Politecnico di Torino

Politecnico di Torino

Master of Science Course
in Electronic Engineering
A.Y. 2023/2024
Graduation Session October 2024

Study of Microring Nonlinearities in Silicon Photonics for Neuromorphic Computing

Supervisors:

Prof. Mariangela Gioannini
Dr. Marco Novarese
Dr. Cristina Rimoldi

Candidate:

Salvatore Salpietro

To my parents, my sister and my aunt,
for helping and supporting me throughout these years of study and
for giving me the opportunity to achieve this dream.

Abstract

The objective of this master's thesis is to develop innovative tools that exploit the nonlinear properties of optical micro-ring resonators, integrated in silicon photonics platform, to create a neural network capable of solving complex tasks such as image and voice recognition, time series prediction, using the reservoir computing (RC) technique.
These micro-ring resonators are part of a photonic integrated circuit in silicon that can be fabricated in the same CMOS foundry of electronic integrated circuits. Photonic implementations of reservoir computing are promising applications due to their ability to achieve very high bandwidth with fast and light training and to exploit the intrinsic parallelism of optical signals.
My thesis has three main objectives: the first is to develop a compact fully optical RC model based on a silicon optical micro-ring resonator to classify datasets into high-speed clusters; the second is to analyze and understand the nonlinear optical effects in micro-rings during these processes; the third is to create an experimental setup to validate this RC method. The experimental work is carried out at the Interdepartmental Center Photonext at PoliTO.
To achieve these goals, I used a silicon micro-ring resonator in add-drop configuration in a photonic integrated circuit, designed at PoliTo and fabricated in IMEC (Belgium).
To demonstrate how the nonlinearities of micro-ring resonators can be exploited within a neural network, I reproduced the experiments and the theory presented in the article [1] by Massimo Borghi, Stefano Biasi, and Lorenzo Pavesi.
I began my study using a compact theoretical model, developed by Dr. Marco Novarese in his PhD thesis [2] that describes the impact of nonlinear (NL) effects on the performance of ring resonators.
These studies are already present in the literature but have not fully explored the potential of silicon photonics for AI applications.
Neural networks are structured to mimic the human brain, imitating how neurons exchange signals. They use training data to learn and improve their accuracy over time. Once the desired accuracy is achieved, neural networks become powerful tools capable of organizing datasets into clusters at high speed. Reservoir computing is an implementation of a recurrent neural network that divides the network into three parts: the input layer, the reservoir, and the output layer. The advantage of this technique is that it is not necessary to train the recurrent part (the reservoir) but only the output layer and this simplifies and speeds up the training process.
In neural networks, the input is mapped into a higher-dimensional space and processed by nonlinear functions while it passes from one layer to the next. I replicate this functionality using the nonlinear response of the micro-ring resonator. By mapping the input into a higher-dimensional, the probability that the data will become separable increases.
In this context, I addressed classification problems such as the delayed XOR task and the recognition of the Iris species. For dataset recognition, I encoded the input into the temporal evolution of the intensity of a pump laser. The input must be injected into the system through the only physically nonlinear node (the micro-ring) and this requires, however, a pre-processing of the

input, defined as masking. The masking procedure serves two purposes: to sequence the input and to maximize the dimensionality of the system that depends by the number of the virtual nodes (specific points in time defined on the output results). The result is a piecewise constant input series, with constant intervals corresponding to the separation between virtual nodes.
Therefore, masking is nothing more than a multiplication by a weight matrix Nv x N with Nv the number of virtual nodes and N the number of features in the dataset. Once encoded and masked, the input is sent into the ring resonator.
In detail, to validate my approach, I developed an experimental setup of pump and probe, already implemented with a similar approach in [1] and [2]. In this experiment I used a laser with a power greater than 10mW as a pump and another low-power laser as a probe. I centered these two lasers at two adjacent resonant frequencies. I modulated the laser input using a Mach Zehnder through a wave, generated by an AWG, with the same trend as the coded input. The pump is then amplified and combined with the probe through a coupler $50\%50\%$ to finally enter in the micro-ring. The pump, coupled to the microring resonance, induces nonlinear phenomena in the device; in fact, two-photon absorption (TPA) occurs, generating a large number of free carriers within the microring. This population of free carriers causes variations in the refractive index of silicon through an effect known as Free Carrier Dispersion (FCD). For this reason, I used the probe to detect the induced variations in the microring.
At the output of the microring, I used a filter to select the wavelength of the probe and directed it through an avalanche photodetector to an oscilloscope from which I acquired and then analyzed the data. Once I acquired the data, using MATLAB, I sampled the virtual nodes and trained the network. So, I performed the training only on the output layer that is a weighted linear combination of the values of the virtual nodes. The goal of the training is to find a matrix of Wout weights such that Y=Wout*X (linear regression) with Y the desired output and X the micro-ring output. The weight matrix was found using the least squares method.
Linear regression has two drawbacks: it is subject to overfitting and cannot easily express nonlinear relationships, therefore a regularization technique was used to calibrate the regression model by adding a penalty on the complexity of the model. The regularization technique used is ridge regression or L2 penalty. To add this regularization, I altered the loss function by adding a lambda penalty term that was found with 5-fold cross validation. To find Wout that minimizes the modified loss function I calculated the gradient with respect to Wout and set it equal to zero. Subsequently, I tested the network both in simulation and experimentally, achieving classification rates comparable to and exceeding 90%.
Testing the xor dataset demonstrated the presence of memory. Furthermore, it has been demonstrated that sampling the probe instead of the pump is more efficient in terms of classification accuracy because the zero-bit encoded by the pump is more subject to noise than that of the probe thanks to the non-linearities of the MRR triggered by the pump signal. I tested different bitrates and found that as the bitrate increases, the accuracy decreases because the duration of the input bit becomes shorter than the free carrier lifetime.
These experiments demonstrated the great potential of this application in reservoir computing systems. This research opens new perspectives for the implementation of high-speed, low-energy optical neural networks, leveraging the potential of silicon photonics.

Contents

Chapter 1

Introduction

1.1 Neuromorphic computing and Silicon Photonics

Neuromorphic computing is a computing technique that attempts to emulate the functioning of neurons and how they communicate in the brain on electronic circuits, overcoming the limitations of systems based on Von Neumann architectures.
In recent years, technological progress has advanced at great speed thanks to increasingly faster, more efficient and powerful digital processing units and increasingly intelligent algorithms. Neural networks are certainly among these algorithms. Neural networks are a subset of machine learning and AI.
The integration of silicon photonics with neuromorphic computing is still in a development phase, but it offers promising prospects in the development of artificial intelligence. Thanks to photonics, we have an alternative to the predominant electronic solutions, offering advantages such as greater bandwidth, intrinsic parallelization thanks to the possibility of exploiting wavelength division (WDM), better energy efficiency, lower latencies, higher throughput, immunity to electromagnetic interference. All these advantages allow to perform complex operations in a short time and with a low energy expenditure in parallel, representing a strong point for accelerating AI processing.
Optical microring resonators can be used as neurons of artificial networks thanks to their nonlinear effects.
The objective of this thesis will be the analysis and use of a silicon microring resonator, in add-drop configuration, employed as a nonlinear element within a neural network. The choice of silicon is motivated by the numerous advantages it offers, including:

- compatibility with CMOS (Complementary Metal-Oxide-Semiconductor) technology, which is the standard process used for the production of electronic circuits. Due to this compatibility, silicon photonic circuits can be manufactured using the same production lines and infrastructure already established for electronics, without the need to invest in new facilities or production technologies.
- high integration density and therefore the construction of compact and low-cost chips that incorporate numerous optical components.
- silicon allows for the production of cheaper photonic circuits compared to technologies based on other materials like gallium arsenide (GaAs) or indium phosphide (InP) because it is a more abundant and low-cost material.

- silicon waveguides offer low signal attenuation and high optical field confinement, thanks to the strong refractive index contrast between silicon and silicon dioxide

However, the use of silicon as an optical medium introduces additional losses due to nonlinear effects that depend on the power and this is especially true in microrings where the input power can increase significantly inside the ring, modifying the spectral response. However, these nonlinear effects can be exploited to our advantage in AI to implement neural networks that are faster and more efficient than neural networks implemented via software in various fields: medicine, telecommunications, scientific researches, finance, robotics, autonomous vehicles, and processors.
Using components such as microring resonators and waveguides, it is possible to perform matrix multiplication operations that are fundamental in high-speed, low-energy dissipation neural networks. For example different input signals can be represented by different wavelengths which can be combined into a single optical waveguide using WDM. Each wavelength passes through a series of tunable microrings that modulate the intensity of the optical signal, weighting them. The weighted signals are then summed, thus emulating the operations that neurons perform to process the input signals.
Nonlinear activation functions, fundamental in artificial neural networks (ANNs), can be implemented in PICs (photonic integrated circuits) via photodiodes, nonlinear materials or optical amplifiers (the latter become nonlinear at high power). In silicon photonics, although Mach-Zehnder interferometers are often employed, this thesis focuses, as already mentioned, on a microring resonator used as a nonlinear node.

1.2 Thesis organization

This thesis is organized as follows. In Chapter 2, the concept of artificial neural networks is introduced, along with their structure and main types. The concept of ”reservoir computing”, technique that simplifies the training of recurrent networks, is also introduced and discussed in depth. A comparison between spatial reservoir architectures and those based on time multiplexing is also presented.
Chapter 3 explores the role of microring resonators in neural networks. The fundamentals of microring behavior are discussed, followed by an exploration of their nonlinearities, which are crucial for their use as nonlinear elements in neural networks. The experimental setup used to test the microring is then described, concluding with a study of the ring response in both linear and nonlinear regimes.
Chapter 4 focuses on the experiments and simulations conducted on a photonic reservoir based on time multiplexing with a single microring resonator. The experimental procedures are described, including data preprocessing, processing, and postprocessing, addressing specific tasks such as the computation of the 1-bit delayed XOR operation and the classification of flowers from the Iris dataset.
Finally, Chapter 5 contains the general conclusions of the work, summarizing the obtained results and discussing potential future directions for microring-based photonic neural networks with time multiplexing.

Chapter 2

Introduction to Artificial Neural Networks

Neural networks are structured to mimic the human brain, imitating how neurons exchange signals. They use training data to learn and improve their accuracy over time. Once the desired accuracy is achieved, neural networks become powerful tools capable of solving complex tasks and organizing datasets into clusters at high speed.
As highlighted in [3], the human brain processes analog signals with high precision and energy efficiency, consuming about 20% of the body's energy equivalent to about 20W and reaching processing speeds estimated to be around 100 teraflops per second. In contrast, digital machines can even exceed 100W during neural network training and are much slower due to the Von Neumann bottleneck which refers to the difference between CPU speed and data transfer speed between the CPU and other components such as memory.
Neural networks find application in various sectors including robotics, in self-driving vehicles for making decisions in complete safety, for the detection of bank fraud, for speech recognition, image recognition and time series predictions. Reservoir computing has made a significant contribution in these sectors but competitive practical applications are still under development.
Neural Networks are made up of nodes, elementary units, connected to each other and behaving similarly to biological neurons. In the conventional McCulloch-Pitts model, each node, or artificial neuron, sums its weighted inputs and applies an activation function as follows:

$$y = f_a \left(\sum_{i=1}^{N_u} w_i u_i + w_0 \right)$$

with f_a a nonlinear activation function, u_i the node inputs, N_u the number of inputs, w_0 the bias and w_i the weights that determine the importance of the connections and can be adjusted to direct the network's output toward a specific goal.
Non-linear activation functions are crucial because a linear function would reduce the network's complexity. Non-linear activation functions enable the neural network to learn and represent complex relationships between inputs and outputs.
In neural networks, the input is mapped into a higher-dimensional space and processed by nonlinear functions as it passes from one layer to the next. I replicate this functionality using the nonlinear response of a micro-ring resonator.
It is easily demonstrable that if the activation function is linear, any number of cascading nodes could be reduced to a single node. A single neuron has limited computational capabilities and cannot solve problems that are not linearly separable, such as the 1-bit delayed xor problem, but a neural network with multiple layers and so nonlinear activation functions can effectively solve the problem. The idea is that, in the final layer, the network will be able to linearly

discriminate the outputs because with more layers the probability that the problem becomes linearly separable increases.
Usually to train a network there are two phases: the feedforward propagation phase and the backpropagation phase. In the feedforward propagation phase, data propagates from left to right through the layers of the network until an output is generated. The backpropagation phase is an algorithm used to train the network to obtain correct outputs. How the training process takes place will be discussed in detail later.

2.1 Structure and Functioning of Artificial Neural Networks

Neural networks can be divided into two main types: feedforward neural networks (FFNNs) and recurrent neural networks (RNNs). In general, both of these types are composed of an input layer, one or more hidden layers and an output layer. The fundamental characteristics of both networks can be described as follows:

- **Input data:**
$$\mathbf{u}(n) = \begin{pmatrix} u_1(n) \\ u_2(n) \\ \vdots \\ u_{N_u}(n) \end{pmatrix} \in \mathbb{R}^{N_u}$$
where N_u is the dimensionality of the input and n the time step.

- **Network states:**
$$\mathbf{x}(n) = \begin{pmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_{N_x}(n) \end{pmatrix} \in \mathbb{R}^{N_x}$$
where N_x is the dimensionality of the network states. The states of the network are responsible for maintaining a memory and processing the information.

- **Output data:**
$$\mathbf{y}(n) = \begin{pmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_{N_y}(n) \end{pmatrix} \in \mathbb{R}^{N_y}$$
where N_y is the dimensionality of the output.

- **Input-to-hidden weights:**
$$\mathbf{W}_u = (w_{u,ij}) \in \mathbb{R}^{N_x \times N_u}$$

- **Hidden-to-hidden weights:**
$$\mathbf{W}_x = (w_{x,ij}) \in \mathbb{R}^{N_x \times N_x}$$

- **Hidden-to-output weights:**
$$\mathbf{W}_y = (w_{y,ij}) \in \mathbb{R}^{N_y \times N_x}$$

- **Optional Output-to-hidden feedback weights:**

$$\mathbf{W}_{\text{fb}} = (w_{\text{fb},ij}) \in \mathbb{R}^{N_x \times N_y}$$

This feedback enhances the memory of the network.

2.1.1 General training procedure for an artificial neural network

There are three types of learning:

1. **Supervised learning**: in this approach, often used for classification tasks as in the case of this thesis, the dataset consists of inputs and their corresponding desired outputs (target). The network is trained trying to minimize the error between the output and the target for each input.
2. **Unsupervised learning**: unlike supervised learning, there are no desired outputs. Therefore, the system must analyze the data to find hidden patterns.
3. **Reinforcement learning**: given a fixed goal, the algorithm learns through a trial-and-error process.

To train neural networks with supervised learning, a widely used technique, during the back-propagation phase, is the "gradient descent", which updates the weights and bias of the network with the goal of minimizing the function loss which measures the error between predicted and actual values. The goal is to adjust the weight vector to iteratively reduce the function loss. At each step i, the weights are updated as follows:

$$w' = w_i - \eta \frac{\partial}{\partial w_i} L(w) \tag{2.1.1}$$

where:

- $\mathbf{w}'$ are the new weights after the update.
- η is the learning rate, which determines how fast the algorithm converges. If η is too small, the convergence will be slow; if it is too large, the algorithm risks to pass beyond the minimum and never converge.
- L is the loss function.

To improve convergence, "Feature Scaling" can be used. There are mainly two methods of feature scaling: min-max scaling and z-score normalization. The first one normalizes the input features in the range between -1 and 1 or between 0 and 1.
Instead, the "Z-Score Normalization" is often used to center the input around zero, using the formula:

$$u_i = \frac{u_i - \mu_i}{\sigma_i} \tag{2.1.2}$$

with μ_i the mean of the feature u_i and σ_i the standard deviation of the feature u_i.
In recurrent neural networks, training is more complex, and the backpropagation through time

(BPTT) algorithm is used, where the neural network is ”unrolled” over time. This training technique requires a lot of time, energy and memory because all the weights of the network must be updated.
A solution to these problems is represented by reservoir computing whose strong point lies in the simplicity and speed of training because only the connections between the readout layer (the layer that collects the states to be read) and the output layer are trained.

2.1.2 Feedforward neural networks

These networks are composed of sequential layers: an input layer, one or more hidden layers, and an output layer, as shown in Figure 2.1.1. In a feedforward network, each node is connected to the nodes of the previous layer through weighted connections, and information flows from left to right through the various layers, never passing through the same node more than once. These networks do not have a state in the sense that they have no memory of past neuron activations, meaning that the output depends only on the current input and thus cannot handle temporal information. The ”state” vector of the neurons in layer $i + 1$ is expressed by the following formula:

$$x^{i+1} = f(W_x{}^i \mathbf{x}^i + b^i) \tag{2.1.3}$$

This formula shows how the state of a neuron in a given layer is a nonlinear function f applied to a weighted linear combination of the outputs of the neurons from the previous layer plus the bias b. The bias vector contains scalars that represent the threshold at which the neurons activate. If the input is inactive, the state will also be inactive, and consequently, so will the output.

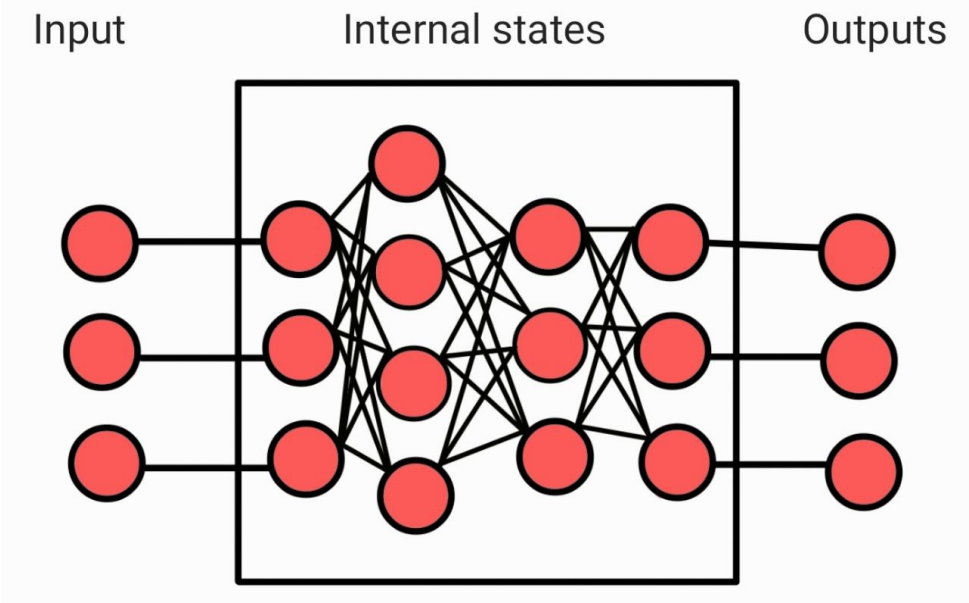

Figure 2.1.1: General scheme of a feedforward neural network. Red circles represent the artificial neurons.

2.1.3 Recurrent neural networks

Recurrent Neural Networks (RNNs) are neural networks characterized by a memory, because, unlike feedforward networks, in recurrent neural networks there can be feedback loops within the same node or directed to a previous node, as shown in Figure 2.1.2. All the connections, shown in black, must be trained. Indeed, if the output becomes inactive, the states of the neu-

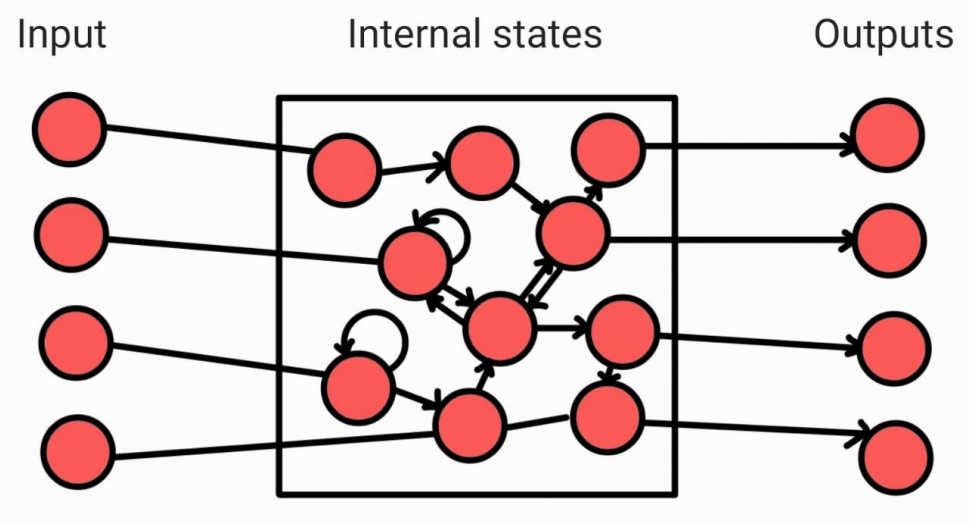

Figure 2.1.2: General scheme of a reccurent neural network.

rons in the subsequent layers can still be active because the internal states depend not only on the input but also on past states and feedback from the output, if present. This provides memory. Consequently, RNNs are particularly powerful for solving problems such as time series prediction or natural language processing.

State Update Equation

The state of a neuron at time step $n+1$ can be expressed as follows:

$$x(n+1) = f\left(W_u u(n+1) + W_x x(n) + W_{\text{bf}} y(n)\right) \tag{2.1.4}$$

where W_x and W_u are the weight matrices for the internal connections of the hidden layer and for the input, respectively. W_{bf} represents the feedback weight matrix from the output layer to the hidden layer. These matrices contain the parameters that will need to be trained. Common choices for f include the hyperbolic tangent function and the logistic sigmoid function.

Output Calculation

The output depends on both the current internal states and the input, through the following equation:

$$\hat{y}(n+1) = g(W_y \cdot \mathbf{x}(n+1) + W_y \cdot u(n+1)) \tag{2.1.5}$$

where $u(n+1)$ represents the input at time step $n+1$, $x(n+1)$ represents the internal states at time step $n+1$, and g is the output activation function.

2.2 Reservoir computing

Reservoir Computing is an approach based on recurrent neural networks. Specifically, in this thesis, the Echo State Network (ESN) will be discussed, a specific type of recurrent neural network that uses a fixed and untrained reservoir. Reservoir computing, being based on recurrent neural networks, has a so-called ”fading” memory, but it differs from them because the connections, and therefore the weights of the hidden neurons in the reservoir, are fixed and randomly assigned, and the training occurs only on the weights of the connections to the output layer.
Figure 2.2.1 shows the general scheme of a reservoir computing system with the connections between the read-out layer neurons and the outputs highlighted in yellow to emphasize that these connections are the only ones being trained.

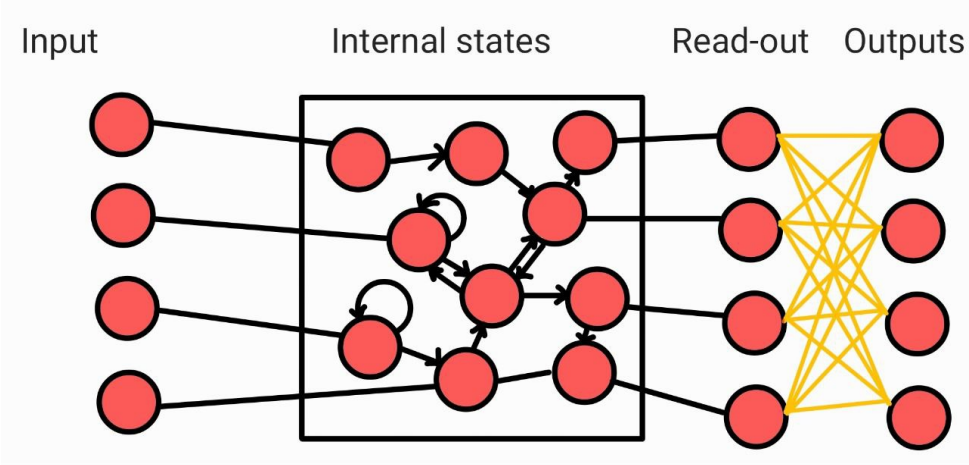

Figure 2.2.1: General scheme of a reservoir computing system. The connections in yellow represent the only trainable weights.

The reservoir must transform the input through a nonlinear function into a higher-dimensional space so that it is capable of distinguishing between different input signals.
If the function is linear, the dimensionality would remain unchanged. By mapping into a higher-dimensional space, the probability that the input data becomes linearly separable increases.
In Figure 2.2.2, graph (a) represents the case where the data in a two-dimensional space are not linearly separable, whereas graph (b) shows that after transformation into a three-dimensional space, the data have become linearly separable by a plane.
This mapping corresponds to a multiplication with a matrix or a vector containing weights, which can occur in space with physical nodes, in time with virtual nodes, or by exploiting WDM. For example, as discussed in [4], the result of the XOR operation cannot be linearly divided in a two-dimensional space, but it can be in a three-dimensional space, making the classification problem simple and linear.
The reservoir can be physically implemented by devices exhibiting intrinsic nonlinear physical phenomena, such as the nonlinear I-V curves of memristors, as demonstrated in [5].
For the reservoir computing technique to function correctly, the effects of previous states and inputs must gradually fade over time. This gradual decay is referred to as fading memory.
A key characteristic of the reservoir is the ”kernel quality,” which indicates the reservoir’s ability to project the input into a high-dimensional space. A reservoir with greater nonlinearity will have higher kernel quality but lower fading memory capacity.
Another characteristic is the ”complexity” or ”expressivity” of the model, which measures the

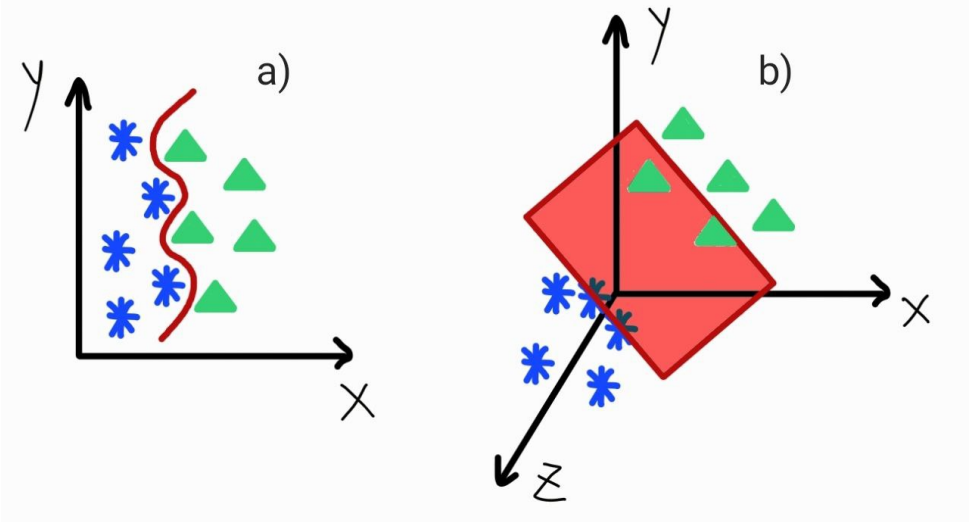

Figure 2.2.2: a) Mapping of data into a two-dimensional space; b) mapping of data into a three-dimensional space.

quality of the model and increases with the size of the reservoir (number of physical or virtual neurons), with nonlinearity, and with memory capacity. However, greater expressivity can reduce robustness to noise. As the size increases, expressivity, memory capacity, and kernel quality improve, but so does the cost of building the network.
As discussed in [3], there is also another type of memory, which is long-term memory, corresponding to the readout function, and it is a procedural memory because, after training, the readout function stores how to calculate the output based on the reservoir's response.
In the context of fading memory, when an input signal enters the reservoir, if it is not amplified, its effect fades over time with a relaxation time constant τ, which indicates how long it takes for the system to forget an input and return to equilibrium. If the system has not yet reached equilibrium and a new input arrives, its response will be influenced by the previous input. This phenomenon, called "reverberation," is referred to as an "echo" when it lasts much longer than the period between inputs. To achieve reservoirs with high expressivity, long echoes are preferred. To obtain a long echo and retain the memory of past inputs in the reservoir, the relaxation time must be longer than the input period. This way, the information from the previous input does not fade before a new input arrives.
Memory in a reservoir and the degree of chaos in the system are closely linked. A very stable reservoir, meaning one that quickly converges to zero in the absence of input, has little memory, whereas a chaotic and therefore less stable reservoir has long but disordered memory. The optimal operating point for the reservoir depends on the application. As mentioned in [4], a chaotic system has excellent separation capabilities because it can distinguish even very similar inputs, but the separation might become so precise that the network could lose its ability to generalize.
As highlighted in [3], there is ongoing research into using reservoir computing with specific substrates not only because they provide complex dynamics but also due to their unique properties. For example, in this thesis, an optical reservoir was employed that uses light to process information, making it suitable for applications requiring high computational speed, while ionic or electrochemical reservoirs, for instance, can be used for their ability to interface directly with living tissues or bioelectronic devices rather than for their speed.
The goal of reservoir computing is to create a mapping from input signals to output signals.
By modeling time continuously, to obtain this mapping, we have training input signals $u_{\text{train}}(t) \in$

$\mathbb{R}^{N_u}$ and corresponding target output signals $y_{\text{train}}(t) \in \mathbb{R}^{N_y}$.
The goal is to create a system that, upon receiving $u_{\text{train}}(t)$, generates an output $\hat{y}_{\text{train}}(t)$ approximately equal to $y_{\text{train}}(t)$. To achieve this goal, five main points can be distinguished:

1. **Reservoir Preparation**: If the task is to classify a dataset, the dataset is divided into two parts: one part that will be used for training and one part that will be used to test the reservoir's ability to generalize on unseen data. Typically, these parts are split $70\% - 30\%$ or $80\% - 20\%$, respectively. The input signal $u_{\text{train}}(t)$ enters the reservoir and generates internal states. Typically, not all states are read, but a subset that constitutes the read-out layer. The states of the read-out layer are denoted by $x_i(t) \in \mathbb{R}^{N_x}$. The states of the nodes at time step n are calculated according to equation 2.1.4.

