Deep learning models for miniaturized silicon photonics sensors

Submitted by

Salah DADDINOUNOU
Master2 Nanodevices and Nanotechnology
Paris Sud University

Under the guidance of

Carlos Ramos
Researcher at Center of Nanoscience and Nanotechnology C2N

Silicon Photonics Group

Acknowledgment

To start, I would like to thank my internship supervisor , Dr. Carlos Alberto Alonso Ramos for allowing me to do my internship within the MINAPHOT team in best conditions. His advice and his daily support allowed me to grow both technically and personally.

I would like also to thank all the researchers in MINAPHOT team who were very supportive, especially Mme Delphine Marris-Morini who was my professor of Nanophotonics this year, thanks to her I steeped into this interesting field.

My thanks goes also to the two PhD students with whom I shared my office during these four months, Dorian and Dinh. Thank you for the fruitful discussion we had together.

My special thanks also to my fellow interns, Makram, Julien, Youen, Lucas, Thomas, Vincent, with whome I had a lot of fun during coffe-breaks.

Contents

Chapter 1

Motivations

Many environmental and diagnostic sensors include miniaturized spectrometers, the working principle of these devices are numerous. Spectrometers can be classified in two categories: dispersive element spectrometers, and Fourier-transform spectrometers (FTS) . In bulk optics, FTS is just a Michelson interferometer with one movable mirror [9].

An interferogram is obtained by taking measurements of output power corresponding to many discrete positions of the movable mirror. Changing the mirror position results in variation of the optical path delay (OPD) leading to alternation between constructive and destructive interference while moving. A monochromatic input results in a sinusoidal interferogram, and a polychromatic one results in superposition of many sinusoids [2].The device is then performing an inverse Fourier-transform (FT) because of this relation between the spatial and the frequency domain. FT spectrometers are preferred for their high resolution, and for their large throughput (ability to capture input light).

In integrated optics, Spacial Heterodyne Fourier Transform Spectrometers (SHFTS) are a modified version of Michelson interferometer [2] where no moving part is included. This advantage makes the SHFTS suitable for miniatirization and integration in photonic chip technology.

In SHFTS design, the interferogram measurement is done in one single shot, the spectrum is then retrieved by performing the Fourier transform. This is a big advantage when fast measurements are required, compared to scanning FT spectrometers where the interferogram is obtained by sweeping through the outputs, so it takes time and it might not be convenient in some applications like the fast moving satellites when doing in-orbit spectroscopy [5].

The disadvantage of this technique is that the fabrication errors and the thermal fluctuations are not taken into account, which results in a distorted

interferogram, this means that the interferogram is no longer a series of sinusoidal functions, and the Fourier transform does not give the exact input frequencies.

The alternative for this concern is to abandon the FT method and describe the relation between the input spectrum and the output interferogram as a linear system. For this purpose, the device is modeled by its transfer matrix, so the output is simply the product of the input by this transfer matrix. Similarly, the input spectrum is obtained by multiplying the interferogram by the inverse of the transfer matrix, the advantage of this method is that all fabrication errors and thermal fluctuations which were omitted on the discreet cosine Fourier transform method are included in the calibration matrix, consequently, they are inverted when calculating the spectrum.

Still one problem when using the calibration matrix method, the spectrum retrieval is temperature-dependent, which means that temperature knowledge is required each time we want to calculate the spectrum, my internship comes as an attempt to solve this problem by proposing a model based on machine learning.

Machine learning is a very large topic which is used in many domains to make accurate predictions, without the need to know all the parameters describing a certain problem. One recent approach of machine learning is called "Deep Learning", despite the fact that it is based in simple maths and and linear algebra, the model has demonstrated a big capabilities and its applications are well known in pattern and speech recognition [10].

Deep learning requires two main steps, training and prediction. This powerful tool is now being used in research in order to build and optimize a model that describes a certain phenomenon which doesn't seem to follow a clear evolution rule.

Chapter 2

Basics and Fundamentals

Over the last two decades, miniaturized photonic devices have had an increasing importance because of the very promising applications in different domains. In telecommunications, modulators have progressed on speed to more than 50 Gb/s[11], the modulators are parts of transceivers, devices which are used in data centers to transport the huge amount of information between servers. In health diagnostics, integrated photonic devices are used to build compact and low cost lab-on-chips [1].

In environmental sensing, spectrometers based on Fourier transformation have been proposed, the configuration of Michelson spectrometer is adapted for integrated photonics[2]. This gives a high resolution devices for gas detection and atmospheric sensing [2]. Further parameters such as the throughput (étendue) are meant to be enhanced with some modifications in the device.

