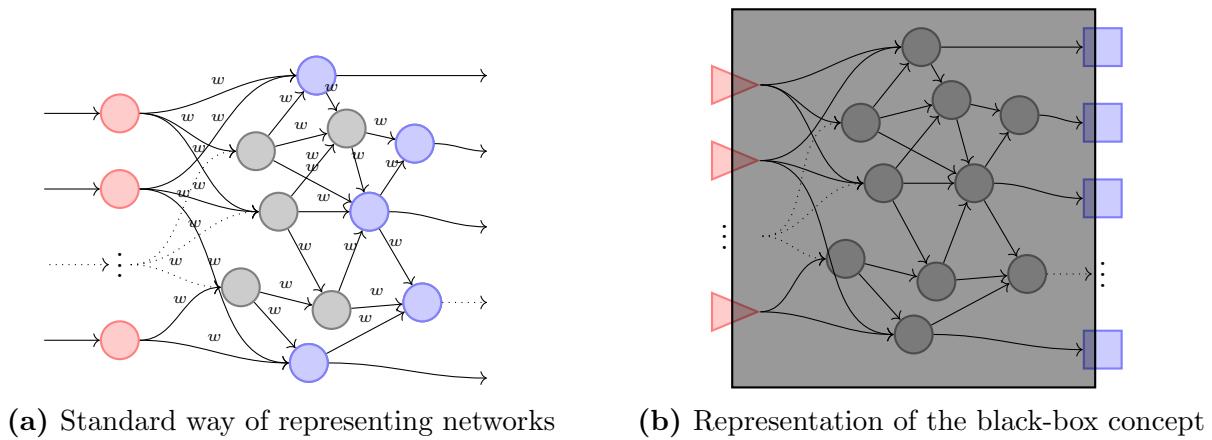


Figure 1.4

This non-standard description is consistent with the idea of functional *black box*, in which input and output are the only visible nodes, while the other are hidden inside, as shown in Figure 1.4b.

1.2.1 Applications of ANNs

Conventional computers are extremely fast and efficient in executing simple algebraic tasks and they can perform more complex activities if provided with the correct series of instructions. Hence algorithms can solve arbitrary difficult problems, provided that the necessary steps are known.

Artificial Neural Networks take a different path in the solution of a given problem. In fact, when ANNs are designed to solve a problem, they learn from the examples that are provided to them. ANNs cannot be programmed, instead they exploit their internal complexity to minimize the error and hence automatically solve the problem. After a network has been prepared, it is able to operate much faster than the complex algorithms on conventional computers.

ANNs have proven to be extremely good at recognizing patterns, which can be used to solve problems of several classes. For example, classification problems are solved by the recognition of attributes of the input element. Likewise, clustering is obtained when a sequence of examples is grouped according to their features. Moreover, regression analysis and time series prediction can be obtained, when a sequence of data is fed to the network.

1.3 Working Principles of ANNs

ANNs are implemented with different topologies in order to solve different problems. However, the behavior of a single network changes depending on the value of the weights applied at each node. For this reason, before a neural network is considered ready to perform a task, it must go through a preparation phase, or training, which it is usually composed by three processes: learning, validation, and test. Each one of these procedure is meant to prepare the network to work as required from the designer.

1.3.1 Learning Process

During the learning process the neural network is run on a set of inputs x , each paired with the correct answer y , or target, in a second set of data. This procedure is called *supervised learning*. The neural network will produce at the output a third set \hat{y} , which should be as close as possible to the correct answers, when the network works properly. Typically weights

among nodes are initialized randomly and the distance of the network outcome from the target function is measured through a loss function. Then weights are modified in order to minimize the loss function. Every time the network is trained on the full dataset, an *epoch* is completed.

Loss function

The loss function $L(y, \hat{y})$ evaluates the difference between the predicted and the correct answer. Usually, this quantity is linked to the geometrical distance between the predicted output and the target $|\hat{y} - y|$.

The most common loss function is the *Mean-Square Error* (MSE). Assuming to have an input set of N examples paired with the same number of targets, and that the outputs and the targets are composed by C values, or classes, the function becomes:

$$L(y, \hat{y}) = f_{MSE}(y, \hat{y}) = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^C (\hat{y}_{n,i} - y_{n,i})^2, \quad (1.2)$$

where each example in the set is subtracted to its target and then squared. Finally the mean of all squares gives the expected result.

Another commonly used function is the *Cross-Entropy Loss* (CEL) (also known as negative log likelihood),

$$L(y, \hat{y}) = f_{CEL}(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^C y_{n,i} \log(\hat{y}_{n,i}), \quad (1.3)$$

which expects positive values at the input. Hence the error $-y \log(\hat{y})$, quantified for each element in each example, is always a positive number. The mean over all examples in the set returns the results.

Depending on the class of the problem, different loss functions are chosen. In fact, since the loss function drives the weights update process, it is important to choose the correct one. For example, in tasks that expect outputs to take continuous values, the MSE is favored. On the other hand, in classification tasks, where the answer of each output node is boolean, the preferred function is CEL.

Weights Update Process

The weights update process is a difficult task, and probably the most computationally expensive one in running a neural network. There is a variety of methods to chose from, depending on the type of ANN and on the resources available.

A widely used algorithm is the gradient descent, from its most simple version to more complex variations such as *Stochastic Gradient Descent* (SGD). This method updates the weights by subtracting a value proportional to the gradient of the loss function in respect to the weights themselves times a positive factor called *learning rate*, as shown below.

$$w_i|_{n+1} = w_i|_n - lr \cdot \frac{\partial L}{\partial w_i|_n} \quad (1.4)$$

where $w_i|_n$ are the current weights, lr is the learning rate, $\frac{\partial L}{\partial w_i|_n}$ is the first derivative of the loss function in respect to the i -th weight at the current step, and $w_i|_{n+1}$ are the updated weights. This method is equivalent to minimize the error on the loss function, by following the gradient $\nabla_w L$. This vector lives in the multidimensional space of the loss function $L : \mathbb{R}^W \mapsto \mathbb{R}$, where W is the total number of weights in the network.

The most used algorithm is called *backpropagation*, or *chain rule*: it computes the first derivative of the loss function L in respect to all the parameters of the network, the weights, starting from the end of the ANN and going backward toward the input, hence its name. Since the number of connections between nodes might be of order of magnitude greater than the number of nodes, it is easy to understand how large networks are computationally expensive to train.

In early days, this algorithm was used together with the Logistic activation function. Due to the saturating behavior of the function, the gradient is repeatedly multiplied by small values at each layer. Hence, the effect of backpropagation becomes negligible for the first layer on deep networks. This problem is known as *vanishing gradient* problem. Thanks to the introduction of non saturating activation functions, the training of deep neural network has become gradually possible [22].

1.3.2 Validation Process

The validation process is carried out at the same time of the learning process and consists on testing the neural network on a new set of examples. Differently from the learning process, during validation the output predicted by the network are compared to the examples, but the weights are not updated. Instead, the loss function is used as a control parameter to prevent overfitting.

Overfitting is the phenomenon in which a neural network recognizes features that are specific of the sampled data instead of the more general ones. It occurs when the network is trained over and over on the same set of data, and it starts to learn details which characterize the test dataset. This typically worsens the ANN prediction capability on other datasets. Therefore, by periodically testing the network on a novel set, i.e. the validation set, one can avoid the problem.

The validation process happens repeatedly throughout the learning process. The learning process is stopped at a minimum of the error on the validation set. Figure 1.5 shows an example of the evolution of the loss function on the three dataset as function of the epochs.

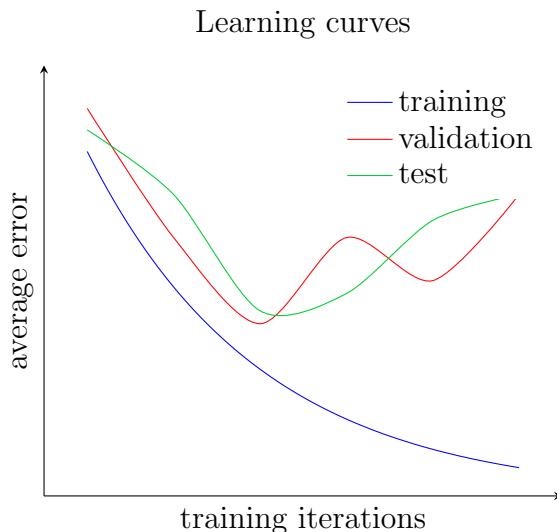


Figure 1.5: The learning curves are the value of the loss function as a function of the training iterations. Both the validation and the test error are usually larger than the errors on the training set. Sometimes, to avoid overfitting, the learning procedure is stopped at a minimum of the validation learning curve [23].

1.3.3 Testing Process

At the end of the learning and validation processes, the ANN have to be tested on never seen dataset: the test data. Predicted outputs are compared to the correct answers and an overall value of the correctness is evaluated, often expressed in percentage.

1.3.4 Datasets

The training process of an ANN is heavily conditioned by the characteristics of the dataset. For example, the number of attributes of the examples in the dataset defines the number of input nodes and similarly the features of the answers define the output nodes. This is a constraint on the network size, which obviously affects the training. Moreover, the size of the dataset influences the length of the training and the presence of effects such as overfitting. For instance, smaller datasets are more prone to overfitting than larger datasets. Hence, the collection of the data and the choice of its attributes are problems that might account for a large part of the time in the development of an ANN.

In addition, considering a given dataset, there is also the problem of defining the subset used for learning, validation, and testing. The most naive division is in three equal parts, since three are the phases of preparation for any ANN. However, as shown later in Subsection 1.6.2, when the resources dictate otherwise, other subdivisions can be implemented. In my case, the decision was to follow the suggestions of the authors of the dataset and divide the examples into a learning set and a test set only, with a ratio between the two close to 50 %.

1.4 Feed Forward ANN

The first and most simple network type is the Feed Forward Neural Network (FFNN). In this kind of ANN, nodes are divided into groups called *layers*. A layer is a collection of nodes that accepts inputs from a preceding group and generate as many outputs as the number of nodes in the layer. Each layer of a FFNN is connected in series with the others, except for the input layer at the beginning and the output layer at the end. As for the single nodes, the inner layer are called hidden, because usually not accessible.

The information travels from the input to the output and gets elaborated from each hidden layer: there are neither connection between nodes of the same layer, nor loops or feedback between layers. The number of hidden layers and the number of nodes they contain depends on the network size. Moreover the connection between the layers might be complete, i.e. each node in the layer accepts each input of the preceding layer, in that case the layer is said to be *fully connected*, or sparse as in the case of convolutional layers (see Section 1.4.2).

1.4.1 Perceptron

The most naive topology of a FFNN is given by the so called *Perceptron*. The Perceptron dates back to the 1957, when the homonym *Perceptron algorithm* was software implemented by Frank Rosenblatt on a computer (IBM 704) and subsequently in hardware as the *Mark 1 perceptron* [2], [15]. The graph of a generic (single layer) perceptron ANN is shown in Figure 1.6.

By adding more than one hidden perceptron layer to the neural network, one obtain the so called *Multi-Layer Perceptron* (MLP). This allows for more computational complexity, e.g. MLP can solve the XOR problem, whereas single layer perceptron cannot ???. When the number of hidden layers is more than two, the network is called *deep*. A deep MLP is shown in Figure 1.7.

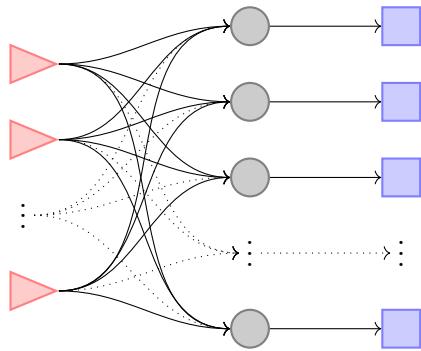


Figure 1.6: Perceptron type neural network: in this representation the perceptron has n inputs and m outputs as well as a hidden layer with m nodes.

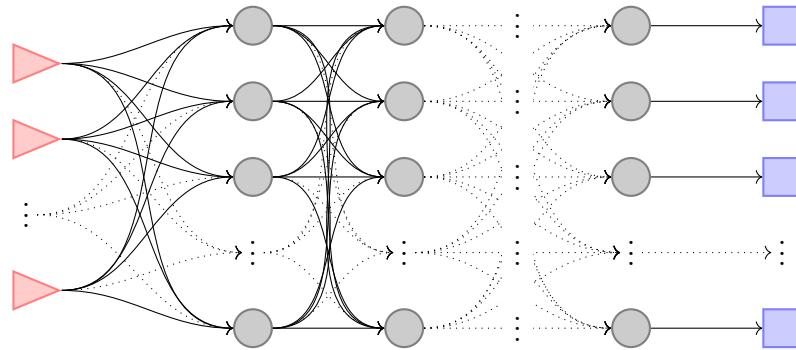


Figure 1.7: Deep Multi-Layer Perceptron (MLP), fully connected.

MLPs might show heavily different topology, i.e. each layer could have a different number of nodes, however often the layers at the beginning are wider than the layer at the end of the network (see following section). Moreover, in literature with the term perceptron one almost always refers to fully connected FFNN ??.

1.4.2 Other Feed Forward ANNs

FFNNs are a large family that includes many other types besides the perceptron one. A few names are autoencoder, time delay, and convolutional neural networks. Autoencoders ANNs are FFNNs with the same number of input and output nodes, whose purpose is to reconstruct its own inputs. For this reason autoencoders employ unsupervised learning. Time delay networks have a FFNN structure and their purpose is to analyze sequence of signals in search of specific patterns.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are inspired to the visual cortex, in which neurons are not fully connected all the inputs but only to a restricted region. Convolutional Neural Networks are a type of FFNN conceived to recognize images without being misled by distortions such as translation, skewing, or scaling. Its input is often represented as a 2D matrix, instead of a 1D vector. This kind of network is usually composed by many layers: the most recurring is prevalent the convolutional one, but other types can be mixed together too.

A convolutional layer performs a two-dimensional convolution over the input matrix of a second 2D matrix of weights, called *feature map*. Thus, each node of the layer operates on a restricted region to understand if a feature is present or not. The operating regions are

commonly overlapping and the feature map is shared among the nodes in the same layer. Often several convolutional layers operate in parallel to extract different features, hence they use different feature maps. Due to the specific operation of convolutional layers, the number of nodes per layer usually decreases with each layer. Figure 1.8 shows the possible scheme of a convolutional layer, operating a 3×3 convolution on the input.

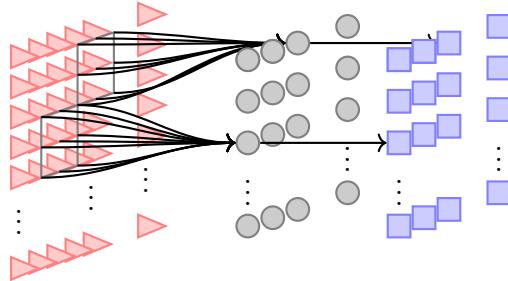


Figure 1.8: Pictorial representation of a convolutional layer. Each node acts on a restricted region of the inputs, in this example a 3×3 region, to extract a feature. Several convolutional layers might be placed side by side to extract different features from the same input.

CNNs are nowadays widely used in image recognition with outstanding results and they improve with a steady pace. A technological application of this kind of network is the real time recognition of obstacles in a vehicle path, for safety and automation purposes in the automotive industry.

1.5 Other Types of ANNs

By changing the topology of the nodes distribution and their connections, one obtains other networks that cannot be catalogued under the class of FFNN.

1.5.1 Recurrent ANN

Recurrent Neural Networks (RNNs) are a kind of network in which a portion of the input of nodes depends on the (past) output of the same nodes or nodes of subsequent layers. That is, information does not propagate only forward like in the FFNN, but can propagate also backward, for example in loops or in feedbacks. As a result, the time-dependent dynamic of the node is usually very important in the study of RNNs. Figure 1.9 shows a recurrent node, in which the output is connected to the input of the node.

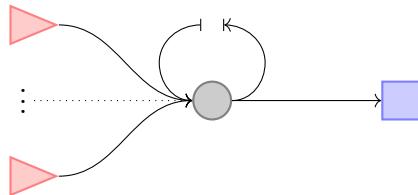


Figure 1.9: Representation of a recurrent node. One of the inputs is given by the output itself. This output to input connection could be mediated by a delay device, so that for example the output at $t - 1$ becomes the input at t . Depending on the structure of the network, there might be recurrent nodes and/or recurrent groups of nodes, i.e. loops.

RNNs have found large use in time series analysis and prediction [23]. However RNNs are occasionally employed to obtain the same classification of FFNN. In this case, the former

are equivalent to the latter, if “unfolded” in time. That is the expansion of the recurrent architecture in time is the same as the topology of the FFNN one in space.

1.5.2 Reservoir Computing

Reservoir Computing (RC) describes a learning method that differs from the deep learning commonly used by MLP (FFNN) and RNN. It claims that it is not necessary to learn all the weights of the network, as in deep learning, but it is sufficient to train only the last (perceptron) layer of the network. RC was initially implemented due to the vanishing gradient problem, for which the last layer was the only one that showed significant change in the weights.

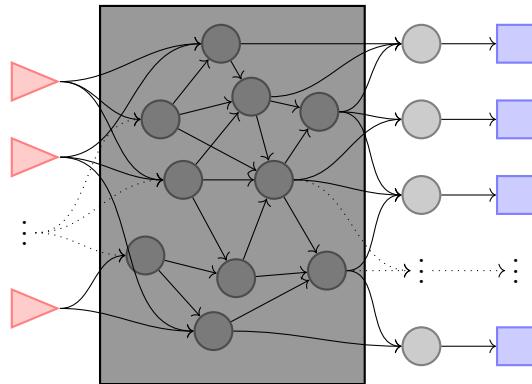


Figure 1.10: A reservoir NN is topologically equivalent to deep networks. However only the last layer is trained. In this picture the trained node are represented outside the box, while those inside are initialized with random weights which are then left unchanged.

Figure 1.10 shows a generic topology that could implement the RC method: only the nodes outside the darker region are trained, whereas those inside provide the computational power. This kind of network can be trained much faster than its respective deep-learning counterpart, i.e. deep FFNN and deep RNN. The question over which training method is more effective is still debated and literature does not provide clear answers yet.

Significant results for RC implemented with photonics have been obtained mainly with the RNN, therefore in this field RC is typically associated to the RNN topology ??.

1.6 FFNN Simulation

In the field of ANN simulation, there are several platforms that are independently developed. Among them there are TensorFlow, Theano, Caffe, Keras, Torch, and PyTorch. Unfortunately, each one of them have its strength and weakness, therefore the choice of the framework becomes very difficult.

To make a choice, the key factors considered were the language used, the features, the flexibility, and the level of diffusion of the framework in the machine learning community. The language used is a crucial factor, because it takes time to learn the peculiarities of a library written for a known language, but even more time to learn to use an unknown programming language. However also the features and the flexibility in the implementation are important traits, because any feature that is not natively implemented needs new coding. In turn, it is difficult to integrate new coding if the framework is not flexible enough to accommodate extensions or custom definitions. Last, but not least, it is easier to find support for a widespread framework in respect to a limitedly used one. In regards to this, Figure 1.11 shows the number of internet searches for some frameworks in the past few years.

In light of the facts above, PyTorch was chosen as framework. PyTorch is an open source machine learning library for python, which development has started only recently and it is based upon the older framework Torch [24]. It provides a high-level platform for the deep learning ecosystem and integrates acceleration libraries that allow fast and lean operation, both on common CPUs and GPUs. It is based on a backpropagation algorithm called *Reverse-mode auto-differentiation*, which allows versatile execution.

To summarize, the library was chosen for its language, its flexibility, and the growing interest for it within the machine learning community. Nevertheless, its features make it a very powerful framework.

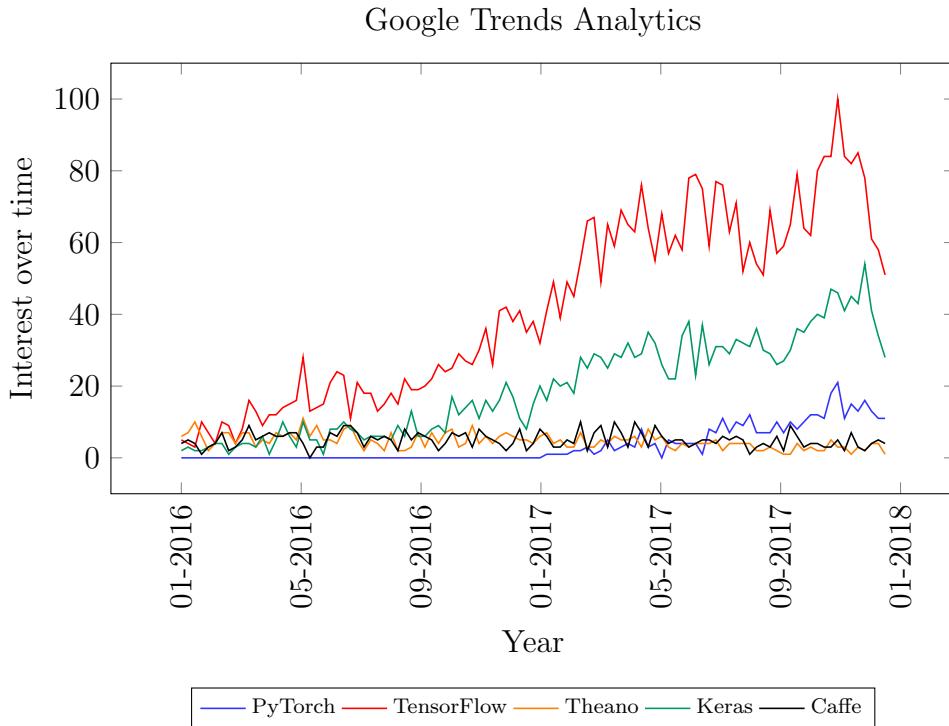


Figure 1.11: Google Search statistics for different keywords in the *machine learning and artificial intelligence* field. Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term.