2. **Execution of the technique**: A training set of input data $u_{\text{train}}(t)$ is used, and the responses $x_i^{\text{train}}(t)$ are read and stored. In the readout layer, the states of the reservoir that need to be read are collected in $x_i^{\text{train}}(t)$, and finally, in the output layer, the outputs will be a weighted linear combination of the readout states. These weights are the only weights in the network that will be updated, i.e., trained. They are the parameters defining a function called the "readout function" G.

3. **Role of G**: The goal is to find the readout function G, which, when applied to the readout layer states, returns the desired output signal. This function G will be trained on the training input data until it can produce an output $\hat{y}(t)$ that closely approximates the target output $y(t)$.

4. **Training**: The objective of training is to find the optimal form of the readout function G, and this is achieved by minimizing the loss function $L(y_{\text{train}}, \hat{y})$, which is the difference between the predicted output $\hat{y}(t)$ and the target output $y_{\text{train}}(t)$ over all training input samples.

5. **Testing**: After training is completed, the test dataset is fed into the system, and the reservoir's ability to predict is evaluated, typically by calculating the accuracy, which is the ratio of correct predictions to the total number of predictions.

2.2.1 Training Process

Typically, the objective of training is to minimize the mean squared error (MSE) or, more generally, the average of the loss function. The average of this loss function is defined as follows:

$$\overline{L}_{\text{train}} = \frac{1}{N_{ts}} \sum_t L\left(y_{\text{train}}(t), G(x_1^{\text{train}}(t), x_2^{\text{train}}(t), \ldots, x_{N_x}^{\text{train}}(t))\right)$$

where N_{ts} is the total number of samples in the training input set. By minimizing the average loss $\overline{L}_{\text{train}}$, we adjust the parameters of the function G. The readout function G can be a linear function of the state variables:

$$\hat{y}(t) = G(x_1(t), x_2(t), \ldots, x_{N_x}(t)) = w_1 x_1(t) + w_2 x_2(t) + \ldots + w_{N_x} x_{N_x}(t)$$

In vector notation, this can be written as:

$$\hat{y}(t) = \mathbf{w_y}^T \mathbf{x}(t)$$

The parameters of G are the weights $w_1, w_2, \ldots, w_{N_x}$ that weight the reservoir states. Finding the optimal G is equivalent to finding the optimal weights $\mathbf{w}$ that minimize the loss function.

2.2.2 Linear Regression

Linear Regression analysis is a statistical process that falls under supervised learning algorithms, used to estimate the linear relationship between two variables. This technique is the one used in this thesis and is employed to estimate a linear function between the states of the readout layer and the desired outputs by finding the optimal parameters (the weights w_i) that minimize the Mean Squared Error between the predicted output $\hat{y}(t)$ and the target output $y(t)_{\text{train}}$, as follows:

$$W_{\text{y}} = \arg\min_{W_{\text{y}}} \sum_{i=1}^{N_x} \left(W_{\text{y}} x_i - y_{i,\text{train}}\right)^2 \tag{2.2.1}$$

which is equivalent to:

$$W_{\text{y}} = (X^T X)^{-1} X^T Y_{\text{train}} \tag{2.2.2}$$

where:

- X is the matrix of state variables of the reservoir (each row represents a state at time t).
- X^T is the transpose of the matrix X.
- $\mathbf{Y}_{\text{train}}$ is the vector of target outputs.
- $(X^T X)^{-1}$ is the Moore-Penrose pseudo-inverse matrix of the product in parentheses.

Squaring gives more weight to larger differences. After the parameters of the function G have been trained on the training data, the function G can be used for new inputs that the network has never seen, and it will be able to predict the correct result with a certain accuracy, which depends on the amount of data the network was trained with.
Linear regression can also be implemented in hardware, as demonstrated in [6].
A typical output encoding for classification problems is one-hot encoding, where the output is represented by a vector filled with values. The position of the values in the vector indicates one of the possible classes, and the value represents the probability of belonging to that class. The higher the value, the more likely it is that the correct class is the one indicated by that position. The technique of selecting the class with the highest value is called ”winner-takes-all” and is the method that will be used in this thesis.

2.2.3 Ridge Regression

The training process using linear regression can encounter two problems: overfitting and underfitting. Overfitting occurs when there are too many features, and the model fits the training data too well, capturing noise and random fluctuations, and losing the ability to generalize to new inputs. Underfitting, on the other hand, happens in the opposite case, where the model makes incorrect assumptions, leading to a reduction in model complexity and, consequently, accuracy. Figure 2.2.3 illustrates an example of overfitting in (a) and a correct model fit in (b). One solution is to reduce the number of features or to use ridge regression, which introduces

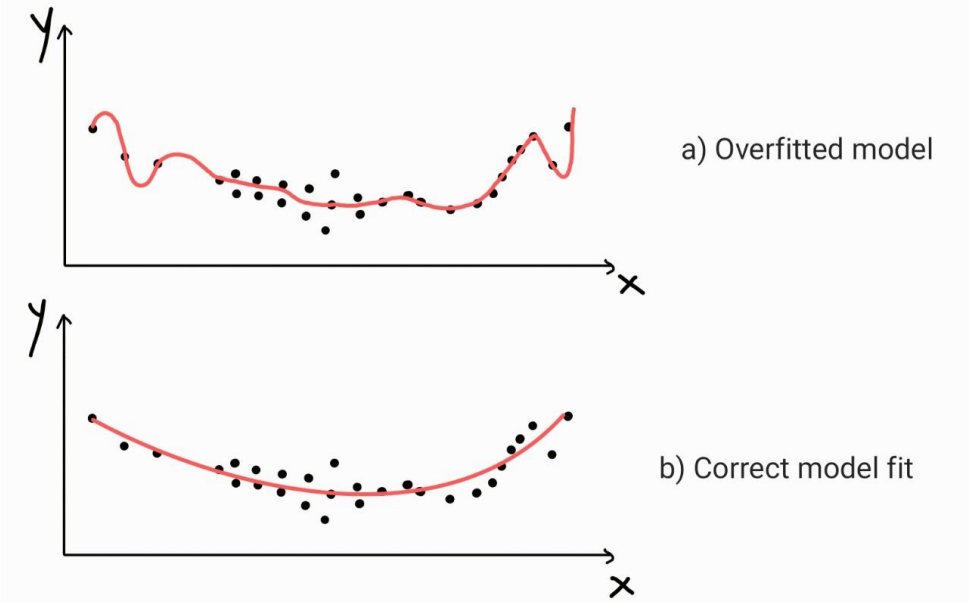

Figure 2.2.3: (a) Overfitting of the model. (b) Correct model fit.

regularization by adding a penalty term λ. With this technique, the number of features remains unchanged, but the weights associated with the features are reduced by λ. A model is considered less complex if the weights are close to or equal to zero. The penalty term is introduced into the loss function. The most common regularization methods used are:

- **LASSO (Least Absolute Shrinkage and Selection Operator)** or **L1 penalty**: This method penalizes the sum of the absolute values of the model's weights. The formula for the loss function in LASSO regression is:

$$\text{Cost Function} = \sum(\text{squared residuals}) + \lambda \sum |w_i|$$

- **Ridge Regression** or **L2 penalty**: In this method, the sum of the squares of the model's weights is penalized. The formula for the loss function in Ridge Regression is:

$$\text{Cost Function} = \sum(\text{squared residuals}) + \lambda \sum |w_i|^2$$

Ridge regression, particularly useful when variables are highly correlated, penalizes the weights based on the square of their value, meaning that larger weights will be penalized more heavily than smaller ones. Thus, unlike simple linear regression, ridge regression seeks to minimize:

$$W_{\text{y}} = \arg\min_{W_{\text{y}}} \sum_{i=1}^{N_x} (W_{\text{y}} x_i - y_{i,\text{train}})^2 + \lambda \|W_{\text{y}}\|^2 \tag{2.2.3}$$

with $||W_{\text{y}}||^2$ being the squared norm of the model's weights, which prevents them from becoming too large. To find W_{y} that minimizes this expression, it is necessary to compute the gradient with respect to W_{y} and set it equal to zero:

$$\frac{\partial}{\partial W_{\text{y}}} \left(||Y_{\text{train}} - W_{\text{y}} X||^2 + \lambda ||W_{\text{y}}||^2 \right) = 0$$

Expanding and simplifying this expression gives:

$$-2Y_{\text{train}}X^T + 2W_{\text{y}}XX^T + 2\lambda W_{\text{y}} = 0$$

Solving for W_{y} yields:

$$W_{\text{y}} = (Y_{\text{train}}X^T)(XX^T + \lambda I)^{-1} \tag{2.2.4}$$

where I is the identity matrix. This is the closed form for W_{y} in linear regression with L_2 regularization. The larger the λ, the greater the penalization. Reducing the weights makes the system less sensitive to the specific characteristics of individual training examples, allowing different versions of the same type of input to be classified into the same target category.

2.2.4 Cross Validation

Reducing the values of lambda can make models more complex, and vice versa. It is necessary to find the right balance to prevent overfitting. Lambda is a positive value and can range from 0 to positive infinity, but it is usually chosen between 0 and 10. A technique that allows the tuning of the hyperparameter λ is cross-validation. There are different variants of cross-validation, but the most common ones are 10-Fold or 5-Fold Cross-Validation. In k-fold cross-validation, the data is split into two sets: a training set and a test set, usually with a division of $70\% - 30\%$ or $80\% - 20\%$, respectively.
The procedure consists of splitting the training set into k subsets, called folds. By choosing a certain lambda, the model is trained on $k - 1$ folds and tested on the remaining fold, calculating the accuracy. This process is repeated k times, each time using a different fold as the test set while the remaining folds are used for training. At the end of all iterations, the average accuracy is calculated. This procedure is then repeated for different values of lambda. In the end, a table will be obtained with the various values of lambda associated with their respective average accuracies. The optimal value of lambda will be the one corresponding to the highest average accuracy.

2.3 Spatially distributed reservoir (spatial multiplexing) and delay-based reservoir (temporal multiplexing)

Previously, the general functioning of reservoir computing was introduced, but it is important to highlight that there are two main implementation modes: spatial reservoirs and delay-based reservoirs.
In a spatial reservoir, neurons are distributed and connected in space, and the larger their number, the greater the dimensionality of the reservoir. The input is distributed among all nodes in parallel. In contrast, in the delay-based system, the inputs are injected sequentially into a single nonlinear node.
Temporal multiplexing was first proposed in [4] for a photonic reservoir. Typically, a delay-based reservoir consists of a nonlinear node and a feedback delay line. Temporal multiplexing allows us to increase the dimensionality of the space by exploiting the concept of virtual nodes and enables the use of less hardware, but we need optoelectronic components that operate at higher frequencies to compensate for the lack of a spatial network. In a reservoir with temporal multiplexing, the delay line expands the system's ability to remember and process past inputs, improving performance on tasks that require extended temporal memory.

τ refers to the fading memory intrinsic to the system, which diminishes over time or in a delay-based reservoir, it also indicates the length of the delay line that provides an explicit and controlled memory through the reintroduction of inputs. Along the delay line, virtual nodes are defined, representing delayed versions of the nonlinearly transformed signal. The temporal separation between one node and the next is indicated by θ. The length and material of the delay line control the delay and, therefore, also the chaos.
In Figure 2.3.1, in (a), the general scheme of a delay-based architecture for a reservoir computing system is shown. As demonstrated in the experiment with a laser subjected to delayed

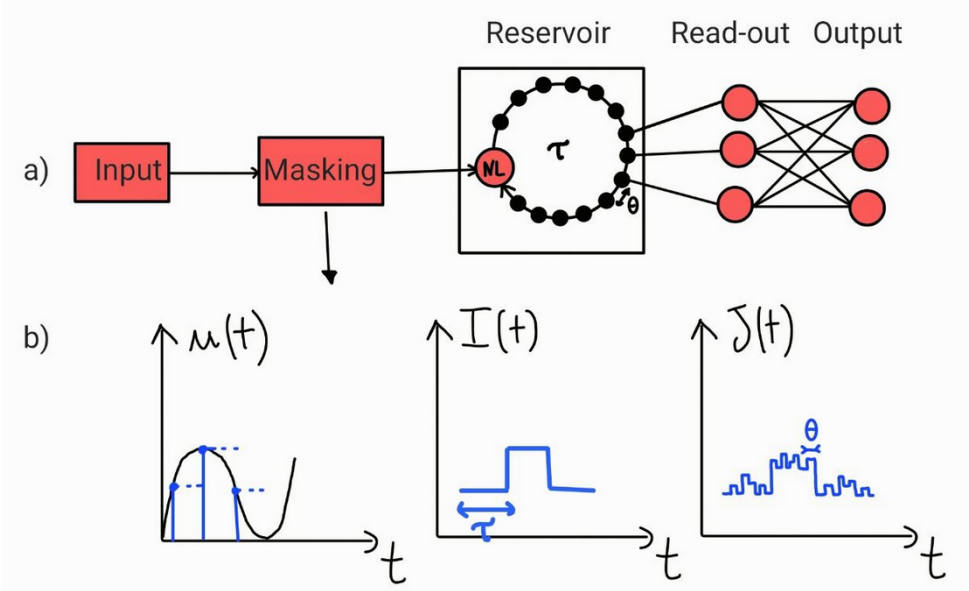

Figure 2.3.1: (a) General scheme of a delay-based architecture in a reservoir computing system; (b) masking procedure.

feedback, described in [3], even a stable laser system can become chaotic when delayed feedback is introduced, clearly showing that delay can have a destabilizing effect on the system.
Below are the steps to implement reservoir computing with temporal multiplexing:

1. **Pre-processing of the Input Signal:** The first step is preprocessing the input, which can be a continuous-time signal $u(t)$ or a discrete-time signal $u(n)$. A sample and hold operation is performed on the input signal, creating a piecewise constant signal $I(t)$ with a duration τ, which is the time it takes for the signal to travel through the delay line.

2. **Application of a Random Time Mask:** Since the input must be injected into the system through the single nonlinear node, pre-processing (temporal multiplexing) of the input is required, known as masking. The masking procedure is shown in (b) in Figure 2.3.1. Specifically, a random mask M is applied to $I(t)$, which serves two purposes: sequentializing the input and increasing the dimensional space used by the system. For each virtual node (i.e., for each interval θ), the mask imposes a random weight that is multiplied by the signal $I(t)$ in that specific interval. The mask corresponds to $M(t) = W_{u,i} \quad \text{for} \quad (i-1)\theta \leq t \leq i\theta$. W_u is a random matrix $N \times Q$, where N is the number of virtual nodes and Q is the input dimension. If the input signal is one-dimensional, the mask will be a vector; otherwise, it will be a matrix. Considering a one-dimensional input, the function of the mask is to alter each constant segment of $I(t)$

into N_x different values. Each of these N_x constant segments will remain constant for a duration of $\theta = \frac{\tau}{N}$. Thus, the final input signal to the single physical node will be [4]:

$$J(t) = \sum_{j=1}^{Q} I_j(t) \cdot M_j(t) \tag{2.3.1}$$

The masking procedure ensures that the system remains in a transient regime to guarantee a rich nonlinear dynamic. After the signal travels through the delay line, it re-enters the nonlinear node.

3. **Reservoir State Update:** After the signal $J(t)$ enters the reservoir, nonlinear dynamics begin, and the internal states of the reservoir evolve according to a nonlinear function of $J(t)$ and the delayed state of the reservoir $x(t-\tau)$ according to the following equation:

$$x(t) = f\left(x(t), x(t-\tau), \gamma J(t)\right), \tag{2.3.2}$$

where γ is a corrective parameter, usually referred to as input gain.

4. **Output Calculation:** The output of the reservoir, $y(t)$, is a linear combination of the values of the N virtual nodes defined along the delay line in a time τ.

$$y(t) = \sum_{i=1}^{N} w_{\text{y,i}}\, x\left(t - \frac{\tau}{N}(N-i)\right) \tag{2.3.3}$$

with $w_{\text{y,i}}$ being the output weight of the virtual neuron i.

The following table illustrates how the relationship between the timescale of the nonlinear node T and the temporal separation of the virtual nodes θ affects the system's behavior.

Table 2.3.1: Relationship between the timescale T of the nonlinear node and the separation of the virtual nodes θ

Relationship between T and θ	**System Behavior**	**Consequences**
$T < \theta$	The nonlinear node reaches a steady state	The diversity of the reservoir states decreases, reducing the system's ability to project the input into a high-dimensional space.
$T > \theta$	The nonlinear node remains in the transient regime	The system continuously processes new inputs, ensuring greater state variety and a richer dynamic response.

There are also variants of reservoir computing that do not use an external delay line. In these configurations, memory and processing capacity depend solely on the nonlinear phenomena of the physical node. Specifically, studies such as those reported in [1] have shown that in photonic reservoirs, a single MRR without a delay line is capable of solving simple tasks that do not require extended memory due to its intrinsic nonlinear inertia.
In this thesis, this variant without a delay line was adopted, successfully implementing a photonic reservoir that was tested on various tasks, as reported in [1]. This approach offers the advantage of reducing hardware complexity while still maintaining good processing capability for various tasks.

Chapter 3

Optical Ring Resonators in Neural Networks

3.1 Basic Concepts of Microring resonators

Silicon microring resonators are devices that play an increasingly important role in the field of silicon photonics, thanks to significant advances in photonic integration and the availability of CMOS-compatible fabrication technologies. These devices have an high refractive index contrast between silicon and its oxide, which allows for the creation of very compact devices with a small bend radius and high performance.
A microring resonator is a type of resonant cavity formed by looping a waveguide onto itself, allowing light to circulate within it either in a clockwise or counterclockwise direction.
The resonance condition is achieved when the roundtrip phase shift is a multiple of 2π, leading to constructive interference and a significant buildup of circulating power P_c. This effect enables microring resonators to be used as spectral filters, particularly useful in applications like wavelength division multiplexing (WDM) or optical sensing. In sensing application they can detect shifts in the resonance wavelength due to changes in the surrounding environment.
Microring resonators can demonstrate multiple resonances, characterized by periodic downward peaks and upward peaks in the spectral response observed at the through port and at drop port respectively. At the through port, destructive interference occurs because the light that remains in the bus waveguide interferes destructively with the light coupled from the microring. This creates a dip in intensity at the resonant wavelengths. At the drop port, constructive interference occurs because the light that remains in the bus waveguide interferes constructively with the light coupled from the microring. This creates a peak in intensity at the resonant wavelengths.
The resonance wavelength λ_0 for a microring is given by:

$$\lambda_0 = \frac{n_{\text{eff},0} L}{m}$$

where $n_{\text{eff},0}$ is the effective refractive index of the waveguide, L is the perimeter of the microring resonator, and m is an integer number.

3.2 Micro ring resonator characteristics

3.2.1 Equations for circulating power versus finesse

In order to achieve the objective of this thesis, which is to create a neural network with a microring resonator, several silicon microrings fabricated by IMEC were tested. Among these, the microring with the best characteristics for the experiments, which will be discussed in the section "Criteria for Choosing the Optimal Microring", was selected.
The microring is in an add-drop configuration, whose general schematic is shown in Figure 3.2.1.
The most important parameters characterizing the microrings are:

- t^2 represents the proportion of power that remains in the bus waveguide after passing the directional coupler.
- k^2 describes how much of the light power traveling along bus waveguide is actually coupled into the microring.
- $a = e^{\frac{-\alpha \cdot L}{2}}$ accounts for losses in the resonator where $\alpha = \alpha_0 + \alpha_{rad}$ with α_0 the linear loss (light scattering, residual doping and single photon absorption) and α_{rad} the loss for ligth irradiated in the cladding.
- E_{in} is the input field. $|E_{\text{in}}|^2 = P_{\text{bus}}$ is the input power.
- $\theta = \beta(w) \cdot L = \frac{2\pi n_{\text{eff}}}{\lambda} \cdot L$ represents the phase shift per round trip, where L is the path length and λ is the wavelength.
- $L = 2 \cdot L_d + 2 \cdot L_c + 2 \cdot r \cdot \pi$. where L_c is the coupler length, L_d the length of the straight waveguide and r the curvature radius.
- η^2 represents the coupling loss. In order to avoid repeatedly writing $1-\eta^2$ in the following formulas, $\zeta = 1 - \eta^2$ will be defined.

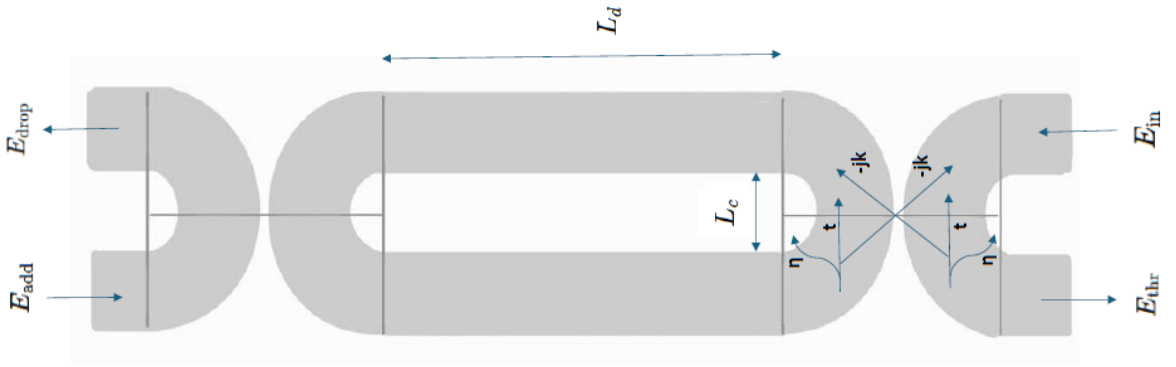

Figure 3.2.1: Schematic of the ring resonators in an add-drop configuration used in this thesis.

The electric field inside the resonator can be expressed as:

$$
\begin{aligned}
E_c(w) = &- j \cdot k \cdot \mathrm{E_{in}} \cdot \zeta - j \cdot k \cdot \zeta \cdot \mathrm{E_{in}} \cdot t^2 \cdot e^{-j\beta(w)L} \cdot e^{-\frac{\alpha L}{2}} \\
&- j \cdot \zeta \cdot k \cdot t^4 \cdot \mathrm{Ein} \cdot e^{-2j\beta(w)L} \cdot e^{-\alpha L} \\
&- j \cdot \zeta \cdot k \cdot t^6 \cdot \mathrm{E_{in}} \cdot e^{-3j\beta(w)L} \cdot e^{-\frac{3\alpha L}{2}} + \ldots \\
&= -j \cdot k \cdot \zeta \cdot E_{\mathrm{in}} \left(1 + t^2 a e^{-j\theta} + t^4 a^2 e^{-2j\theta} + t^6 a^3 e^{-3j\theta} + \ldots\right)
\end{aligned} \tag{3.2.1}
$$

Considering the sum of this geometric series:

$$
\sum_{m=0}^{\infty} (t^2 a e^{-j\theta})^m = \frac{1}{1 - t^2 a e^{-j\theta}}
$$

the power circulating inside the ring becomes :

$$
P_c(\omega) = |E_c(\omega)|^2 = \frac{k^2 \left|E_{\mathrm{in}}\right|^2 (1-\eta^2)^2}{\left|1 - t^2 a e^{-j\theta}\right|^2} \tag{3.2.2}
$$

At resonance, where $\theta = 2\pi m$ and m is an integer number:

$$
1 - t^2 a e^{-j\theta} = 1 - t^2 a
$$

It is therefore possible to express, at resonance, formula 3.2.2 in the following way:

$$
P_c(\omega) = |E_c(\omega)|^2 = \frac{k^2 \left|E_{\mathrm{in}}\right|^2 (1-\eta^2)^2}{\left|1 - t^2 a\right|^2} \tag{3.2.3}
$$

The Finesse (F), is defined as :

$$
F = \frac{\mathrm{FSR}}{\mathrm{FWHM}}
$$

where the FSR (Free Spectral Range) is the distance between two consecutive resonances and the FWHM (Full Width at Half Maximum) is the width of the resonance at half of the maximum intensity.
For an add-drop ring resonator with $t_1 = t_2 = t$:

$$
F = \frac{\pi t \cdot \sqrt{a}}{1 - t^2 \cdot a} \tag{3.2.4}
$$

However, in the case where $k \ll 1$ and both propagation and coupling losses are negligible ($a = 1$, $\eta = 0$), it can be demonstrated that:

$$
t^2 = (1 - k^2)(1 - \eta^2) \approx 1
$$

The resulting finesse F is therefore given by:

$$
F = \frac{\pi\sqrt{t^2 a}}{1 - t^2 a} = \frac{\pi t}{1 - t^2} \approx \frac{\pi}{k^2}
$$

From this, it follows that:

$$k^2 \approx \frac{\pi}{F}$$

So the circulating power at the resonance is:

$$P_c = \frac{P_{\text{bus}} \cdot \frac{\pi}{F}}{(1-t^2)(1-t^2)} = P_{\text{bus}} \cdot \frac{\pi}{F(1-t^2)^2} \cdot \frac{\pi t^2}{\pi t^2} = \frac{P_{\text{bus}} \cdot F}{\pi t^2} = P_{\text{bus}} \cdot \frac{F}{\pi} \tag{3.2.5}$$

Thus, the ratio between the circulating power and the input power is:

$$\frac{P_c}{P_{\text{bus}}} = \frac{F}{\pi} \tag{3.2.6}$$

The expression 3.2.6 is valid when the propagation and coupling losses are negligible. Figure 3.2.2 shows the ratio between the circulating power and the input power as the finesse varies.

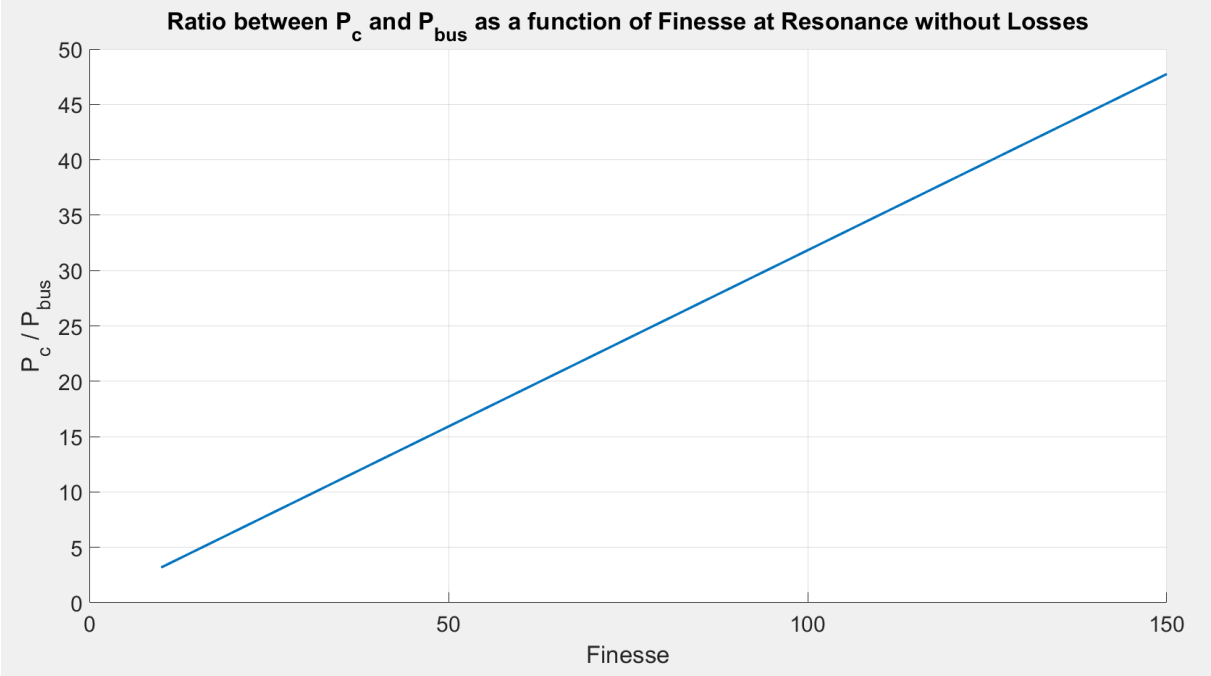

Figure 3.2.2: Ratio between P_c and P_{bus} as a function of Finesse at Resonance without losses.