These devices are based on silicon. Microelectronics industry has developed transistors over 50 years on this material to the point where they reached their limits with Moor's law. Now we benefit from this very mature industry to develop photonic devices in a compact and miniaturized way. That's how the field of silicon photonics was born.

2.1 State of the art

Spectrometers based on dispersion [15] and on interference[12] are known in bulk optics, diffraction gratings and Michelson configuration are used respectively in these two kinds of spectrometers. Michelson interferometer has two perpendicular arms, with one mirror placed at the end of each arm Fig.2.1. When light coming from the source is split in the intersection, it propagates towards the two mirrors, but it does not cross the same distance since one mirror is fixed and the other is movable. This difference in the pathlength

(or the optical path delay OPD) creates a phase shifting between the two propagating waves, which results in interference pattern in the screen once the two beams recombine, Fig.2.1 shows a scheme of Michelson spectrometer.
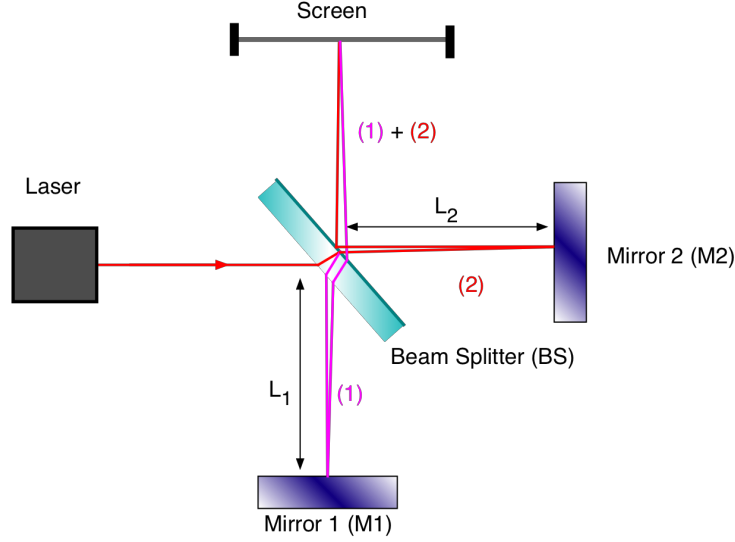

Figure 2.1: Schematic view of the operating principle of Michelson interferometer. M1 is movable, creating phase difference between the two beams coming from M1 and M2, the resulting interference is monitored in the screen after recombining. Image source: [12]

Another type of spectrometer with no moving parts is the Spatial Heterodyne Spectrometer SHS, which is more compact than the last ones. The SHS instrument share structural similarities with the Michelson interferometer [17]. The primary difference between the two is that in the SHS the mirrors are replaced with gratings. The position of the gratings is fixed such that the path difference between the two arms is zero, with the tilt of the gratings set for Littrow wavelength (λ_L). Wavelengths near the Littrow wavelength form fringes in the detector plane because the grating introduces a tilt on the wavefront proportional to the absolute deviation between the wavelength and the Littrow wavelength, as shown in Fig.2.2.

In step towards miniaturization, different types of integrated microspectrometers have been proposed, based on: Mach-Zehnder (MZ) waveguides, echelle gratings and arrayed waveguide gratings (AWG). For the sake of clarity, let's present an overview of the two last ones. The MZ interferometer is the main part of the studied spectrometer in this report, so it will be explained in detail in section 2.3.1.

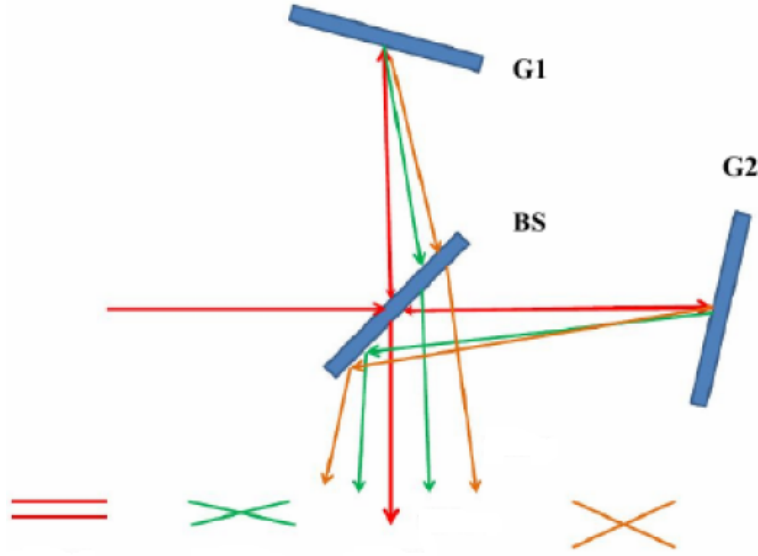The working principle of AWG is illustrated in Fig. 2.3. Incoming light