1.6.1 PyTorch

PyTorch is a Python package which provides a powerful framework for deep learning. It is versatile in the sense that allow customizations at almost every level of operation. For the purposes of this work, a neural network implemented in PyTorch is composed by the definition of three main components and a few lines of code to implement the operation.

The most important part of the network is the so called *model*, which defines the topology of the network by setting parameters such as the number of nodes in each layer and the connections among them. Moreover, it defines also the activation function for each node, separately or in groups. The activation function can be coded from scratch, but the library provides several functions and allows to combine them easily together.

Another important part of the network is the so called *criterion*, which is nothing else than the loss function. The set of functions provided by PyTorch can accommodate almost any necessity. Moreover, their operation can usually be adjusted with some internal parameters.

Similarly to the criterion, the last piece of the system is given by the *optimizer*. Optimizers are a class of algorithms that provides the optimization of the weights during the learning phase. Most commonly used methods are already supported and they provide parameters to fine tune their execution.

I implemented a generic model of a fully connected Multi-Layer Perceptron (MLP), in which I could choose the overall structure of the network by setting some parameters to specific values (see Subsection 1.6.3). The optimizer chosen is called *Stochastic Gradient Descent* (SGD), which is parametrized just by the learning rate, at least in its vanilla implementation. It was then improved with the addition of the momentum and a learning rate scheduler. The former allows to decrease the sensitivity of local variation in the gradient, while the latter decreases the learning rate with the number of iterations. The criterion used was initially the MSE, however it was eventually replaced by the *Cross-Entropy Loss* (CEL), which is more suited to the classification task.

1.6.2 Dataset

Initial tests and trials on artificial neural networks are often made by using randomly generated sets of numbers. Exploiting the finite number of the examples, it is possible to train a network to successfully correlate the input to the output. However, once the system has been successfully set up, to compare different working parameters such as number of nodes and number of layers, a common standardized dataset should be employed. In this way, it is possible to set side by side different network implementations and analyze their performances.

In this preliminary work, the need to keep the system as simple as possible, drove my attention towards datasets of middle to small sizes. With the help of the *UC Irvine Machine Learning Repository* [25], I selected one with slightly less than a thousand entries. The dataset is called *Connectionist Bench (Vowel Recognition - Deterding Data) Data Set*. It is composed by 990 entries of 10 attributes, i.e. input values, grouped in 11 classes.

Each entry is a vector of 10 real numbers plus an integer number between 0 and 10. In view of the fact the activation function of the photonic node is a real-positive-valued function (see Section 3.2), I renormalized the database. The normalization is independently applied to each one of the attributes in such a way that is described by a real number in $[0, 1]$. The normalization is applied before the dataset is divided in the learning and testing sets.

The authors suggest to divide the dataset in a learning set of 528 entries and a testing set of 462 entries. Since the absolute number of entries is low, compared to other datasets, the validation set is not defined and the test set is used in its place.

1.6.3 Simulated ANN operation

The model that I defined allows to modify the number of input (D_{in}) and output nodes (D_{out}), the number of hidden layers (n_H), and the number of nodes in each hidden layer (D_H). Whereas the D_{in} and D_{out} is fixed by the problem, i.e. the dataset, the other are free parameters. Another degree of freedom that I implemented is whether the last hidden layer applies the nonlinear activation function or just the weighted sum, since I observed that in many examples of network the last hidden layer was linear. The scheme of the network is shown in Figure 1.12.

The choice of the structure of the neural network is purely heuristic, since up to now no satisfactory method to design the topology has been proposed in literature. However, given the problem (dataset), at least the input and output layer have a fixed amount of nodes. Specifically, the input nodes are 10, the number of attributes, while the output nodes are 11, the number of classes.

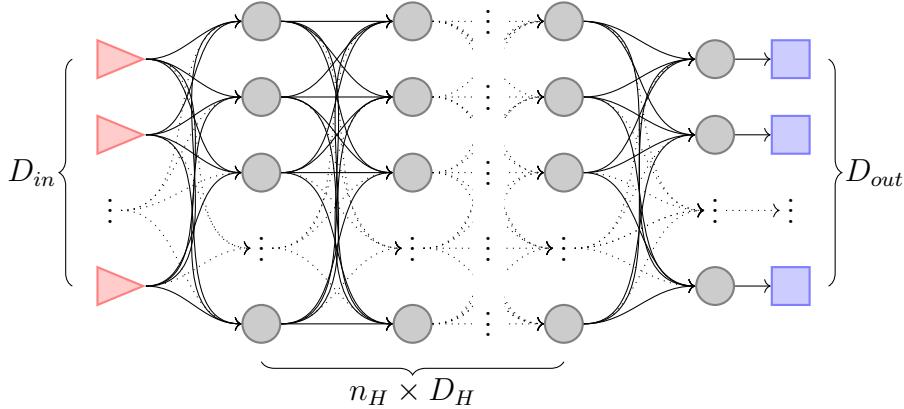


Figure 1.12: Topology of the model implemented in PyTorch. The number of input (D_{in}) and output nodes (D_{out}), as well as the number of nodes in the last hidden layer is determined by the dataset. The number of nodes (D_H) in the other hidden layers and the number of hidden layers itself (n_H) is governed by a parameter.

Using the chosen dataset, I tested the neural network operation on the model defined before, with two different kinds of activation function. Initially I implemented a *Rectified Linear Unit* (ReLU) and a *Logistic* (sigmoid) activation functions. The ReLU function is a standard activation function defined as follows:

$$f_{ReLU}(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x \geq 0 \end{cases}. \quad (1.5)$$

The Logistic function, which has already been shown in Figure 1.3b, is defined by

$$f_{Logistic}(x) = \frac{1}{1 + e^{-k(x-x_0)}}, \quad (1.6)$$

with the parameters $k \in \mathcal{R}^+$ and $x_0 \in \mathcal{R}$. Both functions are widely used in literature, however recently ReLU has become the preferred choice in the ICT community. Unfortunately, the results for the Logistic function were almost the same as a random choice. For this reason in the next section I will describe the results for the ReLU only. I will discuss in Section 3.3 the results of the same network model, with the nonlinear activation function given by the microring resonator.

Learning

After the definition of the model, I had to train the networks. All the examples in the dataset are fed at once into the model, then the overall loss is evaluated, and finally the weights are upgraded. This method is called *batch* operation; others are *online* operation, which evaluate the loss and change the weights at each example, and *minibatch* operation, which uses a random subset of the full dataset.

I provided the Stochastic Gradiend Descent (SGD) algorithm with a learning rates l_r and a momentum m , which should stabilize the evolution. Moreover, I used a scheduler to reduce the learning rate with the iterations. There are several algorithm for the learning rate scheduler: the one I used is called *StepLR* and decays the value of l_r by γ every T epochs. Specifically the learning rate is initially set to $l_r|_{ReLU} = 0.05$ and the momentum to $m = 0.25$, while $\gamma = 0.6$ and $T = 1000$.

I trained the models for 4000 epochs, in order to see the different dynamics of the training procedure: Figure 1.13 shows the typical evolution of the loss on the train and the validation datasets for each model. The network loss (solid lines) on the training dataset decays towards near zero and shows some particular behaviors. During the training, there are times at which the network loss does not improve much, *plateau*, and times at which it improves more rapidly. Moreover, some models show *ripples* during the evolution: this is due either to the learning rate being too high or to the complex shape of the loss function in the parameters space. Note that each time the learning rate scheduler changes l_r the ripples disappear.

The evolution of the loss function on the validation dataset is different. It shows a minimum usually in the second quarter of the total training period, which means that in the second half of the training the network is being overfitted on the data. In fact, I observed that usually the percentage of correct answers decreases slightly toward the end of the 4000 epochs due to overfitting.

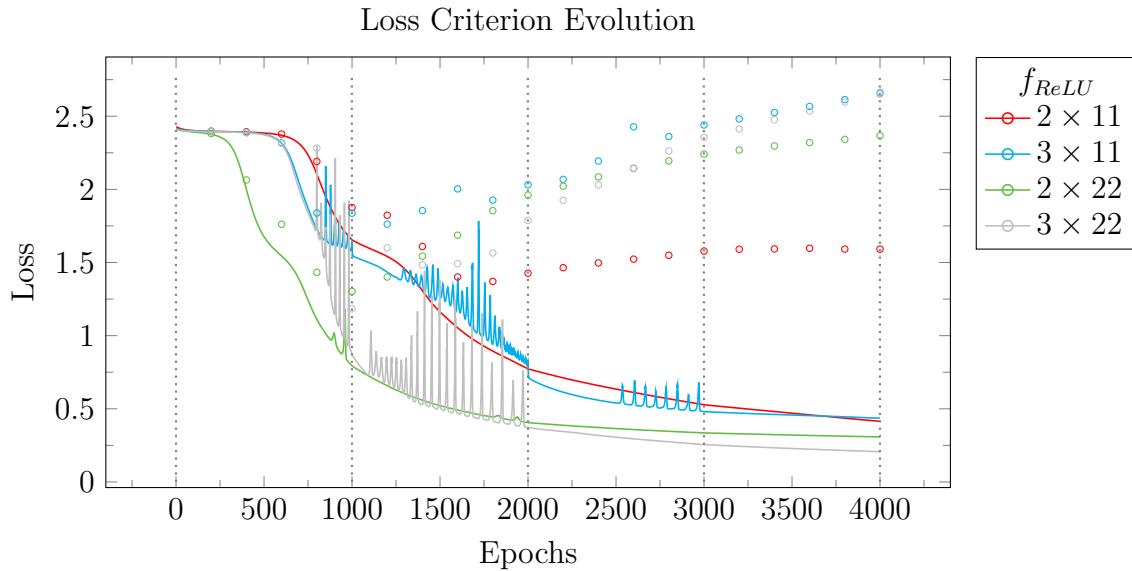


Figure 1.13: Typical evolution of the loss criterion (solid lines) throughout the epochs. Points represent the values of the loss criterion on the validation dataset, carried out repeatedly during the training. Dotted lines are the scheduled change in the learning rate.

Collection of results

Considering that the number of classes is 11, the probability of giving a correct answer at random is $p_{rnd} \sim 9.09\%$. Hence, values of correct answers above that mean that the neural network is working. On the other hand, values around or even below p_{rnd} mean that the current neural network is not working at all.

Since in 4000 epochs the ANNs are become overfitted on the dataset, I select as correctness the maximum value of correct answer given during the training. It is the same as stopping the learning process on the network when a minimum of the validation set is reached.

Table 1.1 collects all the results for the different topologies, where the activation function was provided by a f_{ReLU} function. We can observe that the width of the hidden layers seems to influence the correctness of the network, at least marginally. Moreover the depth of the network, i.e. the number of hidden layers (n_H), seems to be correlated to the deviation of the correctness.

The overall results of these models are satisfactory because they all well over the p_{rnd} threshold.

activation function	no. hidden layers	no. nodes per layer	last hidden layer	Percent correct
f_{ReLU}	2	11	linear	$(48 \pm 6) \%$
f_{ReLU}	3	11	linear	$(40 \pm 11) \%$
f_{ReLU}	2	22	linear	$(54 \pm 2) \%$
f_{ReLU}	3	22	linear	$(51 \pm 4) \%$

Table 1.1: Results of the several network topologies. The statistics is made on 10 different initializations of the same model: the error on the percentage is the standard deviation.

Chapter 2

Integrated Photonics

Photonics is the physical science which studies detection, manipulation, and emission of light. Specifically, integrated photonics is the branch that studies how to reduce and *integrate* macroscopic optical devices in miniaturized structures.

In the past few decades, thanks to the constant improvement of the manufacturing techniques, many productive problems have been progressively resolved. Moreover interest in the field is rising, driven by the growing needs of Information and Communication Technology (ICT), which was in turn following the increase in computational power of electronics. Many integrated devices have been proposed and some of them even commercialized [26].

2.1 Silicon Photonics

Silicon-on-Insulator (SOI) photonics is one of the widest branches of integrated photonics. It pursues the same objectives of all integrated photonics with the addition of following the production steps already developed for microelectronics.

Its framework is silicon, which is transparent in the same range of frequencies used by fiber optics. It is a well-known material, largely studied for microelectronics, and it allows relatively easy manufacturing of high grade structures. The bulk silicon is altered by a series of often numerous production steps. The first step is the burial of an insulator layer of SiO_2 , called *buffer*, which defines the bottom surface of the layer at the top. The pure *Si* volume at the top, the *device* layer, is where integrated optical structures will be built with various processes like photolithography, thermal oxidation, ion implantation, and etching. A schematic representation of the layers is shown in Figure 2.1.

The fact of relying on the SOI framework to build integrated devices is a strength of silicon photonics, because it enables researchers to develop CMOS compatible integrated optical structures. Complementary Metal-Oxide-Semiconductor (CMOS) compatibility is an industry standard created to fabricate microelectronic devices. Therefore by exploiting this standard all the know-how of the manufacturing industries behind the commercial microelectronic products is available to fabricate integrated photonics devices.

Moreover, by sharing the fabrication technology with microelectronic makes it easier to integrate electronic components and obtain hybrid integrated optoelectronic devices or compensate lack of appropriate optical structures.

Hence silicon photonics is a promising technology for a combination of cost and technological reasons. A list of advantages and disadvantages of silicon photonics compared to other integrated photonics technologies is provided in Table 2.1.

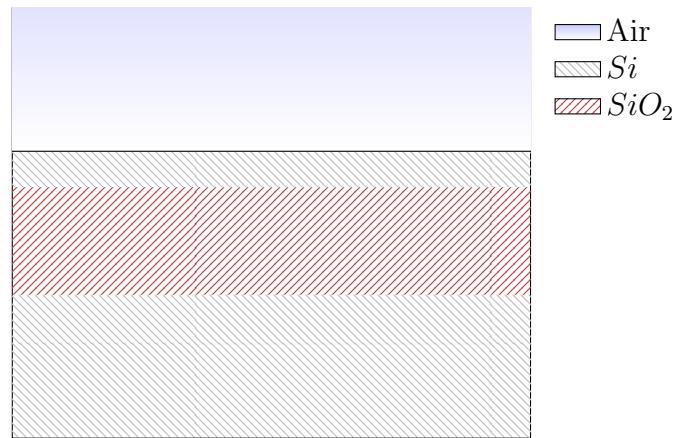


Figure 2.1: SOI productive framework. The buffer layer of SiO_2 is buried underneath the pure crystalline Si device layer which is used to build structures. The top silicon layer will become the volume in which light will be confined. The bottom silicon, underneath the buffer layer, is called substrate.

Advantages	Disadvantages
<ul style="list-style-type: none"> i. Stable, well-understood material ii. Stable native oxide available for cladding and electrical isolation iii. Relatively low-cost substrates iv. Optically transparent at important wavelengths of $1.3\ \mu m$ and $1.55\ \mu m$ v. Well-characterized processing vi. Highly confining optical technology vii. High refractive index means short devices viii. Micro-machining means V-grooves and an effective hybrid technology are possible ix. Semiconductor material offers the potential of optical and electronic integration x. High thermal conductivity means tolerance to high-power devices or to high packing density xi. Carrier injection means optical modulation is possible xii. Thermo-optic effect means a second possibility for optical modulation exists 	<ul style="list-style-type: none"> i. No Pockels effect ii. indirect bandgap means that native optical sources are not possible iii. High refractive index means that inherently short devices which are difficult to fabricate (e.g. gratings) iv. Modulation mechanisms tend to be relatively slow v. Thermal effects can be problematic for some optical circuits

Table 2.1: Advantages and disadvantages of silicon photonics over other integrated photonics technologies. Taken from [27].

2.2 Guided-wave photonics

The most important thing for integrated photonics is the way light is confined and manipulated into microscopic structures. While in conventional optics, light is delivered with bulky mirrors and lenses toward the desired position, in integrated photonics waveguides are the main device for this task. Moreover, they are the fundamental block for building more complex devices, whether they are passive, active, or hybrid optoelectronic devices. A few illustrative devices are the waveguide coupler, the Array Waveguide Grating (AWG), the electro-optic modulators, and integrated Mach-Zehnder Interferometer (MZI).

2.2.1 Waveguides

A waveguide is a path inside a medium, whose volume is defined by a certain number of interfaces between different materials, in which light remains confined and ideally travels with negligible losses. This volume is called *core* of the waveguide, while the medium external to it, if present, is called *cladding*.

One can distinguish two types of waveguides: metallic and dielectric. The former are based on the reflection of the electromagnetic field by the metallic surface. Light is confined in the desired path by a series of metallic mirrors. However, this mechanism works well only for electromagnetic radiation for which metals can still be considered *perfect metals*. For visible and infrared frequencies integrated metallic waveguides are not possible, due to too high absorption of the metals at optical frequencies.

On the other hand, dielectric waveguides are based on the phenomenon of Total Internal Reflection (TIR) at the interface between two dielectric media. To produce TIR, the two materials must have sufficiently different real part of the refractive index. Moreover, they must also be transparent, i.e. they must have low imaginary part of the refractive index, in the required range of frequencies. Hence not all media are suitable to build dielectric waveguides that work at a specific wavelength. For example, silicon is known to be transparent in the near infrared region.

Dielectric Waveguides

Total internal reflection is the well-known phenomenon in which light coming from a material with high refractive index n_H gets reflected at the interface with a material with low refractive index n_L . For this to happen, incident light must be at an angle greater than the critical angle given by the Snell's law:

$$\theta_C = \arcsin \left(\frac{n_L}{n_H} \right).$$

Waveguides have many shapes, however one can initially discriminate them by the dimensionality of the confinement of light. The simplest one is called *slab* and it is composed by two interfaces, which divides the material in three volumes, as shown in Figure 2.2. This type is classified as a 1D-waveguide, because it constrains light in only one dimension.

2D-waveguides on the other hand are more common, to the point that with the term waveguide one usually refers to a device belonging to this category. Since they confine light in two dimension, their core defines a path with which one can conduct light from a place to another. A few types are shown in Figure 2.3.

Obviously, the versatility of straight (2D-) waveguides is limited. Hence more complex structures such as bent waveguides have been developed.

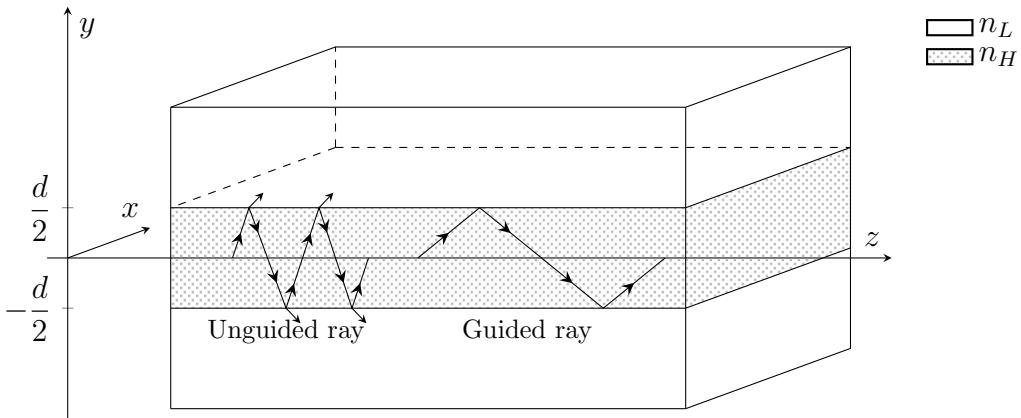


Figure 2.2: Scheme of a dielectric slab waveguide, made by two materials with refractive indexes n_H and n_L , with $n_H > n_L$. Two rays are shown: the unguided one is incident on the interface at an angle smaller than the critical angle. Conversely the guided ray is incident at a greater angle and is therefore totally reflected inside the waveguide [28].

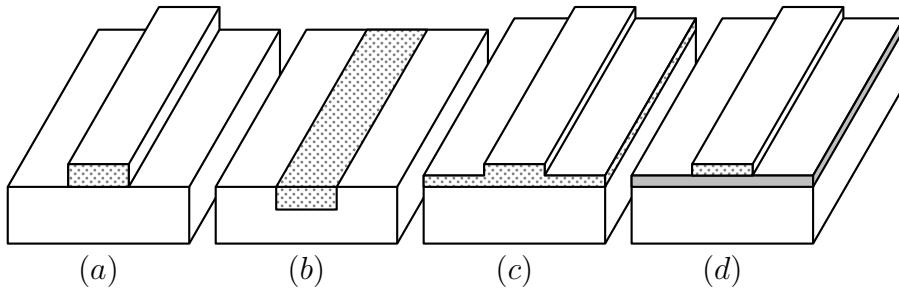


Figure 2.3: Representation of a few types of 2D dielectric waveguides. Darker areas represent higher refractive index. Light is constrained in two directions and can only move forward or backward in the third direction. (a) strip, (b) embedded strip, (c) rib, (d) strip loaded [28].