Considering non-negligible propagation and coupling losses and assuming $t_1 = t_2 = t$, from equation 3.2.4, the self-coupling coefficient can be derived as follows:

$$t = \frac{-\pi\sqrt{a} \pm \sqrt{\pi^2 a + 4F^2 a}}{2F \cdot a} = \frac{-\pi\sqrt{a} \pm \sqrt{a(\pi^2 + 4F^2)}}{2F \cdot a} = \frac{\sqrt{a}}{2F \cdot a}\left(-\pi \pm \sqrt{\pi^2 + 4F^2}\right)$$

So the solutions for t are:

$$t = \frac{1}{2F\sqrt{a}}\left(-\pi \pm \sqrt{\pi^2 + 4F^2}\right)$$

Since t represents a value that must be positive, the real solution is the formula with the positive sign:

$$t = \frac{1}{2F\sqrt{a}}\left(\sqrt{\pi^2 + 4F^2} - \pi\right)$$

By squaring both sides, the following result is obtained:

$$t^2 = \frac{1}{4F^2 a}\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2$$

Consequently, the coupling coefficient can be expressed as follows:

$$k^2 = 1 - \frac{t^2}{1-\eta^2} = 1 - \frac{1}{4F^2 a\,(1-\eta^2)}\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2$$

The formula of the circulating power P_c at resonance is provided in equation 3.2.3, but for convenience, it is also provided here:

$$P_c = k^2 \cdot P_{\text{bus}} \cdot \frac{(1-\eta^2)^2}{|1 - t^2 a|^2}$$

Substituting k^2 and t^2 into formula 3.2.3:

$$P_c = \left(1 - \frac{\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2}{4F^2 a(1-\eta^2)}\right)^2 \cdot P_{\text{bus}} \cdot \frac{(1-\eta^2)^2}{\left(1 - \frac{\left(\sqrt{\pi^2+4F^2}-\pi\right)^2}{4F^2}\right)^2}$$

By Expanding the term $\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2$, the following is obtained:

$$\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2 = 2\pi^2 + 4F^2 - 2\pi\sqrt{\pi^2 + 4F^2}$$

Thus, the term in the denominator becomes:

$$\left(1 - \frac{2\pi^2 + 4F^2 - 2\pi\sqrt{\pi^2 + 4F^2}}{4F^2}\right)^2 = \left(\frac{4F^2 - 2\pi^2 - 4F^2 + 2\pi\sqrt{\pi^2 + 4F^2}}{4F^2}\right)^2 =$$
$$= \left(\frac{2\pi\sqrt{\pi^2 + 4F^2} - 2\pi^2}{4F^2}\right)^2 = \left(\frac{\pi(\sqrt{\pi^2 + 4F^2} - \pi)}{2F^2}\right)^2 = \frac{\pi^2(\sqrt{\pi^2 + 4F^2} - \pi)^2}{4F^4}$$

Replacing this result into the expression for P_c:

$$P_c = \left(1 - \frac{\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2}{4F^2 a(1-\eta^2)}\right)^2 \cdot P_{\text{bus}} \cdot \frac{(1-\eta^2)^2}{\frac{\pi^2(\sqrt{\pi^2+4F^2}-\pi)^2}{4F^4}}$$

Further simplifying:

$$P_c = \left(1 - \frac{\left(\sqrt{\pi^2 + 4F^2} - \pi\right)^2}{4F^2 a(1-\eta^2)}\right)^2 \cdot P_{\text{bus}} \cdot \frac{4F^4(1-\eta^2)^2}{\pi^2(\sqrt{\pi^2+4F^2}-\pi)^2} =$$

$$= P_{\text{bus}} \cdot \frac{4F^4(1-\eta^2)^2}{\pi^2(\sqrt{\pi^2+4F^2}-\pi)^2} - P_{\text{bus}} \cdot \frac{\left(\sqrt{\pi^2+4F^2}-\pi\right)^2 \cdot 4F^4(1-\eta^2)^2}{\pi^2(\sqrt{\pi^2+4F^2}-\pi)^2 \cdot 4F^2 a(1-\eta^2)} =$$

$$= P_{\text{bus}} \cdot \frac{4F^4(1-\eta^2)^2}{\pi^2(\sqrt{\pi^2+4F^2}-\pi)^2} - P_{\text{bus}} \cdot \frac{F^2(1-\eta^2)}{a\pi^2}$$

The expression can be rewrite in the following way:

$$P_c = \frac{P_{\text{bus}}(1-\eta^2)F^2}{\pi^2}\left(\frac{4F^2(1-\eta^2)}{2\pi^2+4F^2-2\pi\sqrt{\pi^2+4F^2}} - \frac{1}{a}\right) \tag{3.2.7}$$

This is the expression for the circulating power P_c as a function of the finesse F, the coupling loss η^2, and the intrinsic losses a. Thus, the ratio between the circulating power and the input power is:

$$\frac{P_c}{P_{\text{bus}}} = \frac{(1-\eta^2)F^2}{\pi^2}\left(\frac{4F^2(1-\eta^2)}{2\pi^2+4F^2-2\pi\sqrt{\pi^2+4F^2}} - \frac{1}{a}\right) \tag{3.2.8}$$

The expression 3.2.7 accounts for the impact of losses on the circulating power while the expression 3.2.5 is valid for zero losses. A comparison between the two expressions is shown in Figure 3.2.3. In this figure, as well as in all subsequent figures in this chapter, the parameters used that are not subject to variations correspond to those of the microring selected for the experiments, which will be presented later in this thesis. For clarity, the key parameters are also listed below:

- resonant wavelength: $\lambda_0 = 1534.53\,\text{nm}$
- Power coupling coefficient: $\kappa^2 = 0.035622$
- Coupling losses: $\eta^2 = 0.0073$
- Intrinsic losses: $\alpha_0 = 2\,\text{dB/cm}$
- Group index: $n_g = 4.27032$
- Radiative losses: $\alpha_{\text{rad}} < 0.01\,\text{dB/cm}$
- Effective refractive index: $n_{\text{eff},0} = 2.32$
- Confinement factor: $\Gamma = 1$
- Path length: $L = 79.42\ \mu\text{m}$
- Effective area: $A_{\text{eff}} = 0.075\,\mu\text{m}^2$

While the ideal case without losses provides a simple relationship between circulating power and finesse, the inclusion of propagation and coupling losses introduces corrections that become increasingly important as the finesse increases. In fact, as can be seen in 3.2.3, the ratio between the circulating power and P_{bus} with losses is lower than the case without losses and the

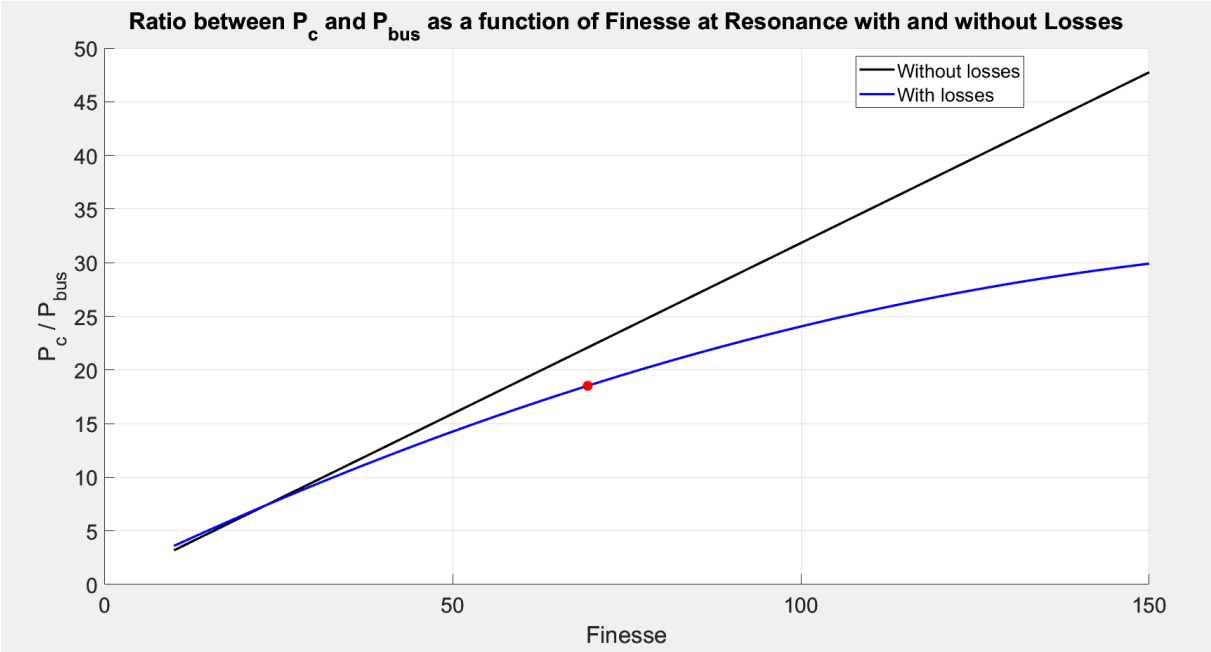

Figure 3.2.3: Comparison of ratio between P_c and P_{bus} as a function of Finesse at Resonance with and without losses. The black line shows the ideal case without losses while the blue line shows the case with losses with $a = 0.9982$, $\eta^2 = 0.0073$ and $L = 79.42\ \mu m$.

gap increases more and more as the finesse increases. With the losses, the circulating power Pc is lower compared to the case in which the losses are negligible, for the same values of finesse. As the finesse increases, the circulating power increases in both cases. A higher finesse indicates that the resonator is more selective, with a narrower resonance bandwidth, allowing the resonator to store more energy over multiple round-trips.
The red point in the graph 3.2.3 on the curve with losses represents the ratio between circulating power and input power for the microring selected for the experiments presented later on in this thesis. Similarly, in the subsequent graphs, the red point will consistently refer to the selected microring.
The ratio between the circulating power P_c and the input power P_{bus} was also analyzed as a function of finesse for different values of α, which represents the sum of the intrinsic losses and the losses due to light radiated in the cladding: $\alpha_0 + \alpha_{\text{rad}}$.
As shown in Figure 3.2.4, the analysis demonstrates that as α increases, the ratio between P_c and P_{bus} decreases because with higher losses, the amount of power retained in the microring is reduced.

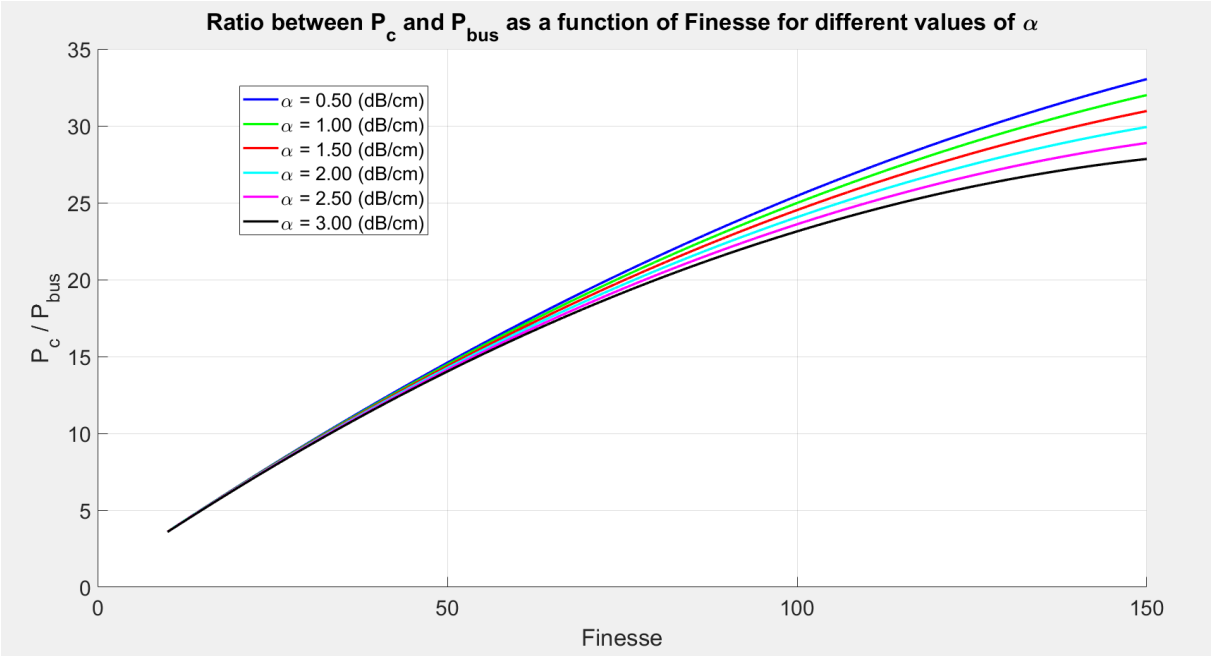

Figure 3.2.4: Ratio between P_c and P_{bus} as a function of finesse for different values of α with $\eta^2 = 0.0073$ and $L = 79.42\ \mu m$.

3.2.2 Equations for circulating power versus quality factor Q

Expressing the finesse in terms of the quality factor Q as $F = \frac{Q\lambda_{\text{res}}}{n_g L}$, the equations 3.2.5, 3.2.7 can be rewritten in terms of Q.
The equations 3.2.5 becomes the following:

$$P_c = P_{\text{bus}} \cdot \frac{\frac{Q\lambda_{\text{res}}}{n_{\text{g}} L}}{\pi} = \frac{P_{\text{bus}} \cdot Q\lambda_{\text{res}}}{\pi n_{\text{g}} L} \tag{3.2.9}$$

The equations 3.2.7 becomes the following:

$$P_c = \frac{P_{\text{bus}} \cdot (1-\eta^2) \cdot \left(\frac{Q\lambda_{\text{res}}}{n_{\text{g}} L}\right)^2}{\pi^2} \left(\frac{4\left(\frac{Q\lambda_{\text{res}}}{n_{\text{g}} L}\right)^2 (1-\eta^2)}{2\pi^2 + 4\left(\frac{Q\lambda_{\text{res}}}{n_{\text{g}} L}\right)^2 - 2\pi\sqrt{\pi^2 + 4\left(\frac{Q\lambda_{\text{res}}}{n_{\text{g}} L}\right)^2}} - \frac{1}{a} \right)$$

Simplifying:

$$P_c = \frac{P_{\text{bus}} \cdot (1-\eta^2) \cdot Q^2\lambda_{\text{res}}^2}{\pi^2 n_{\text{g}}^2 L^2} \left(\frac{4Q^2\lambda_{\text{res}}^2(1-\eta^2)}{n_{\text{g}}^2 L^2 \left(2\pi^2 + \frac{4Q^2\lambda_{\text{res}}^2}{n_{\text{g}}^2 L^2} - 2\pi\sqrt{\pi^2 + \frac{4Q^2\lambda_{\text{res}}^2}{n_{\text{g}}^2 L^2}}\right)} - \frac{1}{a} \right) \tag{3.2.10}$$

The graph in Figure 3.2.5 shows the ratio between circulating power P_c and P_{bus} as a function of the quality factor Q for the cases described by equations 3.2.9 and 3.2.10.

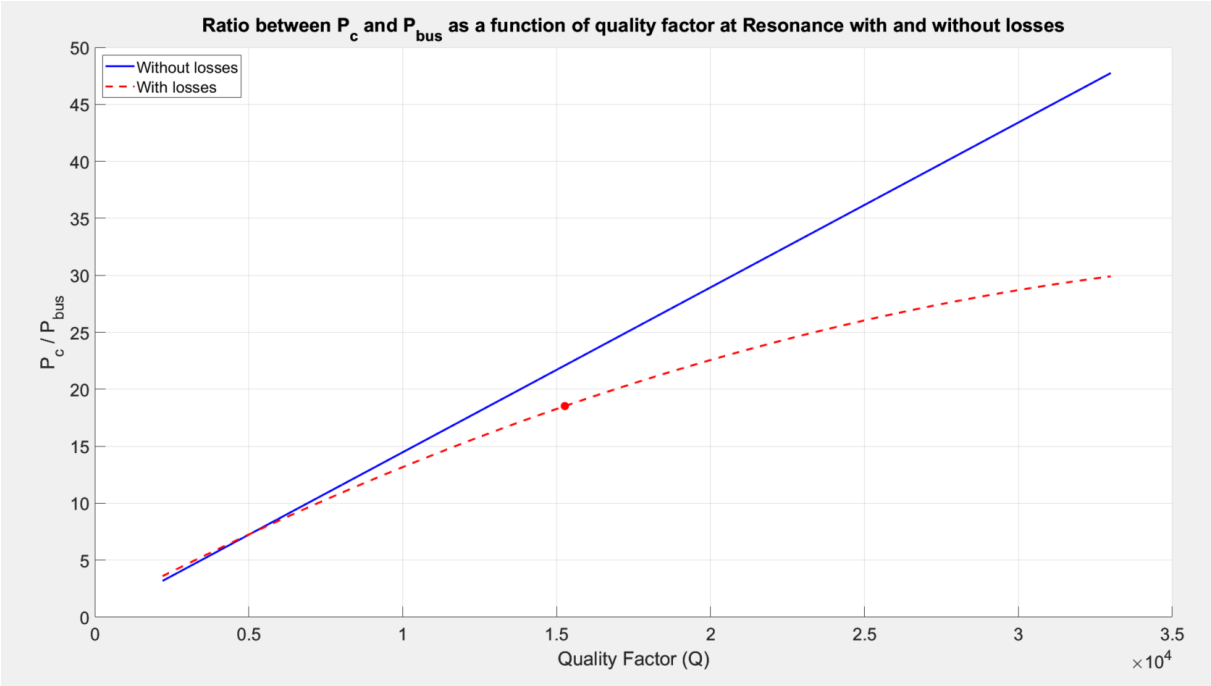

Figure 3.2.5: Ratio between circulating power P_c and P_{bus} as a function of Quality Factor Q with and without losses, obtained for $a = 0.9982$, $\eta^2 = 0.0073$ and $L = 79.42\ \mu m$.

3.2.3 Equations for circulating power in terms of ring length, k^2 and a

Circulating power versus ring length

To obtain the ratio between the circulating power P_c and the input power P_{pump} as a function of the ring length L, the value of L was varied from 10 μm to 500 μm. For each value of L, a was calculated with $\alpha_0 = 2$ dB/cm and $\alpha_{\text{rad}} < 0.001$ dB/cm. Subsequently, the corresponding finesse F was calculated for each value of a with $k^2 = 0.035622$. The ratio P_c/P_{pump}, shown in figure 3.2.6, was determined with and without losses using the formulas 3.2.8 and 3.2.6, respectively.
In the case with losses, the greater the value of L, the greater the losses, resulting in a lower finesse. If the finesse decreases, it means that energy dissipation within the microring increases, and consequently, the circulating power decreases. Therefore, a higher value of L leads to a lower ratio, as the losses increase.
In the ideal case without losses, the microring is able to retain more energy, and thus the ratio remains consistently higher compared to the case with losses.

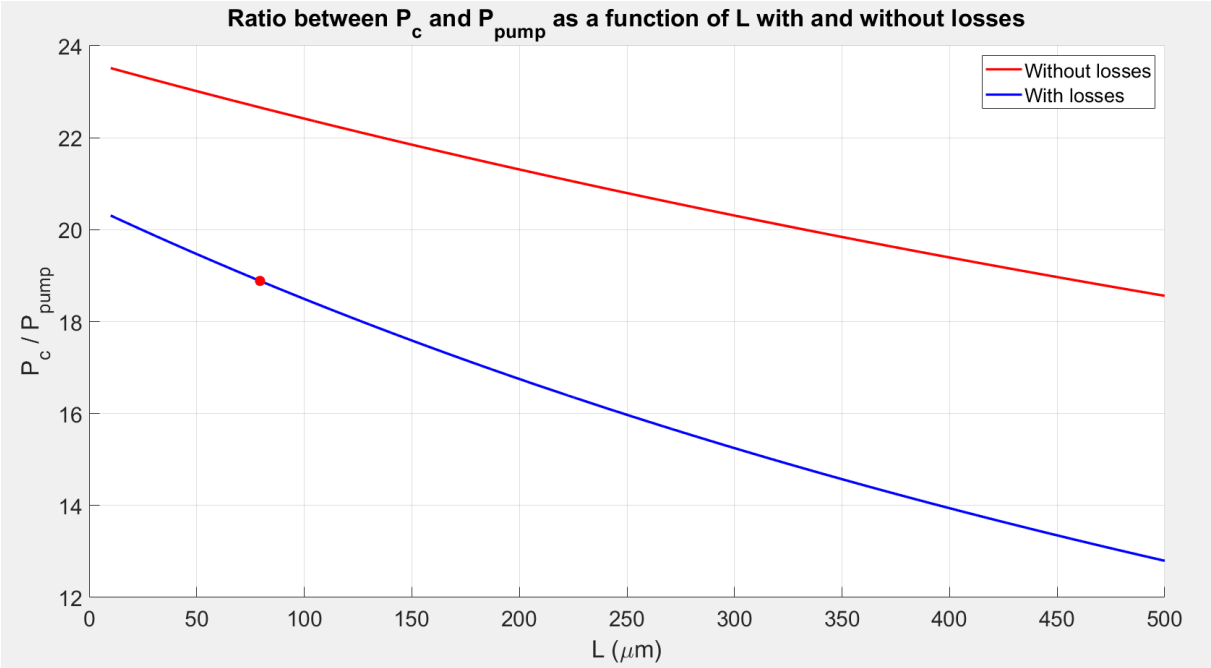

Figure 3.2.6: Ratio between circulating power P_c and P_{bus} as a function of the ring length L with and without losses at resonance, assuming $\eta^2 = 0.0073$ and $k^2 = 0.035622$.

Circulating power versus k^2

The ratio between the circulating power P_c and the input power P_{bus} at resonance, with negligible losses, as a function of k^2 is given by:

$$\frac{P_c}{P_{\text{bus}}} = \frac{F}{\pi \cdot t^2} = \frac{1}{\sqrt{1-k^2} \cdot k^2}$$

This expression is derived by considering the following relationships:

$$t = \sqrt{1-\kappa^2}; \quad F = \frac{\pi \cdot t}{1-t^2}$$

With losses the equation 3.2.8 is considered, with the following considerations:

$$t = \sqrt{(1-\kappa^2)(1-\eta^2)};$$
$$F = \frac{\pi \cdot t \cdot \sqrt{a}}{1-a \cdot t^2}$$

In Figure 3.2.7, the ratio P_c/P_{bus} is shown as a function of κ^2 for the cases with and without losses at resonance with κ^2 varying from 0 to 0.3.
In Figure 3.2.7, an initial increase in the ratio between the circulating power and the input power is observed because a greater amount of power is coupled into the resonator, reaching a maximum value beyond which it begins to decrease. The decrease occurs because, as κ^2 becomes larger, the energy does not have enough time to ”circulate” effectively within the resonator to create significant power buildup, as the coupled power exits almost immediately. This reduces the time the power remains inside the resonator, and as a result, the circulating power decreases, which leads to a reduction in the P_c/P_{bus} ratio.

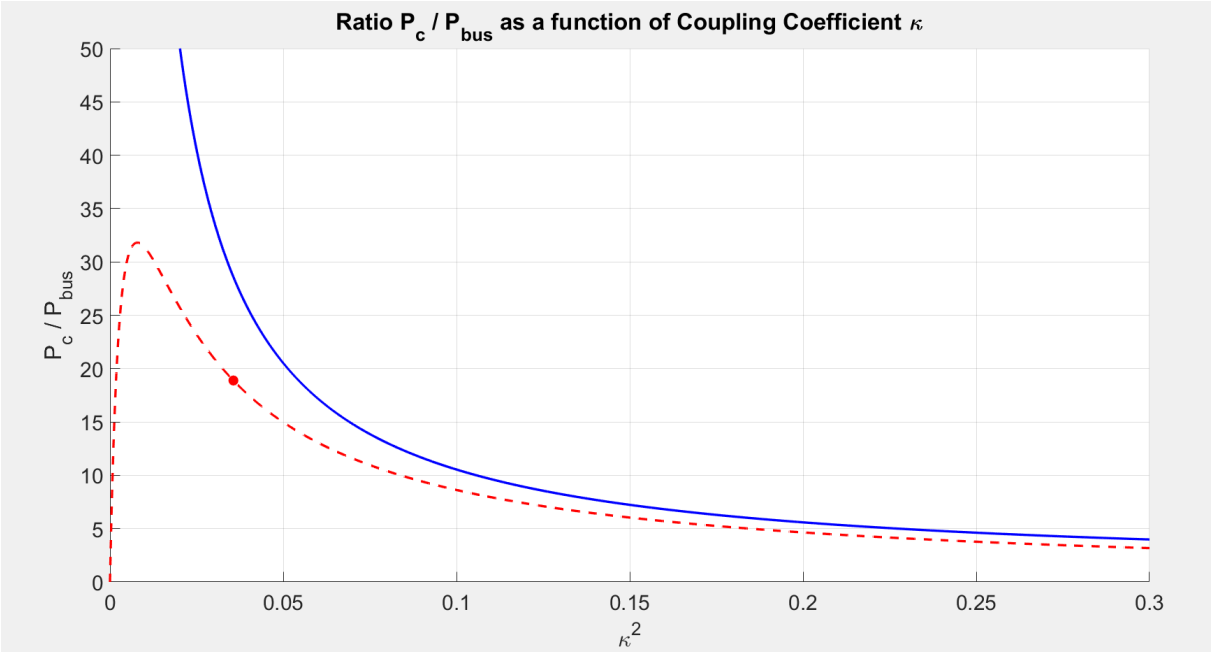

Figure 3.2.7: Ratio P_c/P_{bus} as a function of coupling coefficient κ^2 for the different cases with and without losses at resonance, with $\eta^2 = 0.0073$, $a = 0.9982$.

Circulating power versus a

The figure 3.2.8 illustrates the relationship between the ratio P_c/P_{bus} and a. The circulating power ratio P_c/P_{bus} increases with a approach 1. This means that the less losses there are, the more power is trapped in the ring.

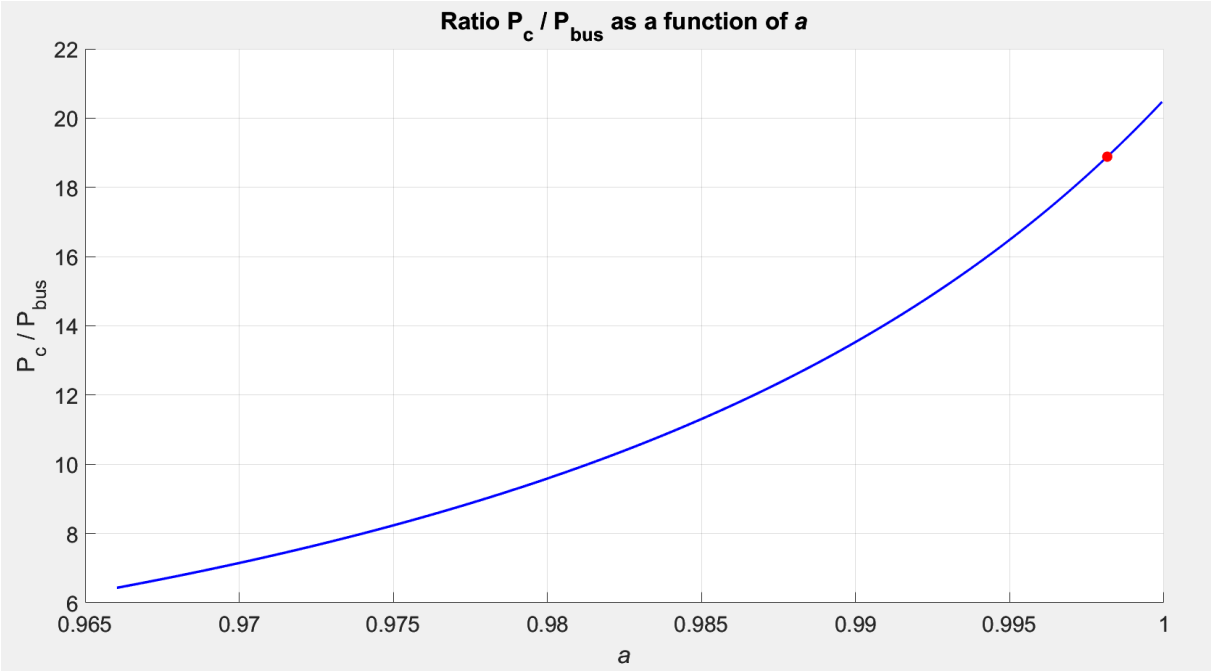

Figure 3.2.8: Ratio P_c/P_{bus} as a function of a, assuming $\eta^2 = 0.0073$ and $k^2 = 0.035622$. The graph shows that as a increases, the ratio increases, i.e. the power circulating inside the ring increases.

3.2.4 Transmission coefficient at the through and drop ports.

In a microring resonator in an add-drop configuration, the general expression of the electric fields at the through and drop ports are given by the following formulas:

Expression for E_{th}

$$E_{\text{th}} = E_{\text{in}} t_1 + \frac{E_{\text{in}} a e^{i\theta} |k_1|^2 t_2^* + E_{\text{add}} \sqrt{a} k_1 e^{i\theta/2} k_2^*}{a e^{i\theta} t_1^* t_2^* - 1} \tag{3.2.11}$$

This equation represents the total field at the through port, accounting for the input field E_{in}, the drop port input field E_{add}, and the round-trip loss factor $ae^{i\theta}$.