Figure 2.2: The concept of fringe formation in SHS. The Littrow wavelength in red doesn't create any fringes because the waves emerge parallel to each other from the two arms of the interferometer. Wavelengths other than Littrow form fringes on the detector because they are tilted by the gratings. Image source: [17]

is distributed over many delay lines in a star coupler. The individual contributions are then recombined in a second star coupler[16]. The star couplers function as an imaging system. When all the contributions of the delay lines are in phase, the field profile of the input aperture is imaged in the plane of the output apertures. The quality of that image depends on the number of contributions (delay lines), but it is never perfect. When the wavelength changes, there will be a constant phase difference between every two arms, and this will tilt the phase front in the output star coupler. This, in turn, will shift the image. Therefore, there is a one-of-one mapping of the position in the image plane to the wavelength within a single FSR. The mode-to-mode response of input to output is determined[16].

Echelle gratings operate in a similar way as AWGs, by combining a number of optical delays with a free-space imaging system. The key difference with AWGs is that echelle gratings implement the delay lines as distances in the free propagation region set by the location of an array of reflecting facets. As shown in Fig.2.4, the light traverses the same free propagation region twice. Like an AWG, an echelle grating can separate many wavelength channels into their own output waveguide.

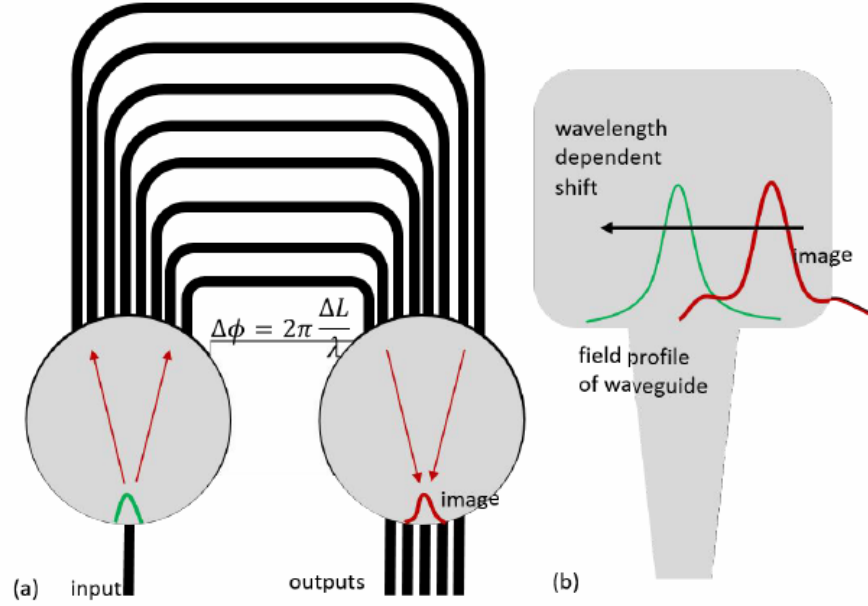Planar waveguide spectrometers based on echelle gratings and AWG can

Figure 2.3: (a) The working principle of an AWG, two star couplers at the input and output where nterferences take place. (b)The position of the maximum field profile depends on the input wavelength . Image source: [16]
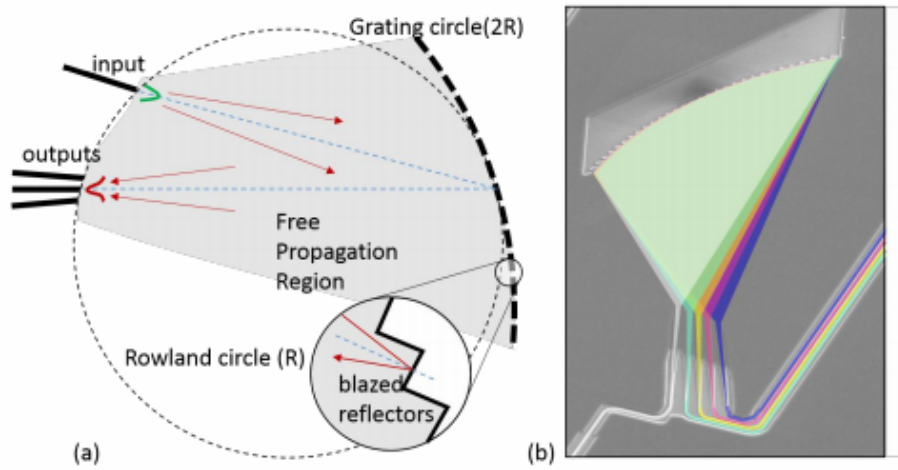