Propagation of light inside waveguides

A more complete description of guided modes is given by wave physics, where waveguides modes are described as superposition of planar Transverse Electro-Magnetic (TEM) waves. The result of the superposition is the so called *mode*, which is expressed mathematically by

$$\phi(\mathbf{r}) = \phi_m(x, y) e^{i(\beta_m z - \omega t)}. \quad (2.1)$$

$\phi_m(x, y)$ is the transverse field distribution, z is the direction of propagation, β_m is the propagation constant of the m -th mode, and $\omega = 2\pi\nu$ is the angular frequency of light. Since light has two independent polarization, there are different modes for different polarizations. Modes are therefore separated in *Transverse Electric* (TE) and *Transverse Magnetic* (TM), which propagates differently due to geometrical asymmetries.

Each mode, identified by the m index, travels inside the waveguide with a certain propagation constant β and maintaining a field distribution ϕ_m . Both of them depend on the materials and the geometry of the waveguide. However, while β_m decrease with increasing m indexes, the features of the function ϕ_m grows in number. Light often propagates in a manner described by a mode or superposition of them. Usually other field distributions are rapidly scattered away.

The field distribution of some of the first modes, in the simplest case of a slab waveguide, is represented in Figure 2.4. The function inside the core is characterized by maxima, minima and nodes with a certain periodicity and the number of nodes is strictly linked to the index m . Outside, the distribution of the field decays exponentially. For waveguide with a 2D core,

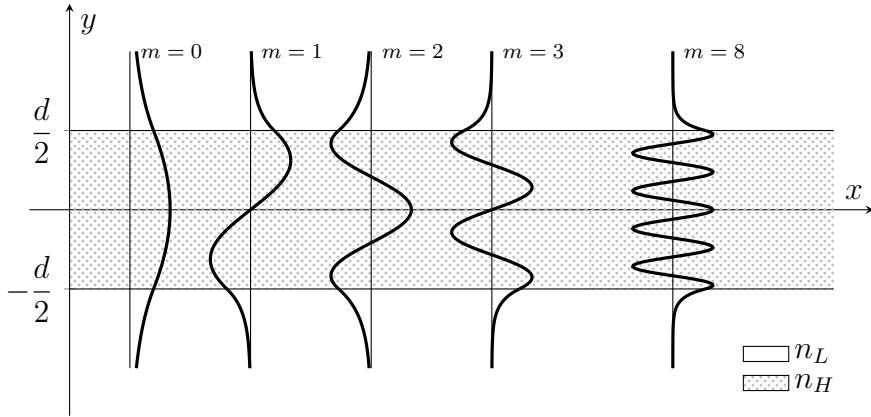


Figure 2.4: Field distribution inside a slab waveguide.

the field can be expected to be a multivariate version of similar functions. However, differently from the analytical solution for slab waveguides, they are almost always found numerically.

The propagation constant can be written as

$$\beta_m = n_{eff}k_0 + i\alpha_{eff}/2 \quad (2.2)$$

where k_0 is the wavevector in vacuum, $n_{eff} := c_0/v_{ph}$ is the *effective refractive index* of the mode, and α_{eff} is *effective absorption coefficient*. The effective refractive index is the ratio between the speed of light and the *effective phase velocity* v_{ph} at which each wavefront propagates in the core. It is a pure number and its value is between the core refractive index and the cladding refractive index $n_L < n_{eff} < n_H$. Given a waveguide geometry, larger n_{eff} means a greater confinement of the electromagnetic field inside the core and a slower propagation speed along the waveguide. The effective absorption coefficient instead is defined by the ratio between the input and output powers for a material of depth L , such that $I_{out}/I_{in} = e^{-\alpha_{eff}L}$, and has units of m^{-1} . A low absorption coefficient is synonym of transparency and low power absorption per unit length.

Each one of these parameters, which describe the propagation of light along the waveguide, depends both on the materials and the geometry of the core and the cladding, but also on the frequency of light. For dielectric waveguides, the fundamental mode ($m = 0$) is always supported while higher modes might not be, depending both on the waveguides and on the frequency of light. For this reason one distinguishes *single-mode* waveguides, which allow propagation for only the fundamental mode in a certain operative range of frequencies, from *multi-mode* waveguides, which allow higher order modes to propagate.

Equation (2.1) describes the propagation of a monochromatic light wave, which has ideally the same mode amplitude from $t = -\infty$ to $t = +\infty$. This is obviously not a good physical representation, as light is generated and absorbed. Eventually light is described to travel in wavepackets of finite duration, which are inherently non-monochromatic.

When light becomes non-monochromatic, the frequency dependence of the propagation constant β has to be considered. Usually, if working on a restricted range of frequencies, this dependence is characterized by a Taylor expansion at the second order.

$$\beta(\omega) \simeq \beta(\omega') + \frac{\partial\beta(\omega)}{\partial\omega}\Delta\omega + \frac{1}{2}\frac{\partial^2\beta(\omega)}{\partial\omega^2}\Delta\omega^2 \quad (2.3a)$$

$$\simeq \frac{1}{v_{ph}}\omega' + \frac{1}{v_g}\Delta\omega + \frac{1}{2}\text{GVD } \Delta\omega^2 \quad (2.3b)$$

$$\simeq \frac{n_{eff}}{c_0}\omega' + \frac{n_{eff}^g}{c_0}\Delta\omega + \frac{1}{2}\text{GVD } \Delta\omega^2 \quad (2.3c)$$

where $\Delta\omega = \omega - \omega'$ and ω' is the frequency around which β is expanded. Equation 2.3b is the same expansion, but the derivative are substituted by new quantities: the *phase velocity* v_{ph} , which describes the propagation speed of the wavefront of each frequency, the *group velocity* v_g , which describes the speed at which the ensemble of frequencies, the wavepacket, propagates, and finally the *group velocity dispersion* GVD, which describes the relative behavior of the different frequency components in a single wavepacket. Values of $\text{GVD} \neq 0$ the wavepacket will either compress or dilate in time. Those quantities are defined as follows:

$$v_{ph} := \left(\frac{\beta}{\omega'} \right)^{-1} \quad (2.4a)$$

$$v_g := \left(\frac{\partial \beta}{\partial \omega} \right)^{-1} \quad (2.4b)$$

$$\text{GVD} := \frac{\partial^2 \beta}{\partial \omega^2} = \frac{\partial}{\partial \omega} \frac{1}{v_g}. \quad (2.4c)$$

On the other hand, it could be useful to express the propagation constant in terms of the effective refractive index, where $n_{eff} = c_0/v_{ph}$. Then, similarly to the meaning of n_{eff} , one can define the *effective group index* n_{eff}^g as the ratio between the speed of light in vacuum c_0 and the effective group velocity v_g of a packet of light inside a waveguide

$$n_{eff}^g := \frac{c_0}{v_g} = n_{eff}(\omega) + \omega \frac{\partial n_{eff}}{\partial \omega} \simeq n_{eff}(\lambda) - \lambda \frac{\partial n_{eff}}{\partial \lambda}, \quad (2.5)$$

thus obtaining Equation 2.3c.

Nevertheless, our aim is to work in the quasi-static regime, i.e. to work with a time scale such that all measurable physical quantities converge to a finite value and do not show any time dependence. Thus the monochromatic wave propagation description is still a valid model, because it is a very good approximation of slowly varying and almost monochromatic waves, such as the one I will use in my experiments.

Coupling of light between waveguides

As shown in Figure 2.4, the field distribution of modes propagating inside a slab waveguide is mainly contained inside the core. However, the evanescent wave, i.e. the electromagnetic field outside the core, is non zero but decays exponentially from its value at the interface.

Evanescent coupling exploits the exponential decay of evanescent field outside the core to transfer light between two waveguides [27]. This means that when two waveguides are separated by a sufficiently small distance, the field distribution of one waveguide cannot decay sufficiently fast and instead extends itself over the core of the second waveguide. When this happens, optical power is transferred between the waveguides.

The simpler case is obtained when two parallel straight waveguides are separated by a small distance, a gap, and the field is non zero only at the input of one of them. Therefore light propagates along the waveguide, its evanescent field overlaps with the core of the second waveguide. Light is then *coupled* inside the second waveguide as it propagates along the first waveguide and power is transferred between the waveguides.

The maximum value of power transferred per unit length is determined mainly by the distance between the waveguides. The larger the gap, the lower will be its value. Moreover, the exchange of optical power is a periodic function of the propagating distance. Depending on the specific materials and on the geometry of the coupling region, a known amount of light can be coupled between the waveguides.

2.2.2 Microring optical cavity

In integrated photonics the microring is an optical cavity made by bending a waveguide on itself. To insert light into the cavity, the microring is often side coupled with one or two straight waveguides. Light inserted from the different ports into the cavity interfere with itself, after a round trip. When the waves are in phase, they generate constructive interference and the energy stored in the cavity increases exponentially. On the other hand, when the signals are out of phase, the interference is destructive and no energy is stored in the cavity.

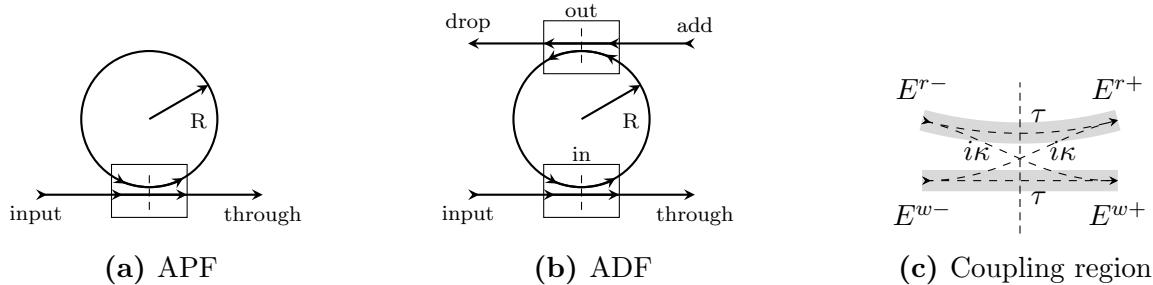


Figure 2.5: Schematic representation of microring configurations and of their coupling regions. APF has two channels, *input* and *through*, and one coupling region. ADF has four channels, *input*, *through*, *add* and *drop*, and two coupling regions. The coupling constants (κ and τ) are in principle different for each coupling region.

In the simpler case when there is only one waveguide side coupled to the resonator the optical cavity has only two ports, which are usually called *input* and *through*. This configuration, shown in Figure 2.5a, is called *All-Pass Filter* (APF), because in the ideal case, where no losses happen, all the signal passes from the input to the through channels.

By adding a secondary waveguide coupled to the ring, one obtains the so called *All-Drop Filter* (ADF) configuration, shown in Figure 2.5b. Its name derives from the fact that the two additional ports are called *add* and *drop* respectively. This simple structure can be readily used as a signal mixer: signals at the resonance wavelengths of the cavity are directed from the input to the drop channel or from the add channel to the through port. Additionally signals out of resonance travel straight from the input to the through channel and from the add to the drop port.

In both configurations light is transferred from the waveguides to the microring within the *coupling regions*, as shown in Figure 2.5c. The simplest model considers the region as infinitely small compared to the other dimensions. However the model can be expanded to include coupling regions with a finite length, which transform the microring in the so called *racetrack resonator*.

The theoretical model usually employed to describe resonators analytically decomposes their structure in a series of simpler substructures. Both the APF and the ADF configurations are studied as ordered combinations of straight waveguides, coupling regions, and bent waveguides. Other used descriptions, such as numerical simulation obtained with Finite Element Methods (FEM), achieve more precise results, however they are more dependent on the specific geometry of the problem. Moreover the approximation given by the analytical model is sufficiently accurate to make quantitative prediction of the coupling and it gives a clear description of the physics behind it. Furthermore it can be expanded to include non-trivial phenomena (see Section 2.3). In the following section I will go through the necessary steps to solve the case of the ADF configuration in the theoretical model.

Add-Drop-Filter theory

ADF configuration is obtained when a microring is coupled to two waveguides. Such structure is composed by three different basic structures: four straight waveguides, two bent waveguides which together form the whole ring, and two coupling regions between the waveguides and the microring. Each of these pieces transfers light to or from the outside or another piece.

Since our experiments are carried out in a time scale such that physics phenomena can be considered quasi-static, the field propagating in the device is described by a scalar complex function which, assuming a monochromatic continuous EM wave, loses any temporal dependence:

$$E(z) = |E(z_0)|e^{i\beta z}, \quad (2.6)$$

where z is the direction of propagation, loosely defined to accommodate propagation both along straight and bent waveguides ($z \sim r\theta$), t is time, $\omega = 2\pi\nu$ is the angular frequency, and $\beta = n_{eff}k_0 + i\alpha_{eff}/2$ is the propagation constant. Both dependence of the effective index $n_{eff} = n_{eff}(\omega)$ and the effective loss factor $\alpha_{eff} = \alpha_{eff}(\omega)$ will be expanded only when necessary.

Generally the propagation along the straight waveguides is considered lossless (i.e. $\alpha_{eff} \approx 0$), while the propagation along the two bent regions has a non negligible loss (i.e $\alpha_{eff} > 0$). The coupler between the resonator and the waveguides is considered like a beamsplitter. In this approximation, light does not propagate in this part as it was in the previous ones. Its operation is reduced to the exchange of power between the two input ports and two output ports, which is described by the following matrices:

$$\begin{pmatrix} E_{ch}^{w+} \\ E_{ch}^{r+} \end{pmatrix} = \mathbf{M} \begin{pmatrix} E_{ch}^{w-} \\ E_{ch}^{r-} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \tau & i\kappa \\ i\kappa & \tau \end{pmatrix}, \quad (2.7)$$

where E is the field amplitude of the ' ch ' channel either in the waveguide ' w ' or in the ring ' r ', before ' $-$ ' and after ' $+$ ', the coupling. An explanatory diagram is shown in Figure 2.5c. The matrix \mathbf{M} is characterized by two real valued parameters, τ and κ , between 0 and 1. Specifically, they must verify the following constraint:

$$\det(\mathbf{M}) = |\tau|^2 + |i\kappa|^2 = 1. \quad (2.8)$$

which represents the conservation of energy.

At this point one can solve the problem by putting all the pieces together. Neglecting the propagation inside the waveguide, the full system of equation that describe the problem is:

$$E_{in}^{w+} = \tau E_{in}^{w-} + i\kappa E_{in}^{r-} \quad (2.9a)$$

$$E_{in}^{r+} = i\kappa E_{in}^{w-} + \tau E_{in}^{r-} \quad (2.9b)$$

$$E_{out}^{r-} = E_{in}^{r+} e^{i\beta\pi R} \quad (2.9c)$$

$$E_{in}^{r-} = E_{out}^{r+} e^{i\beta\pi R} \quad (2.9d)$$

$$E_{out}^{w+} = \tau E_{out}^{w-} + i\kappa E_{out}^{r-} \quad (2.9e)$$

$$E_{out}^{r+} = i\kappa E_{out}^{w-} + \tau E_{out}^{r-} \quad (2.9f)$$

The first two equations (2.9a) and (2.9b) describe the exchange of optical power between the first channel and the microring resonator (see *in* box in Figure 2.5b). Then the third and fourth equations (2.9c) and (2.9d) delineate the propagation of light in the two halves of the resonator, from one coupling region to the other. Lastly the remaining two equation

characterize the transfer of light between the resonator and the add-drop channel (see *out* box in Figure 2.5b).

Using these equations it is easy to define new quantities of interest: the first one is the transmittance from the *input* to the *through* port.

$$\eta_T(\omega) := \frac{E_{in}^{w+}}{E_{in}^{w-}} = t \frac{1 - e^{i\beta 2\pi R}}{1 - \tau^2 e^{i\beta 2\pi R}} \quad (2.10)$$

Similarly, one can also define the transmittance from the *input* to the *drop* port.

$$\eta_D(\omega) := \frac{E_{out}^{w+}}{E_{in}^{w-}} = \frac{-\kappa^2 e^{i\beta \pi R}}{1 - \tau^2 e^{i\beta 2\pi R}} \quad (2.11)$$

However, these quantities have complex values and therefore are difficult to study. For this reason, one usually defines the transmission between the same ports as the square modulus of the transmittance. Thus it follows that

$$T(\omega) := |\eta_T|^2 = \tau \frac{(1 - \gamma)^2 + 4\gamma \sin^2(n_{eff} k_0 \pi R)}{(1 - \tau^2 \gamma)^2 + 4\tau^2 \gamma \sin^2(n_{eff} k_0 \pi R)} \quad (2.12)$$

and

$$D(\omega) := |\eta_D|^2 = \frac{\kappa^4 \gamma}{(1 - \tau^2 \gamma)^2 + 4\tau^2 \gamma \sin^2(n_{eff} k_0 \pi R)} \quad (2.13)$$

where the notation is simplified by loss parameter $\gamma = e^{-\alpha_{eff}\pi R}$. The dependence of $T = T(\omega)$ and $D = D(\omega)$ from the frequency (or wavelength) of light is obtained by making explicit the wavevector dependence $k_0 = \frac{\omega}{c_0} = \frac{2\pi}{\lambda}$ in which c_0 is the speed of light in vacuum.

Apart from a simpler description, transmission is directly connected with the optical power as it represents the ratio between the input and output power. In fact, the optical power is defined as the square modulus of the optical field, except for a constant factor:

$$I = \frac{1}{2} n_0 c_0 |E|^2 \propto |E|^2$$

Both transmission spectra, as shown in Figure 2.6, have either peaks or dips: at each microring resonance, the through spectrum shows dips while the drop spectrum shows peaks. The depth of the dips, the height of the peaks, and the width of both of them is defined by few parameters: the coupling constant τ^1 and the half round trip loss factor $\gamma = e^{-\alpha_{eff}\pi R}$.

The frequency of each resonance is identified by a positive integer number, that verifies the following equation:

$$\omega_m = m \frac{c_0}{n_{eff}(\omega) R}, \quad (2.14)$$

where the dependence of $n_{eff} = n_{eff}(\omega)$ has been explicitly shown.

Each resonance is spaced from the next one by a quantity called *Free Spectral Range* (FSR). By exploiting the Taylor expansion of β seen in equation (2.3c) one obtains

$$FSR_{\omega_m} \simeq \frac{c_0}{n_{eff}^g(\omega_m) R}. \quad (2.15)$$

where the *effective group index* $n_{eff}^g(\omega)$ has been defined in equation (2.5). Another important quantity is the width of the peaks or dips of each resonance. Ordinarily one evaluates the so called *Full-Width-Half-Maximum* (FWHM) of the transmission in the drop channel:

$$FWHM_{\omega_m} = \frac{c_0}{n_{eff}^g(\omega_m)} \frac{1 - \tau^2 \gamma}{\pi R \tau \sqrt{\gamma}} \quad (2.16)$$

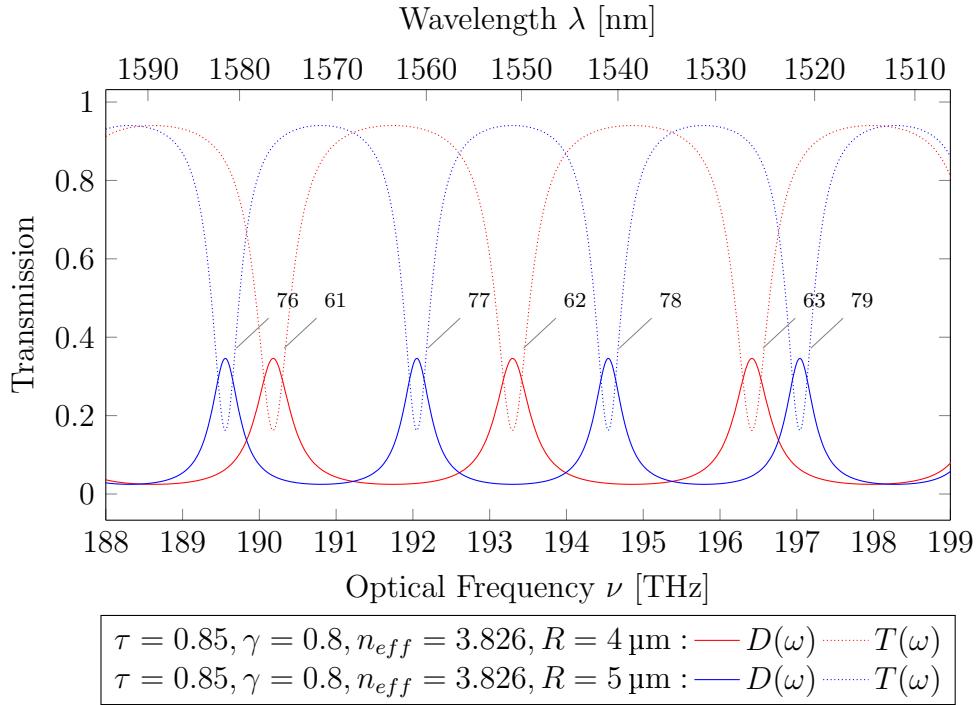


Figure 2.6: Transmission spectra of microring resonators in ADF configuration for the through $T(\omega)$ and drop ports $D(\omega)$. The resonators considered differ in radius ($R = 4 \mu\text{m}$ and $R = 5 \mu\text{m}$), but share the coupling constant $\tau = 0.85$, the half round trip loss factor $\gamma = 0.8$ and the effective index $n_{eff} = 3.826$.