Expression for E_{d}

$$E_{\text{d}} = -\frac{t_2 \left(-E_{\text{add}} t_2 + E_{\text{add}} a e^{i\theta} |k_2|^2 t_1^* + E_{\text{add}} a e^{i\theta} |t_2|^2 t_1^* + E_{\text{in}} \sqrt{a} k_2 e^{i\theta/2} k_1^* \right)}{t_2 - a e^{i\theta} |k_1|^2 t_1^*} \tag{3.2.12}$$

This expression represents the electric field at the drop port.
Assuming $E_{\text{add}} = 0$, $t_1 = t_2$, and $k_1 = k_2$, the expressions simplify to:

$$\frac{E_{\text{th}}}{E_{\text{in}}} = t_1 + \frac{a e^{i\theta} |k_1|^2 t_1^*}{a e^{i\theta} (t_1^*)^2 - 1} \tag{3.2.13}$$

$$\frac{E_{\text{d}}}{E_{\text{in}}} = -\frac{\sqrt{a} t_1^2 e^{i\theta/2} |k_1|^2}{t_1^2 - a e^{i\theta} |t_1|^4} \tag{3.2.14}$$

So the power transmission coefficient at the through and drop ports, including coupling losses, is:

$$T_{\text{thr}} = t^2 \frac{\left| 1 - (1 - \eta^2) a e^{j\theta} \right|^2}{\left| 1 - t^2 a e^{j\theta} \right|^2} \tag{3.2.15}$$

$$T_{\text{drop}} = \frac{k_1^4 \left(1 - \eta^2\right)^2 a}{\left| 1 - t^2 a e^{j\theta} \right|^2} \tag{3.2.16}$$

3.3 Criteria for Choosing the Optimal Microring

Figure 3.3.1 shows the layout of the tested chip, designed at the Politecnico di Torino.
The choice of the right microring for reservoir computing experiments requires important considerations of the coupling coefficient κ, the quality factor Q, the ratio between the circulating power P_c and bus power P_{bus} and the output power. The ratio P_c/P_{bus} initially increases with κ^2 as more power is coupled into the resonator, but beyond a certain point, the ratio decreases because the power exits too quickly from the ring. Where the ratio is higher, the nonlinear effects will be stronger and this is very useful for reservoir computing experiments where the ring is used as a nonlinear element. However, attention must also be paid to the transmission coefficient. As shown in figure 3.3.2 on the left, it is necessary to avoid a too low k because it leads to a relatively high transmission coefficient T_{thr} at resonance, which is not ideal. At

Figure 3.3.1: chip layout.

resonance, the lowest possible transmission at the through port is desired for the purposes of this thesis because when T_{thr} is low at resonance, it means that more optical power is stored in the ring, thereby maximizing the nonlinear effects.
Additionally, the quality factor Q increases with decreasing κ, as shown in figure 3.3.3. Q defines the ability of the microring to mantain energy within itself. Thus, as k increases, more energy is coupled into the microring, but at the same time, the energy will more easily escape from the microring, leading to a decrease in Q.
The relationship between Q and κ depends by FWHM, which, in the case of weak coupling, is expressed [7] as:

$$\text{FWHM} = \frac{\kappa^2 \lambda^2}{\pi n_{\text{eff}} L}$$

A lower κ results in a lower FWHM, which leads to an higher Q, increasing the sharpness of the resonance. In Figure 3.3.3, the relationship between Q and k is illustrated, based on the implementation of the formula 3.3.1.

$$Q\text{-factor} = \frac{\pi n_g L \sqrt{t^2 a}}{\lambda_{\text{res}} (1 - ta)} \tag{3.3.1}$$

In the figure 3.3.2 on the right side, the transmission coefficient as a function of losses is shown. It can be observed that the transmission at resonance at the through port decreases as the losses

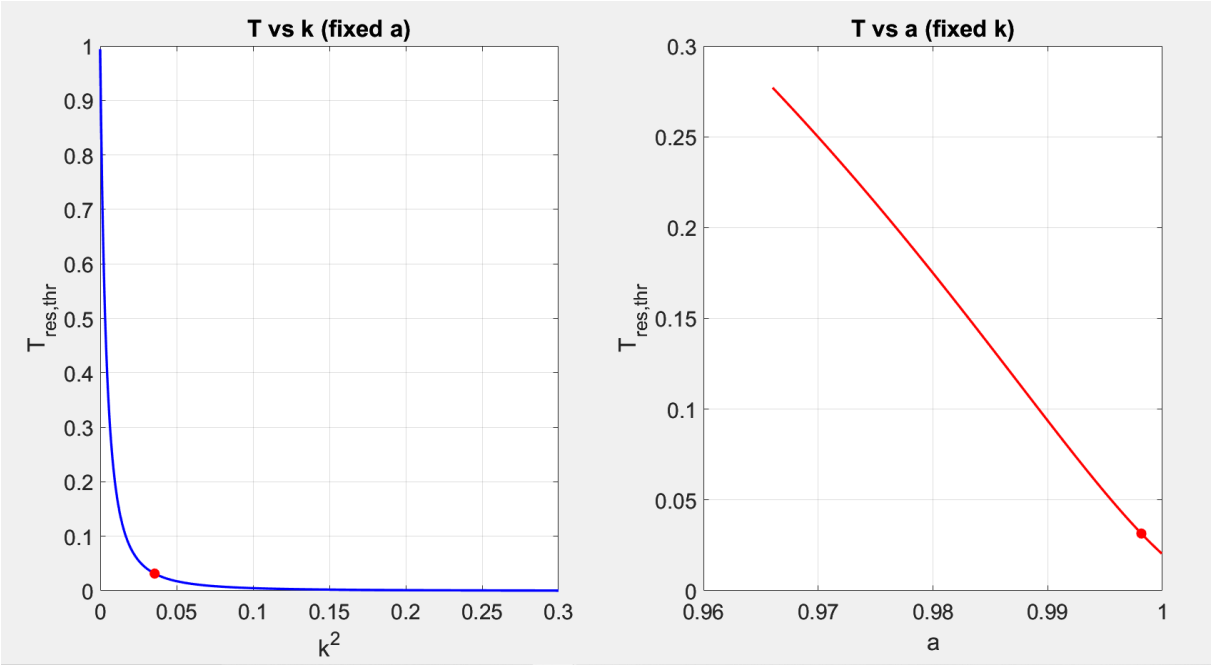

Figure 3.3.2: The transmission coefficient T_{thr} at resonance as a function of κ (left) and a (right). The graph as a function of κ was obtained by fixing $a = 0.9982$, while the one as a function of a was obtained by fixing $\kappa^2 = 0.035622$.

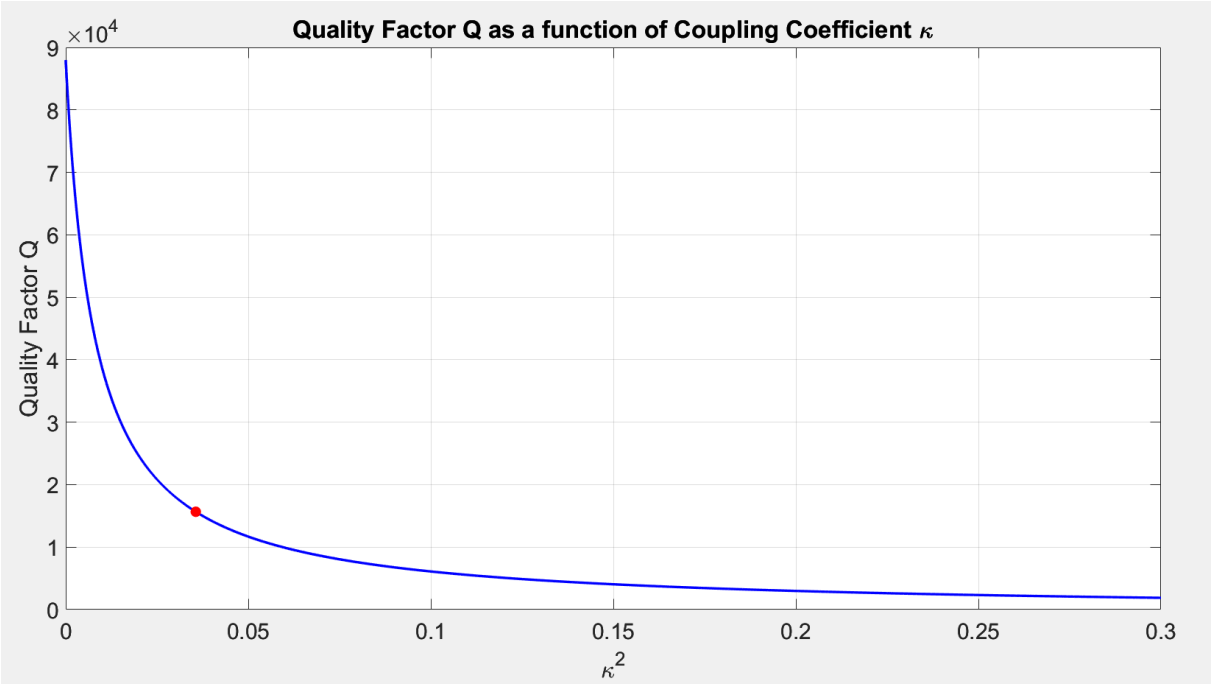

Figure 3.3.3: The relationship between the quality factor Q and the power coupling coefficient k^2, as described by Equation 3.3.1 with $a = 0.9982$ and $L = 79.42\ \mu m$.

decrease.
Therefore, the optimal microring is achieved by balancing P_c/P_{bus}, T, and Q, selecting a κ^2 that is small enough to maximize the circulating power, but not too small, otherwise too little power is coupled into the resonator, leading to excessive transmission at the through port at resonance. It is also necessary to have a high quality factor, which decreases with increasing κ^2. An op-

timal range of κ^2 is between the peak and the minimum of the red curve in Figure 3.2.7, as in this region the circulating power is sufficiently high, the transmission T at resonance is low enough, and Q is high enough for the experiments conducted in this thesis. In fact, the chosen microring, indicated by the red point in figure 3.2.7, is located in the center of this region.
For this reason, various rings were tested, and the one that provided the best compromise between these characteristics was selected.
I tested a total of 6 microrings, 3 from chip 16 and 3 from chip 19 of column D, all manufactured by IMEC. The microring D8D9 from chip 19 was chosen because it shows the best compromise between k, Q, and T.
At the cold resonance wavelength (1534.53 nm) the ring has a Q factor of 15267.52, a k^2 of 0.0356, intrinsic losses α = 1.8769 dB/cm, and radiation losses of -21.1876 dB.
It is also interesting to show the relationship between k and the coupler distance. To achieve this, the transmission spectrum at the through port of different rings was measured, and the coupling coefficient k was determined through a fitting process.
To better demonstrate this relationship, since the six rings tested had only three different coupler distances, I used the measured spectra from Dr. Marco Novarese on other rings to create a curve that represents the relationship between k and the coupler distance with more data points. In the figure 3.3.4, it is shown the coupling coefficient k as a function of the coupler distance. The plot demonstrates the inverse relationship between k and the coupler distance. k describes the

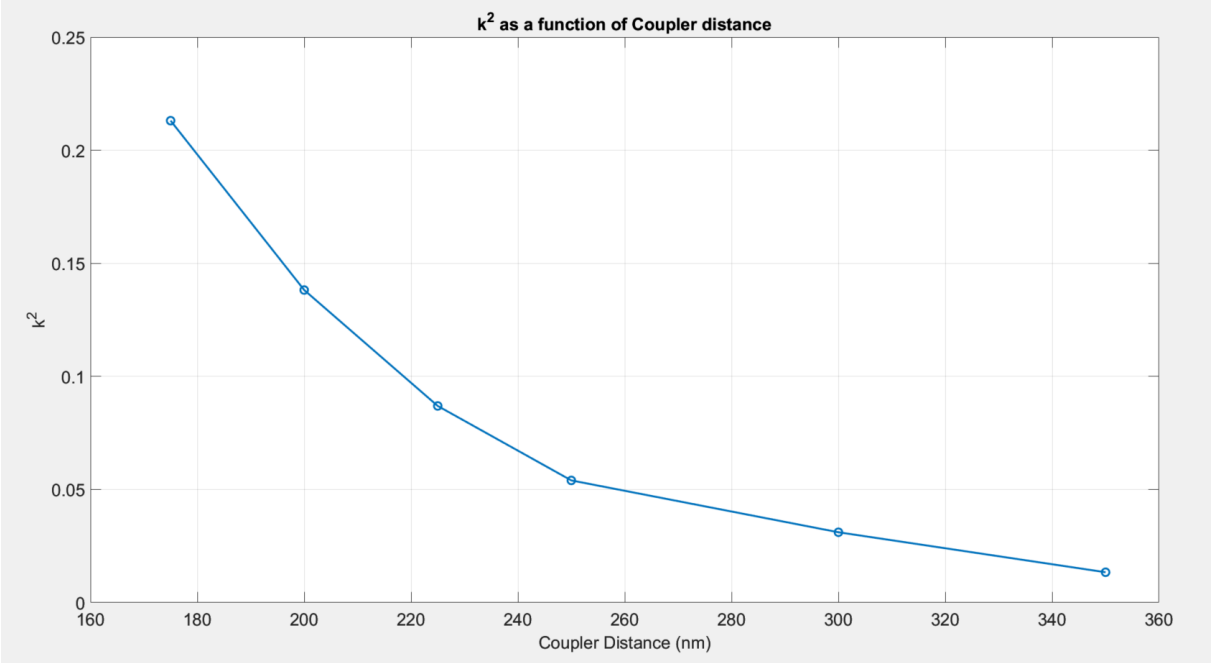

Figure 3.3.4: Experimental plot of the coupling coefficient k as a function of the coupler distance.

efficiency with the energy is transferred from the bus to the microring. The smaller the coupler distance, the larger the value of k , and therefore, more energy is transferred from the bus to the microring.

3.4 Nonlinear model of the microring resonator

In microring resonators, nonlinear effects are of fundamental importance because they modify the optical response of the ring, especially at high powers. The main nonlinear effects are two-photon absorption (TPA) and the thermo-optic effect. The process of two-photon absorption (TPA) occurs when the combined energy of two photons is sufficient to excite an electron from the valence band to the conduction band. This excitation leads to the generation of free carriers in both the conduction and valence bands. As a result, the nonlinear phenomena of free-carrier absorption (FCA) and free-carrier dispersion (FCD) are induced. FCA consists of the absorption of a photon that excites one of the free carriers just generated from an excited state to another with higher energy. This affects the imaginary part of the refractive index.
The generated free carriers cause a variation in the refractive index, which is FCD. FCD alters the real part of the refractive index.
FCD also contributes to increased absorption, increasing the overall optical loss and reducing the quality factor of the resonator.
The thermo-optic effect derives from release of the absorbed photon energy into heat, which leads to thermalization and recombination through Shockley-Read-Hall (SRH) processes. These are two phenomena that generate heat (self-heating), thus increasing the temperature and altering the refractive index of the silicon waweguide.
The interaction between nonlinear effects can cause optical bistability, meaning that the system can have two stable transmission states for the same wavelength or input power, depending on whether the power or wavelength has been increased or decreased over time [8]. This is caused by the different time scales of the slower thermo-optic effect (on the order of hundreds of ns to reach thermal equilibrium) and the much faster FCD (on the order of tens of ns). The thermo-optic effect causes a resonance shift towards longer wavelengths, while FCD shifts it towards shorter wavelengths. Optical bistability typically manifests as a hysteresis cycle, where the system's response depends on its history. The double path in the hysteresis cycle is due to the fact that the heating and cooling of the resonator occur more slowly compared to the change in optical power. The combination of nonlinear effects can also lead to complex behaviors such as self-oscillation of the output signal, considering a high constant input at a wavelength near resonance.
To analyze the impacts of nonlinear effects on the ring resonator and how these affect the reservoir computing experiments, the model developed and validated in [2] was used.
In the following subsections of this section all the equations of the above mentioned model are presented.

3.4.1 Differential equations governing thermal dynamics in the system

The self-heating phenomenon can be modeled through the concept of thermal impedance Z_T.The thermal impedance represents a property of the system that describes how effectively the device can dissipate heat and is defined as:

$$Z_T = \frac{\Delta T}{P_{\text{abs}}}$$

where $\Delta T = T_{\text{max}} - T_0$ represents the temperature increase due to the heat generated, P_{abs} the absorbed optical power, and T_0 is the ambient temperature.
In [2], to model the transient thermal response of the microring , several thermal time constants were considered, each associated with a different node of the equivalent thermal model of the

microring, called the Foster model. The Foster model describes the temperature variation as the sum of the contributions of the different thermal nodes, each characterized by its own thermal impedance and thermal capacitance. These thermal nodes are thus equivalent to electrical RC circuits.
The general equation that describes the temperature variation over time is given by:

$$\Delta T(t) = \sum_{i=1}^{n} \Delta T_i \left(1 - e^{t/\tau_{\text{th},i}}\right) \tag{3.4.1}$$

where ΔT_i represents the temperature variation at each node and $\tau_{\text{th},i} = Z_{T,i}C_i$ is the thermal time constant of node i.
In this thesis, a pump-probe experiment was conducted to perform reservoir computing experiments. The pump signal is used to generate free carriers and induce nonlinear effects in the silicon waveguide, while the probe signal is used to measure the effects caused on the system by the pump.
The temperature dynamics at different nodes for the pump signal are described by the following differential equations:
Node 1:

$$\frac{dT_1}{dt} = -\frac{\Delta T_1}{\tau_{\text{th},1}} + \frac{P_{\text{abs,pump}}}{C_1} \tag{3.4.2}$$

Node 2:

$$\frac{dT_2}{dt} = -\frac{\Delta T_2}{\tau_{\text{th},2}} + \frac{P_{\text{abs,pump}}}{C_2} \tag{3.4.3}$$

Node 3:

$$\frac{dT_3}{dt} = -\frac{\Delta T_3}{\tau_{\text{th},3}} + \frac{P_{\text{abs,pump}}}{C_3} \tag{3.4.4}$$

Node 4:

$$\frac{dT_4}{dt} = -\frac{\Delta T_4}{\tau_{\text{th},4}} + \frac{P_{\text{abs,pump}}}{C_4} \tag{3.4.5}$$

The total temperature variation caused by the pump is the sum of the contributions from each node. The same equations hold for the probe signal, with the absorbed power of the probe signal $P_{\text{abs,probe}}$ instead of the absorbed power of the pump signal $P_{\text{abs,pump}}$. Thus, the total temperature variation ΔT_{tot} is the sum of the contributions of all the temperature variations caused by both the pump and the probe.
In the table 3.4.1, the relationships between Z_T and various parameters are described. Z_T is a physical constant that depends on the material and geometric properties of the device. If Z_T increases, the device dissipates heat with greater difficulty.
The silicon waveguide has a high refractive index contrast with the surrounding silicon dioxide. This high refractive index contrast helps to confine the light within the waveguide. However, if the SiO_2 layer is too thin, the optical field can penetrate and extend into the underlying substrate, which is made of silicon, causing light to be radiated in the substrate.

Table 3.4.1: Relationship between Z_T and physical parameters

Parameter	**Effect on Z_T**	**Effect on thermal dissipation**
Thickness of SiO_2 layer	If it increases, Z_T increases	Thermal dissipation worsens because SiO_2 has a lower thermal conductivity than silicon.
Ring resonator length	If it increases, Z_T decreases	Thermal dissipation improves because there is more area for heat dissipation.

3.4.2 Generation of free carriers and differential equations for electron and hole density

In this subsection, the differential equations describing the dynamics and generation of free carriers in microring resonators, in the presence of *pump* and *probe* signals, are analyzed, and how carrier generation is influenced by *Single Carrier Absorption* (SCA), TPA, as well as by *cross two-photon absorption* (XTPA) between the *pump* and *probe* signals, is discussed.
In semiconductors, there are different types of recombination that can occur, including:

- **Radiative recombination**: an electron in the conduction band recombines with a hole in the valence band, emitting a photon. This type of recombination occurs in direct-gap semiconductors, such as GaAs, but does not occur in silicon microrings due to the nature of silicon, which is an indirect-gap semiconductor.
- **Auger recombination**: the recombination between an electron and a hole generates energy that is transferred to another electron (or hole), which is promoted to a higher energy state. This type of recombination becomes significant at high free carrier densities.
- **Trap-assisted recombination (Shockley-Read-Hall, SRH)**: This is the main recombination mechanism in silicon microrings. In this case, free carriers are captured by defects or impurities in the crystal lattice, which act as traps. Electrons and holes recombine through the energy levels introduced by these traps, releasing energy in the form of heat rather than light. SRH theory provides an accurate explanation of the nonlinear carrier dynamics.

Free Carrier Recombination Rate Equation

Often, the rate equation for free carriers is simplified into the following form:

$$\frac{dN}{dt} = G_{\text{gen}} - R_{\text{rec}}$$

Where:

- G_{gen} is the generation rate discussed in the following.

- R_{rec} is the total recombination rate, given by the sum of spontaneous, non-radiative, and stimulated recombination contributions, i.e., $R_{\text{rec}} = R_{\text{sp}} + R_{\text{nr}} + R_{\text{st}}$. In the case of silicon microring resonators, since silicon is an indirect-gap semiconductor, only non-radiative recombination (R_{nr}) occurs.

In detail:

$$R_{\text{sp}} = BN^2 \quad \text{(spontaneous recombination)}$$

$$R_{\text{nr,SRH}} = AN \quad \text{(non-radiative recombination via Shockley-Read-Hall)}$$

$$R_{\text{nr,Auger}} = CN^3 \quad \text{(non-radiative recombination via Auger)}$$

where N is the free carrier density.
As optical power increases, the free carrier density rises, and the recombination rate increases because there are more electrons and holes available for the recombination process. This affects the rate at which carriers recombine. The carrier lifetime τ can decrease significantly when there are many free carriers because the probability of an electron or a hole being captured by a trap increases. This means that the traps fill up more quickly.

Calculation of the Total Generation Rate G

The total carrier generation rate for the pump, G_{pump}, is defined as the sum of the contributions from two-photon absorption and cross two-photon absorption:

$$G_{\text{pump}} = G_{\text{pump,TPA}} + G_{\text{pump,XTPA}} \tag{3.4.6}$$

where contributions are defined as:

$$G_{\text{pump,TPA}} = \frac{\alpha_{\text{TPA}} P_{\text{c,pump}}^2}{2\frac{hc}{\lambda_{\text{pump}}} A_{\text{eff}}} \tag{3.4.7}$$

$$G_{\text{pump,XTPA}} = \frac{2\alpha_{\text{TPA}} P_{\text{c,pump}} P_{\text{c,probe}}}{\frac{hc}{\lambda_{\text{pump}}} A_{\text{eff}}} \tag{3.4.8}$$

where:

- α_0 represents the linear losses of the system,
- η_{SCA} is the efficiency of the SCA process,
- $P_{\text{c,pump}}$ is the circulating power of the pump,
- $P_{\text{c,probe}}$ is the circulating power of the probe,
- $\hbar\omega_0 = \frac{hc}{\lambda}$ represents the photon energy at the operating wavelength,
- λ_{pump} is the pump wavelength,
- A_{eff} is the effective area,
- α_{TPA} is the two-photon absorption coefficient.

The same equations apply to the probe signal, replacing $P_{\text{c,pump}}$ with $P_{\text{c,probe}}$ and λ_{pump} with λ_{probe}.
XTPA (*cross two-photon absorption*) is the simultaneous absorption of one photon from the pump signal and one photon from the probe signal, and this causes additional losses in the system.
Thus, the total carrier generation rate is the sum of all the contributions:

$$G = G_{\text{pump,TPA}} + G_{\text{probe,TPA}} + G_{\text{XTPA,pump}} + G_{\text{XTPA,probe}} \tag{3.4.9}$$

Differential Equation for the Excess Electron Density n_e

Excess carriers refer to the density of electrons or holes generated beyond the levels present under thermodynamic equilibrium conditions. The variation of the excess electron density over time can be described by the following differential equation[9]:

$$\frac{dn_e}{dt} = G - \frac{1}{\tau_{n_0}} \left(\frac{(n_0 + n_1 + n_e)(n_e - p_e)}{N_f} + \frac{n_e n_1}{n_e + n_0} \right) \tag{3.4.10}$$

where:

- n_e is the excess electron density,
- p_e is the excess hole density,
- n_0 and n_1 are the electron concentrations under equilibrium conditions,
- N_f is the trap density,
- τ_{n_0} is the recombination lifetime for electrons.

The first term at RHS is the generation rate, while the second term is the SRH recombination.

Differential Equation for the Excess Hole Density p_e

Similarly, the variation of the excess hole density over time [9] is described by:

$$\frac{dp_e}{dt} = G - \frac{1}{\tau_{p_0}} \left(\frac{(p_0 + p_1 + p_e)(p_e - n_e)}{N_f} + \frac{p_e p_1}{p_e + p_0} \right) \tag{3.4.11}$$

with:

- p_0 and p_1 are the hole concentrations under equilibrium conditions,
- τ_{p_0} is the recombination lifetime for holes.

3.4.3 Differential equations for the circulating field

Before presenting the differential equations for the circulating field, it is essential to discuss the losses and the variation in the refractive index, caused by nonlinear effects.

Microring losses

The absorption coefficient in the microring depends on TPA, FCA, XTPA, linear losses, and bending losses.
For the pump field, it is:

$$\alpha_{\text{eff,pump}} = \alpha_0 + \alpha_{\text{rad}} + \frac{\alpha_{\text{TPA}} P_{\text{c,pump}}}{A_{\text{eff}}} + \frac{2\alpha_{\text{TPA}} P_{\text{c,probe}}}{A_{\text{eff}}} + \Gamma \Delta\alpha_{\text{FCA}}$$

with α_{TPA} being the TPA absorption coefficient and $\Delta\alpha_{\text{FCA}}$ the FCA losses, which depend on the electron and hole density according to the following empirical expression [10]:

$$\Delta\alpha_{\text{FCA}}(n_e, p_e) = 8.88 \cdot 10^{-21} n_e^{1.167} + 5.84 \cdot 10^{-20} p_e^{1.109}$$

The same equation applies to the probe signal, replacing $P_{\text{c,pump}}$ with $P_{\text{c,probe}}$ and vice versa. Thus, the formula that accounts for losses in the resonator is:

$$a_{\text{abs,pump}} = e^{-\left(\alpha_{\text{eff,pump}} \frac{L}{2}\right)}$$

The same equation applies to the probe signal, replacing $a_{\text{abs,pump}}$ with $a_{\text{abs,probe}}$.

Total Refractive Index Variation

The total variation in the refractive index is mainly due to the contributions from FCD, the thermal effect, and the Kerr effect.
The total variation in the refractive index caused by the pump is:

$$dn_{\text{eff,pump}} = dn_{\text{eff,FCA}} + dn_{\text{eff,T}} + dn_{\text{eff,Kerr,pump}}$$

with [10]:

$$dn_{\text{eff,FCA}} = -\Gamma \left(5.4 \cdot 10^{-22} n_e^{1.011} + 1.53 \cdot 10^{-18} p_e^{0.838}\right)$$

and:

$$dn_{\text{eff,T}} = \Gamma \frac{dn_{\text{Si}}}{dT} \Delta T_{\text{tot}}$$

$$dn_{\text{eff,Kerr,pump}} = \Gamma n_2 \left(\frac{P_{\text{c,pump}}}{A_{\text{eff}}} + 2\frac{P_{\text{c,probe}}}{A_{\text{eff}}} \right)$$

with $\frac{dn_{\text{Si}}}{dT}$ being the silicon thermo-optic coefficient and $n_2 = 4.4 \cdot 10^{-18}\ [m^2/W]$ being the Kerr coefficient in silicon. For the probe field, the equation is the same but with the usual differences.

Differential Equations for the Circulating Field

The equation for the circulating field $E_{\text{c,pump}}(t)$ in the microring for the pump [2] is the following:

$$\begin{aligned} \frac{\partial E_{\text{c,pump}}}{\partial t} = \Bigg(& - \left(\kappa_{\text{pump}} \cdot \sqrt{1-\eta^2} \cdot \frac{1}{t_{\text{pump}}^2 a_{\text{pump}}(t)} \right) E_{\text{bus,pump}} \\ & + E_{\text{c,pump}}(t) \left(\frac{1}{t_{\text{pump}}^2 a_{\text{pump}}(t)} - 1 \right) \\ & - i \left(\frac{2\pi c_0}{\lambda_{\text{pump}}} \cdot \Delta n_{\text{eff,pump}} \cdot \frac{L}{c_0} + w2_{\text{pump}} \cdot \frac{L}{c_0} \cdot n_g \right) \cdot E_{\text{c,pump}}(t) \Bigg) / \tau_g \end{aligned} \tag{3.4.12}$$

with:

- $w2_{\text{pump}} = 2\pi \cdot \Delta f_{\text{pump}}$: is the frequency detuning term of the pump with respect to the cold resonance of the ring, where $\Delta f_{\text{pump}} = \frac{c_0}{\lambda_{\text{pump}}} - \frac{c_0}{\lambda_{\text{pump}}+\Delta\lambda_{\text{pump}}}$, with $\Delta\lambda_{\text{pump}}$ being the pump detuning in nanometers. The wavelengths of the pump and probe, without detuning, are at two adjacent cold resonances of the microring.

- $w2_{\text{probe}} = 2\pi c_0 \cdot \Delta f_{\text{probe}}$: is the same quantity as the previous term but for the probe.

The equation for the circulating field $E_{\text{c,probe}}(t)$ in the microring has the same form as that for the pump, but with the specific probe variables.

3.5 Schematic Overview of the Experimental Setup

To demonstrate that reservoir computing experiments are valid not only in theory but also in practice, an experimental setup based on a pump-probe experiment has been developed, similar to what was done in [1] and [11].
The schematic of the experimental setup is illustrated in Figure 3.5.1.
For the pump-probe experiment, two different tunable laser sources were used: the Hewlett-Packard 8168F generated the probe laser, while the Agilent N7714A provided the pump laser. The wavelengths of both lasers were carefully selected to match two adjacent resonant cold wavelengths of the microring resonator. The pump laser was modulated using a Mach-Zehnder modulator (JDS Uniphase N-334711A) driven by a waveform generated by an Arbitrary Waveform Generator (AWG, Agilent 33220A), following the same pattern as the encoded input. The Agilent 33220A supports a maximum frequency of 6 MHz, a sample rate of 50 MSa/s, and a typical rise/fall time of 35 ns. The modulated pump signal was then amplified using an Erbium-Doped Fiber Amplifier (EDFA) to achieve the desired power level. The pump was then filtered (JDS Uniphase TB9) and combined with the probe through a 50:50 optical coupler before entering the microring. The pump, coupled to the microring resonance, induces nonlinear phenomena in the device; specifically, two-photon absorption (TPA) occurs, generating a large number of free carriers within the microring plus the other nonlinear effects previously discussed. This population of free carriers causes variations in the refractive index of silicon through an effect known as Free Carrier Dispersion (FCD).
For this reason, the probe was used to detect the induced variations in the microring. At the output of the microring, a filter (JDS Fitel TB9) was used to select the wavelength of the probe, which was then directed to an oscilloscope (Agilent DSO5034A), which has a 300 MHz bandwidth and a 2 GSa/s sample rate, through an avalanche photodetector (APD Lab Buddy, BW $=$ 7 GHz), from which data were acquired and subsequently analyzed. Additionally, a temperature controller (ILX Lightwave LDC-3744B) was employed to maintain the microring at a stable temperature, using a Peltier cell to prevent thermal fluctuations that could affect the experiment. Three polarizers were strategically placed along the optical path to maintain the correct polarization state of the lasers, optimizing their interaction with the microring. The experimental setup also incorporated an Exapod to align the microring with the fibers and so to minimize coupling losses as much as possible.