Figure 2.4: (a) basic design of the echelle gratings, the imaged profile depends on the input wavelength. (b) Example of a fabricated device demultiplexing a polychromatic input after diffraction in the grating part. Image source: [16]

be designed and fabricated with sub-nanometer spectral resolution and very compact chip sizes [2]. However, the étendue or the throughput, which is an important parameter giving the amount of input light that can be captured, is small in echelle grating configuration. In AWG, higher spectral resolution implies reducing the width of the imaged slit, which means also a small throughput [2]. Increased étendue was obtained by interleaving two AWG [13]. The working principle is based on Fourier transform spectroscopy (FT) known by its high resolution and large étendue. The FT configuration was then demonstrated with an array of Mach-Zehnder waveguides [2]. We focus in our study in this last type of spectrometers.

2.2   Fourier transform spectroscopy

Fourier transform spectrometer FT offers a wide spectral range, high resolution, large optical throughput [2]. The classical FT spectrometer include moving mirrors and complex design making them very sensitive to vibrations, and they need to be calibrated before measurement [3]. On the flip side, spacial heterodyne spectrometer (SHS) are based on Michelson configuration, but mirrors are replaced by non-moving dispersion gratings, providing spatially dispersed interferogram of light. The SHS spectrometers are less sensitive to vibrations and thermal fluctuations. The miniaturized version of SHS spectrometers are proposed via the FT-AWG (two interleaved AWG) [2] . Despite the fact that the throughput is increased in this device compared to classic AWG spectrometers, it stills relatively small. However, the SHS spectrometer design based on an array of Mach-Zehnder (MZ) waveguide interferometers provide multiple inputs which allow the optical throughput to be increased [2]. In this report we investigate exclusively SHS spectrometers based on MZ array. In such devices, the output power is collected along the array by increasing the OPD, the obtained data are called an interferogram, which is simply a power amplitude as function of MZ position. It is noticed that the spectrometer performs a cosine Fourier transform of the given spectral input, to get a spacial interferogram at the output :

$$I(x_i) = \int_0^{FSR} B(\bar{\sigma}) cos 2\pi\bar{\sigma} x_i d\bar{\sigma} \qquad (2.1)$$

where B is the incident spectral density, and $\bar{\sigma} = \sigma - \sigma_L$ is the shifted wavenumber, relative to the Littrow wavenumber σ_L, at which a monochromatic input produces a constant spatial power distribution at the outputs of the MZI array [4]. The spatial coordinate x_i has discrete values and serves to label the optical path delay. The term Fourier-transform spectroscopy

reflects the fact that a Fourier transform is required to turn the raw data into the actual spectrum.

SHS spectrometers based on FT have been implemented in satellites in some space missions in order to measure for instance the mesospheric radicals in the UV spectral range, or for space-born heterodyne observation of water at the infrared range [3]. Spectrometers should not be bulky nor heavy in order to be implemented in space missions. Since the cost of the mission depends on the size of the spectrometer, the miniaturization is the solution for this problem.

2.3   Device description

The simulated device consists on an array of Mach-Zehnder interferometers fabricated using the lithographic technology in silicon-on-insulator platform(SOI). This section describes the geometry of the device and its functionalities, we start by the building block of the device which is a single MZ interferometer, than we scale up to the whole spectrometer design by assembling a number of MZI together to build the spectrometer, the working principle will then be explained.

2.3.1   The building block: MZI

The Mach-Zehnder device is an interferometer which can be made of two directional couplers. The directional coupler allows light to couple from one weveguide to another one which is very close. In the first directional coupler -the splitter-, only one input port is used, the light is equally split between the two arms. Then, the light propagates independently in the two arms, one of them is longer than the other. This difference in the optical path creates a phase difference between light propagating in the two arms. In the second directional coupler -the combiner- the optical power coming from the two arms are added, which results in constructive or destructive interference because of the phase delay. So the signal exists in the two output ports, a constructive interference in one port corresponds to destructive interference in the second.