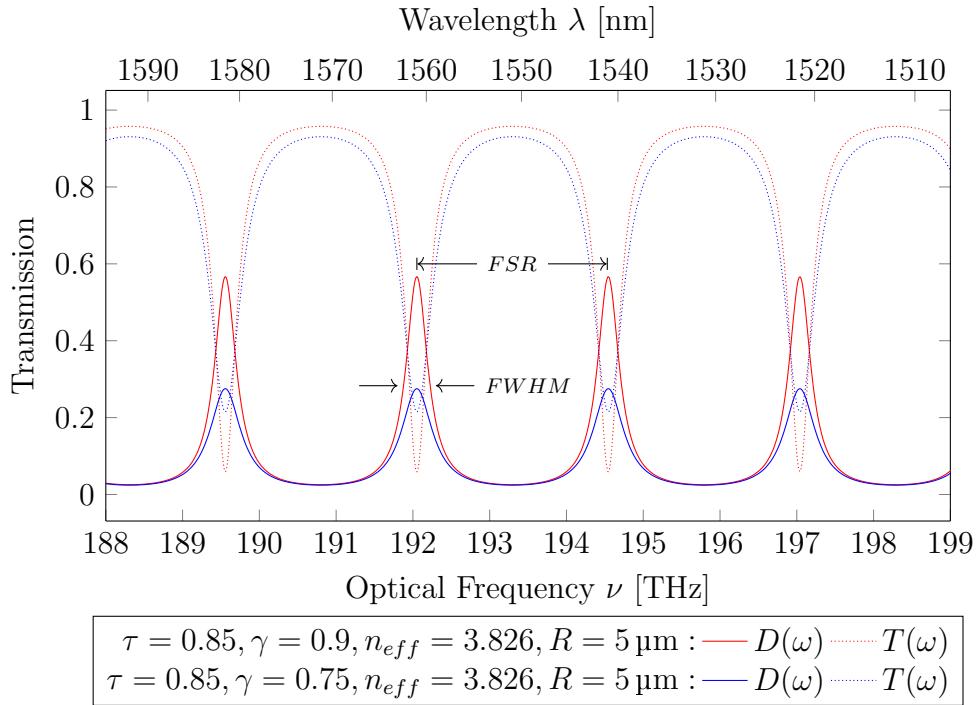


Figure 2.7: Transmission spectra of microring resonators in ADF configuration for the through $T(\omega)$ and drop ports $D(\omega)$. The microring has radius $R = 5 \mu\text{m}$ and effective index $n_{eff} = 3.826$. The coupling constant is $\tau = 0.85$, while the loss factor γ takes two values: 0.9 (red) and 0.75 (blue). The arrows indicate a $FSR_\nu \approx 2.49 \text{ THz}$ and a $FWHM_\nu \approx 0.35 \text{ THz}$, which gives a quality factor of around $Q \approx 550$. The blue curve has $FWHM_\nu \approx 0.5 \text{ THz}$ and $Q \approx 390$. Each resonance is identified by its positive integer number.

Figure 2.7 shows both the *FSR* and the *FWHM* for a family of resonances.

Values such this are often expressed in the wavelength domain:

$$FSR_{\lambda_m} \simeq \frac{\lambda_m^2}{n_{eff}^g(\lambda_m) 2\pi R} \quad \text{and} \quad FWHM_{\lambda_m} \simeq \frac{\lambda_m^2}{n_{eff}^g} \frac{1 - \tau^2 \gamma}{2\pi^2 R \tau \sqrt{\gamma}} \quad (2.17)$$

where $n_{eff}^g(\lambda_m)$ is the appropriate redefinition of the effective group index in the wavelength domain, shown in equation (2.5).

Another important quantity to describe an optical resonator response is the *enhancement factor EF*. It is defined as the ratio between the incident optical power I_{inc} and the power circulating inside the cavity I_{ins} :

$$EF := \frac{I_{int}}{I_{inc}} \simeq \frac{|E_{in}^{r+}|^2}{|E_{in}^{w-}|^2} = \frac{\kappa^2}{(1 - \tau^2 \gamma)^2 + 4\tau^2 \gamma \sin^2(n_{eff} k_0 \pi R)}, \quad (2.18)$$

where in the last equality, I assumed low loss waveguides $\gamma \rightarrow 1$. This assumption allows to consider the field at an arbitrary position along the ring.

Finally, an additional figure of merit is the important quantity is the *quality factor Q-factor*, or simply Q . It has many similar definitions, depending on the field of study. The most physical, but less operative one defines Q as 2π times the ratio between the energy stored in the cavity and the energy lost each cycle.

$$Q := 2\pi \times \frac{\text{energy stored}}{\text{energy lost per cycle}} \quad (2.19)$$

However, for my purposes this definition is too cumbersome, thus I used a similar one instead:

$$Q := \frac{\omega_m}{FWHM_{\omega_m}} = \frac{\lambda_m}{FWHM_{\lambda_m}} \quad (2.20)$$

Both are sensible definitions and as Q becomes larger they become approximately equivalent.

Physically, the quality factor is linked to the photon lifetime in cavity t_m by the relation

$$Q \simeq \frac{\omega_m}{FWHM_{\omega_m}} \propto \omega_m t_m. \quad (2.21)$$

From this description one can observe that higher quality factors mean longer photon lifetimes. This is due to the fact that the photon lifetime is strictly dependent on the losses of the cavity and the Q factor is inversely related to them. However, in an optical cavity losses are given by absorption, but also by the coupling of the input and output channels. Thus, given a certain waveguide loss (defined by both the propagation and radiative contributions), the same Q -value can be obtained by different sets of loss factors. One can identify three distinct regimes: over-coupling, critical-coupling, and under-coupling regimes. Figure 2.8 shows contour plots of the maximum transmission in the drop channel and the quality factor as functions of the cavity losses (γ) and the coupling coefficients ($\kappa^2 = 1 - \tau^2$).

Coupling regimes

Microring resonator and more in general optical cavities are able to store very large quantities of optical energy. This energy, however, is continuously removed by the cavity in two main ways. The first one is the coupling mechanism between the waveguides and the microring, characterized by κ and τ , with extract light from the cavity in the same way that injects

¹one can choose also κ , but it does not matter which one, since $\tau^2 = 1 - \kappa^2$

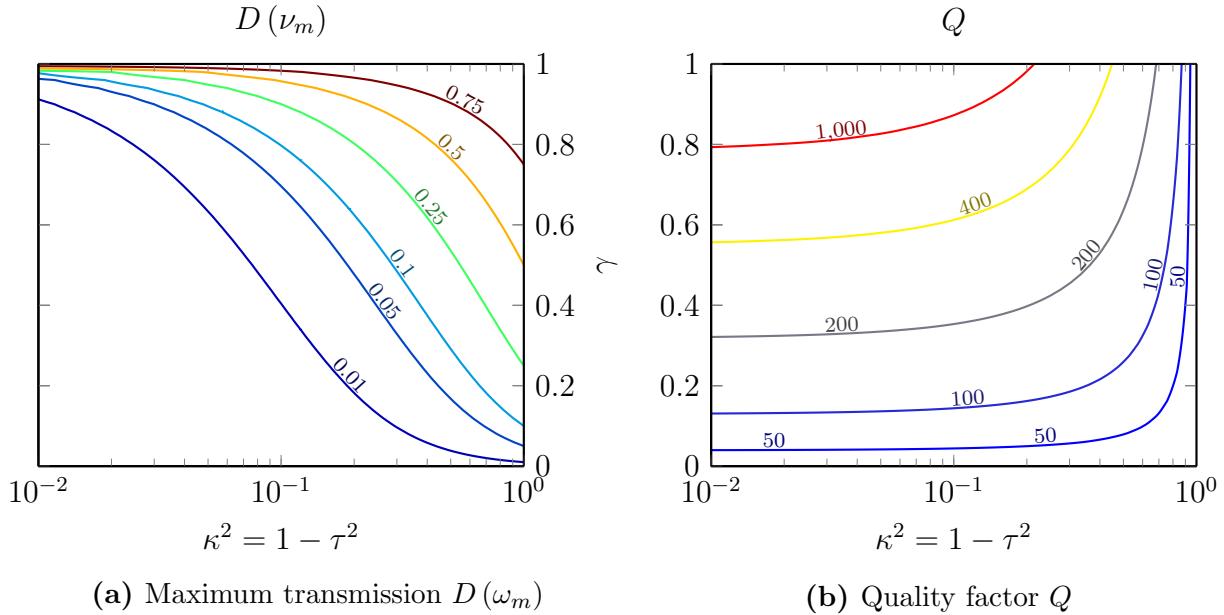


Figure 2.8: Contour plots of the maximum transmission $D(\omega_m)$ and of the quality factor Q , parametrized for the coupling coefficient τ^2 and the half round trip loss factor γ . The Q -factor is obtained for a resonance near $v_m \approx 200$ THz (arbitrarily chosen) and an effective group index $n_{eff}^g \approx 4$.

the input signals in it. The second one is the intrinsic loss, described by the half round trip loss factor γ , which in turn is composed by absorption of the material and radiative induced scattering. Absorption loss is a given property of the material, whereas radiative scattering is linked to the roughness of the surfaces and the presence of imperfections.

The quality factor can be written as

$$\frac{1}{Q_{tot}} = \frac{1}{Q_{intr}} + \frac{1}{Q_{coupl}} \quad (2.22)$$

where Q_{coupl} is dependent on the geometry and the material, whereas Q_{intr} is dependent on the quality of the manufacturing processes.

A microring resonator in the ADF configuration has two coupling regions and therefore in principle we can have symmetrical or asymmetrical coupling mechanisms. I shortly address here the case of asymmetric coupling ($\tau_1 \neq \tau_2$) to define the three coupling regimes. The coupling constants κ_i and τ_i and the loss factor γ can add up in different ways to obtain the same quality factor. Specifically, when $\tau_1 = \gamma^2 \tau_2$ is verified, the resonator is in the *critical-coupling regime*. In this particular condition, which requires asymmetrical coupling, the transmission on the through channel at the resonant frequencies goes to zero and $Q = Q_{intr}/2$. The values of the parameters for which this condition is verified split the space of parameters in the other two regimes. When $\tau_1 < \gamma^2 \tau_2$, and $Q > Q_{intr}/2$, the resonator is in *under-coupling regime*, while for $\tau_1 > \gamma^2 \tau_2$, and $Q < Q_{intr}/2$, it is in *over-coupling regime*. In both cases, the transmission on the through channel is always greater than zero.

2.3 Nonlinear Optics

In the previous section, the material of waveguides and resonators have been considered as media with linear response to electromagnetic radiation. This is, however, only an approximation, because any dielectric material shows nonlinearities if probed with a high enough electromagnetic field.

In integrated photonics, due to the lateral confinement of the field inside the waveguide, much higher field amplitude is reached in comparison to free space. Moreover, inside an optical cavity, the field reaches even more higher amplitudes thanks to its large enhancement factor and nonlinearities might arise.

Despite silicon shows a small third order nonlinear refractive index, significant nonlinear responses are obtained through a proper microresonators optical design.

One can distinguish two kinds of optical nonlinearities: electronic nonlinearities and thermal nonlinearities.

2.3.1 Electronic nonlinearities

When electromagnetic radiation propagates inside a dielectric, atoms become polarized. This polarization is usually considered linear, because the applied fields are normally weak in comparison to the inter-atomic electric fields. However, when they grow sufficiently large, e.g. 10^5 to 10^8 V m^{-1} [28], the relation between the total polarization vector and the applied optical field becomes nonlinear. Then the following generalized equation is used instead:

$$\mathbf{P} = \epsilon_0 \left(\chi^{(1)} + \chi^{(2)} : \mathbf{E}\mathbf{E} + \chi^{(3)} : \mathbf{E}\mathbf{E}\mathbf{E} + \dots \right) \quad (2.23)$$

which in principle is a vectorial relation, where the $j - th$ order susceptibility term $\chi^{(j)}$ is a tensor of rank $j + 1$ [29]. Often, the spatial dependences simplify and the equation can be reduced to be scalar only. This is the case of linearly polarized TEM waves, when the electric field is directed only along one direction.

Equation 2.23 allows to expand the definition of refractive index and absorption coefficient to:

$$\left(n + i \frac{c_0}{2\omega} \alpha \right)^2 = 1 + \chi \simeq 1 + \chi^{(1)} + \chi^{(2)} E^2 + \chi^{(3)} E^3 + \dots, \quad (2.24)$$

where the vectorial form has already been reduced to the scalar one.

Silicon is a centrosymmetric material and thus does not exhibit optical nonlinearities of even orders, i.e. $\chi_{Si}^{(2)} = 0$. The lowest order nonlinear effects in silicon belong to the third order $\chi_{Si}^{(3)} \neq 0$. In this category - third order nonlinearities - two are the fundamental effects that modifies the refractive index and the absorption coefficient.

Kerr effect

The Kerr effect is given by the real part of the third order nonlinear susceptibility $\chi^{(3)}$ and it produces a change in the refractive index characterized by:

$$\Delta n_{Kerr} \simeq n_2 I, \quad (2.25)$$

where I is the intensity of light beam and n_2 is the second-order (or intensity dependent) nonlinear refractive index, defined by

$$n_2 = \frac{3}{4\epsilon_0 c_0} \operatorname{Re} [\chi^{(3)}] \stackrel{Si}{\simeq} 4.5 \times 10^{-18} \text{ m}^2 \text{ W}^{-1}. \quad (2.26)$$

The value of n_2 is verified for light at $\lambda = 1.54 \mu\text{m}$ in silicon [30].

Two photon absorption

The other fundamental effect is the Two Photon Absorption (TPA), which is linked instead to the imaginary part of the third order nonlinear susceptibility. It does not produce a change to the refractive index but to the absorption coefficient instead.

$$\Delta\alpha_{TPA} = \beta_{TPA}I, \quad (2.27)$$

where β_{TPA} is the TPA coefficient which is experimentally found to be [30]

$$\beta_{TPA} = \frac{3}{2\varepsilon_0 c_0^2 n^2} \text{Im} [\chi^{(3)}] \stackrel{Si}{\simeq} 7.5 \times 10^{-12} \text{ m W}^{-1}. \quad (2.28)$$

As a consequence to the absorption of two photons, a free carrier is generated in the conduction band of silicon. The presence of free carriers alters both the refractive index and the absorption coefficient.

2.3.2 Charge carrier nonlinearities

When free carriers are generated in the conduction band, they can either move around until they thermalize or they can absorb an incoming photon. The first case is the *Free Carrier Dispersion* (FCD) effect, which affects the refractive index. The second case is the *Free Carrier Absorption* (FCA) effect and affects the absorption coefficient. Both effects are usually characterized as first order expansions, as follows:

$$n(N) = n(N_0) + \left. \frac{dn}{dN} \right|_{N_0} \Delta N \quad (2.29)$$

$$\alpha(N) = \alpha(N_0) + \left. \frac{d\alpha}{dN} \right|_{N_0} \Delta N \quad (2.30)$$

where the value of the Free Carrier Dispersion (FCD) and Free Carrier Absorption (FCA) coefficients are [31], [32]

$$\left. \frac{dn}{dN} \right|_{N_0} \stackrel{Si}{\simeq} -1.73 \times 10^{-21} \text{ m}^3 \quad \text{and} \quad \left. \frac{d\alpha}{dN} \right|_{N_0} \stackrel{Si}{\simeq} 1.1 \times 10^{-15} \text{ m}^2 \quad (2.31)$$

respectively, and ΔN is the variation in free carrier population per volume. The free carrier dynamics is usually defined as

$$\frac{d\Delta N}{dt} = -\gamma_{FC}\Delta N + G \quad (2.32)$$

where γ_{FC} is the empirical population loss factor and G is the total generation rate per unit volume [33]. The generation rate G is defined as

$$G = \frac{P_{abs}^{TPA}}{2\hbar\omega V_{ring}} \quad (2.33)$$

In our case the free carrier are mainly generated by the two photon absorption process, hence G is linked to the power absorbed by the material by the TPA effect only. Therefore the absorbed power can be explained as a function of optical power:

$$P_{abs}^{TPA} = A (1 - \gamma_{TPA}^2) I \simeq A 2\pi R \Delta\alpha_{TPA} I = V \Delta\alpha_{TPA} I \quad (2.34)$$

where $A = V/2\pi R$ is the area of the microring.

2.3.3 Thermal nonlinearities

The optical nonlinearities of the refractive index due to the changes in a material being heated is called *Thermo-Optic Effect* (TOE). This happens when heat source is, for example, a resistance, but also when light propagates inside a medium, because a portion of the photons is absorbed. To characterize this change in the refractive index caused by the temperature, a first order expansion is usually made:

$$n(T) = n(T_0) + \left. \frac{dn}{dT} \right|_{T_0} \Delta T, \quad (2.35)$$

where $\frac{dn}{dT}$ is called *Thermo-Optic Coefficient* (TOC) and ΔT is the temperature shift.

The Thermo-Optic Coefficient (TOC) of silicon at 300 K is [32] :

$$\left. \frac{dn}{dT} \right|_{300\text{ K}} = 1.86 \times 10^{-4} \text{ K}^{-1}. \quad (2.36)$$

Moreover, due to the fact that in silicon integrated structures the cladding is usually made by SiO_2 , a thermally insulating material, the heat generated by the beam confined in the cavity will stay within the core, thus amplifying the effect even further.

The temperature dynamics of a silicon microring resonator is driven by the optical power absorbed and by the heat lost to the surrounding medium. Specifically the Fourier law reads

$$\frac{d\Delta T(t)}{dt} = -\gamma_{TH}\Delta T(t) + \frac{P_{abs}}{M_{ring}C_p} \quad (2.37)$$

where M_{ring} is the mass of the microring, C_p is the heat capacity at constant pressure of silicon, and γ_{TH} is a phenomenological quantity which describes the heat loss rate.

$$P_{abs}^{tot} = A(1 - \gamma^2) I \simeq V(\alpha + \Delta\alpha_{TPA} + \Delta\alpha_{FCA}) I \quad (2.38)$$

The power absorbed is caused both by linear and nonlinear processes. Hence the total value is given by the linear absorption coefficient α , and by the nonlinear shift of absorption coefficient produced by the two photon absorption $\Delta\alpha_{TPA}$ and by the free carrier absorption $\Delta\alpha_{FCA}$.

2.3.4 Nonlinear perturbation of microring response

When applied to the response of a microring resonator, the optical nonlinearities are particularly effective because of the enhancement factor of the cavity. Not all nonlinearities alter the behavior of the optical resonator in the same way. From the point of view of timescales, the fastest effects are the electronic and free carrier effects. Nonlinear responses given by such effects reach a steady state in less than 1 ps for the first and 1 ns for the second. On the contrary, the timescale of thermal perturbations is of the order of the 1 μ s. Each one of these values, however, depends strictly on the geometry of the structure.

On the other hand, also the magnitude of the different effects plays an important role. Having similar effects on the refractive index, at short timescales Kerr and FCD win over the thermal one, however, at longer timescales the thermal effect is the predominant one. Hence ultrashort light pulses would have shown a completely different dynamics than the one described in this thesis..

In this discussion I will consider only the sum of all effects at steady state and the thermo-optic effect is the prevalent effect at medium-high intensities, as shown in Figure 2.9.

When injecting light into a resonator, its transfer function ultimately depends on two factors: the frequency of the light compared to the frequency of the resonance and the optical power

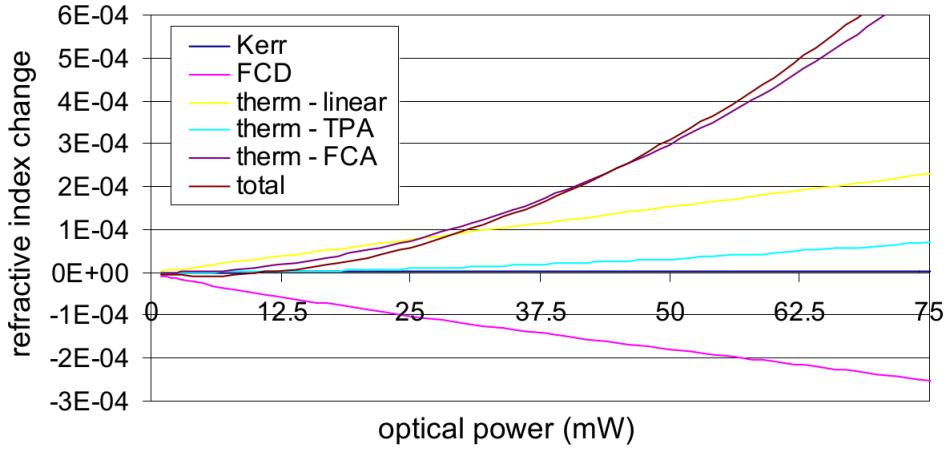


Figure 2.9: Nonlinear contributions to the refractive index change inside the ring resonator as function of the cavity power, taken from [34].

injected. At low intensities the response of a microring to frequencies around its resonance is symmetrical, as was shown in the preceding section. At higher intensities, however, the nonlinear phenomena change the refractive index of the material so much that the resonance frequency shifts. Not all nonlinear mechanisms act in the same way: for example, thermal and FCD phenomena have opposed and competing effects, which shift the resonance in opposite directions. However, as already stated, on long timescales thermal nonlinearities predominates over all the others. Hence, after the system has reached the quasi-static regime, as the intensity of the input beam grows, the cavity heats up and the refractive index increases. This produces a shift of the *cold* resonance towards smaller frequencies, as shown in Figure 2.10.