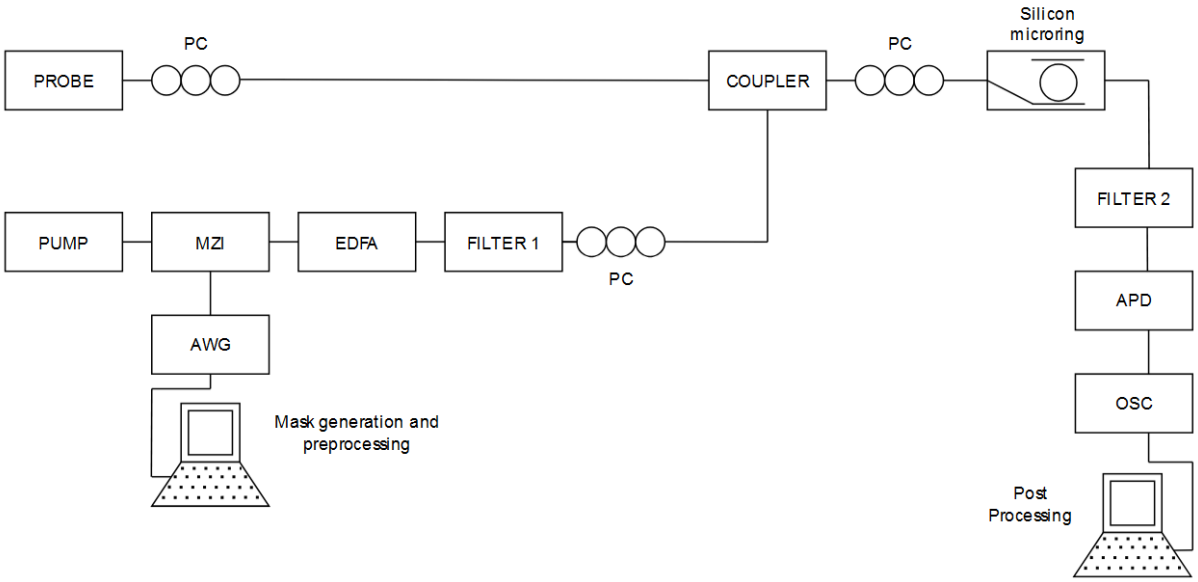

Figure 3.5.1: Experimental setup for the pump-probe measurement with the silicon photonics microring resonator. The configuration includes: Probe laser source, Pump laser source, Mach-Zehnder Interferometer (MZI), Erbium-Doped Fiber Amplifier (EDFA), various filters (FILTER 1 and FILTER 2), coupler, polarization controllers (PC), Avalanche Photodiode (APD) and Oscilloscope (OSC) with a computer for mask generation, preprocessing and post-processing.

3.6 Devices under test

The experimental setup employed a silicon microring resonator and a Mach-Zehnder interferometer, devices that will be described in detail in the following sections.

3.6.1 Description of the chosen silicon microring

As discussed in the previous sections, among the various microring resonators tested, the racetrack microring D8D9 on chip 19 was chosen for conducting the reservoir computing experiments. The selected microring has the following characteristics:

- Coupler length: $L_c = 5\,\mu m$
- Longer side of the microring: $L_d = 19\,\mu m$
- Radius: $r = 5\,\mu m$
- Coupler distance: $250\,\mu m$

The silicon microring resonator is fabricated on a Silicon-On-Insulator (SOI) platform with a silicon device layer and a buried oxide bottom layer. The device uses grating couplers, which couple light from an optical fiber into the waveguide on the chip surface and vice versa. Additionally, the device incorporates directional couplers to enable efficient power transfer between waveguides.

Grating Couplers

Grating couplers are periodic structures created by etching on SOI or through selective deposition of amorphous silicon on SOI. The grating couplers of the tested photonic chip are optimized for a wavelength of 1550 nm, with a fiber-to-waveguide (FtW) insertion loss, as reported in [12], of approximately 3 dB. These have been also designed to work only for TE polarization. They use the diffraction of light to couple light from a waveguide to an optical fiber or vice versa. The grating's periodic structure consists of alternating regions with different refractive indices, created by the pattern of etched and unetched areas. When the grating's period is larger than the wavelength of the light, the diffracted light will result in efficient coupling. The diffraction behavior of a grating coupler (GC) can be explained using the Bragg condition (or phase-matching condition), which is:

$$k_0 \sin \theta_{\text{air}} + mG = \beta_m,$$

where $G = \frac{2\pi}{\Lambda}$ is the grating vector and m is the diffraction order of the grating [13].
As shown in Figure 3.6.1, in the grating couplers on SOI, trenches are etched in the silicon layer with the following definitions:

- Trench length: L_A
- Trench depth: L_c
- Grating periodicity: Λ
- Length of the unetched section: L_B

The grating fill-factor F is defined as the ratio between the length of the unetched section L_B and Λ:

$$F = \frac{L_B}{\Lambda}$$

The effective index of the grating n_{eff}[14] is given by:

$$n_{\text{eff}} = F \cdot n_A + (1 - F) \cdot n_{clad},$$

where n_s and n_{ai} are the effective indices of the silicon areas and the etched areas, respectively. The periodic variation of the refractive index allows the light propagating within the waveguide to be diffracted outwards. This process is fundamental to the functioning of grating couplers, as it enables the transfer of energy from the waveguide to the optical fiber and vice versa. According to the first-order Bragg condition, the grating period Λ is given by:

$$\Lambda = \frac{\lambda}{n_{\text{eff}} - \sin \theta_{\text{air}}},$$

where λ is the wavelength of the coupling light, and θ_{air} is the angle of diffraction.

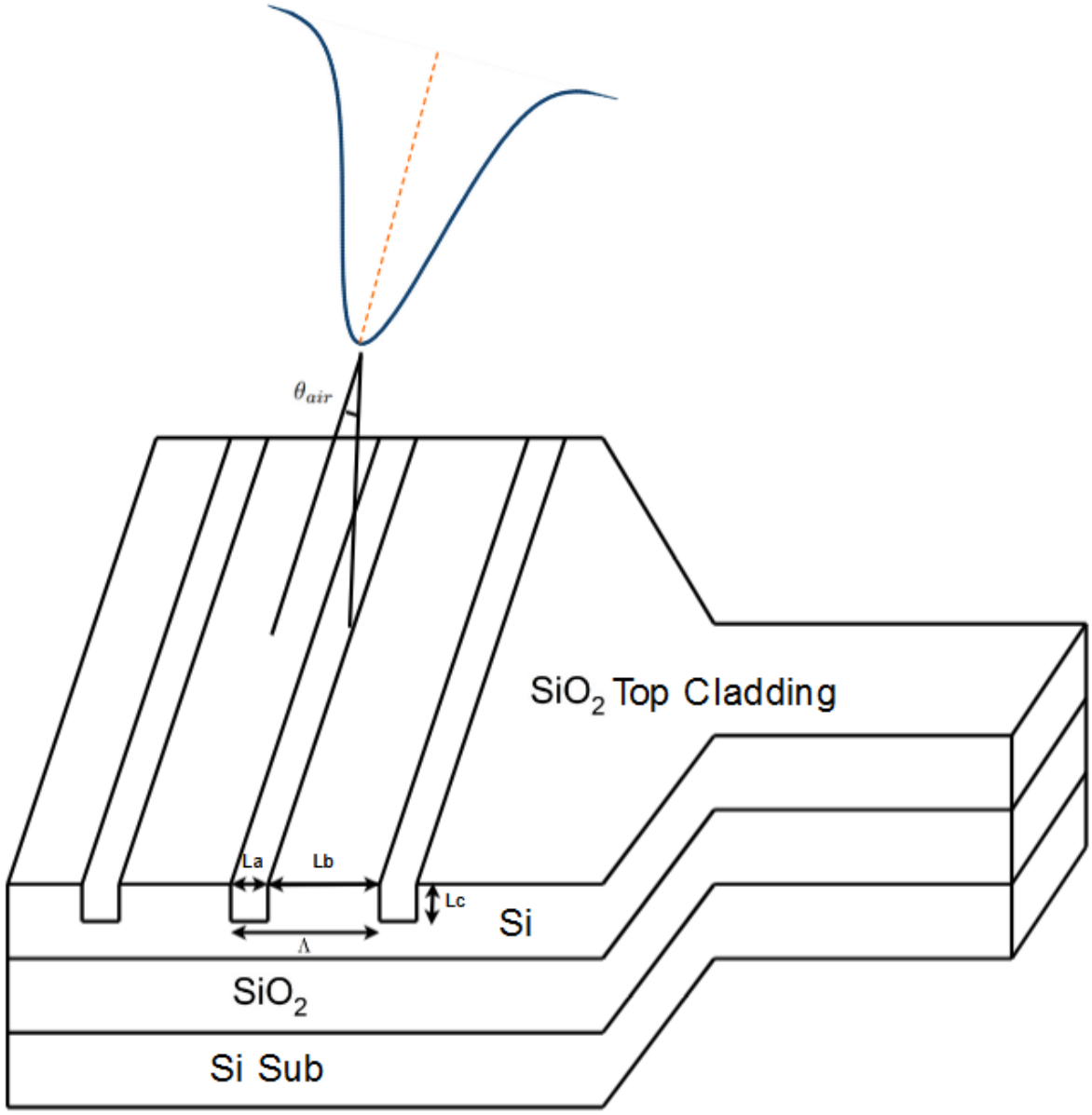

Figure 3.6.1: Grating coupler structure

Directional Couplers

A directional coupler is a structure that enables the transfer of power between two adjacent waveguides. The coupling coefficient, k, varies sinusoidally with the length of the coupler. The amount of power transferred between two waveguides is described by the following equation [12]:

$$K(\lambda) = k(\lambda)^2 = \sin^2(\kappa(\lambda) \cdot L_c + \kappa_0(\lambda))$$

with κ $[1/\mu m]$ the coupling strength between two straight Si strip waveguides and κ_0 the coupling between two bending SWGs. The longer the wavelength, the higher the value of k, as the confinement within the waveguide decreases.

3.6.2 Mach-Zehnder modulator

The Mach-Zehnder modulator (MZM) used is the JDS Uniphase model, with a maximum extinction ratio of 25 dB and a bandwidth up to 12 GHz. The Mach-Zehnder modulator is an optical device that modulates the intensity of light. It does so by exploiting the interference between the two arms of the modulator, caused by the change in the refractive index achieved by applying a bias voltage. The graph shown in Figure 3.6.2 displays the output power of the MZM as a function of the bias voltage for two offset voltage levels: 0V (in blue) and 5V (in red), with an input power of 10 mW. This experiment was conducted in preparation for a light beam modulation experiment where a square wave is encoded, for the delayed 1-bit XOR experiment. The square wave, generated by the AWG, has two voltage levels: 0V and 5V (later set to 0V and 10V). To optimize the delayed XOR experiment, where the beam intensity must follow the square wave, it is essential to understand how the output light intensity of the MZM varies as a function of the bias voltage for the two offset voltage levels.
From the graph, it can be observed that maximum extinction occurs at the minimum of the blue curve at a bias voltage of approximately 3.33V, with an output power of -10.51 dBm. This value is significant because, by fixing the bias voltage at the minimum of the blue curve, when the square wave assumes the low level, the output will be -10.51 dBm. Conversely, when the square wave assumes the high level (with an offset voltage of 5V), the output power will be approximately 7.26 dBm. Consequently, the extinction ratio will be around 18 dB.
To ensure the maximum ER, the bias was initially set at the minimum of the curve, but due to thermal variations that affect the modulator's behavior over time, it was necessary to periodically monitor the output signal with an oscilloscope. Using the oscilloscope, the bias voltage was adjusted to keep the ER constant and, when necessary, slightly shifted from the minimum to ensure operation in the more linear region of the curve. Figure 3.6.3 shows the evolution of

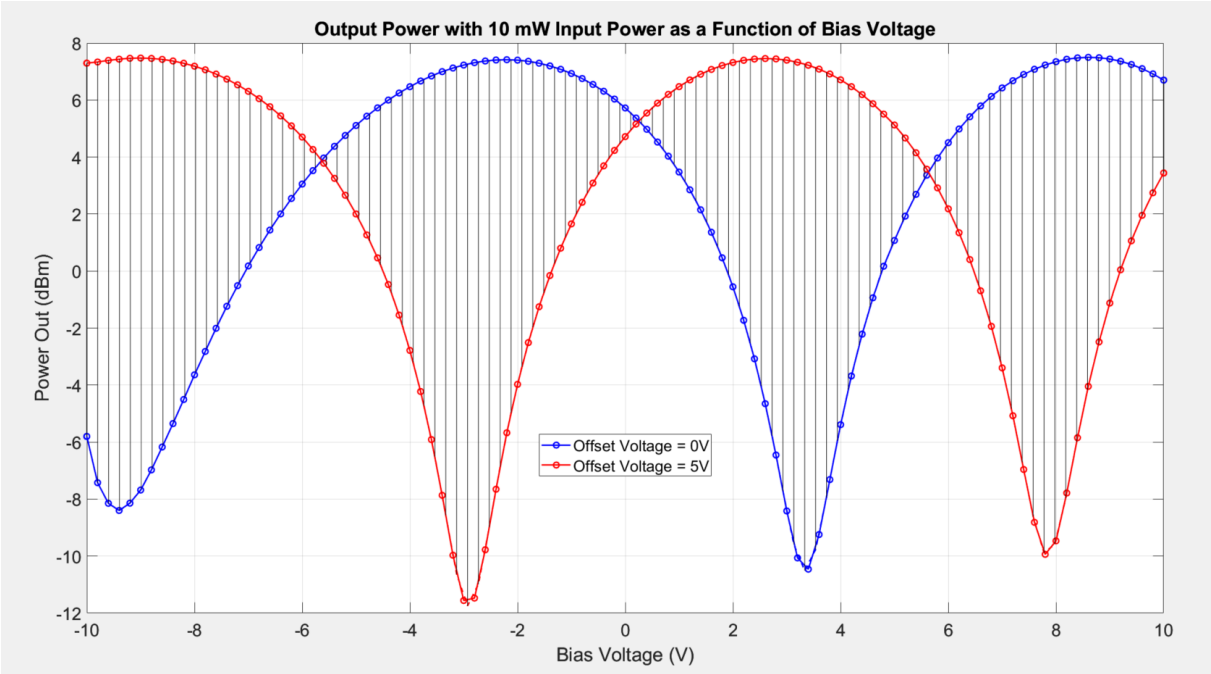

Figure 3.6.2: Graph of the output power of the Mach-Zehnder modulator as a function of the bias voltage, for two offset voltage levels summed over the bias voltage: 0 V and 5 V. The black vertical lines highlight the ER for each value of the bias voltage.

the Mach-Zehnder modulator's output power as a function of the input power, while keeping

the bias voltage constant at the value of V_π, which corresponds to the minimum transmission. It can be observed that, as the input power increases, the output power increases linearly. This test was done to check that the MZI does not introduce additional non-linearities.

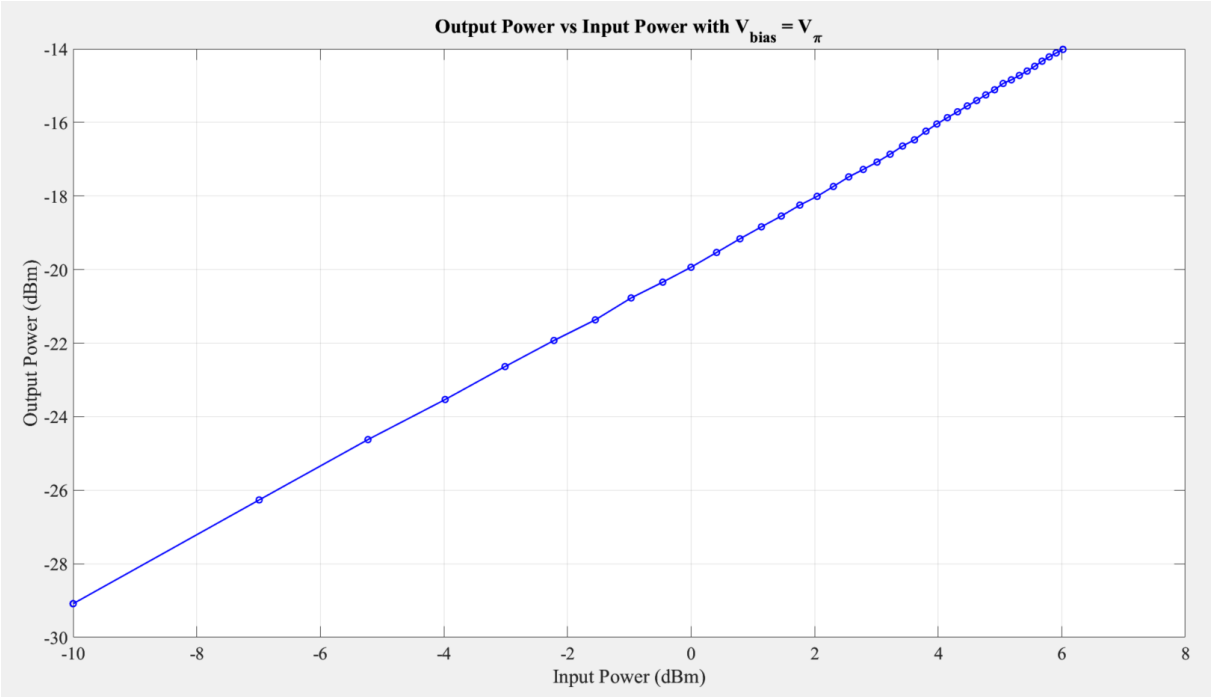

Figure 3.6.3: Output power as a function of input power, with the bias voltage set to V_π.

3.7 Variables of the microring derived in Linear Regime

To correctly set up the pump-probe experiment, it is necessary to first derive the parameters of the microring in the linear regime, which means in the absence of nonlinear effects. The measurements were performed at the through port. The racetrack resonator has the same coupling coefficients for both the upper and lower waveguides, i.e., $\kappa_1 = \kappa_2$.
In Figure 3.7.1, the transmission spectrum at the through port is shown, displaying three adjacent resonances at $\lambda_0 = 1534.53$ nm, $\lambda_0 = 1541.5176$ nm, and $\lambda_0 = 1548.5607$ nm. In the linear regime, the values of λ_0, k^2, η^2, n_{eff}, and n_g were extracted by fitting the experimental data using Equations 3.2.15 and 3.2.16. The linear losses α_0 considered are those specified by the foundry. For the values of α_{rad}, Γ, and A_{eff}, those from [2] were used, as they are also valid for this microring, given that it belongs to the same chip used in [2] and has the same radius.
Coupling losses depend on the distance between the bus waveguide and the resonator and therefore do not depend on the wavelength, making them the same for all resonances. The fitting results yield the following coupling coefficients and quality factors:

- For $\lambda_0 = 1534.53$ nm: $\kappa^2 = 0.035622$, $Q = 15267.5197$
- For $\lambda_0 = 1541.5176$ nm: $\kappa^2 = 0.041776$, $Q = 13788.2608$
- For $\lambda_0 = 1548.5607$ nm: $\kappa^2 = 0.042329$, $Q = 13551.7731$

These results indicate that the quality factor of the resonator decreases with increasing wavelength due to dispersion effects.
Consequently, in the pump-probe experiments, the pump was placed at the lowest resonance wavelength (1534.53 nm) while the probe was placed at the next resonance wavelength (1541.52 nm). The pump was positioned at the lowest resonance wavelength to achieve higher circulating power and thus generate a greater number of free carriers.

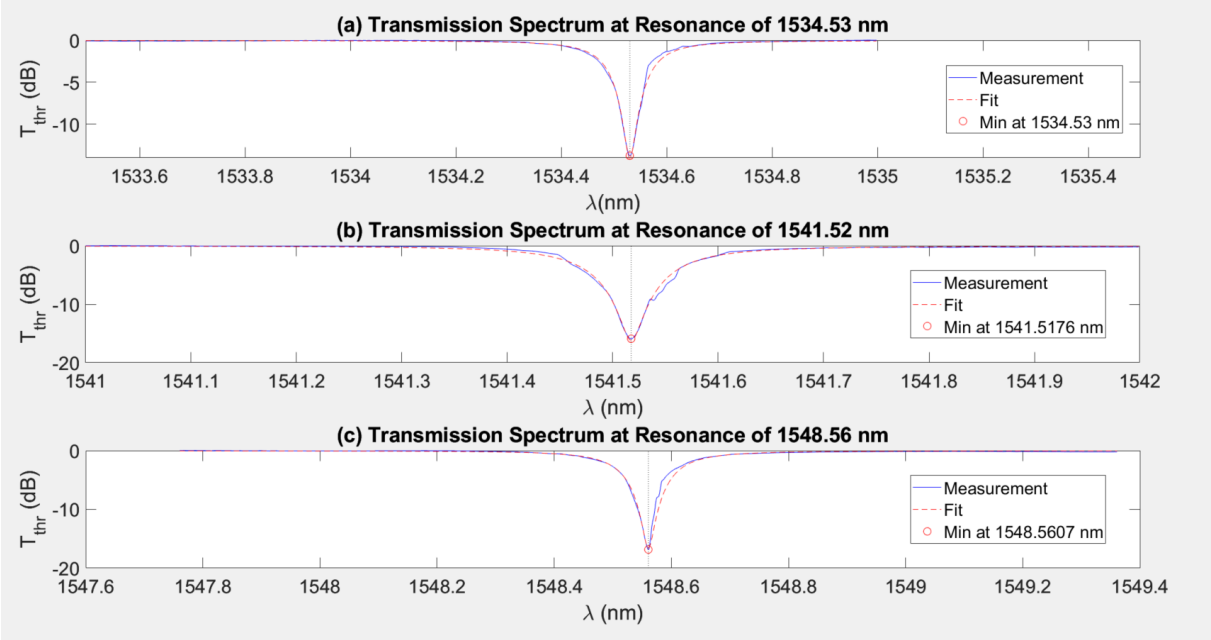

Figure 3.7.1: Transmission spectrum at different cold resonances

In Table 3.7.1, all the key parameters characterizing the microring under test in the linear regime are reported.

Table 3.7.1: Summary of Parameters in the Linear Regime for Racetrack Resonator under test

Parameter	**Ring Pump**	**Ring Probe**
λ_0	1534.53 nm	1541.5176 nm
κ^2	0.035622	0.041776
η^2	0.0073	0.0060
α_0	2 dB/cm	2 dB/cm
n_g	4.27032	
α_{rad}	$<$ 0.01 dB/cm	
$n_{\text{eff},0}$	2.32	2.33
Γ	1	
A_{eff}	0.075 μm^2	

In Figure 3.7.2, the transmission coefficient measured at the through port of the microring resonator is shown and compared with the theoretical model. The measurements were taken at low input power, so that no nonlinear effects are present. As can be observed, the two curves closely match, indicating good agreement between the measured data and the model. This demonstrates that the parameters calculated in the linear regime are accurate.

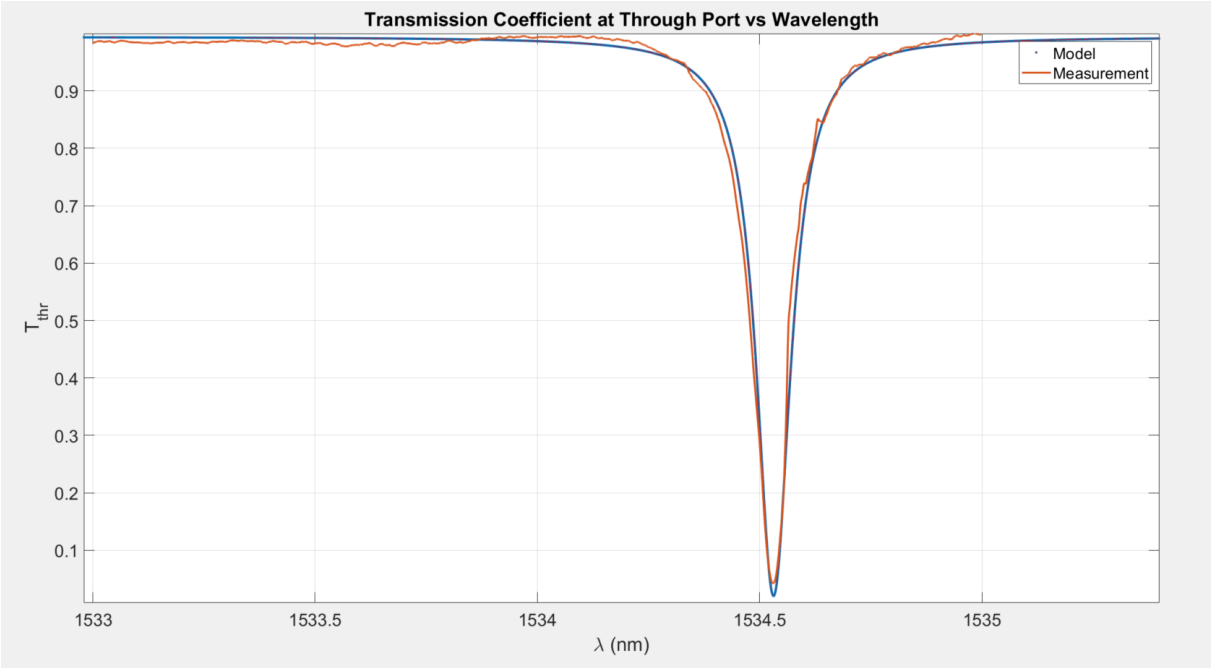

Figure 3.7.2: Comparison of the measured transmission coefficient and the theoretical model at the *through* port.

3.8 Nonlinear measurements

To ensure that reservoir computing experiments function correctly, it is essential to study the nonlinear behavior of the microring. For this purpose, a series of transmission spectra measurements were performed by varying the input power. Figure 3.8.1 illustrates the transmission spectra captured at the through port of the racetrack resonator for increasing input power values, ranging from very low to very high levels where nonlinear effects become increasingly significant.
For each grating coupler, there are losses of approximately 2.8 dBm. Consequently, a value of 2.8 dBm was added to the input powers.

3.8.1 Analysis of Nonlinear Effects

As shown in Figure 3.8.1, with varying input power, the transmission spectra exhibit both a blue shift and a red shift in the resonant wavelength. Initially, for low power values, a slight blue shift can be observed due to the dominant effect of the change in the refractive index, known as free carrier dispersion (FCD). However, as the power increases further, thermal effects take precedence, leading to a red shift in the resonance wavelength.
The relationship between input power (P_{bus}) and the shift in resonance wavelength ($\Delta\lambda_{\text{cw, res}}$) is depicted in Figure 3.8.2. A slight initial negative decrease can be observed due to FCD, and for input powers above approximately 1 mW, there is a growing positive increase due to self-heating instead.
Figure 3.8.3 shows the variation in the transmission coefficient at resonance ($T_{\text{thr, res}}$) with increasing input power. The graph indicates a nonlinear increase in transmission at resonance, and this means that the ring can be used as a non linear element in a neural network. This increase is due to the combined effects of free carrier absorption (FCA) and thermal-induced

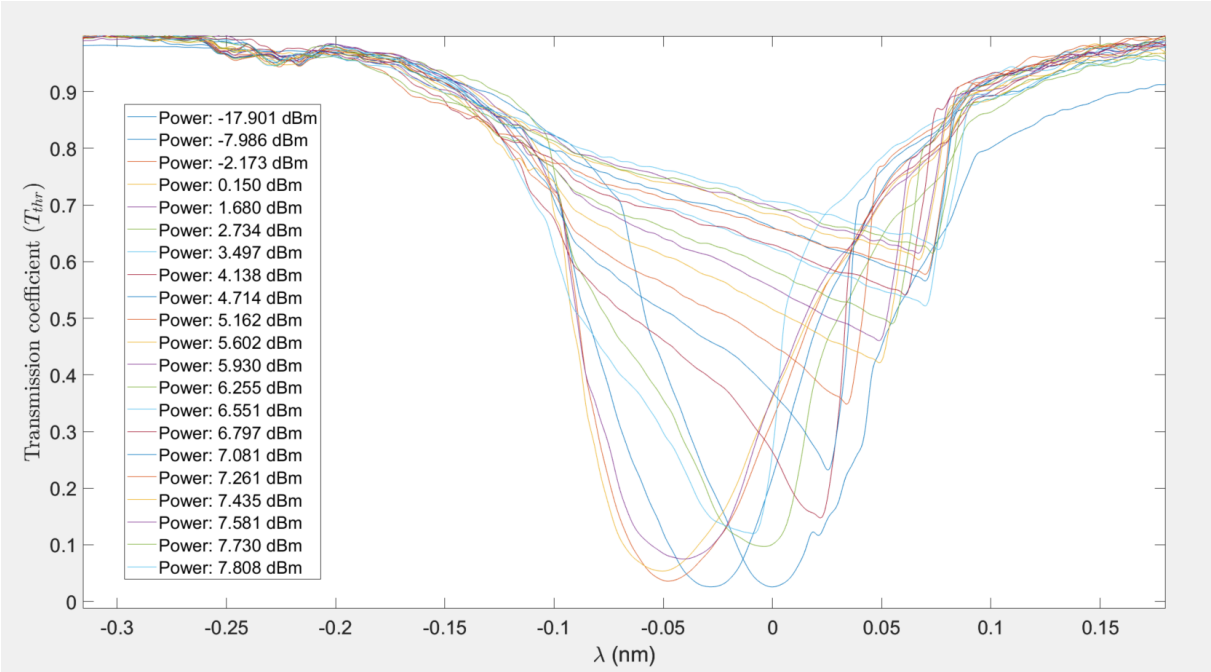

Figure 3.8.1: Transmission spectra at the through port of the racetrack resonator for increasing input power levels.

index changes, as described previously.

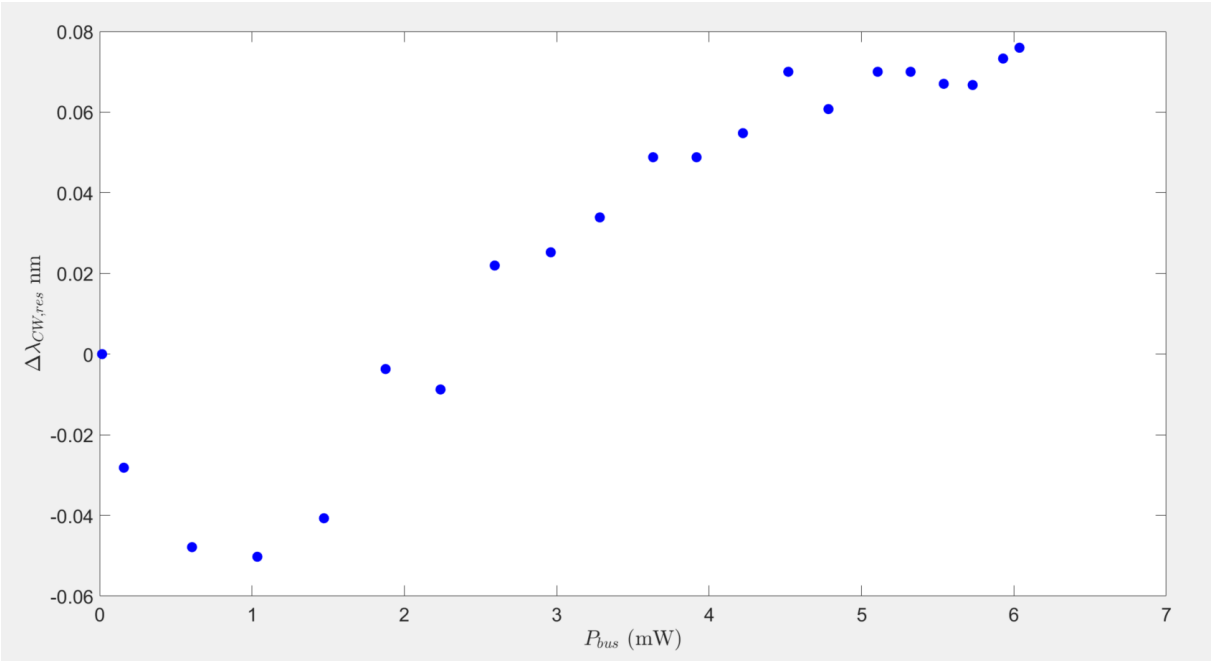

Figure 3.8.2: Resonant wavelength shift ($\Delta\lambda_{\text{res}}$) as a function of input power (P_{bus}).

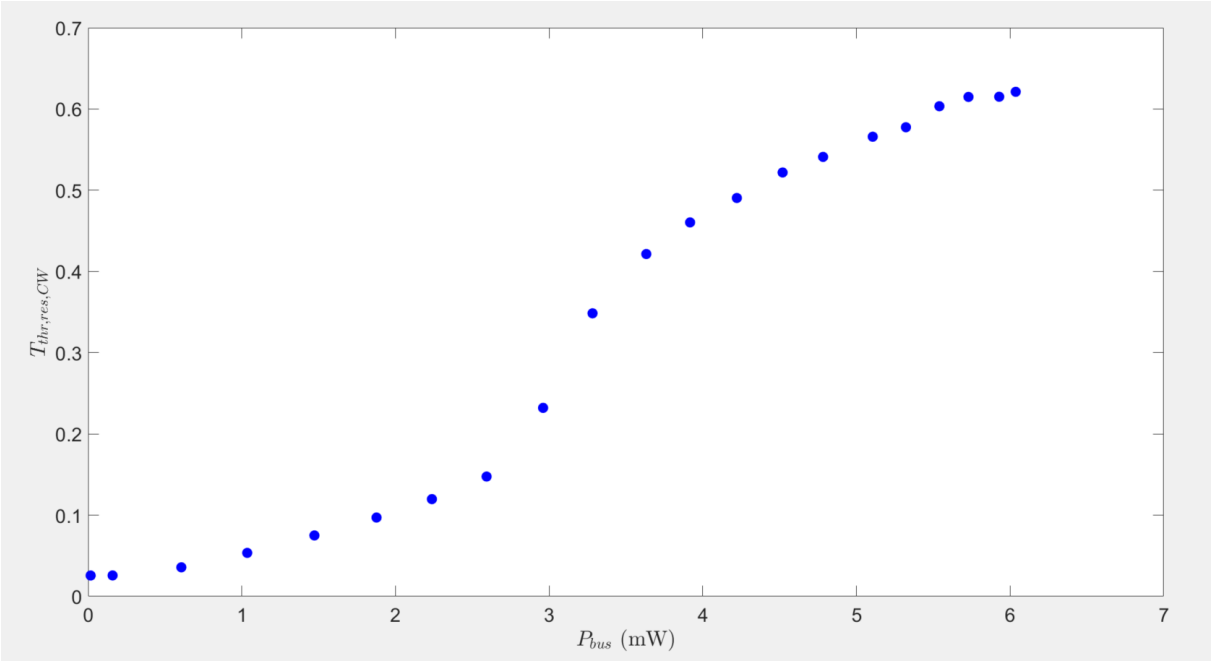

Figure 3.8.3: Transmission coefficient at resonance ($T_{\text{thr,res}}$) as a function of input power (P_{bus}).

3.8.2 Double Resonance

A brief note should be made regarding another ring, different from the one selected for the experiments shown so far, which exhibits a very particular characteristic: the double resonance. As shown in Figure 3.8.4, the graph represents the output power from the ring at the through

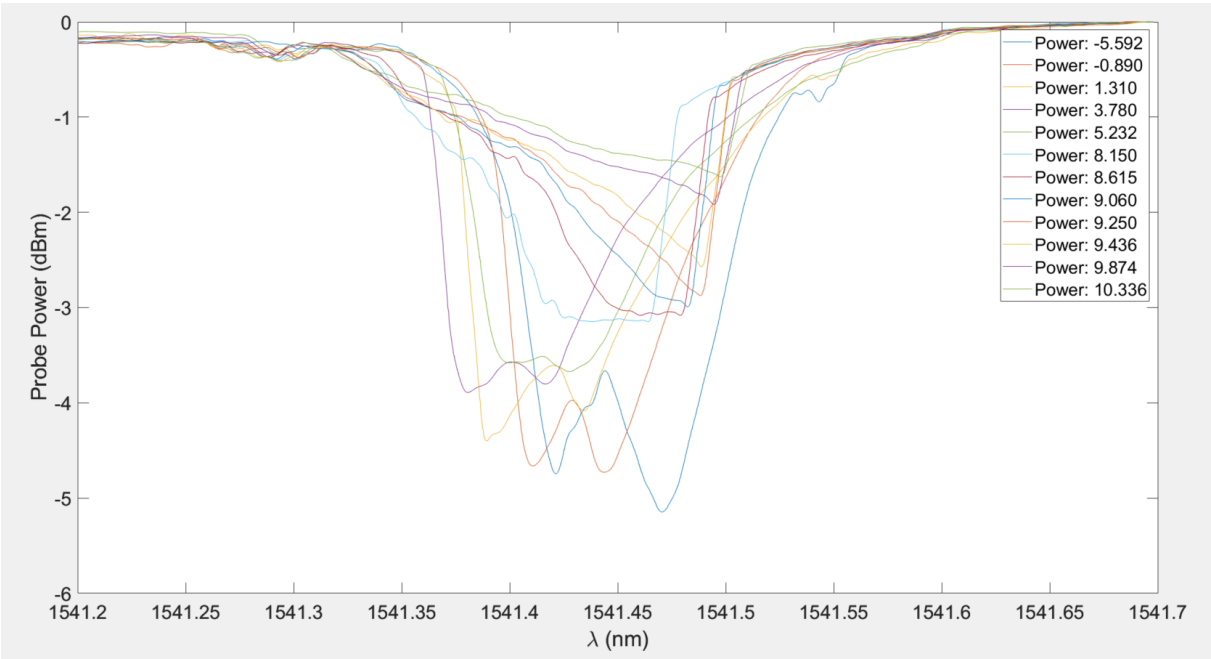

Figure 3.8.4: Output power from the ring at the through port as a function of wavelength, varying the input power.

port as a function of wavelength λ, varying the input power. The different curves correspond to

different input power levels, and it can be observed that the depth of the curves decreases as the input power increases.
Each grating coupler has losses of approximately 2.75 dBm. As a result, a value of 5.5 dBm was added to the input powers. The ring shows a double resonance, visible in the two distinct minima in each curve of the graph.
In a microring resonator, each mode can travel in two directions: forward, excited specifically when power is injected into the microring, and backward. When a portion of the power carried by the forward-propagating mode is coupled into the backward-propagating mode, the so-called contradirectional coupling occurs, which causes a distortion of the ideal spectrum, visible as a double dip in the spectrum.
As discussed in [15], this effect can be caused by external perturbations or imperfections in the fabrication of the device, such as the roughness of the waveguide sidewalls. Due to this double resonance, the ring shown in Figure 3.8.4 was discarded.

3.8.3 Analysis of the Nonlinear Response to Power Steps

To better understand the non-linear response, further numerical simulations were done.
In this subsection, the behavior of the transmission coefficient at resonance at the through port of the microring resonator is analyzed as the pump signal power varies. Two different situations are reported: the response to a constant power step of 2 dBm and one of 10 dBm. Figures 3.8.5 and 3.8.6 show, respectively, the system behavior for the two power levels.
In Figure 3.8.5, with a pump signal of 2 dBm, the transmission coefficient stabilizes after a transient phase, reaching a constant value.
In contrast, Figure 3.8.6 shows the behavior of the transmission coefficient for a pump signal of 10 dBm. In this case, periodic oscillations (*self-oscillation*) are clearly observed.

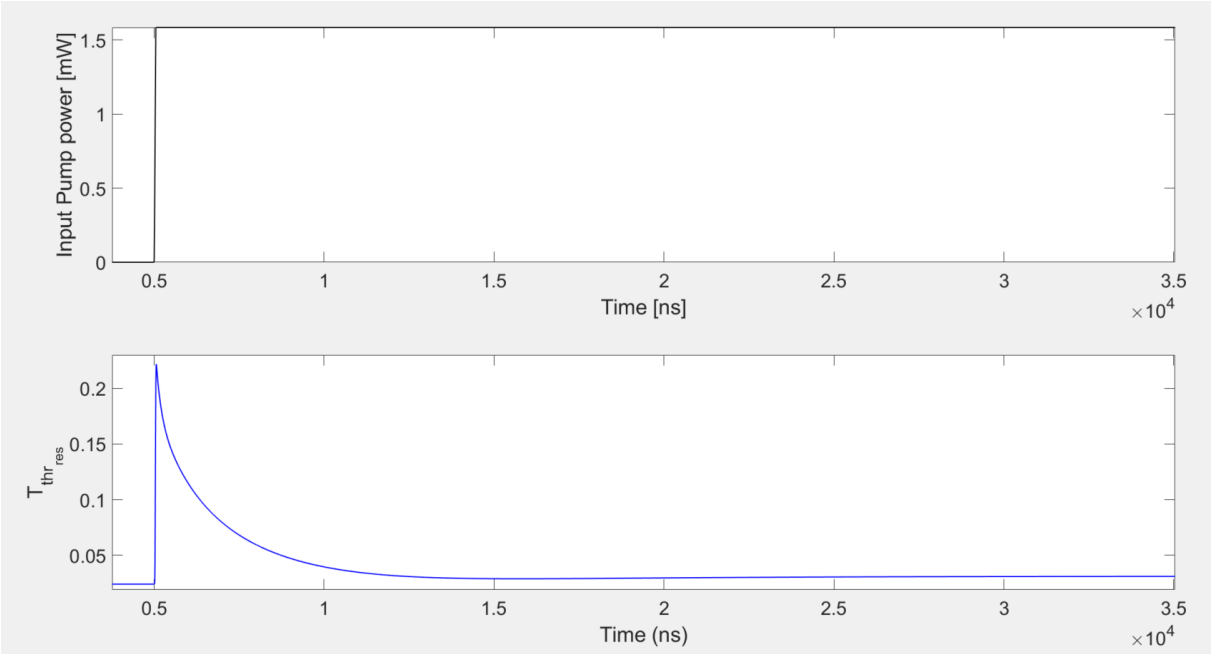

Figure 3.8.5: Simulated behavior of the transmission coefficient at the resonance at the *through* port ($\lambda = 1541.5176$) for a constant power step of 2 dBm.

To explain the behavior of the transmission coefficient at the *through* port for a power step of

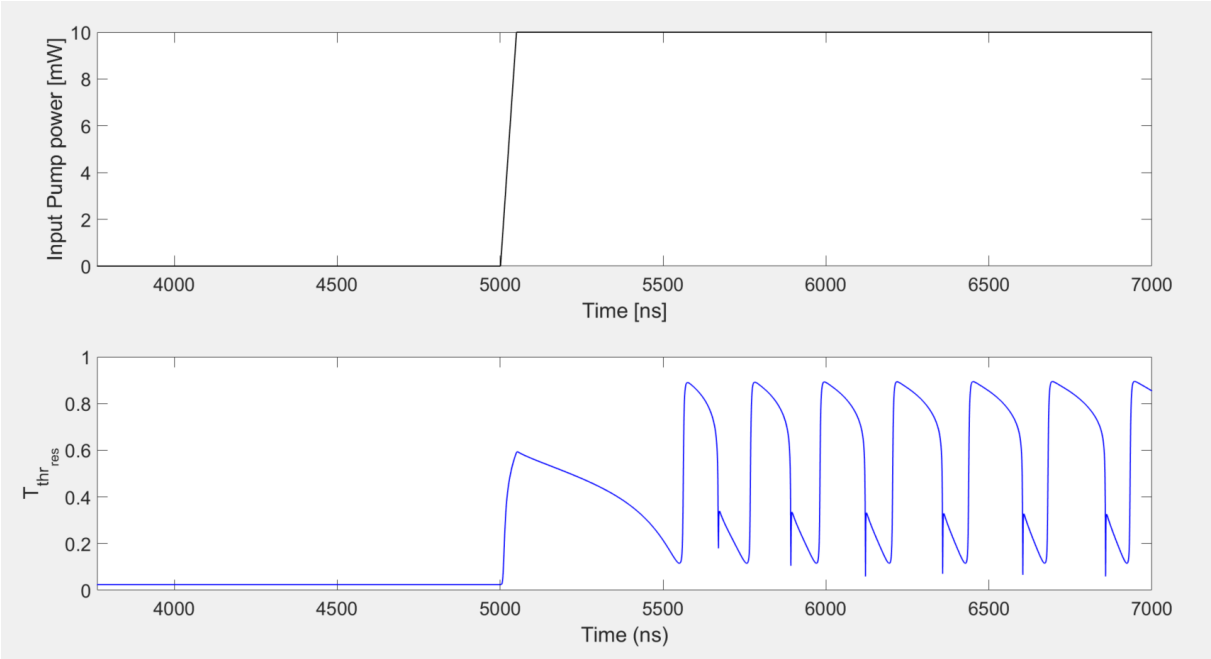

Figure 3.8.6: Simulated behavior of the transmission coefficient at the resonance at the*through* port ($\lambda = 1541.5176$) for a constant power step of 10 dBm, where the self-oscillation phenomenon is observed.

2 dBm, figure 3.8.7 shows the variation of the resonance wavelength for three different effects: the thermal effect, the free carrier dispersion (FCD) effect, and the combination of both.

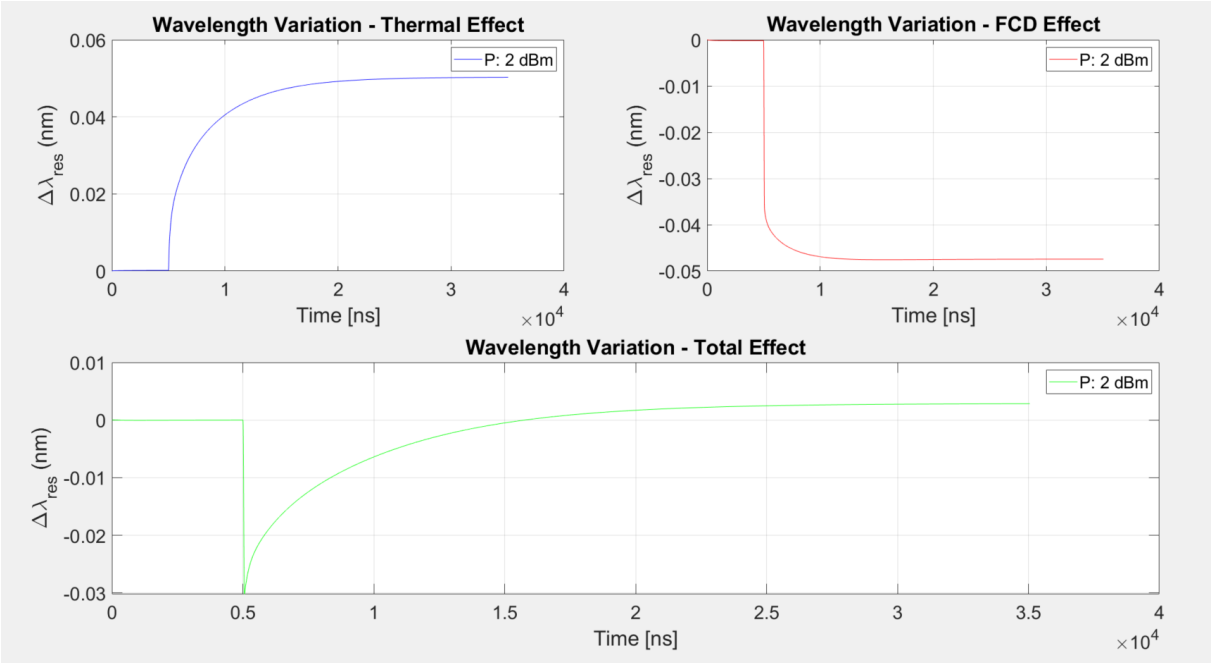

Figure 3.8.7: Resonance wavelength variation due to thermal effects, FCD, and the combination of both for a 2 dBm step.

At the beginning of the power step application, the optical power inside the microring generates a moderate number of free carriers through the two-photon absorption (TPA) process. These free carriers cause a free carrier dispersion (FCD) effect, which shifts the resonance wavelength

towards the blue. This shift is rapid and initial, reflecting the microring's immediate response to the generation of free carriers.
After the initial blue shift, the thermal effect starts to become more dominant. The heating of the microring shifts the resonance wavelength towards the red.
Thus, the initial rapid response is dominated by free carrier dispersion, followed by a slower and more gradual response dominated by thermal effects, finally reaching an equilibrium between the two effects.
Similarly to what is observed for a 2 dBm step, the behavior of the transmission coefficient at resonance at the *through* port of the microring in response to a constant step of 10 dBm can be explained by analyzing the graphs in Figure 3.8.8. In Figure 3.8.8, the variation of the resonance wavelength due to the thermal effect, the free carrier dispersion (FCD) effect, and the combination of both is illustrated.
In this case, the "self-oscillation" phenomenon occurs for high input power. Self-oscillations occur when the circulating power inside the microring becomes sufficiently high to generate a significant number of free carriers.
In this situation, the free carrier lifetime becomes comparable to the thermal time constant of the system. Initially, the high signal power generates a large number of free carriers through the two-photon absorption (TPA) process. This increase in free carriers causes an initial blue shift of the resonant wavelength, due to free carrier dispersion (FCD). As the resonance wavelength shifts, the circulating power decreases, and consequently, the generation of free carriers decreases, because less circulating power means less TPA and therefore fewer free carriers. Simultaneously, the recombination of free carriers begins, generating heat. This causes the resonant wavelength of the microring to shift back toward the "cold" resonance wavelength. When this happens, the circulating power increases again, along with the number of free carriers.
When the resonant frequency of the ring aligns again with the cold resonance (λ_0), the maximum circulating power no longer reaches the initial levels. This is because a large number of free carriers have been generated, and these act as impurities within the microring, causing additional absorption (FCA). FCA introduces additional losses, reducing the energy in the microring. In this context, the quality factor, which represents the ring's ability to retain energy, is degraded. The carriers that are absorbed generate more heat, shifting the resonance wavelength toward the red, and thus the circulating power and free carriers decrease. After the ring begins to cool down, shifting the wavelength back toward the blue, the circulating power increases again, generating a large number of free carriers once more, and thus the oscillation begins.
The analysis of the transmission coefficient behavior at resonance at the through port of a microring in response to power steps is crucial for evaluating the nonlinear dynamics of the system. This study will help better understand reservoir computing experiments.
In Figure 3.8.9, it is shown how T_{thr} at resonance varies as the optical power applied to the microring increases.
On the horizontal axis, the input powers are shown, ranging from 2 dBm to 12 dBm, and for each input power, the final value of $T_{\text{thr,res}}$, measured at the last moment of the simulation, is shown on the vertical axis.
For low powers, up to 6 dBm, a stable behavior is observed, meaning the transmission coefficient assumes a single final value. However, when the input power reaches 8 dBm, self-oscillations occur in the system, as evidenced by the fact that the single line splits into three separate lines, representing the maximum, average, and minimum values of the transmission coefficient during the oscillations.
The general trend of the average transmission coefficient at the cold resonance wavelength shows that as the power increases, the coefficient approaches 1, degrading the quality factor.

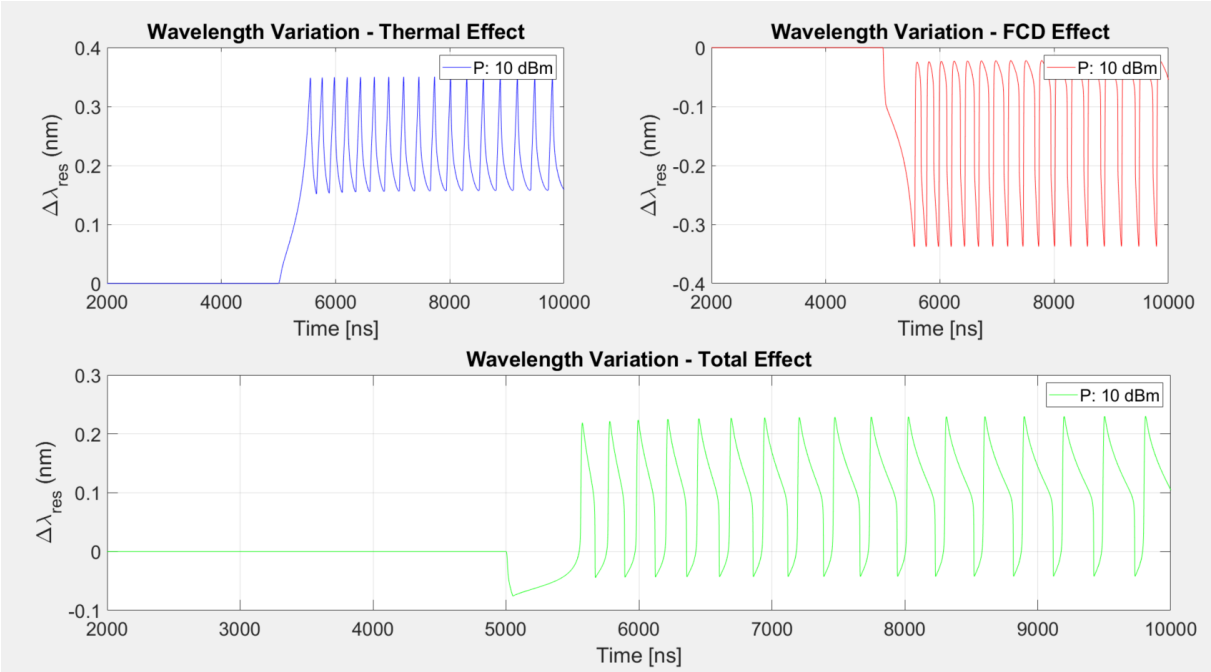

Figure 3.8.8: Resonance wavelength variation due to thermal effects, FCD, and the combination of both for a 10 dBm step.

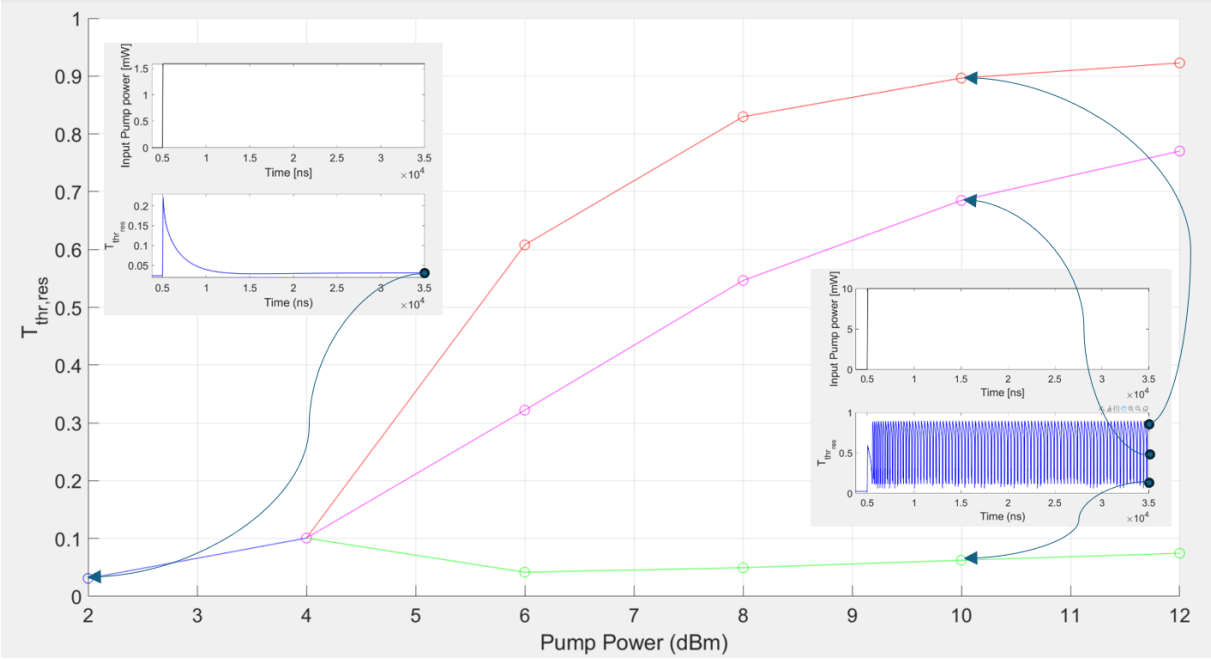

Figure 3.8.9: Final value, or in the presence of self-oscillations, the maximum, average, and minimum values of the transmission coefficient at resonance at the through port of the microring ($\lambda = 1541.5176$), as a function of input power. The graph includes two subfigures showing the temporal evolution of the input power and the transmission coefficient $T_{\text{thr,res}}$ for two different power levels, 2 dBm and 10 dBm.

Chapter 4

Photonic reservoir with a silicon microring and time multiplexing

Research is focusing on the study of photonic reservoirs to combine their high-speed optical signal processing with the predictive capabilities offered by reservoir computing.
As previously discussed, one possible method to create a photonic reservoir is to use microrings as nonlinear nodes, as demonstrated in [16].
For example, in [17], an artificial feedforward photonic neural network based on the spatial distribution of nodes was developed for image classification. In the network, optical signals propagating through the waveguides are weighted by p-i-n junctions, captured by photodiodes, summed, amplified, and finally sent to an MRR that implements an activation function.
The problem with these systems is that they have many physical nodes, making their implementation complex.
In contrast, in delayed feedback photonic systems, the reservoir implementation is simpler because it consists of only one microring and, in some cases, a delay line. This approach overcomes the issues associated with spatial networks, such as high losses due to physical connections and technological complexity in fabrication.
In [1], a neural network with a single silicon microring without a delay line was presented and tested with two experiments: the 1-bit delayed XOR operation and the classification of the iris flower. In this thesis, both experiments were reproduced, both experimentally and through simulation, to test the network.
In [18], a delay line was added to the microring, which resulted in greater memory and improved the system's nonlinear behavior.

4.1 Implemented photonic reservoir

As mentioned earlier, two experiments were conducted: the 1-bit delayed XOR and the recognition of the Iris flower. To perform these tests, a pump and probe experiment was designed, in which the pump, injected into the microring, causes the two-photon absorption phenomenon that generates free carriers (electrons and holes) in the microring. These free carriers induce Free Carrier Dispersion (FCD), which shifts the resonance frequency of the microring. These nonlinear phenomena non-linearly influence the probe's response.
In the experiments, temperature plays a key role. In fact, when the power in the ring is low,

self-heating in the MR is negligible. However, when the power reaches a couple of mW, the temperature greatly influences the carrier dynamics. The procedure for conducting the two experiments is outlined below.

4.2 Operational Description of the Experiments

For both experiments, the procedure consists of three distinct phases: preprocessing, signal processing in the microring, and postprocessing.