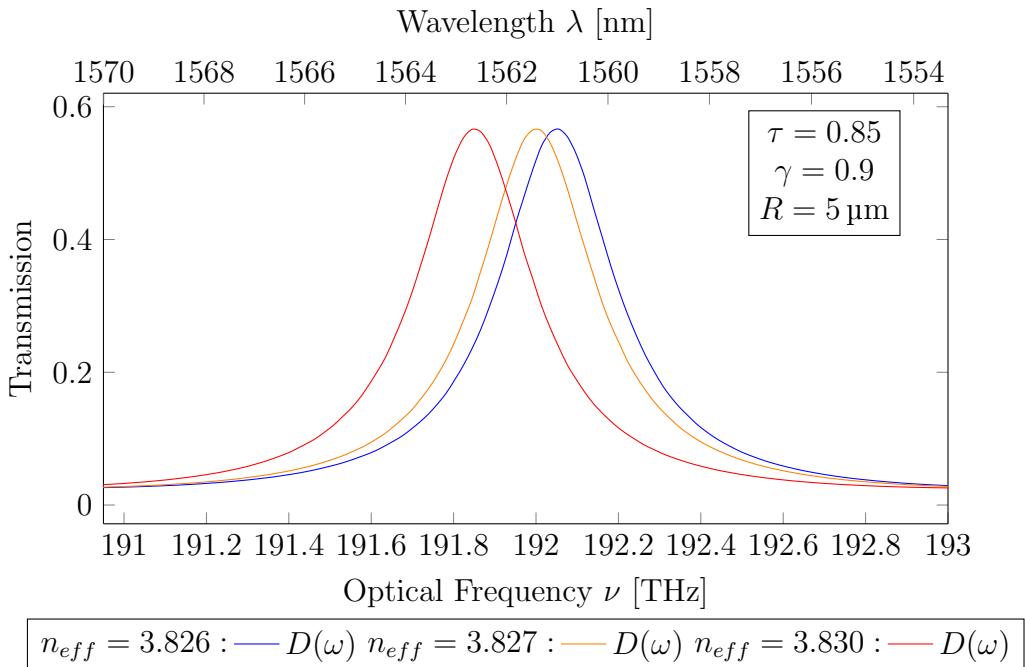


Figure 2.10: Microring resonator optical response in three different conditions: unaltered (blue), slightly increased $\Delta n = 1 \times 10^{-3}$ (orange), and substantially increased $\Delta n = 1 \times 10^{-2}$ refractive index (red). Both changes simulates the nonlinear effect of (mostly) thermal nonlinearities.

This means that by injecting a certain (high) optical power at higher frequency than the resonance frequency, or on the blue shoulder of the resonance, the obtained transmission will

be higher than the one described by the linear equations. On the other hand, by injecting the same amount of optical power at lower frequency, or on the red shoulder of the resonance, the resulted transmission will be lower than the one described by the linear equations.

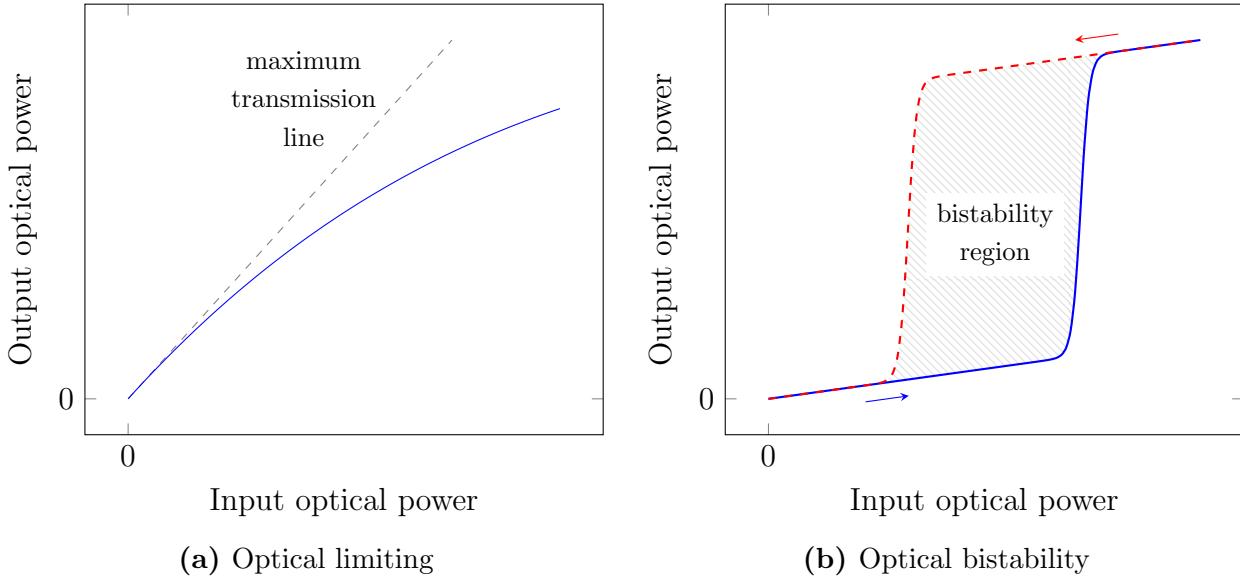


Figure 2.11: Examples of (a) optical limiting and (b) optical bistability. These curves are just a qualitative example of the phenomena. More accurate representation will be shown in the Subsection 2.4.3.

These features can be exploited to obtain two different effects. By choosing a signal with a fixed frequency on the blue shoulder of the *cold* resonance, the effect obtained is *optical limiting*. Optical limiting is the progressive decrease in transmission with the increase of the input signal. Hence, the effective transmission produced is nonlinearly dependent on the value of the input signal, as shown in Figure 2.11a. On the other hand, by choosing a signal on the red shoulder of the *cold* resonance, the effect obtained is *optical bistability*. Optical bistability is the existence of two distinct steady states of the transmission function for the same interval of input signals. In an optical microring resonator this produces abrupt *jumps* from one steady state to the other, when at the edges of such intervals. Specifically, at the right edge of the bistability region, the system switches from the lower to the higher transmission state, whereas at the left edge, the system switches from the higher to the lower transmission state. Moreover, the state inside the bistability interval is determined by the past states of the optical cavity. This defines a hysteresis cycle around the bistability region. All these features are shown in Figure 2.11b.

The empirical values of the nonlinear constants introduced in Sections 2.3.1 to 2.3.3 are collected in Table 2.2.

Parameter	Value	Unit	Source
n_2	4.5×10^{-18}	$\text{m}^2 \text{W}^{-1}$	[30]
β_{TPA}	7.5×10^{-12}	m W^{-1}	[30]
$\frac{dn}{dN}$	-1.7×10^{-21}	m^3	[32]
$\frac{d\alpha}{dN}$	1.1×10^{-15}	m^2	[35]
γ_{FC}	300×10^{-9}	PHz	[33]
$\frac{dn}{dT}$	1.86×10^{-4}	K^{-1}	[32]
γ_{TH}	12×10^{-9}	PHz	[33]

Table 2.2: Collection of the empirical values of nonlinear constants.

2.4 Integrated photonics applied to ANNs

Even though the vast majority of hardware architectures for artificial neural networks are developed within the electronic platform, integrated photonics is emerging as a fast and efficient alternative. The way in which information is elaborated inside neural networks seems to be more favorable, in comparison to the boolean logic, to the development of photonic integrated processors [36]. The aim of photonics is to match the performances of electronics in the development of neuromorphic networks by exploiting its inherent suitability to parallel computing, its bandwidth and speed, and also its power efficiency.

Nevertheless building a neural network within a photonic integrated chip is a demanding task. For this reason, the results achieved until now have been obtained by using a small number of integrated devices. This fact inevitably limits the maximum complexity of the neural networks built. However, many neuromorphic networks are based on the Reservoir Computing (RC) approach or the Recurrent Neural Network (RNN) type or both. The first one is a learning method that allows to have little to no control on the majority of the network while employing all efforts on the training of a single output layer [37], [38]. Hence very fast and complex system can be exploited without having real control on the local effects that happen inside the *reservoir*. On the other hand, a recurrent neural network can be implemented with only one computational node [38], [39], as it exploit time to distribute the computation. In this type of network, the lack of physical complexity is compensated by the extension in time of the required task.

The idea of this work is to tackle the problem head on and develop a framework for feed-forward artificial neural network based on optical effects.

As shown in Section 1.2, the principal parts of an artificial neural network are its nodes, which in turn are built by two separate subcomponents: the weighted sum and the nonlinear activation function. In the next sections I will describe how the nonlinear response of a microring could be exploited in an all-optical node of a feedforward ANN.

2.4.1 Weighted sum of inputs

The difficulty in the weighted sum of inputs does not lie in the type of operation, i.e. sum and multiplication, but in the fact that each node has many inputs. Excluding the cases in which the weighted summation has been carried off by online or offline elaboration on standard computers, there are two important examples of all-optical weighted sum.

The first one [40] is based on a cascade of Mach-Zehnder Interferometer (MZI), which transforms all the outputs of a whole layer to the set of inputs of the next one. This happens by a sequence of constructive and destructive interferences that alter the starting coherent signal.

The second example [41] is given by a series of microring resonator, parametrically tuned to transmit a portion of light. In this work, different signals are multiplexed on many wavelengths and are weighted by consecutive rings built in such a way that they affects only a certain wavelength. Moreover, the weight is implemented by thermal finetuning of the resonator cavities. This example, however, exploit an optoelectronic nonlinear stage composed of two photodiodes and a light source nonlinearly dependent to the current produced by the photodiodes.

2.4.2 Nonlinear Activation Function

The development of novel optical mechanisms for the nonlinear activation function in ANNs has not progressed as much as the development of the weighted sum of the node inputs has. In fact, many proposed PICs by means of integrated optical structures realize the weighting part

of the inputs only. The nonlinear activation function is then implemented either by electronic devices [41] or even by computer aided evalutation [40].

An interesting proposal is to employ a saturable absorber to obtain a nonlinear activation function [39]. Saturable absorbers are materials that have a reduced absorption of light for high optical intensities.

The idea of my work is to employ the optical bistability in a microring resonator as a mean to obtain a nonlinear activation function, that simulate the commonly used sigmoid one. Optical bistability, as seen in Subsection 2.3.4, is an optical effect that occurs when injecting light in a microring resonator with a lower frequency in respect to the resonance frequency.

In a quasi-static regime, optical bistability is mainly due to the therm-optic effect, which depends from the temperature. However, temperature depends in turn from the optical power circulating inside the cavity, which is obviously determined by the input intensity. This allows to obtain a total output transmission that depends nonlinearly on the input intensity.

2.4.3 Simulations

In order to better understand the physical mechanisms underneath the microring behavior, I used numerical methods to simulate the response of a microring resonator. Since the experiments are carried out in the quasi-static regime, the system can be described by a set of three coupled algebraic equations: one for the enhancement factor, one for the free carrier dynamics, and one for the thermal dynamics.

$$EF(\omega) = \frac{\kappa^2}{(1 - \tau^2\gamma\Delta\gamma)^2 + 4\tau^2\gamma\Delta\gamma \sin^2 \left((n_{eff} + \Delta n_{eff}) \frac{\omega}{c_0} \pi R \right)} \quad (2.39)$$

$$\Delta N = \frac{1}{\gamma_{FC}} \frac{\beta_{TPA}}{2\hbar\omega} I^2 \quad (2.40)$$

$$\Delta T = \frac{1}{\gamma_{TH}} \frac{1}{M_{ring}C_p} V \left(\alpha_{eff} + \beta_{TPA}I + \frac{d\alpha}{dN} \Delta N \right) I \quad (2.41)$$

where

$$I = EF I_{input} \quad (2.42)$$

$$\Delta\gamma = e^{-(\beta_{TPA}I + \frac{d\alpha}{dN}\Delta N)\pi R} \quad (2.43)$$

$$\Delta n_{eff} = n_2 I + \frac{dn}{dN} \Delta N + \frac{dn}{dT} \Delta T \quad (2.44)$$

This set of equations allows a simple but complete description of the physical quantities of interest in the system.

To decrease numerical errors, a reduced variable approach has been used. This means that every factor in the equations has been divided by the order of magnitude typical of the problem, to reduce or increase its numerical representation towards unity. For examples, distances have been reduced to μm and powers to mW and so on.

In addition, important values that are not explicitly computed, such as the transmission of the through and drop channels, can be extracted with just a simple multiplication. Moreover, adding another light source at a different wavelength (or frequency) is very simple. It is obtained by adding a second equation for the enhancement factor $EF(\omega_2)$ and by redefining the optical intensity circulating inside to $I = EF(\omega_1) I_{input}(\omega_1) + EF(\omega_2) I_{input}(\omega_2)$. This neglects interference effects between the two beams, but then again the whole evaluation is a first-order-approximation.

To obtain the correct values that verifies the set of equations, I employed an iterative algorithm. First of all I assume a certain set of initial conditions for EF , ΔN , and ΔT , which I will call $EF|_0$, $\Delta N|_0$, and $\Delta T|_0$ respectively. For example one could choose to start from a *cold* system by setting them to zero. Then I evaluate the values the same variable at the next step by parametrically replacing the variables with their newly discovered values.

- 1) $EF|_1 = EF(\omega, EF|_0, \Delta N|_0, \Delta T|_0)$
- 2) $\Delta N|_1 = \Delta N(EF|_1, \Delta N|_0, \Delta T|_0)$
- 3) $\Delta T|_1 = \Delta T(EF|_1, \Delta N|_1, \Delta T|_0)$

which is easily generalized by putting $|_0 \rightarrow |_n$ and $|_1 \rightarrow |_{n+1}$. These three steps must be repeated until they converge to a steady state value.

At low input intensities nonlinearities are negligible, thus the output has the shape of the *cold* resonance, as shown in Figure 2.12.

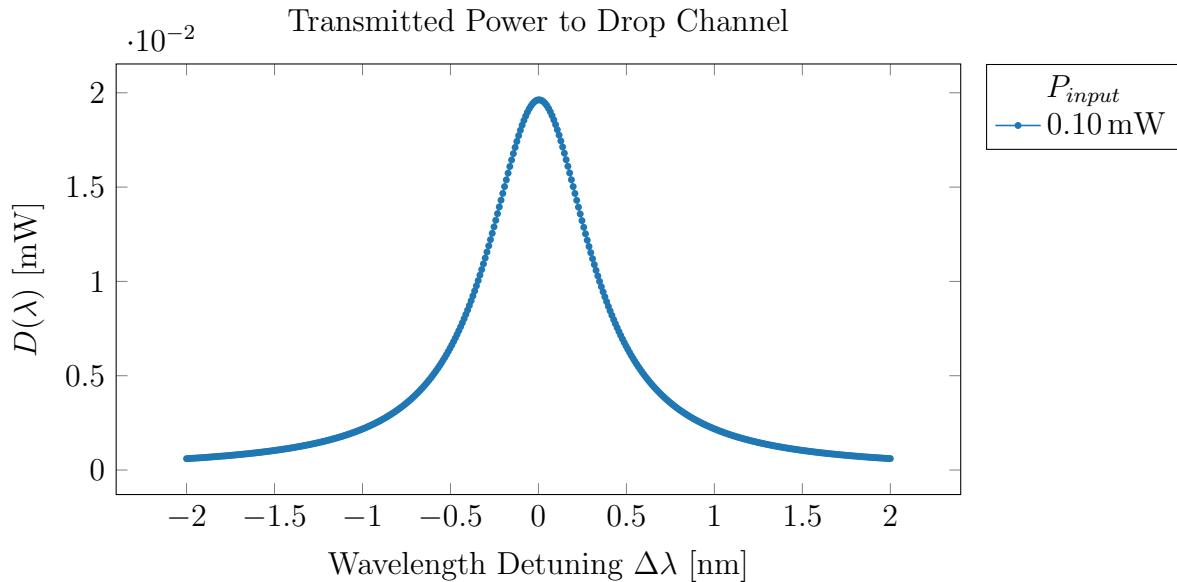


Figure 2.12: Shape of the resonance at 1537.417 nm, when probed with a low intensity signal. As expected is a lorentzian-type of curve, like in Figure 2.6.

At higher intensities the nonlinear effects become non negligible and therefore the number of iteration required increases. The algorithm evolution follows the same steps of the physical system: the system start off in a *cold* state and the gets progressively heated by the light that gets coupled inside the cavity.

However, the optical bistable region is also a region of instability for the algorithm, because there are two stable and infinite unstable states. As for the physical system, which selects the state depending on the past condition of the cavity, the same happens for the algorithm. By choosing as initial conditions the output values of the the variables obtained by the evaluation of the previous set of parameters, proved to be a successful approach to the problem.

Still, this method is not enough at high intensities, when the algorithm could not reach a steady state value due to the high number of iterations required. This is due to the fact that in certain regions, the values of consequent iterations oscillate around the steady state without any tangible progress toward convergence. This drastically increases the number of iteration required. For this reason I modified the algorithm from the vanilla iteration above to a more complex one. In this revised algorithm the variables at each new step are evaluated with a mix of the last and second to last values instead of just the last one. This method contrasts the phenomenon of uncontrolled value oscillation.

The results of the algorithm, for input parameters given by a sequence of signals of increasing optical power at a precise detuning from the resonance wavelength, map half of the hysteresis cycle. It is important in this case to set the initial conditions as the output values of the last evaluated state. Similarly, by setting the input parameters with a sequence of signals of decreasing optical power, the results map the other half of the bistability hysteresis cycle. The collection of the outcomes, obtained by repeating this evaluations for a set of different wavelengths on the red shoulder of the resonance, is shown in Figure 2.13.

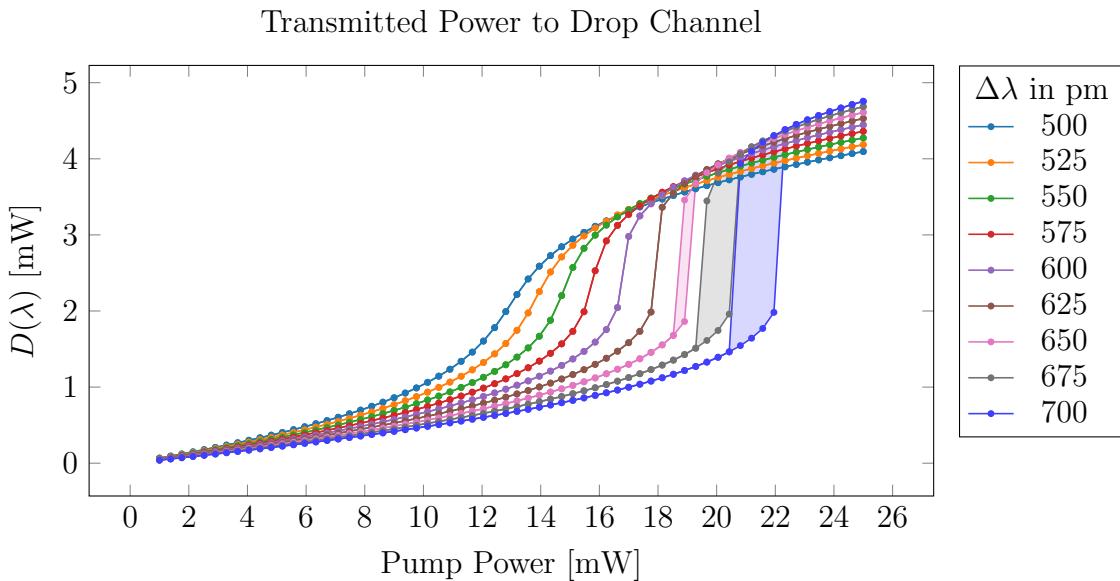


Figure 2.13: Hysteresis cycles described by the increase and decrease cycle of power, for input beams at different detunings from the *cold* resonance wavelength, in pm. All cases have longer wavelengths in comparison to the resonance wavelength.

The simulations predict that, for wavelengths near the resonance, the hysteresis cycle will disappear in a single curve, similar to a sigmoid. Forward and backwards sweeps will coincide in this case. On the other hand, inputs at farther wavelength from the resonance will show an increasing area of bistability and the jumps at the edges of the bistability regions are more sharp for higher input powers.

Another interesting way to map the bistability effect is to set the initial conditions for a fixed input optical power at a shorter wavelength than the resonance wavelength. Then evaluate a certain number of sample wavelengths in a progressive sweep toward longer wavelengths until the whole shape is mapped. Also in this case it is important to set the initial conditions of each run as the output values of last evaluation. Similar results can be obtained by inverting the sweep, thus showing another type of hysteresis cycle. Figure 2.14 shows an example of this behavior.