4.2.1 Preprocessing

To explain the preprocessing, the well-known Iris dataset will be taken as a reference, which contains four features: sepal length, sepal width, petal length, and petal width of 150 flowers from three different species: Setosa, Virginica, and Versicolor. Only the Setosa flower is linearly separable from the others. The role of the reservoir is to attempt to make the entire dataset as linearly separable as possible by increasing the input space dimension from 4 to N_v. In the preprocessing, the goal is to transform the input data into a format suitable for modulating the pump through the Mach-Zehnder modulator. The output of the Mach-Zehnder, after being amplified and filtered, is injected into the ring.
The first step is to load and shuffle the dataset to ensure that the model is more robust and capable of generalizing during training. Next, it is necessary to construct an **input matrix** U_{in} to organize the data effectively.
In the case of the Iris dataset, each flower has four properties representing the input features. Therefore, the matrix U_{in} will be filled with the values of these features and will have dimensions equal to the number *N* of properties (4) multiplied by the number of flowers *M* (150).
Subsequently, a **masking matrix** W_{u} is applied to the data matrix U_{in}, with dimensions $N_v \times N$. This matrix W_{u} represents a random mask that defines how the input properties are distributed across the virtual nodes in the system. It is a weight matrix with values ranging from -1 to 1. This masking process is essential because it sequentializes the input and projects it into a higher-dimensional space, transforming the data into a format that allows for better linear separability at the output. Thus, each column of the product between the two matrices represents a specific flower, composed of N_v elements, and each of these values is held constant for a period of $\Delta t = \frac{1}{B}$, where B is the bitrate. Each flower therefore has a total duration of $T = N_v \cdot \Delta t$.
The larger the value of N_v, the greater the number of values characterizing each individual flower, which increases the processing time. However, the ring dynamics become more complex, enhancing the system's ability to detect hidden connections in the data.
This data matrix is then sent to and processed by the AWG, which applies an amplitude of 10V between the minimum and maximum values of the matrix. The AWG then sends the signal to the Mach-Zehnder modulator to modulate the pump laser light passing through it.
In simulation only, the power is scaled within a range between P_{min} and P_{max}, representing the minimum and maximum power injected into the ring, respectively, to achieve the desired average power and extinction ratio.
In both the experiment and the simulation, a gap of 100 ns is introduced between one flower and the next, during which the pump laser power is set to zero to ensure that the response of one flower does not influence the response of the next one.
Additionally, only for the experimental setup, a large interval was inserted at the beginning of

the entire sequence that encodes the whole dataset, during which the pump power is set to zero. This is because the AWG generates a periodic signal with a frequency equal to $\frac{B}{n}$, where *B* is the bitrate and *n* is the number of values that encode the entire dataset.
Therefore, the initial interval where the pump power is off serves only to distinguish between one period and the next on the oscilloscope. After the pump signal, which encodes the dataset information, exits the Mach-Zehnder, it is amplified, filtered, and combined with the probe signal before finally entering the ring.

4.2.2 Postprocessing

This subsection describes the training procedure of the neural network, performed in MATLAB. The training process is supervised and begins with proper data preparation.
The output power at the through port is sampled at time intervals of Δt, and these sampling instances define the virtual nodes. Then, the samples corresponding to the intervals where the pump power is zero, i.e., between each flower, are discarded. The remaining data are organized into a state matrix X with dimensions $N_v \times M$.
Next, the data in this state matrix are split into a training set and a test set in proportions of 70% and 30%, respectively.
The training set, therefore, includes the first 105 flowers, while the remaining flowers are reserved for testing. The labels, i.e., the targets corresponding to each flower, are also divided into training and test sets in the same proportions.
To evaluate the accuracy of the network's predictions, the labels are compared with the predictions.
To improve the network's generalization ability, regularization is applied during the training phase, as explained in the first chapter. The regularization parameter λ is obtained using the five-fold cross-validation technique. Various values of the regularization parameter were tested to identify the one that minimizes the error and, consequently, maximizes accuracy. The error is defined as the number of incorrect predictions out of the total predictions made, while accuracy is the opposite, i.e., the number of correct predictions out of the total predictions.
Therefore, to find the best value of the regularization parameter, the error rate is calculated for each value of λ, and the one corresponding to the lowest error rate is chosen. Finally, the network is trained using the optimal λ on the entire training set.
For the training, the weight matrix W_y is first calculated using equation 2.2.4, and then this matrix is multiplied by the reservoir state matrix *X*. After the network has been trained, it is tested on the test set by multiplying W_y by the test states. The result will be a 3 x N_{test} matrix, with N_{test} being the number of flowers in the test set.
Finally, the predictions are obtained by analyzing this matrix through the "winner takes all" procedure, explained in the first chapter. This procedure identifies the flower's class based on the position of the maximum value in each column.
The predicted outputs are then compared with the targets to calculate the overall accuracy of the network.

4.2.3 Experimental tests performed on the IRIS dataset

The first test performed on the IRIS dataset is shown in Figure 4.2.1, which displays the waveforms of the pump and probe power output from the microring with 50 virtual nodes, obtained experimentally. The average input pump power is 4.8 dBm. The figures on the left show the

pump power (top) and the probe power (bottom), both normalized between 0 and 1, with a bitrate of 10 Mbps, while the figures on the right show the same signals, but with a bitrate of 30 Mbps. The circles in the graphs indicate the sampled points at intervals of Δt, corresponding to the virtual nodes. These values will be inserted into the state matrix X to be used in the training of the neural network.
By sampling the output probe power, the experimentally obtained accuracy is 97.73% with a bitrate of 10 Mbps, while it slightly drops to 95.45% for bitrates of 20 and 30 Mbps. This slight decrease is due to the increased signal speed, as with a higher bitrate, the duration of each time step decreases, becoming shorter than the carrier lifetime. As a result, the transfer of information from the pump to the probe within the microring becomes inefficient.
In contrast, when sampling the input pump power, the accuracies are significantly lower: 84.1%, 77.27%, and 86.36% for 10 Mbps, 20 Mbps, and 30 Mbps, respectively. The greater dynamic complexity of the probe signal compared to that of the pump, due to the nonlinearities, improves the classification performance. This comparison highlights the usefulness of the microring. Finally, it is important to note that the use of a Mach-Zehnder introduces a non-linear transformation of the type $\sin^2$ or $\cos^2$, which further facilitates the training process. However, the microring remains essential as it introduces additional non-linearities and dynamic memory, improving the system performance and thus the ability to solve complex tasks.

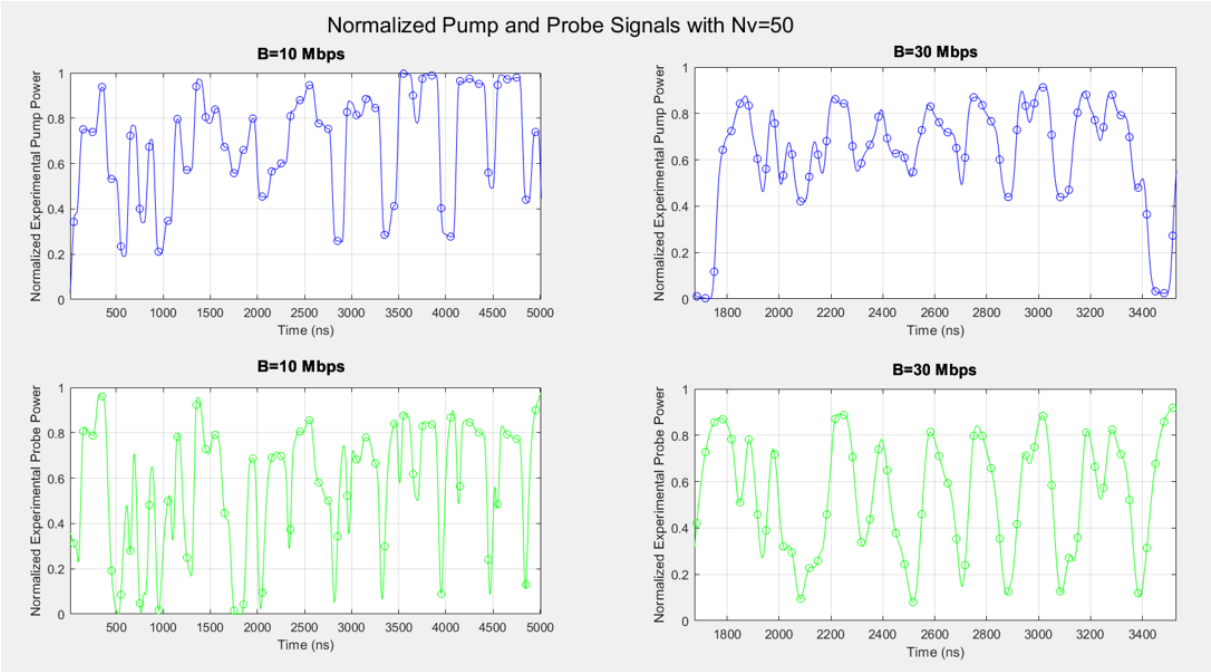

Figure 4.2.1: Normalized pump in and probe out signals with $N_v = 50$. The figures on the left show the "pump in" power (top) and the "probe out" power (bottom) for a bitrate of 10 Mbps, while those on the right show the same signals for a bitrate of 30 Mbps. The circles indicate the virtual nodes.

Table 4.2.1 and Figures 4.2.2, 4.2.3, 4.2.3, 4.2.4 illustrate the experimental results obtained by sampling the probe signal output from the microring and performing training on this signal with a bitrate of 10 Mbps, 50 virtual nodes, and an average power of 4.8 dBm.
Table 4.2.1 reports the results of the five-fold cross-validation used to determine the optimal value of λ. As shown in the table, the optimal value of λ is 0.1 because it achieves an average accuracy of 94.29% across all folds. The table 4.2.1 shows, for each value of λ, the number of incorrect bits for each fold, the error rate for each fold, the average of the latter two, the

accuracy for each fold, and the average accuracy.

Table 4.2.1: Results of the five-fold cross-validation for different values of λ.

Lambda	**Number of incorrect samples per fold**	**Error rate per fold (%)**	**Accuracy per fold (%)**	**Average number of incorrect samples**	**Average error rate (%)**	**Average accuracy (%)**
0.000	3.00 1.00 5.00 2.00 2.00	14.29 4.76 23.81 9.52 9.52	85.71 95.24 76.19 90.48 90.48	2.60	12.38	87.62
0.001	2.00 1.00 4.00 2.00 2.00	9.52 4.76 19.05 9.52 9.52	90.48 95.24 80.95 90.48 90.48	2.20	10.48	89.52
0.010	2.00 1.00 3.00 2.00 1.00	9.52 4.76 14.29 9.52 4.76	90.48 95.24 85.71 90.48 95.24	1.80	8.57	91.43
0.100	2.00 1.00 2.00 1.00 0.00	9.52 4.76 9.52 4.76 0.00	90.48 95.24 90.48 95.24 100.00	1.20	5.71	94.29
1.000	2.00 2.00 3.00 2.00 1.00	9.52 9.52 14.29 9.52 4.76	90.48 90.48 85.71 90.48 95.24	2.00	9.52	90.48
10.000	2.00 3.00 3.00 2.00 1.00	9.52 14.29 14.29 9.52 4.76	90.48 85.71 85.71 90.48 95.24	2.20	10.48	89.52

In figure 4.2.2, the network's prediction results for the test set, consisting of 44 flowers, are shown with blue diamonds, and the respective targets are shown with black circles. The network made only one error, confusing a flower of the Versicolor class with one from the Virginica class, highlighted by a red oval. This result, obtained with the optimal λ (0.1), derived from Table 4.2.1, highlights the high accuracy of the reservoir model in these types of applications. Figure 4.2.3 represents the confusion matrix of the experiment, showing the classification accuracy for each flower species, as well as the overall accuracy of 97.73%. The Setosa class was recognized with an accuracy of 100%, the Versicolor class with 90.9%, and the Virginica class with 100%.

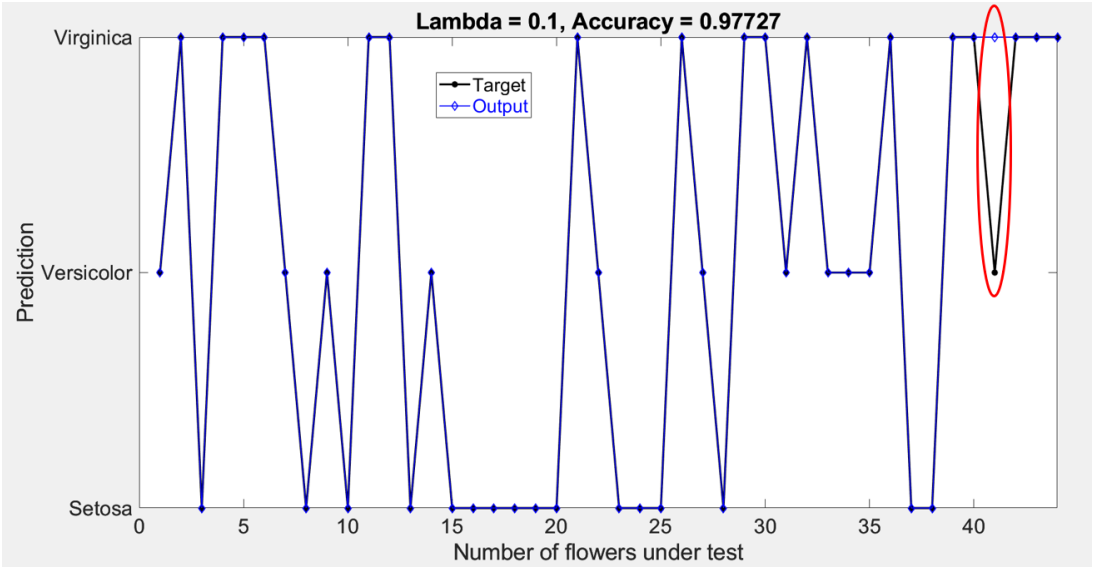

Figure 4.2.2: Network prediction results on the test samples. The targets are indicated by black circles, while the network's predictions are represented by blue diamonds. The λ used is the one found through the five-fold cross-validation.

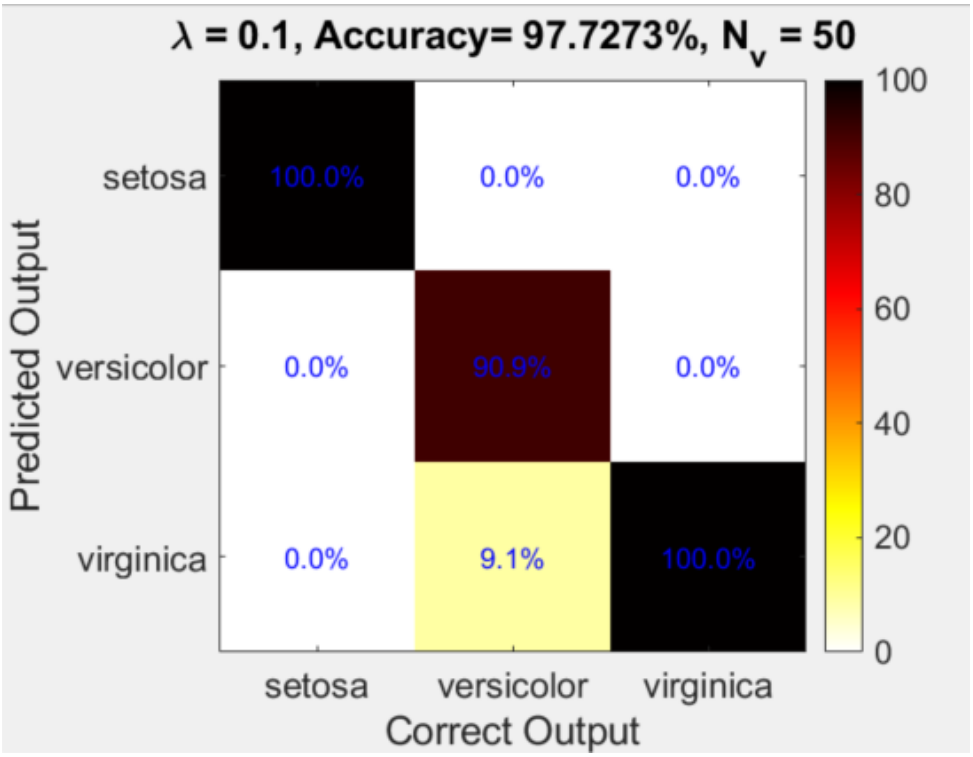

Figure 4.2.3: Confusion matrix showing the classification accuracies. The overall accuracy is 97.73% with $\lambda = 0.01$ and $N_v = 50$.

Figure 4.2.4 shows the evolution of the intensity of the transmission coefficient at resonance at the through port for each sequence of flowers. Each row of the map represents the values of the transmission coefficient sampled at each virtual node and encodes a different flower. The colors indicate the intensity of the transmission coefficient. Vertically, the flowers are arranged with all Setosa flowers first, followed by Versicolor, and finally Virginica.

A wide variability in color intensity can be observed, reflecting the diversity in the virtual nodes' responses. This diversity helps improve classification accuracy because it allows the system to map the input into a higher-dimensional space, making the classes more easily separable.

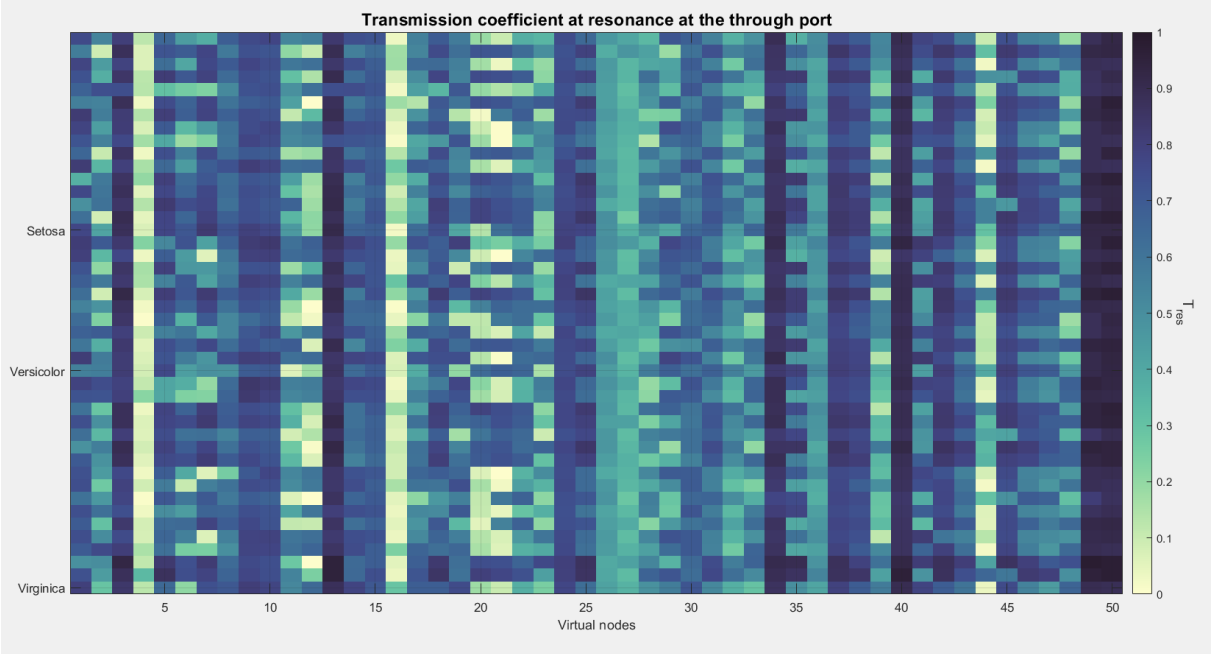

Figure 4.2.4: Map of the intensity of the transmission coefficient at resonance measured at the through port for each flower. Each row represents a flower, and each column corresponds to a virtual node. The transmission coefficient was calculated by dividing the power at the through port of the microring, measured at the wavelength of the probe, by the probe input power to the microring.

4.2.4 Simulations of neuromorphic experiments on the IRIS dataset

The results presentes in this subsection were obtained through MATLAB simulations for practical reasons. Figure 4.2.5 shows the confusion matrices obtained by sampling the output probe power for different configurations of the number of virtual nodes N_v with a bitrate of 30 Mbps and an average pump power of 7 dBm. From Figure 4.2.5, it can be observed that as N_v increases, the classification accuracy improves, but the simulation time also increases.
With 50 virtual nodes, the accuracy is 97.73%, while with 25 and 10 nodes, it decreases to 95.45%, and with 5 nodes, the accuracy drastically drops to 65.91%. As the number of nodes decreases, each flower is encoded with fewer values equal to N_v, which results in a reduced memory capacity and a reduction in the complexity and number of nonlinear relationships between the virtual nodes, leading to a decrease in the predictive performance of the model.
Another test, conducted through simulation, analyzed the power of the probe signal output at 30 Mbps while varying the average pump power. From this test, it was found that accuracy decreases at both low and high power levels. This phenomenon is understandable by analyzing the graphs in Figure 4.2.6, which show the transmission coefficient at resonance at the through port for three different pump powers: 2, 6, and 15 dBm.
In the first map (2 dBm), the intensity of the virtual nodes varies within a very narrow range, with the distribution predominantly concentrated near resonance, resulting in a very "flat" color

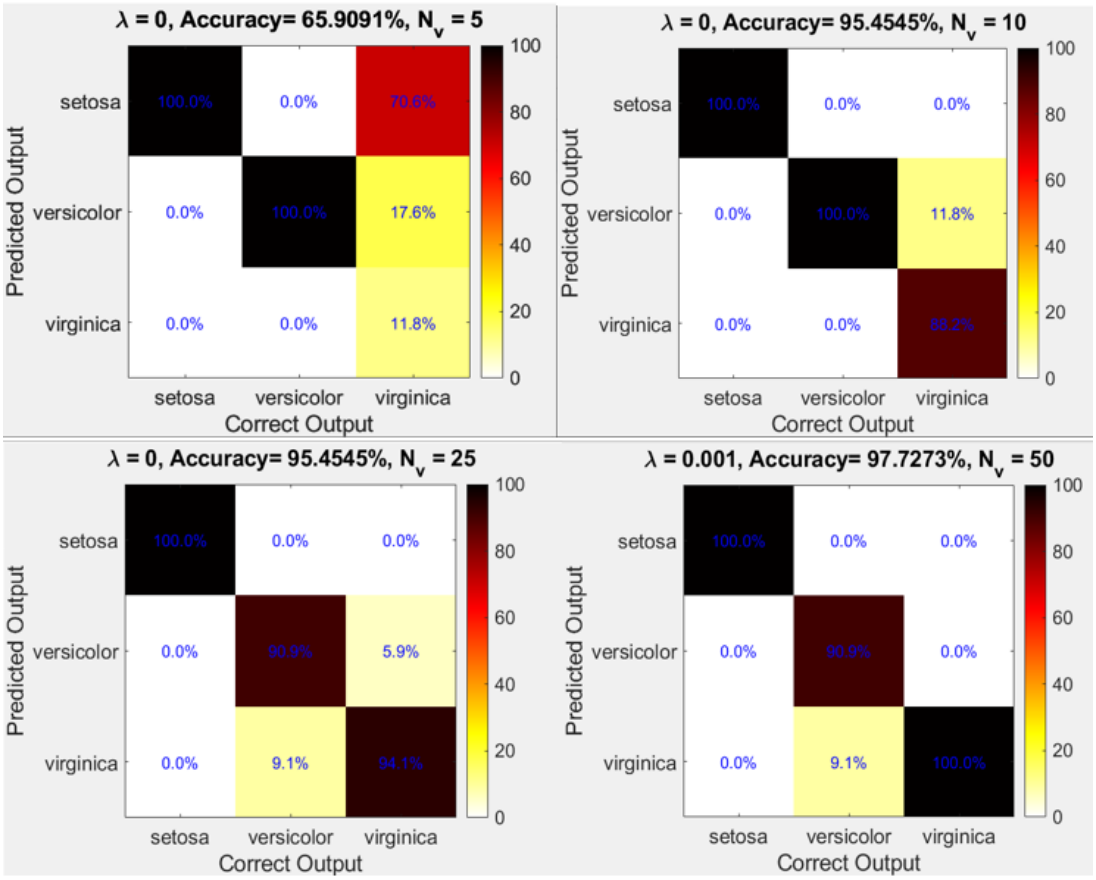

Figure 4.2.5: Confusion matrices obtained through MATLAB simulation for the recognition of Iris species with a bitrate of 30 Mbps and an average power of 7 dBm. Different numbers of virtual nodes were considered: $N_v = 5, 10, 25$, and 50.

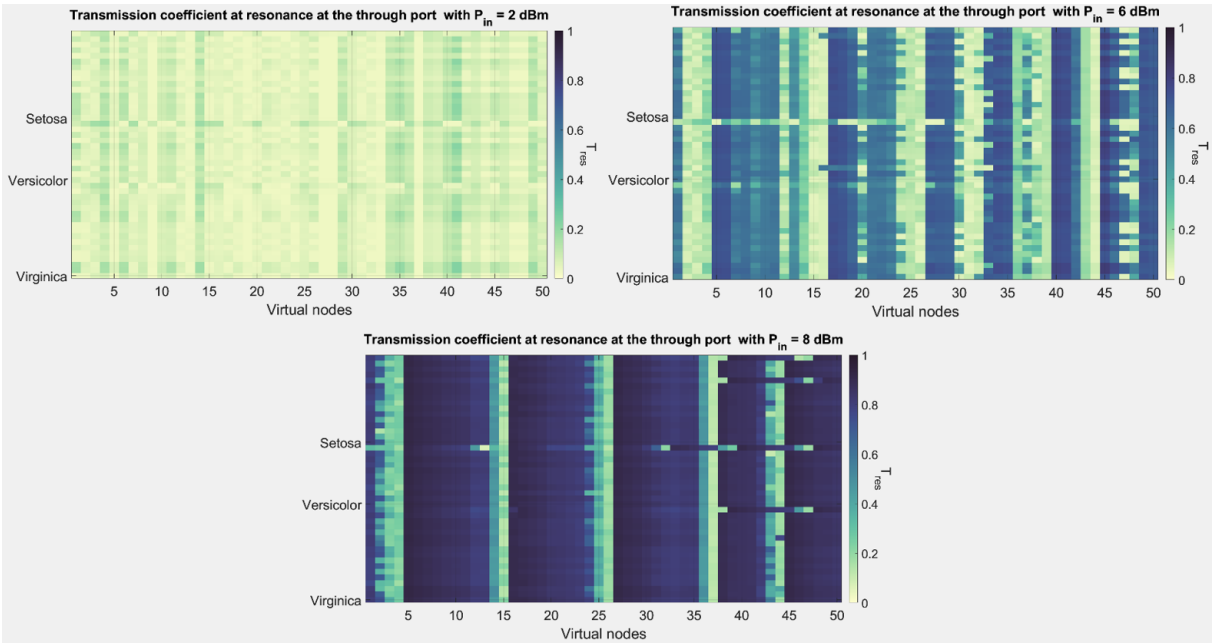

Figure 4.2.6: Simulated transmission coefficient at resonance at the through port for 3 different pump powers: 2, 6, and 15 dBm, respectively left, right, and below center.

map. Despite this, the accuracy remains high at 97.73%. However, further reducing the power makes this effect more pronounced, and the accuracy drops to 93.18%. The graph for this last accuracy is not shown due to poor visibility.
The graph on the right (6 dBm), showing an accuracy of 97.73% (the highest observed), highlights variability in the intensity of the virtual nodes that spans a larger range of values, fully exploiting the nonlinear properties of the microring.
In the graph below (15 dBm), similar to the first one, the variability of the nodes is limited to a narrow range, mostly concentrated far from resonance, causing a reduction in accuracy (93.18%). As the power increases further, this effect becomes more pronounced, with accuracy dropping to 88.63%, as the microring stays out of resonance for longer periods due to the high power. In fact, as observed from the maps, bands form, and the higher the power, the wider and more evident these bands become, indicating that the microring remains out of resonance for longer periods.
In Figure 4.2.7, four graphs obtained through simulation are shown, illustrating the various physical phenomena involved during the experiment. The graphs show the behavior of different parameters for three average pump power levels: 2, 4, and 8 dBm, with a bitrate of 30 Mbps and 50 virtual nodes.
In detail, the transmission coefficient, electron and hole density, and temperature variation were analyzed.

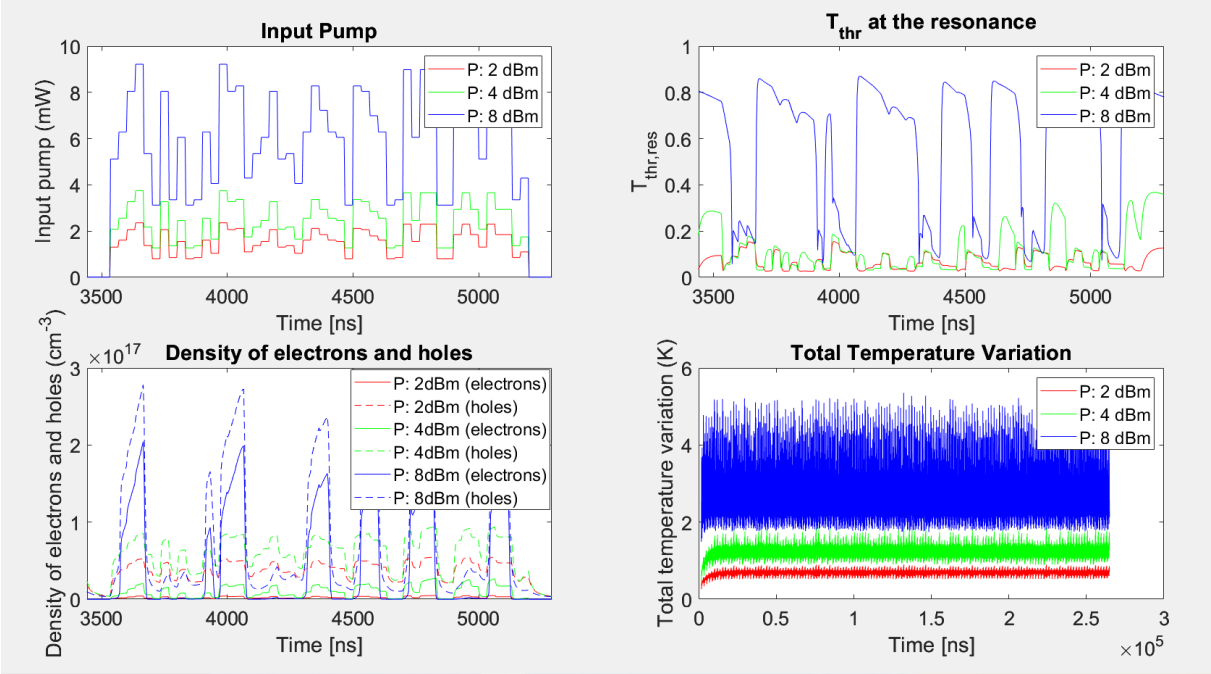

Figure 4.2.7: Behavior of the transmission coefficient at the through port at resonance, the charge carrier density (electrons and holes), the wavelength variation due to thermal effects and free carrier dispersion, and the overall temperature for different pump powers. The graph of the total temperature variation is shown over a longer time range compared to the other three graphs.