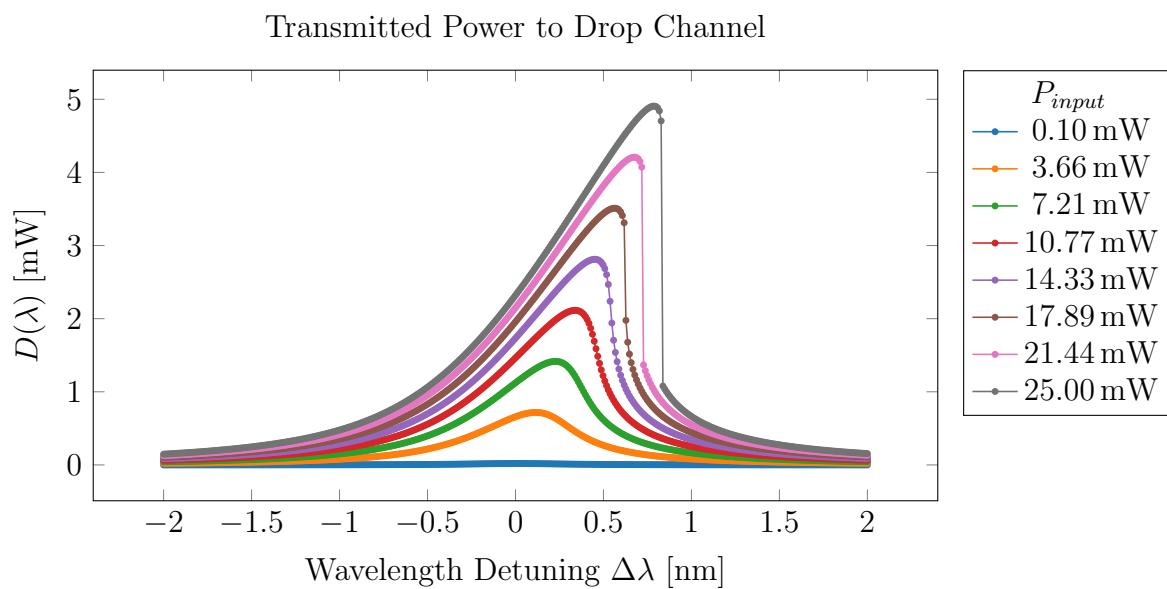


Figure 2.14: Behavior of wavelength sweeps around 1537.417 nm for a set of fixed input optical powers. The lower curve (blue), is the same portrayed in Figure 2.12.

Chapter 3

Samples, setup and experiments

All the experiments have been carried out on PIC developed within a FP7 project (IRIS, [42]) that aimed to develop an optical router. An optical router is a matrix of microresonators whose resonances are tuned at will to (re)direct an optical signal along a predefined path. We exploit the nonlinear optical response of these nodes to demonstrate the feasibility of an all optical nonlinear activation function made on PIC.

3.1 The samples

The IRIS project studied the design and the implementation of an integrated reconfigurable silicon photonics switch matrix, a routing device, as a replacement for electronic devices used in the telecommunication industry. The completed integrated photonic circuit consists of a matrix of waveguides crossing each other and linked by couples of racetrack resonators, thermally-controlled. At the ends of the waveguides other structures, interleavers and AWGs, allowed many signals at different wavelengths to be multiplexed/demultiplexed onto/from the same waveguide.

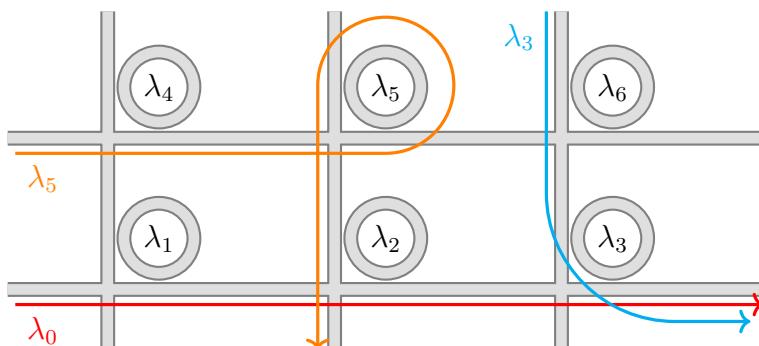


Figure 3.1: Scheme of a routing node of the full PIC of IRIS. λ_0 travels from the input port to the through port; λ_5 is redirected from the input port to the drop port; λ_3 is redirected from an add port to the through port.

Each node of the matrix can be modeled by a microring resonator in ADF configuration (see Subsection 2.2.2). Any signal that enters a node detuned from the resonance of the microring is directed toward the through port of the node. On the other hand, any signal that enters a node in resonance with the microring is redirected toward the drop port. Figure 3.1 shows the scheme of a node in the full PIC, where the signals is routed.

The development of the final IRIS chip has required several steps of improvement and few generation of test samples were fabricated with different optical properties. Thus the first step of the thesis has been to choose, among the many different devices available, the most promising ones. The structure selected (nicknamed *mini-matrix*) is shown in Figure 3.2 and it is composed by: a waveguide, coupled to eight drop channels through ring resonators. In this family of devices, there were mini-matrices built with single microrings, double microrings, single racetracks, or double racetracks. Figure 3.3 shows a simplified scheme of the *mini-matrix*.

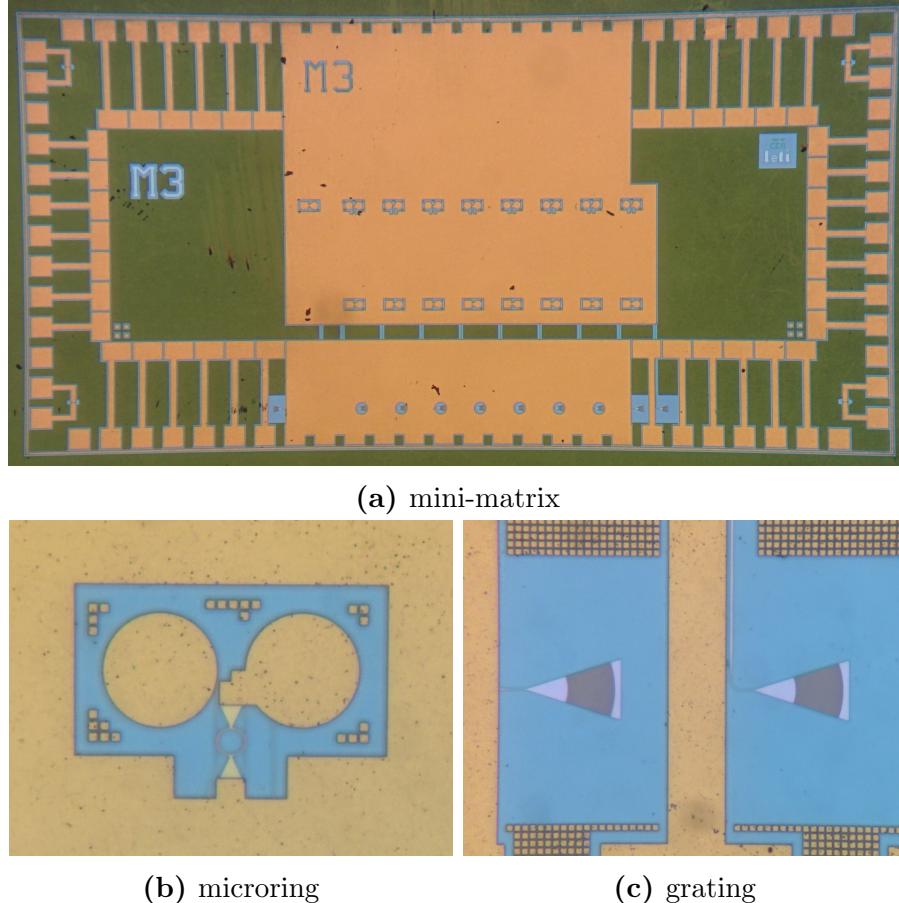


Figure 3.2: Magnified photograph of the *mini-matrix* device chosen (a) and two important structures: (b) microring resonator and (c) output grating couplers.

The resonance of each microring is strictly defined by the design of the structure, however it is ultimately changed by imperfections and uncertainties of the manufacturing process. Hence, since the optical router requires precise positioning of the resonances, their fine tuning is obtained by controlling the temperature of the material. Each one of the microring resonators in the structure has been manufactured with a thermo-electric heater, which is used to produce the thermo-optic effect (see Subsection 2.3.3) in order to induce the required resonance tuning [42]. Figure 3.4 shows the scheme of the microring resonator geometry, without any non-optical structures (e.g. heaters, metallic contacts).

3.1.1 Initial characterization

To obtain an initial characterization of the sample, I arranged an experimental setup to measure its fundamental attributes, such as the Free Spectral Range (FSR), the Full-Width-Half-Maximum (FWHM), and the Q factor of the resonances. These features can all be obtained through a study of the transmitted intensity in the through and drop channels, shown in Figure 3.3. A scheme of the experimental setup used is showed in Figure 3.5.

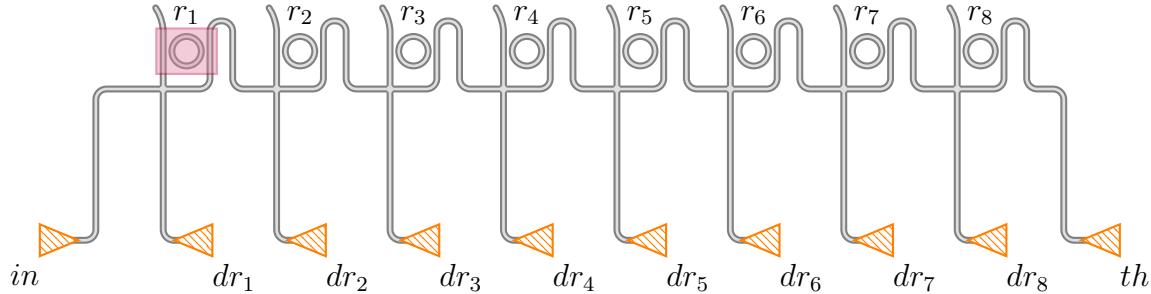


Figure 3.3: Simplified scheme of the *mini-matrix* integrated photonic circuit (structures are not in scale). Triangles (orange) are the grating couplers (in , dr_{1-8} , th). Waveguides and rings (r_{1-8}) are shown as thick grey lines. The bigger (purple) highlighted area is a coupling region, shown in Figure 3.4.

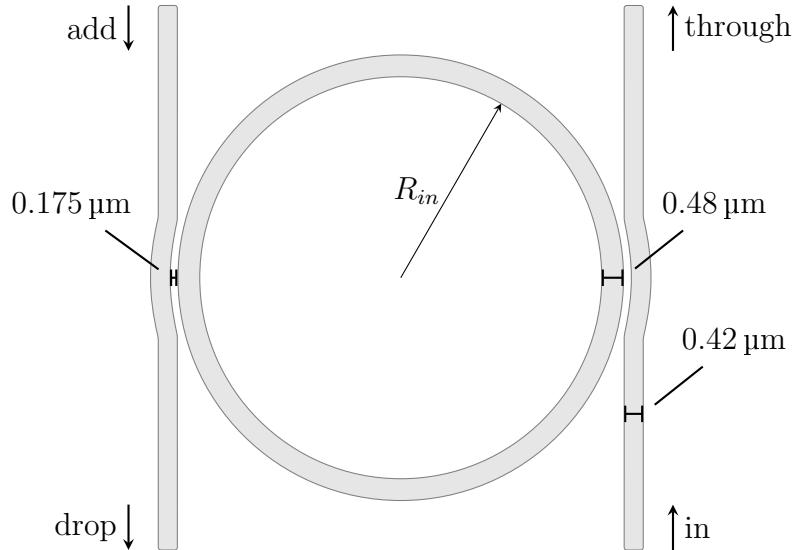


Figure 3.4: Microring optical cavity region, structures are in scale. The ring has an internal radius of $R_{in} = 4.43 \mu\text{m}$ and its width is of $0.48 \mu\text{m}$. The waveguides side coupled to the ring have a width of $0.42 \mu\text{m}$ and distance from the ring of $0.175 \mu\text{m}$.

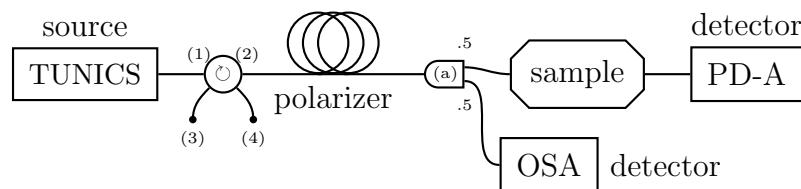


Figure 3.5: Scheme of the setup used to obtain an initial characterization of the optical nodes. The source is an external cavity tunable laser that produces a laser beam in the range 1500 nm to 1580 nm. The second element is an optical circulator and the third is a fiber polarizer. After the signal is split into two arms (50%:50%): one goes to an OSA and is used as reference during the measurements, while the other is fed into the sample. Finally the intensity of the signal passed through the node is read by a Germanium PD (PD-A).

The laser source used in this measures is an external cavity Yamatsu TUNICS T-100S tunable infrared laser. It produces a linearly polarized continuous wave laser beam of optical power up to 8 mW in the wavelength range 1500 nm to 1580 nm, with a minimum wavelength step of 1 pm. The generated light is injected into an optical circulator to protect the laser cavity from back-reflections. A circulator is a device that works similarly to a roundabout: input signals from the first port are directed only toward the second port. Similarly, signals injected in the second port are guided in the third port only and so on. The circulator used in the setup has four ports. Afterwards the beam is divided into two equal parts by means of a fiber coupler: one is analyzed by an Anritsu 9710C ?? Optical Spectrum Analyzer (OSA) and the other is brought to the sample. The output of the sample is coupled in a fiber and taken to the (PD-A) infrared photo-detector (PD)

The presence of an OSA after the source is due to the fact that the self-referencing mechanism of the laser that occurs during its initialization was not reliable. In fact, I observed that measures of the resonance taken in different days (after reinitializing the source) showed different values for the resonance. Hence I analyzed the wavelength emitted from the source with the OSA and observed that the actual wavelength emitted deviated rather sensibly from the supposed one. In order to correct this problem, I sampled the supposed-actual wavelength relationship with the OSA, and then I implemented an online correction by exploiting the its inverse function. In doing so, I assumed the OSA as a reference. Figure 3.6 shows a typical calibration, where the emitted wavelength is greater than the actual one at shorter wavelengths but smaller at longer ones.

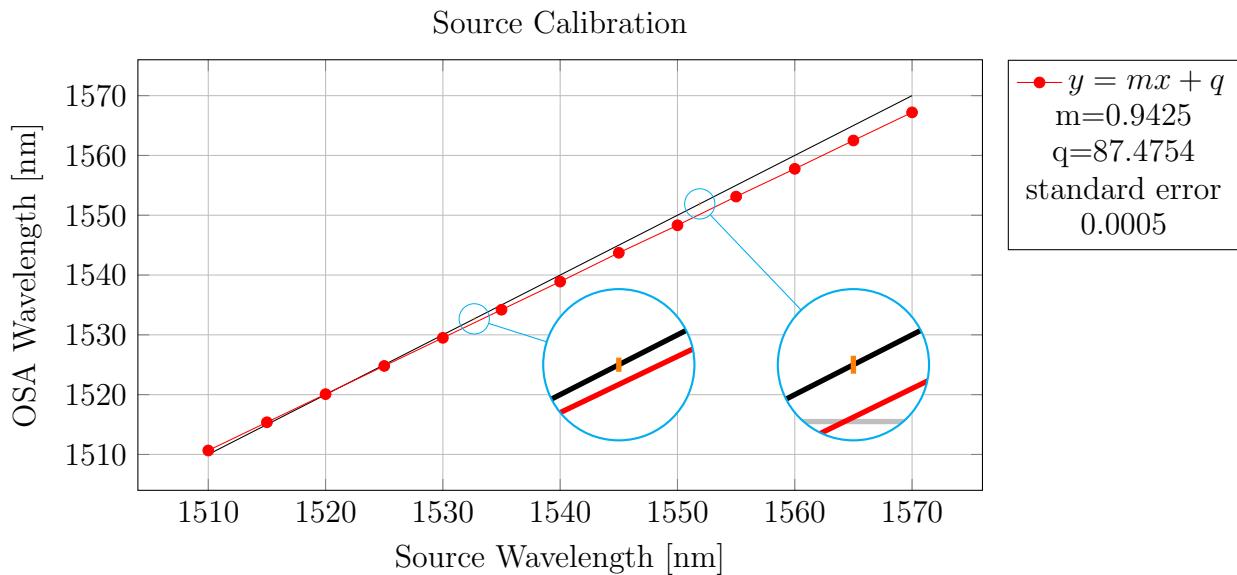


Figure 3.6: Typical wavelength dependence calibration. The black line shows the identity. The orange segments in the magnified frames represent the FWHM of the resonances.

All the structures built on the test chip are accessible via grating couplers and are designed to support modes in the TE polarization only. Grating couplers are planar periodic structures engineered to couple proper light bandwidth from free space into integrated waveguides and vice versa. In my case, the grating transmission is maximum for wavelengths from 1540 nm to 1550 nm and decreases outside. The coupling loss per grating is almost 5 dB at the maximum transmission, i.e. a third of the signal power is lost at each coupling.

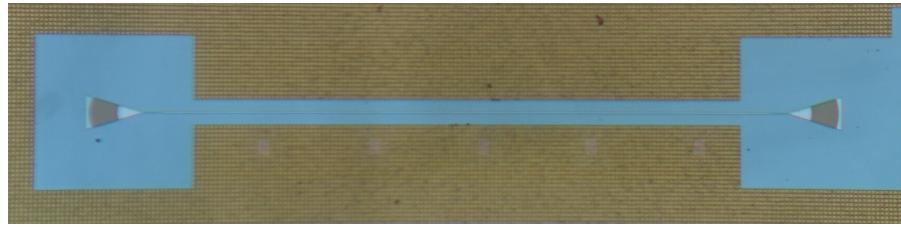


Figure 3.7: Grating transmission efficiency test structure.

To properly characterize the coupler transmission, I used a structure manufactured for this purpose. Depicted in Figure 3.7, it consists of an input grating coupler, a straight waveguide, and an output grating coupler. Its transmission efficiency is shown in Figure 3.8.

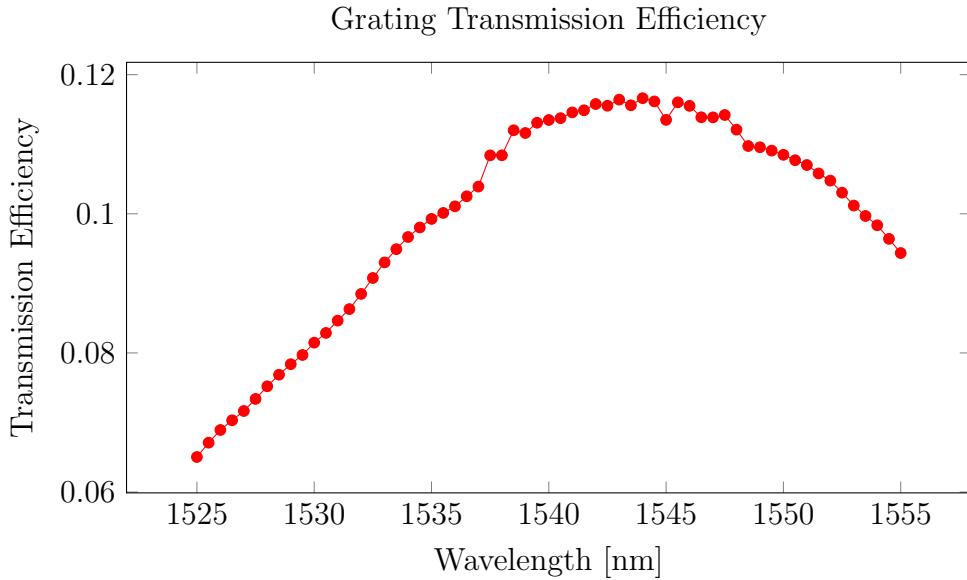


Figure 3.8: The maximum transmission efficiency is around 0.117 and the coupling loss per grating is almost 5 dB, average over five consecutive measures.

The sample is adjusted on a 2-DOF linear stage in between two 3-DOF linear stages that holds the input and output coupling fibers. Figure 3.9 shows the respective positions of the input and the output fibers in respect to the grating couplers.

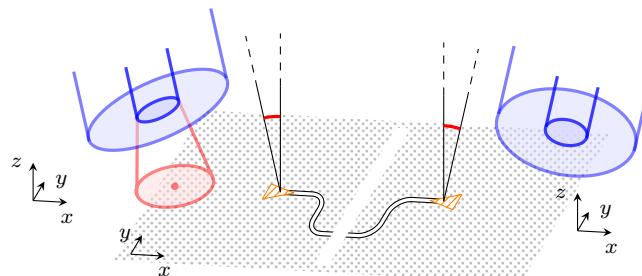


Figure 3.9: Alignment phase of the input and output fibers over the corresponding grating couplers. The optical fibers are aligned at an angle of 14° over the grating couplers in such a way to inject and collect as much light as possible. The sample was placed in a specific slot of a copper sample holder, which was secured onto a 2-DOF manual stage. The two fiber holders were placed on a 3-DOF manual stage each, to align them properly.

Once secured the input and output optical fibers in place on the linear stages, the alignment phase begins. First the source light is positioned at a wavelength that transmit on the correct channel, for example the resonance wavelength transmit in the drop channels but not in the through channel and vice versa. Then the optical fibers are carefully moved in the respective positions that maximize the coupled signal. The alignment have to be checked periodically, many times a day, to compensate for mechanical relaxations of the setup.

To obtain the best results, I also designed a new modular fiber holder with fixed angle (Figure 3.10) that was manufactured twice, one for each fiber. The main body has been manufactured in aluminum, except for the actual part that sustains the fiber, which was produced in iron. The metallic tip of the holder, appropriately grooved to house the optical fiber, can hold still the fiber with the help of little magnets. This technique aims at reducing as much as possible the vibrations of the fibers ends, which causes fluctuations in the amount of light coupled to the gratings.

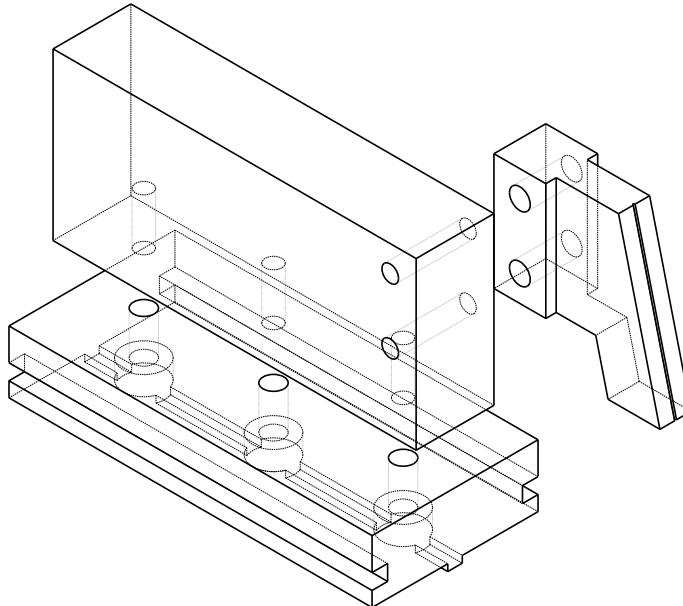


Figure 3.10: Exploded-view drawing of the fiber holder design. The holder is composed by three pieces. The tip is made of iron, in order to be used together with small magnets.

Since the PICs developed for IRIS support TE modes only, after the positions of the optical fibers have been optimized, the coupled signal is maximized by the use of a polarizer. The polarizer, placed between the fiber coupler and the sample, is a device composed by a long optical fiber. Such fiber is coiled in three separate loops which have the effect of a half-wave plate, one quarter-wave plate, and another half-wave plate respectively. Each coil can be manipulated, in a similar manner to the bulky quater-/half-wave plates, to change the output (linear) polarization of the light.

In order to further reduce the impact of noise on measures, I also boxed the three linear stages with a rigid enclosure. This decreased the noise, due to the movement of the fiber tips in air by at least one order of magnitude.

For the initial characterization, I studied the response of each output at input signals with low optical power and in the working range of the grating couplers, to identify the position of the resonances and their shape. Figure 3.11 shows the normalized transmission of the eight drop channel (dr_1 to dr_8) and the through channel (th). The first drop channel (dr_1) has the expected lorentzian shape, with a free spectral range of $FSR = (19.2 \pm 0.1)$ nm, while the shapes of all the other drop channels are highly distorted because of the light collected by dr_1 .

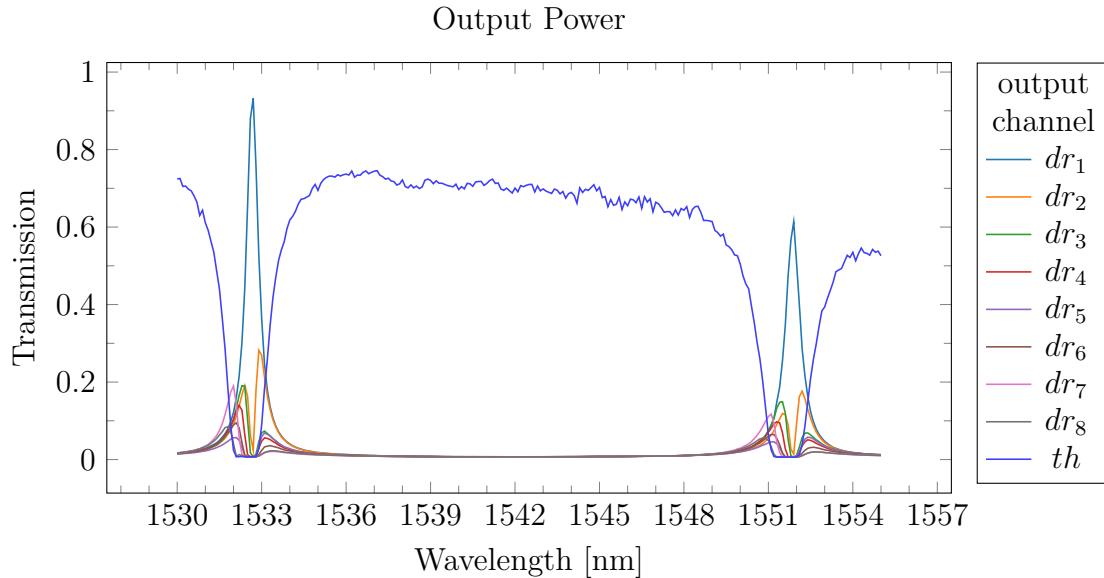


Figure 3.11: Transmission spectra of the drop channels (dr_1 to dr_8) and through channel th , average over five consecutive measures. The dr_1 channel shows the expected shape, whereas the other drop channels are clearly disturbed by the first one. The through channel shows a shape similar to the expected one, but it is actually the results of all the drop channels.

Another interesting fact is that the peak transmission of dr_1 channel is higher than the maximum transmission of th . This is due to two factors: primarily the coupling maximization process with the linear stages has a repeatability error of about 10 %, secondarily the th channel is a longer structure and hence its losses are expected to be higher than the 'shorter' dr_1 channel.

Since dr_1 is the only channel with an unperturbed transmission, I focused my attention on its response only. Specifically, I studied the resonance near 1551.9 μm : Figure 3.12 shows such resonance and highlights its center $\lambda_{res} = (1551.9 \pm 0.1)$ nm and width $FWHM = (0.60 \pm 0.05)$ nm. The estimated quality factor is $Q = 2600 \pm 200$.

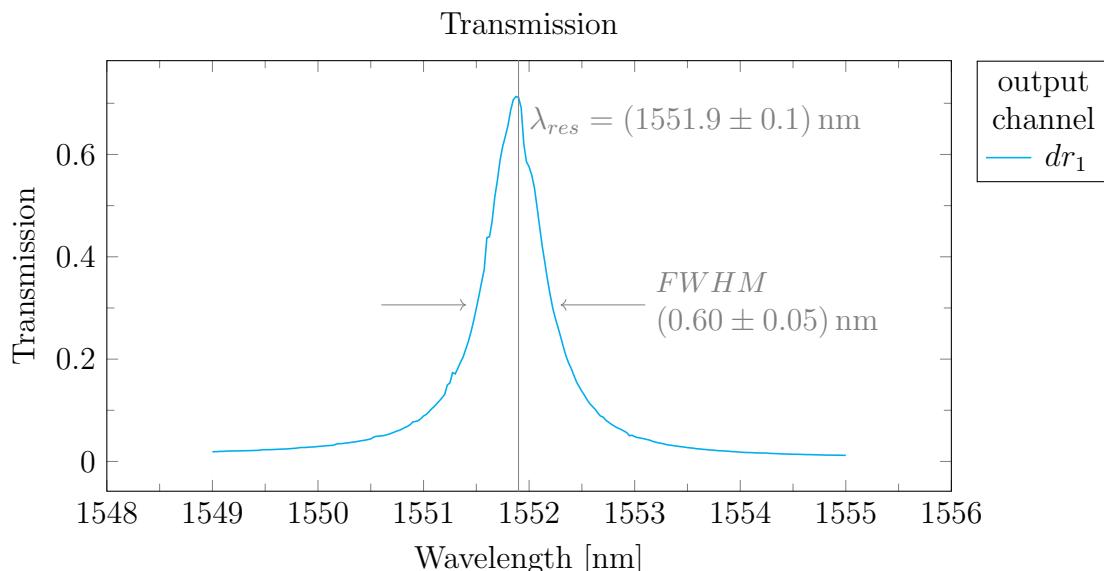


Figure 3.12: Resonance at 1551.9 nm, transmission to the drop port dr_1 , average over five consecutive measures.

3.2 Characterization of the Activation Function

To characterize the nonlinear response of the microring resonator the setup used for the initial measures has been modified. To induce the thermal bistability effect in the microring the signal of the TUNICS has been amplified using an Erbium-Ytterbiym-Doped Fiber Amplifier (EYDFA). In this configuration, the role of the circulator placed between the sources and the rest of the devices acquires even more importance, as back-reflected signals that enter optical amplifier in the wrong direction can cause serious damage to the equipment.

In addition to the amplifying stage, a remotely controlled Variable Optical Attenuator (VOA) was added between the polarizer and the sample, to quantitatively characterize the nonlinear response of the microring. The VOA employed is controlled by a voltage signal in the range 0 V to 5 V: it provides full transparency for 0 V and full attenuation for 5 V and, in between, the attenuator behaves similarly to a sigmoid function, shown in Figure 3.13. This curve has been identified with its closest analytical polynomial, to obtain an inverse formula that links transparency values in the range [0, 1] to the respective correct voltage value.

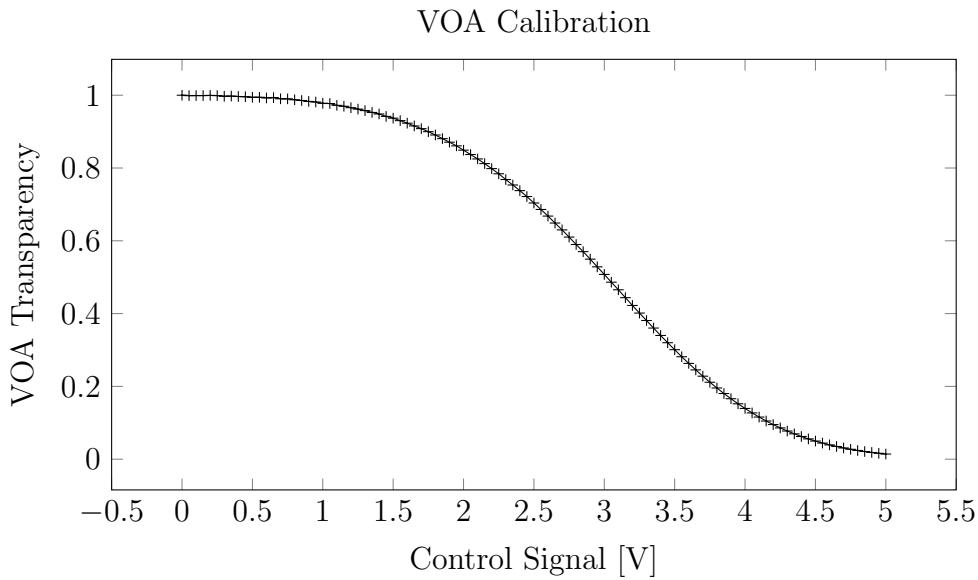


Figure 3.13: Typical calibration curve of the attenuation curve of the VOA. The data has been fitted with a spline to implement online control over the attenuation value.

This setup successfully induces the nonlinear response in the microring for detunings up to values comparable to half $FWHM$. However, instabilities of the source and the amplification stages in the optical power and in the wavelength of the signal increase the uncertainty on the measures. Specifically, the system shows a drift in optical power immediately after the source and amplification stages are switched on and random noise (e.g. 5 μ W of noise over 1 mW at the detector). Moreover, the wavelength of the light generated shows a slow drift toward shorter wavelengths, which becomes important for long measurements. The typical wavelength stability of the source is rated at ± 3 pm/h (from instrument datasheet).

The first corrective measure is obtained by allowing the system to thermalize. The most critical instrument is the EYDFA, which might take up to half an hour to reach a steady state. This step reduces the slow drift in the output power of the source, but does not suppress the instantaneous noise in power and wavelength of the signal.

In order to provide a correction for the instantaneous noise, a small part of the signal is collected from the main path, before the VOA, via a fiber coupler (99%:1%) and it is the equally distributed via a second fiber coupler (50%:50%) to the OSA and to a second infrared

PD (PD-B). The completed setup is shown in Figure 3.14.

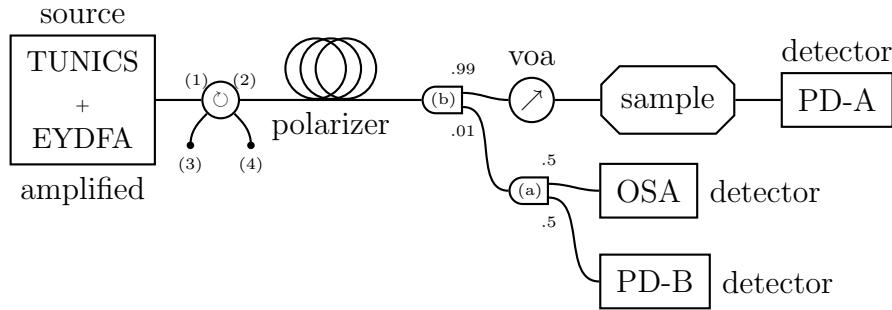


Figure 3.14: Scheme of the setup used to obtain characterization of the thermal bistability in the microring resonator. The source (TUNICS) is amplified by an EYDFA and is followed by the circulator and by the polarizer. Then a small part of the signal is collected by a fiber coupler (99%:1%) and it is equally distributed between the OSA and an additional infrared PD (PD-B). The signal in the main path passes through a Variable Optical Attenuator (VOA), then it is injected into the sample, and finally extracted to be read by the first PD (PD-A).

In this configuration, PD-B measures the instantaneous power in input at the VOA and, together with the value of transparency selected on the VOA, provides control over effective power coupled into the PIC and correction over the noise in the power of the source. The sampling rate of PD-B is the same of PD-A and is limited by the acquisition board (1 kHz).

On the other hand, the OSA provides a check over the wavelength of the source, but no correction can be implemented. Moreover, the measures obtained by means of the OSA, due to intrinsic limitations of the instrument and to the high complexity of the measure itself, have a much lower sampling rate in comparison to the measures given by the PDs.

Since this setup has been employed to measure several times the hysteresis cycle, for the same input wavelength, the measurement could run for several minutes consecutively. In order to collect a significant number of points, without important drifts in the signal wavelength, the sampling time had to be reduced as much as possible. For this reason, the OSA has been employed as a check on the signal wavelength, before and after each measure of the hysteresis loops.

To summarize, the data collected is composed by the optical power measured before the VOA, by the transparency of the VOA, and by the optical power measured at the output. This three-point measurement allows more robustness against power fluctuations.

3.2.1 Bistability wavelength dependence

The first feature studied is the dependence of the overall shape of the bistability hysteresis loop on wavelength detuning. As seen in Subsection 2.4.3, the form of the bistability changes with the detuning from the resonant wavelength. This behavior is also observed in the transmission spectra experimentally measured, as shown in Figure 3.15.

As expected, the region of bistability disappears into a sigmoid for small detunings, while becomes larger for increased detunings. The range of wavelengths with which the sample has been probed is upper limited due to the growing pump power required to activate the bistability.

Even though the behavior observed experimentally is similar to the one obtained with the simulations, there are some differences: first of all, in the experiments the transmission for high input power saturates at the same value for different detunings. Moreover, in the experiment all wavelengths shows a peak and a decrease in transmission from a certain level of input power.

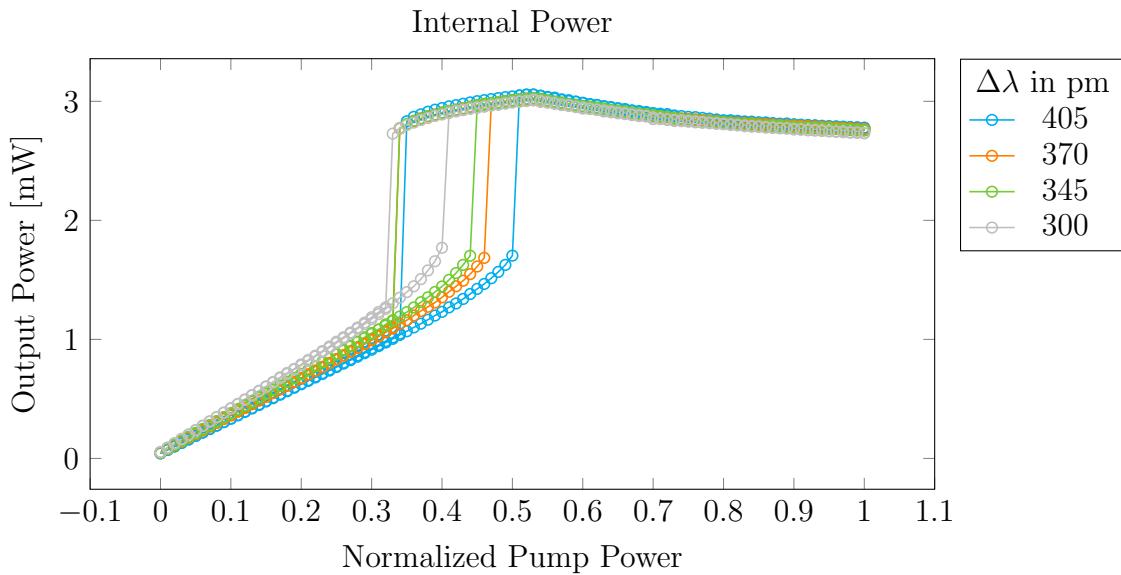


Figure 3.15: Many bistability loops at different wavelengths. The loops that are closer to the microring resonance have smaller bistability regions or they have not one at all. Loops that are farther from the resonance have larger bistability regions.

On the contrary, in the simulations the output power does not saturate and shows different transmission efficiency for each detuning (see Figure 2.13).

Several are the factors that might produce these discrepancies between the experiments and the simulations. First of all, the setup and especially the sample are composed by many different structures. Each of them has its unique response to high values of optical power. Secondly, the parameters employed in the simulation are an estimate of the actual parameters that define the microring resonator and its coupling with the waveguides. Lastly, the theoretical model implemented in the numerical calculations is an approximation of the physical phenomena that occur in the system.

Bistability region edges

In order to precisely characterize the abrupt jumps on the edges of the bistability region, the stability of the system is crucial. Hence I implemented a specific series of measurements, which samples only the nearest part of the edge: the amount of points sampled is smaller, however the complete measurement is faster and the sampling near the bistability edges is more dense.

The first step of each measurement is setting the source wavelength at a specific value, which is measured by the OSA. Then the serial measurements begins: for ascending loops, at the beginning VOA is initialized at full attenuation and then is used to sample the edge with small increasing steps. For descending loops, the VOA is initialized at full transparency and the steps are decreasing instead. After the predefined number of loops is completed, the wavelength of the source is measured with the OSA again.

The data collected included twenty loops for each detuning, which are five almost equally spaced in a wavelength range from 1552.2 nm to 1552.3 nm. Each set of data has been analyzed to find the abrupt jumps from a state to the other. A simple study of the discrete derivative was employed to achieve that. The same procedure has been applied both to the rising as well as to the descending loops. The results are five data points for the first kind of loops and just as much for the descending one, as shown in Figure 3.16. The wavelength of each loop is measured by checks with the OSA before the start and after the completion of each loop.

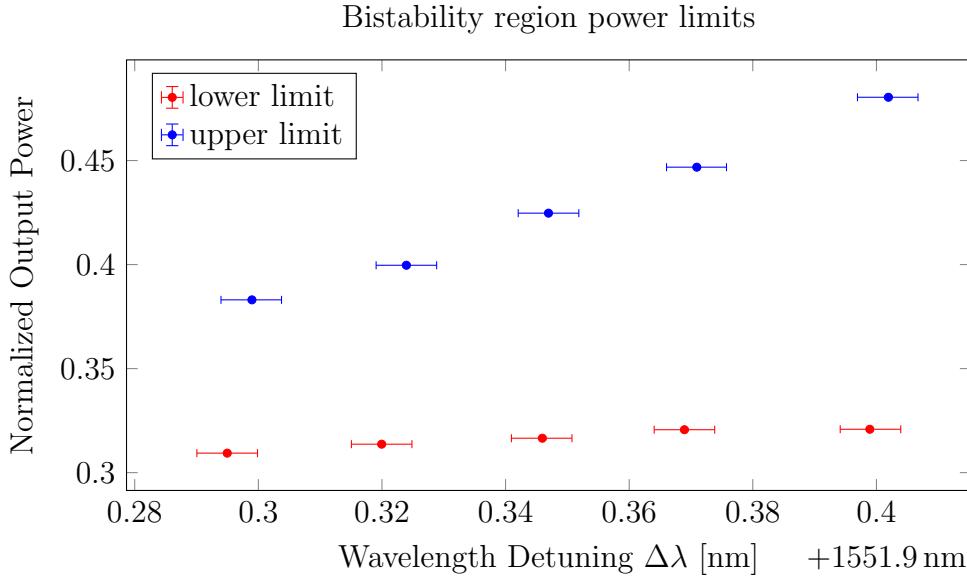


Figure 3.16: Wavelength dependence of the bistability region edges. Upper limits are shown as blue dots, while lower limits are shown as red dots. The wavelength is estimated by checks in between consecutive measures obtained with the OSA. Its uncertainty is assumed to be 5 pm (TUNICS Wavelength Setting Repeatability) for all the points. The power uncertainty is given by the statistic of the data collected.