The first graph shows the temporal behavior of the pump power, encoding a single flower, at the input for the three different average power levels. The second graph illustrates the behavior of the transmission coefficient T_{thr} at resonance at the through port as a function of time. The third graph shows the electron and hole density, which at low power levels follows the pump

quite well. However, as the power increases, the effect of TPA becomes increasingly dominant, creating very steep peaks. The curves for the holes have a more amplified response than those for the electrons. So, the electron and hole densities are not perfectly symmetric.
The fourth graph shows the temperature behavior.
In the second graph at higher power levels, bands out of resonance can be observed, as already seen in Figure 4.2.6. For example, at 8 dBm, these bands are evident, where the transmission coefficient is out of resonance with a slight variation. This means that in these bands, the microring is less efficient at transferring information from the pump to the probe.
This behavior can be better understood by observing Figure 4.2.8. In this figure, it can be seen that when the electron density rapidly decreases, the shift due to free carrier dispersion quickly cancels out, while the temperature decreases more slowly compared to the recombination of the carriers. This difference in speed drives the microring quickly out of resonance. Additionally, the higher the pump power, the longer it will take for the system to return to resonance, as the cooling process will take more time. This explains why at high power levels, the out-of-resonance bands are wider and persist for longer.

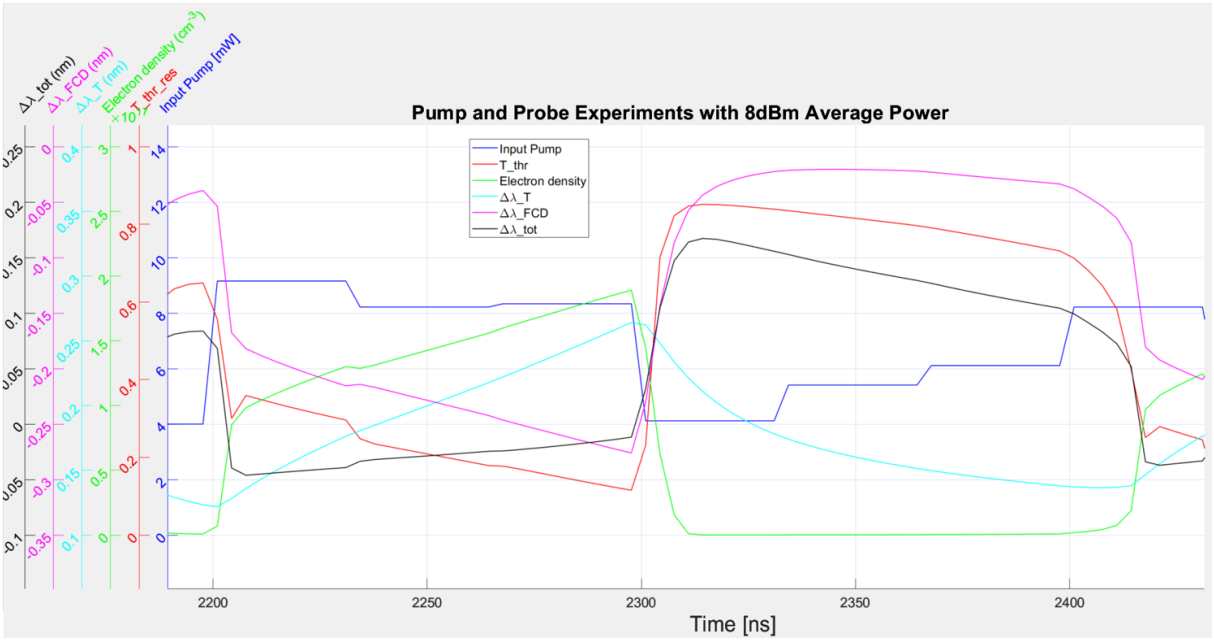

Figure 4.2.8: Time evolution of the input pump power (blue line), the transmission coefficient T_{thr} at resonance (red line), the electron density (green line), the resonance wavelength shift due to temperature variation $\Delta\lambda_T$ (magenta line), the shift due to free carrier dispersion ($\Delta\lambda_{\text{FCD}}$, cyan line), and the total wavelength shift $\Delta\lambda_{\text{tot}}$ (black line).

At lower power levels, this phenomenon is reduced because the temperature effect is less significant, allowing the pump to transfer information to the probe more accurately. However, at very low power levels, the amount of nonlinearities caused by the pump are negligible, therefore the transmission coefficient varies so little that the pump cannot effectively transfer information to the probe.
This detailed analysis reveals that the accuracy in classifying the IRIS species decreases as the pump power increases, due to the reduced variability in the microring's response. When the temperature increase, caused by the self-heating, is high, the red shift of the resonance causes the transmission coefficient becomes less sensitive to input changes, reducing the pump's ability to transfer information to the probe. In conclusion, the goal of accurately discriminating the

entire dataset was achieved, managing to obtain performance equivalent to that achieved with more complex systems with many nonlinear nodes.

4.3 1-bit delayed XOR

The second experiment conducted is the 1-bit delayed XOR, which aims to predict the XOR operation between a bit and the previous one. As with the IRIS experiment, the methodology adopted for the experiment is the same with some differences. In this case, during the preprocessing phase, a random sequence of 5000 bits is generated, forming the matrix U_{in}. As in the IRIS experiment, a matrix $\mathbf{W}_{\text{u}}$ is created, consisting of N_v rows and a column filled with ones, because in this task the challenge is to exploit the fading memory effect, which occurs when the response to one bit influences the response to the next bit. The product of the two matrices returns the input matrix.
For each bit in the sequence, the XOR operation with the previous bit is calculated to generate a dataset with three columns. The first column contains the result of the XOR operation, while the second and third columns indicate the type of result. Specifically, if the XOR operation produces a "0," then a "1" is placed in the second column and a "0" in the third column. Conversely, if the XOR operation returns a "1," a "0" is placed in the second column and a "1" in the third. Therefore, the last two columns serve as labels for training the model.
In the postprocessing phase, the data from matrix X, which contains the values of the virtual nodes sampled from the output signal of the ring, must be transformed and reorganized.
Initially, the matrix X with dimensions 3x4998 (where the first two bits, each composed of three virtual nodes, have been removed) is reorganized into a matrix with dimensions 6×2499, called X_{transf}, in such a way that two consecutive bits (hence 6 virtual nodes) are paired on the same column. This matrix is then transformed a second time into one with dimensions 6×4997, where the data is arranged in a specific way: each column contains the three current virtual nodes at the top and below them the three virtual nodes corresponding to the previous bit. This will be the matrix on which the training will be carried out, as was done for the IRIS experiment.
For the 1-bit delayed XOR task, fading memory is necessary, meaning that each state of the system must remember the previous one. For this reason, the relationship between the bit duration (T_b) and the carrier lifetime (τ_r) plays a crucial role.
In the XOR task, to achieve high accuracy, the duration of each bit must be similar to the carrier lifetime so that the response of the nonlinear node to a new bit is influenced by the previous one.
A positive wavelength detuning was applied to the probe signal because, when placing the probe exactly at resonance, if the average pump power is high, the power variation could push the system's response back and forth across the resonance peak, causing instability.

4.3.1 Tests performed for the 1-bit delayed XOR

The first test conducted is shown in Figure 4.3.1, where a comparison is made between the results obtained in simulation and the experimental results with a bitrate of 10 Mbps, an average pump power of 6 dBm, and an ER=13. On the left side of the figure, the simulated signals for the pump in and probe out are displayed, while on the right, the experimental results for the same signals are reported. It can be observed that the model simulates the behavior obtained experimentally quite well. All curves have been normalized between 0 and 1. The small circles

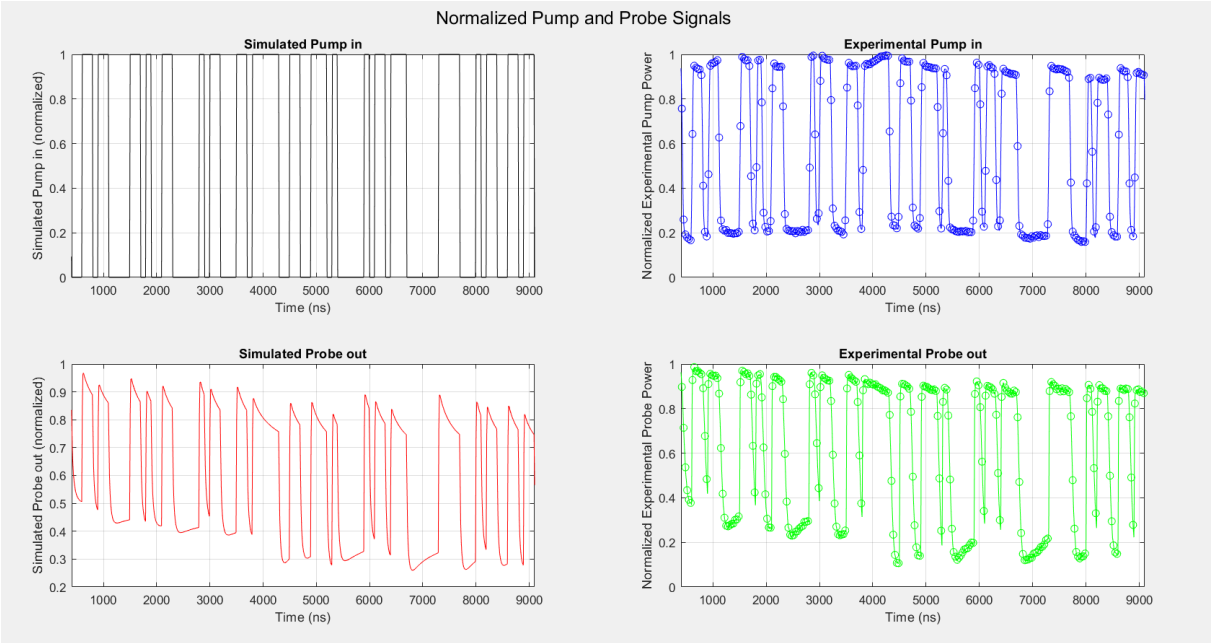

Figure 4.3.1: Comparison between simulated (left) and experimental (right) ”pump in” and ”probe out” signals for the 1-bit delayed XOR experiment. The curves are normalized between 0 and 1.

on the curves represent the virtual nodes, i.e., the points that will be sampled for use in the model's training, both in simulation and experimentally.
From the probe out curves, a memory effect can be observed, visible during the transition from one state to the next. In fact, it can be seen that the responses are in a transient regime, meaning that the current state also depends on the previous state, making the XOR task solvable by a linear classifier.
By analyzing the probe signal, it can be noted that when the pump signal is high, the probe signal is also high but shows slight attenuation over time, whereas when the pump signal is low, the probe signal is also low and, if the period is long, it undergoes a slight increase over time.
These behaviors can be better understood by referring to figure 4.3.2, which was generated from the data obtained through simulation.
When the ”pump in” signal increases, the free carrier density rapidly grows, leading to a significant shift of the micro-ring resonance wavelength relative to the probe wavelength (on which a positive detuning was applied). During this phase, the shift caused by the FCD becomes predominant over the temperature variation, causing an overall shift of the resonance toward shorter wavelengths (blue shift). As a result, a rapid increase in the transmission coefficient is observed.
While the ”pump in” signal remains high, the carrier density continues to increase, but at a much slower rate, and the carriers begin to recombine. As a consequence of the generation of all these carriers, the temperature of the microring starts to rise, becoming more influential than the FCD. This results in a slight decrease in the transmission coefficient during the period in which the pump in signal stays high.
Similarly to the IRIS experiment, figure 4.3.3 shows the confusion matrix obtained by sampling the probe out signal with a bitrate of 10 Mbps, an average pump power of 6 dBm, and an ER=13, experimentally derived. From the chart, it can be observed that the model achieved 100% accuracy with $\lambda = 0.1$, meaning that all samples were correctly classified. A 100% ac-

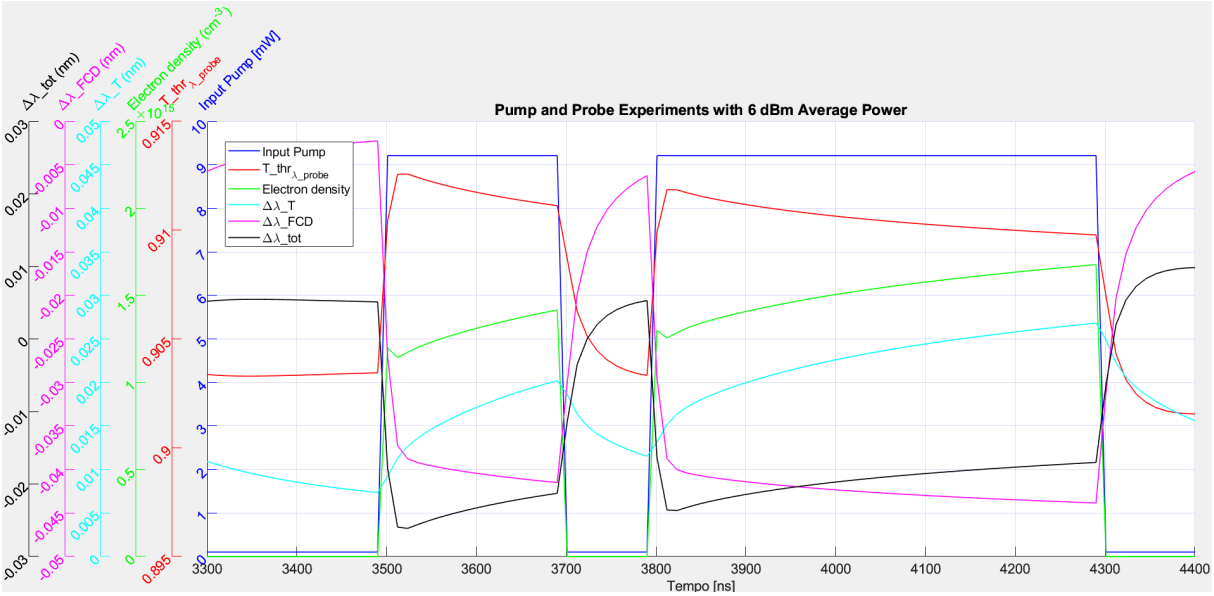

Figure 4.3.2: Behavior of ”pump in”, transmission coefficient at the through port at the probe wavelength, electron density, resonance wavelength shift due to FCD, temperature effects, and their combination with an average pump power of 6 dBm and a bitrate of 10 Mbps.

curacy was also obtained in simulation.

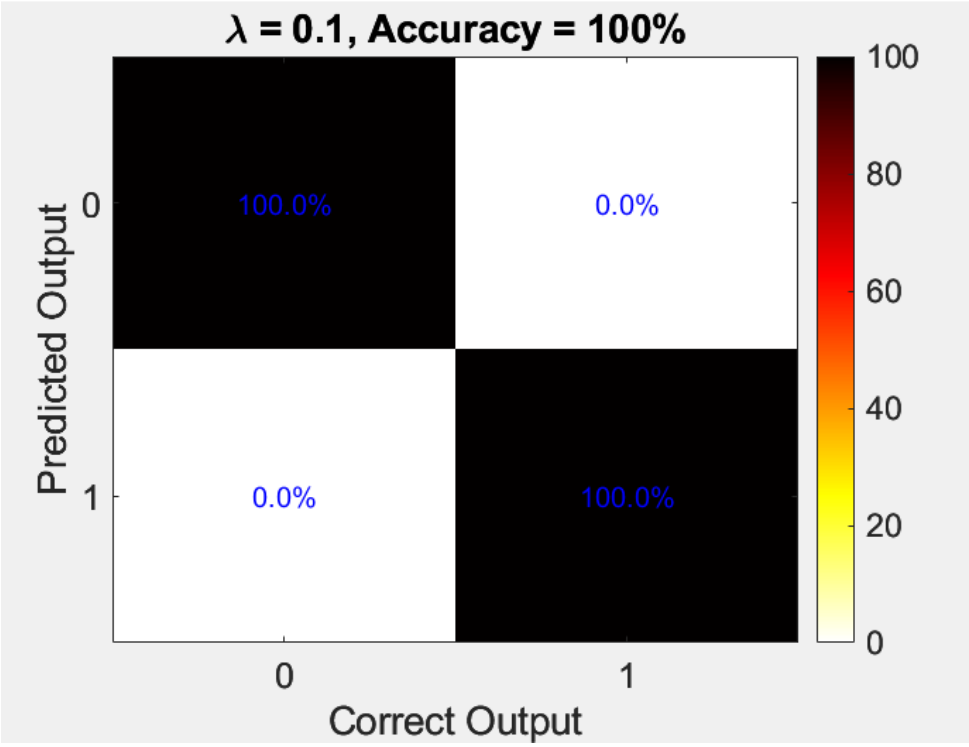

Figure 4.3.3: Confusion matrix obtained from the sampled probe signal, experimentally derived with a bitrate of 10 Mbps, an average pump power of 6 dBm, and an ER=13. The matrix shows 100% accuracy.

Figure 4.3.4 shows the comparison between the network’s predictions and the targets for the same signal, and as can be seen, they match perfectly for all samples under test.

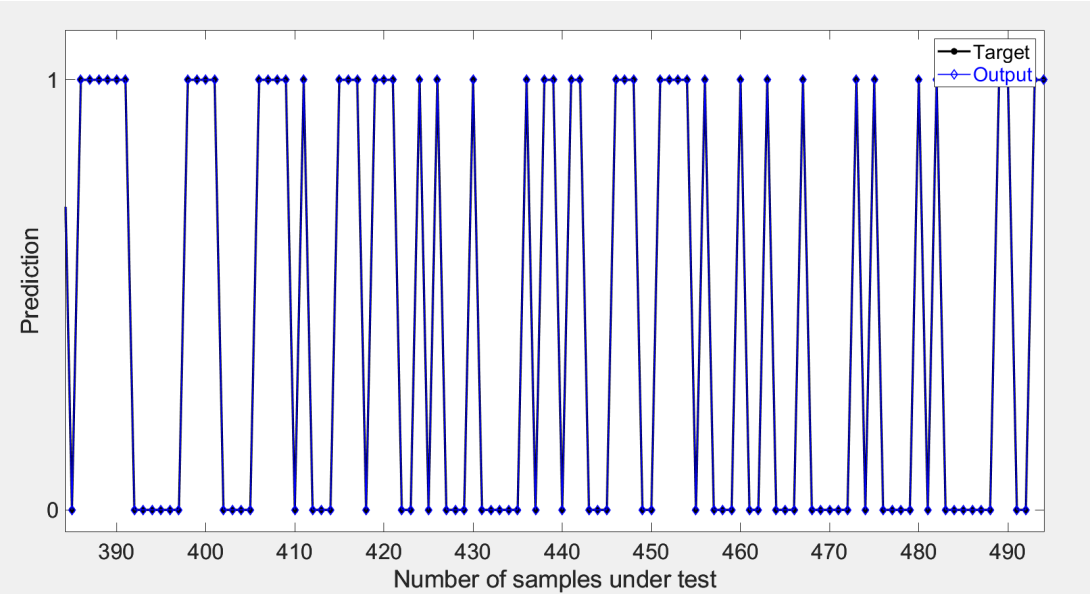

Figure 4.3.4: Comparison between the targets and the network's predictions. The black dots represent the target, while the blue diamonds represent the predicted output.

Finally, Figure 4.3.5 shows the behavior of the accuracy obtained from experimental data as a function of the bitrate. The horizontal axis shows the bitrate (in Mbps), while the vertical axis

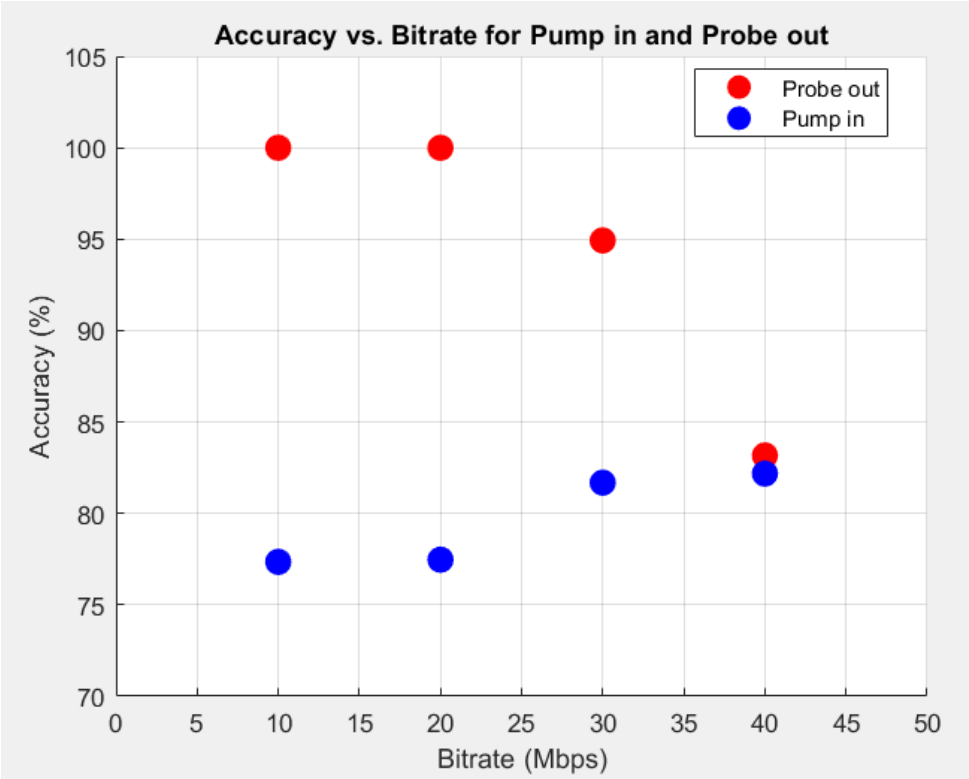

Figure 4.3.5: Behavior of experimental accuracy as a function of bitrate for the Pump in and Probe out signals with an average power of 6 dBm.

shows the accuracy in percentage. The red circles indicate the accuracy obtained by sampling the "probe out", while the "blue circles" indicate the accuracy obtained by sampling the pump

in. As can be seen, as in the IRIS experiment, sampling the probe out yields higher accuracy than sampling the pump in. Moreover, it can be observed that as the bitrate increases, the accuracy of the probe out decreases because the duration of each bit becomes too short compared to the carrier lifetime.

Chapter 5

Conclusions

In this thesis, a *reservoir computing* system based on a silicon microring with time multiplexing was implemented, capable of solving two complex experiments: the classification of the IRIS dataset and the one-bit delayed XOR task. The results obtained were excellent, demonstrating that the microring significantly improves classification accuracy thanks to the nonlinearities generated by free carrier dispersion, two-photon absorption, and temperature variations.
In the case of the IRIS experiment, it was shown that by sampling the output probe signal power from the microring at 10 Mbps (with an average input pump power of 4.8 dBm), an accuracy of 97.73% was achieved. However, by increasing the bitrate to 20 and 30 Mbps, the accuracy slightly decreased to 95.45%, due to the increased signal speed, which makes the information transfer from the pump to the probe less efficient.
Another important result concerns sampling the input power without passing through the microring. In this case, a significantly lower accuracy compared to that obtained by sampling the output probe signal from the microring was achieved, equal to 84.1%, 77.27%, and 86.36% for bitrates of 10, 20, and 30 Mbps, respectively. This demonstrates that the nonlinearities introduced by the microring are essential for optimal dataset classification.
It was also shown that as the number of nodes decreases, the accuracy drops due to the reduction in complexity and the number of nonlinear relationships between the virtual nodes, falling to an accuracy of 65.91% with 5 nodes.
In addition, simulation results showed that the system accuracy decreases at both very low and very high power levels. At very low power, the variability of the transmission coefficient is limited, while at very high power, the microring remains mostly out of resonance due to the high temperature generated by the *pump*. In both cases, the transfer of information from the *pump* to the *probe* is consequently inefficient.
Regarding the one-bit delayed XOR experiment, both experimental and simulated results have shown that the system is capable of solving the task with high accuracy. It was demonstrated that by sampling the output probe signal from the microring, an accuracy of 100% was achieved with a bitrate of 10 Mbps and an average *pump* power of 6 dBm. As in the IRIS experiment, it was shown that sampling the output probe signal results in higher accuracy compared to sampling the input pump. Moreover, it was shown that as the bitrate increases, the accuracy decreases because the duration of a single bit becomes too short compared to the carrier lifetime.
In conclusion, the results obtained demonstrate that the microring resonator, with its nonlinearities introduced by FCD, TPA, and temperature, as well as its memory capability, allows mapping data into a higher-dimensional space, making them linearly separable and thus enabling the model to solve even very complex tasks, provided these do not require an excessively

extended temporal memory. In the latter case, it will be necessary to introduce a delay line.
The future applications of this study include image recognition of numbers from 0 to 9, testing its limits and potential.
Photonic processing offers significant advantages in terms of energy efficiency since it enables higher speeds compared to traditional systems based on conventional electronics. However, the technological difficulty in realizing more complex photonic neural networks and integrating them with electronic systems represents a significant challenge in achieving the same performance as software-based neural networks.

Bibliography

[1] M. Borghi, S. Biasi, and L. Pavesi. Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations. *Scientific Reports*, 11:15642, 2021.

[2] Marco Novarese. *Modelling and characterisation of microrings for semiconductor lasers integrated in the Silicon Photonics platform*. Doctoral dissertation, Politecnico di Torino, Turin, Italy, 2023.

[3] Matteo Cucchi, Steven Abreu, Giuseppe Ciccone, Daniel Brunner, and Hans Kleemann. Hands-on reservoir computing: a tutorial for practical implementation. *Neuromorphic Computing and Engineering*, 2:032002, 2022. Topical Review.

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

[5] B. Penkovsky, L. Larger, and D. Brunner. Efficient design of hardware-enabled reservoir computing in fpgas. *Journal of Applied Physics*, 124(16):162101, 2018.

[6] Zhen Sun, Gianluca Pedretti, Andrea Bricalli, and Daniele Ielmini. One-step regression and classification with cross-point resistive memory arrays. *Science Advances*, 6(11):eaay2378, 2020.

[7] Patrick Steglich, Dominik G. Rabus, Cinzia Sada, Martin Paul, Michael G. Weller, Christian Mai, and Andreas Mai. Silicon photonic micro-ring resonators for chemical and biological sensing: A tutorial. *IEEE Sensors Journal*, 2021.

[8] Vilson R. Almeida and Michal Lipson. Optical bistability on a silicon chip. *Optics Letters*, 29(20):2387–2389, 2004.

[9] Marco Novarese, Sebastian Romero Garcia, Stefania Cucco, Don Adams, Jock Bovington, and Mariangela Gioannini. Study of nonlinear effects and self-heating in a silicon microring resonator including a shockley-read-hall model for carrier recombination. *Optics Express*, 30(9):14341–14357, 2022.

[10] M. Nedeljkovic, R. Soref, and G. Z. Mashanovich. Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1–14-m infrared wavelength range. *IEEE Photonics Journal*, 3(6):1171–1180, 2011.

[11] Marco Novarese, Sebastian Romero-Garcia, Jock Bovington, and Mariangela Gioannini. Dynamics of free carrier absorption and refractive index dispersion in si and si/polysi microrings. *IEEE Photonics Technology Letters*, 2022.

[12] imec. *Library Handbook — ISIPP50G 2.2.0*, 2022.

[13] Lirong Cheng, Simei Mao, Zhi Li, Yaqi Han, and H.Y. Fu. Grating couplers on silicon photonics: Design principles, emerging trends and practical issues. *Micromachines*, 11(7):666, 2020.

[14] Riccardo Marchetti and Ali e Chen Xia e Cristiani Ilaria e Richardson David J e Reed Graham T e Petropoulos Periklis e Minzioni Paolo Lacava, Cosimo e Khokhar. High-efficiency grating-couplers: demonstration of a new design strategy. *Scientific Reports*, 7(16670), 2017.

[15] Wim Bogaerts, Peter De Heyn, Thomas Van Vaerenbergh, Katrien De Vos, Shankar Kumar Selvaraja, Tom Claes, Pieter Dumon, Peter Bienstman, Dries Van Thourhout, and Roel Baets. Silicon microring resonators. *Laser & Photonics Reviews*, 6(1):47–73, 2012.

[16] C Mesaritakis, V Papataxiarhis, and D Syvridis. Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system. *J. Opt. Soc. Am. B*, 30:3048–3055, 2013.

[17] S. Biasi, R. Franchi, D. Bazzanella, and L. Pavesi. On the effect of the thermal cross-talk in a photonic feed-forward neural network based on silicon microresonators. *Frontiers in Physics*, 10:1093191, 2022.

[18] G. Donati, C.R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris. Microring resonators with external optical feedback for time delay reservoir computing. *Optics Express*, 30(1):522–537, 2022.

List of Figures

List of Tables