The uncertainty on the wavelength position is assumed to be the 5 pm for all the points, which is the wavelength setting repeatability of the TUNICS. The uncertainty on the power is evaluated from the statistic of all the loops measured at the same wavelength. Hence it is defined as the RMS error of twenty data points for each wavelength.

It is interesting to observe that the experimental data, much like the simulations, show a steeper dependence on wavelength for the upper limits than for the lower limits. Nevertheless, both quantities seem to increase with wavelength more distant from the resonance. Moreover, the position of the edges of the bistability region seem to be precisely defined by the parameters of the system.

3.3 Optical Activation Function

The first step to implement an optical activation function is to test the nonlinear response of the microring inside a simulated ANN. In order to do so, I fitted the nonlinear response to obtain an approximated analytical description. The analytical description was then inserted as activation function in the same models used in Subsection 1.6.3. Also the training and validation/test datasets are the same, hence it is possible to compare the results of the two activation functions: the Rectified Linear Unit (ReLU) and the nonlinear optical response.

3.3.1 Optical bistability as nonlinear activation function

Having obtained a description of the response of the microring resonator in a wide range of wavelengths, I implemented the shape of such response in a simulated ANN. I chose a representative set of data: specifically, the data belongs to the increasing half of the bistability loop for input light at 1552.300 nm. Since the response is defined apart of a multiplication factor, I normalized the function such that its domain and codomain were represented by the interval $[0, 1]$.

As seen in the previous Section (3.2.1), the nonlinear response shows a decay in transmission after a certain input value. In the fit I neglected all the data after the maximum, because the cause of such behavior is still unknown: the theoretical model does not predict it. Moreover, this choice simplifies the analytical description of the activation function, which is helpful because simpler activation functions are less computational intensive on the simulations. Faster simulations, in turn, help me in the heuristic test of the many parameters of each network model.

The functions employed to fit the data is a combination of function that are already implemented in PyTorch library. This allows a simple implementation in the simulations. Specifically the function is composed by *ReLU* functions only

$$f_{fit} := a f_{ReLU}[x] + b f_{ReLU}[x - x_0] + c f_{ReLU}[x - x_1], \quad (3.1)$$

where the parameters a , b , x_0 , c , and x_1 are fixed by the fit.

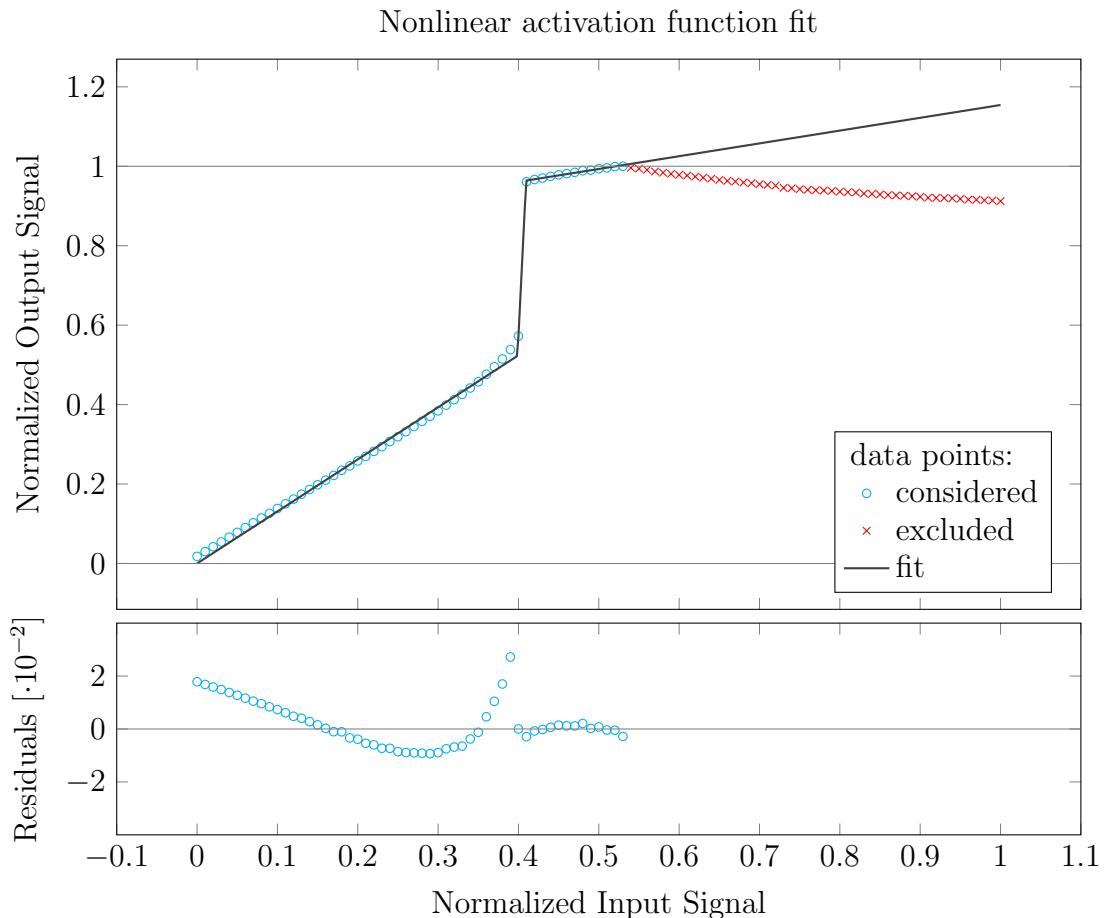


Figure 3.17: Curve fitting and residuals on a set of data. The nonlinear curve is the increasing half of a bistability loop. The fit is quite close to the original values, except near the bistability edge.

Figure 3.17 shows the nonlinear response function, the fit, and the residuals. The data are distinguished in the data points considered and those excluded from the fit. The curve fit is very close to the original data points, with a maximum residual of 0.027 near the bistability edge.

3.3.2 Optical Nonlinear Response in ANN

In Section 1.6 I trained several models of ANN, in which the activation function was the ReLU. The networks had 10 inputs (D_{in}) and 11 outputs (D_{out}) and the last layer was composed by 11 nodes as well. The number of hidden layers (n_H) and the number of nodes per hidden layer (D_H) were free parameters.

Using f_{fit} as activation function, I tested models with $n_H = [2, 3]$ and $D_H = [11, 22]$ and obtained correctness up to $\sim 50\%$. In this case, the last hidden layer implemented both the weighted sum and the activation function. I trained the models for 4000 epochs, in order to be consistent to what I did for the ReLU based ANNs. The optimizer is the SGD, parametrized by the learning rate l_r and the momentum m . The learning rate scheduler is *StepLR* (γ , T) and the criterion used is Cross-Entropy Loss (CEL).

The simulation parameters are: $l_r = 6 \times 10^{-3}$, $m = 0.5$, $\gamma = 0.8$, and $T = 500$.

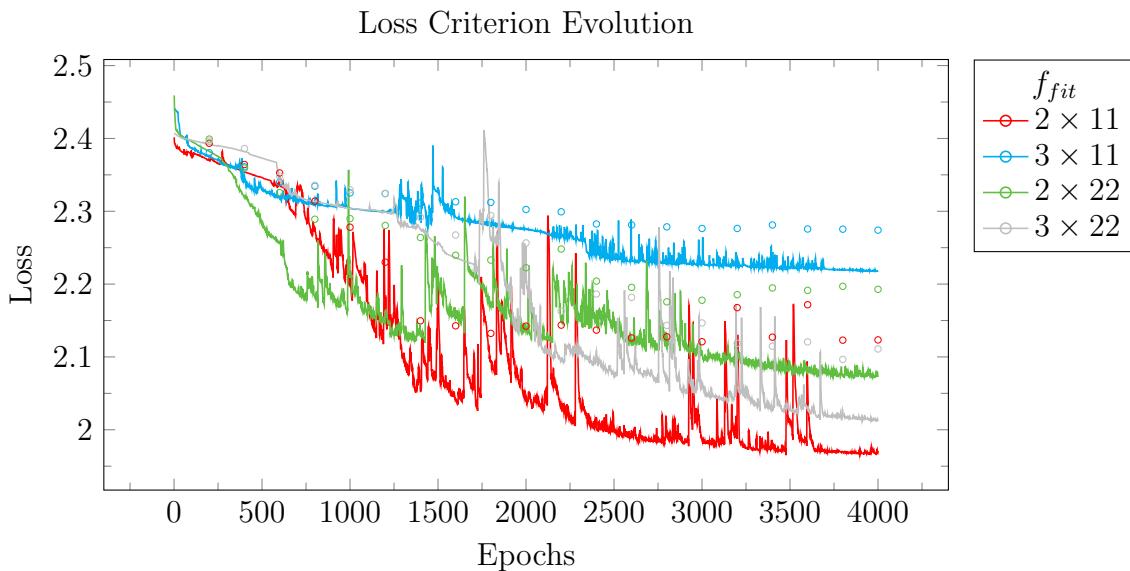


Figure 3.18: Evolution of the loss criterion (solid lines) throughout the epochs. Points represent the values of the loss criterion on the validation dataset, carried out repeatedly during the training.

Figure 3.18 shows the typical evolution of the loss on both the train and the validation datasets, for each model. Despite the models are trained for 4000 epochs, the lowest loss reached is near 1.9, whereas in Section 1.6 loss values below 0.5 were reached. This factor it is connected to the lower learning rate employed and together they points towards the conclusion that it takes longer to train f_{fit} based models in comparison to f_{ReLU} based ones. Moreover, overfitting does not appear yet inside the 4000 epochs.

Another important behavior is the instability of the evolution: the improvement of the loss function is scattered with spikes. This is due, at least partially, to the complex shape of the surface of loss function of the f_{fit} based models. Moreover, higher values of l_r induce higher instability, whereas lower values are not train fast enough.

The results for the distinct models that implements the optical bistability activation function are collected in Table 3.1. The percentage of correct answers on the test dataset is lower than the results obtained for the ReLU based ANNs. This might be caused by the fact that f_{fit} based models work better with lower l_r and are therefore harder to train. However, it is important to note that the optical node in the most promising configuration (2×22) gives near the 30% of correct answers, which is above the $p_{rnd} = 9.09\%$ threshold of random choice.

activation function	no. hidden layers	no. nodes per layer	other parameters	Percent correct
f_{fit}	2	11	linear	$(24 \pm 5) \%$
f_{fit}	3	11	linear	$(19 \pm 5) \%$
f_{fit}	2	22	linear	$(30 \pm 6) \%$
f_{fit}	3	22	linear	$(26 \pm 6) \%$

Table 3.1: Results of the f_{fit} activation function on several network topologies. The statistics is made on 10 different initializations of the same model: the error on the percentage is the standard deviation.

Conclusion

Summary

This work was focused on using an optical microring resonator as nonlinear node of a feedforward artificial neural network. The problem has been tackled simultaneously by two directions: the experimental measurement of the nonlinear response of the node and its simulation by means of software libraries.

The first part of the thesis aimed to introduce the fundamental concepts of artificial neural network and machine learning. It was followed by the arrangement of a simulated neural network on the PyTorch software library.

The central part of the thesis explained the physical effects related to the working principles of an optical microring resonator in the All-Drop Filter configuration. These effects initially have been studied in their analytical linear approximation. Subsequently, nonlinear perturbations have been introduced to the resonator response. The physical effect produced have been analyzed numerically.

In the last part, my work has been to assemble an experimental setup and a related measurement methodology. This setup has been used to gather experimental data on the actual response of an optical microring resonator.

Afterwards, the gathered data proved to be useful in the implementation, inside the simulated artificial neural networks, a nonlinear activation function resembling the response of the resonator.

Accomplishments

During the period of time of this work I assembled the experimental setup to study the integrated photonic device. The experimental measures have been carried out with a continuous enhancement of the setup.

The properties and the physical mechanisms that regulate operation of the resonator have been characterized. To obtain a more profound knowledge of the physical phenomena, I successfully carried out numerical simulations.

I implemented a small neural network in PyTorch, which is a software library used by the machine learning community. The code was initially used to solve a benchmark problem, with standard artificial neural network parameters and functions. Finally I implemented also a new activation function that mimics the response of a microring resonator in All-Drop Filter configuration. This allowed me to confront the operation of the standard function with the one based on the optical bistability.

Improvements

The setup should be enhanced by working on its bottlenecks, which are, in particular, the light source and the photodetection system. The light source should be stabilized as much as possible, both in power and in wavelength. This is worth because the microring resonators are quite sensitive to the parameters of the input signals.

On the other hand, the system of the photodetectors could be improved in the response time as well. In fact, due to both hardware and software limitations, a full measure can take a long

time, depending on the number of sampling points. The photodetectors should be rearranged in a way such that their control could be simplified. Moreover, the control software has been developed to be adaptive rather than fast in its operation. Working on both the software and the hardware of the detectors should improve the single measurement speed. Hence, the time spent on the full measure will benefit too.

A better experimental setup would allow to acquire more data in less time and therefore to obtain a superior statistics. In turn this data, collected with a robust method, will consent also to complete a quantitative analysis. From this new work elaboration speed, energy efficiency, and requirements for node cascadability could be estimated. Eventually this will help the design of a class of new ad hoc devices, whose characteristics will be defined suit neural network nodes operation.

Future perspective

As much important as the accomplishments, the futures studies which can start from this work or make use of part of it are many and interesting.

First of all, the natural continuation on this work on feedforward neural networks can start with the improvements suggested above. Improve the experimental setup, explore new zones in the parameters space and realize a quantitative analysis. Hopefully, this will lead to the definition of an all-optical framework for feedforward neural network development and operation.

Moreover, starting from the same framework with or without some modifications, an interesting development will be to research on the methods employed by reservoir computing. All-optical reservoir computing could put together the efficiency of the algorithm to the inherent complexity of structures based on multiple coupled optical cavities.

Furthermore, the development of a time-dependent theory would allow the analysis of physical effects in different regimes other than the quasi-static one. This will consent to investigate new phenomena and test their properties against the requirements for the neural network node operation. By exploiting new optical effects, we will be able to examine other implementations, such as spiking neurons. The output for this type of neurons depends both on the signals intensity and rate. In this perspective, for example, the presence of a hysteresis loop could be exploited to implement some kind of short-term memory in the node.

In common to all these projects, there must be a simultaneous development of the software simulation, based on standard libraries. Specifically, work must be done to implement physical characteristics in the simulations, such as stochastic noise on the input and output data and upper and lower boundaries to the signals values. The development of the simulations on standard libraries is important because it allows performance comparison between the software functions and the proposed physical effects.

Bibliography

- [1] M. D. Godfrey and D. F. Hendry, “The computer as von neumann planned it”, *IEEE Annals of the History of Computing*, vol. 15, no. 1, pp. 11–21, 1993.
- [2] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.”, *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [3] J. E. Stone, D. Gohara, and G. Shi, “Opencl: A parallel programming standard for heterogeneous computing systems”, *Computing in science & engineering*, vol. 12, no. 3, pp. 66–73, 2010.
- [4] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. J. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio, “Theano: New features and speed improvements”, *CoRR*, vol. abs/1211.5590, 2012. arXiv: 1211.5590. [Online]. Available: <http://arxiv.org/abs/1211.5590>.
- [5] R. A. Nawrocki, R. M. Voyles, and S. E. Shaheen, “A mini review of neuromorphic architectures and implementations”, *IEEE Transactions on Electron Devices*, vol. 63, no. 10, pp. 3819–3829, 2016.
- [6] *Humanbrainproject.eu*. [Online]. Available: <https://www.humanbrainproject.eu/en/silicon-brains/>.
- [7] H. Markram, “The human brain project”, *Scientific American*, vol. 306, no. 6, pp. 50–55, 2012, ISSN: 00368733, 19467087. [Online]. Available: <http://www.jstor.org/stable/26014516>.
- [8] M. Foedisch and A. Takeuchi, “Adaptive real-time road detection using neural networks”, in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, Oct. 2004, pp. 167–172. DOI: 10.1109/ITSC.2004.1398891.
- [9] S. Soman, M. Suri, *et al.*, “Recent trends in neuromorphic engineering”, *Big Data Analytics*, vol. 1, no. 1, p. 15, 2016.
- [10] T. F. de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, “Progress in neuromorphic photonics”, *Nanophotonics*, vol. 6, no. 3, pp. 577–599, 2017.
- [11] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [12] D. O. Hebb *et al.*, *The organization of behavior: A neuropsychological theory*, 1949.
- [13] B. Farley and W. Clark, “Simulation of self-organizing systems by digital computer”, *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 76–84, 1954.
- [14] N. Rochester, J. Holland, L. Haibt, and W. Duda, “Tests on a cell assembly theory of the action of the brain, using a large digital computer”, *IRE Transactions on information Theory*, vol. 2, no. 3, pp. 80–93, 1956.

- [15] F. Rosenblatt, “The perceptron a perceiving and recognizing automaton”, *tech. rep., Technical Report 85-460-1*, 1957.
- [16] M. L. Minski and S. A. Papert, “Perceptrons: An introduction to computational geometry”, *MA: MIT Press, Cambridge*, 1969.
- [17] C. Clabaugh, D. Myszewski, and J. Pang, *Neural networks - history*, 2000. [Online]. Available: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/index.html>.
- [18] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network”, in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, IEEE, 2014, pp. 844–848.
- [19] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool, “Wespe: Weakly supervised photo enhancer for digital cameras”, *arXiv preprint arXiv:1709.01118*, 2017.
- [20] D. D. Johnson, “Generating polyphonic music using tied parallel networks”, in *International Conference on Evolutionary and Biologically Inspired Music and Art*, Springer, 2017, pp. 128–143.
- [21] J.-P. Briot, G. Hadjeres, and F. Pachet, “Deep learning techniques for music generation-a survey”, *arXiv preprint arXiv:1709.01620*, 2017.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [24] Pytorch.org. [Online]. Available: <http://pytorch.org/about/>.
- [25] Uc irvine machine learning repository. [Online]. Available: <http://archive.ics.uci.edu/ml/index.php>.
- [26] Corning optical cable. [Online]. Available: <https://www.corning.com/optical-cables-by-corning/worldwide/en/products/thunderbolt-optical-cables.html>.
- [27] G. T. Reed, *Silicon photonics: The state of the art*. John Wiley & Sons, 2008.
- [28] B. E. Saleh, M. C. Teich, and B. E. Saleh, *Fundamentals of photonics*. Wiley New York, 1991, vol. 22.
- [29] G. Agrawal, *Applications of nonlinear fiber optics*. Academic press, 2001.
- [30] S. Chen, L. Zhang, Y. Fei, and T. Cao, “Bistability and self-pulsation phenomena in silicon microring resonators based on nonlinear optical effects”, *Optics Express*, vol. 20, no. 7, pp. 7454–7468, 2012.
- [31] T. S. Moss, G. J. Burrell, and B. Ellis, *Semiconductor opto-electronics*. Butterworth-Heinemann, 2013.
- [32] T. J. Johnson, M. Borselli, and O. Painter, “Self-induced optical modulation of the transmission through a high-q silicon microdisk resonator”, *Optics express*, vol. 14, no. 2, pp. 817–831, 2006.
- [33] M. M. Mancinelli, “Linear and non linear coupling effects in sequence of microresonators”, PhD thesis, University of Trento, 2013.
- [34] G. Priem, P. Dumon, W. Bogaerts, D. Van Thourhout, G. Morthier, and R. Baets, “Optical bistability and pulsating behaviour in silicon-on-insulator ring resonator structures.”, *Optics express*, vol. 13, no. 23, pp. 9623–9628, 2005.

- [35] M. Borghi, “Linear, nonlinear and quantum optics in silicon photonics”, PhD thesis, University of Trento, 2016.
- [36] B. J. Shastri, T. F. de Lima, A. Tait, M. A. Nahmias, and P. R. Prucnal, “Emergence of neuromorphic photonics”, in *Photonics in Switching*, Optical Society of America, 2017, PTu3C-4.
- [37] G. Van der Sande, D. Brunner, and M. C. Soriano, “Advances in photonic reservoir computing”, *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017.
- [38] N. D. Haynes, M. C. Soriano, D. P. Rosin, I. Fischer, and D. J. Gauthier, “Reservoir computing with a single time-delay autonomous boolean node”, *Physical Review E*, vol. 91, no. 2, p. 020 801, 2015.
- [39] A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, “All-optical reservoir computer based on saturation of absorption”, *Optics express*, vol. 22, no. 9, pp. 10 868–10 881, 2014.
- [40] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, *et al.*, “Deep learning with coherent nanophotonic circuits”, *Nature Photonics*, vol. 11, no. 7, p. 441, 2017.
- [41] A. N. Tait, T. F. Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, “Neuromorphic photonic networks using silicon photonic weight banks”, *Scientific Reports*, vol. 7, no. 1, p. 7430, 2017.
- [42] F. Testa, C. J. Oton, C. Kopp, J.-M. Lee, R. Ortuño, R. Enne, S. Tondini, G. Chiaretti, A. Bianchi, P. Pintus, *et al.*, “Design and implementation of an integrated reconfigurable silicon photonics switch matrix in iris project”, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 6, pp. 155–168, 2016.

Acknowledgements