Each MZ interferometer has two input and two output ports and can be represented in terms of three functional sections which perform signal splitting, provide a differential phase delay, and recombine the two split signals, respectively Fig.2.6. In order to describe the mode propagation of light inside the two arms, the transfer matrix formalism allows us to represent the MZI as a product of three 2x2 transfer matrices, each corresponding to one
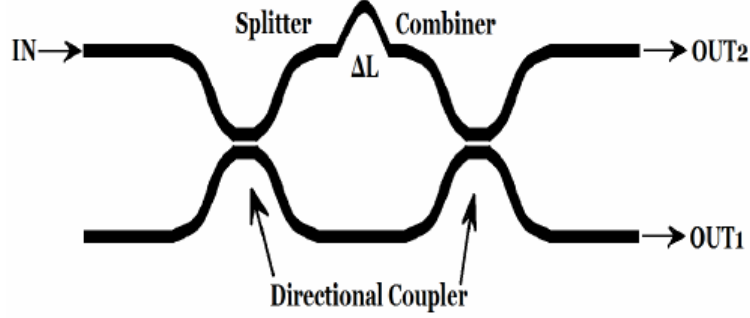
Figure 2.5: schematic view of one MZ interferometer with the three functional parts: splitting, propagating, combining. Image source:[14]

functional section. <mark>The physical quantities which are used to describe signal propagation are the two components of the modal amplitude $a_{1,i}$ and $a_{2,i}$ of either the magnetic or electric field</mark>, the first index refers to the input port 1 or 2, the second index i refers to the $i-th$ MZI. The optical signal through the different parts of one MZI is represented by a 2×1 column vector.
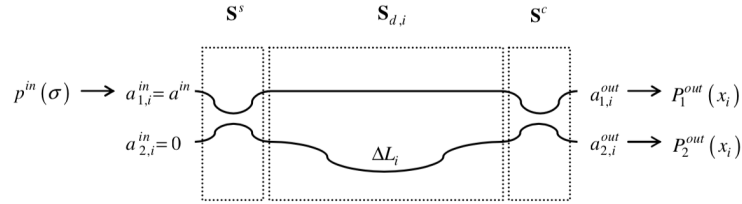

Figure 2.6: The MZI is represented by the three functional matrices, only one input is used. Coupling insures that power comes out from both output ports, but only one is necessary to retrieve the input spectrum. Image source:[2]

$$\begin{bmatrix} a_{1,i}^{out} \\ a_{2,i}^{out} \end{bmatrix} = S_i \cdot \begin{bmatrix} a_{1,i}^{in} \\ a_{2,i}^{in} \end{bmatrix} = S_c \cdot S_{d,i} \cdot S_s \cdot \begin{bmatrix} a_{1,i}^{in} \\ a_{2,i}^{in} \end{bmatrix}$$

Transfer matrix of the directional coupler for the splitter and the combiner are similar, and each one of them contains the power transmission factor γ and the power coupling ratio κ.

$$S_s = \gamma_s \begin{bmatrix} \sqrt{1-\kappa_s} & -i\sqrt{\kappa_s} \\ -i\sqrt{\kappa_s} & \sqrt{1-\kappa_s} \end{bmatrix} S_c = \gamma_c \begin{bmatrix} \sqrt{1-\kappa_c} & -i\sqrt{\kappa_c} \\ -i\sqrt{\kappa_c} & \sqrt{1-\kappa_c} \end{bmatrix}$$

If no losses take place in the splitter and in the combiner, the power transmission factors are $\gamma_s=\gamma_c=1$. Additionally, if we consider that splitting

and recombining are performed equally between the two waveguides, we can write for the coupling factors $\kappa_c = \kappa_s = 1/2$. The simplified matrix for both splitter and combiner writes:

$$S_s = S_c = \begin{bmatrix} \frac{1}{\sqrt{2}} & -i\frac{1}{\sqrt{2}} \\ -i\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

The third part is the delay matrix. First, we know that there is no coupling in this intermediate part because the two waveguides are not close enough to allow coupling, the matrix is then diagonal. However, propagation function $e^{-\beta Li}$ and loss factor $e^{-\alpha L_i}$ must be considered this time in both propagation arms. The delay matrix is then given by:

$$S_d = \begin{bmatrix} e^{-\alpha L1,i}e^{-\beta L1,i} & 0 \\ 0 & e^{-\alpha L2,i}e^{-\beta L2,i} \end{bmatrix}$$

Where α is the linear propagation loss coefficient, and β is the propagation constant $\beta = \frac{2\pi}{\lambda}n_{eff}$ of the mode inside the silicon waveguides, n_{eff} is the effective index of the propagating mode. We can rewrite the delay matrix as:

$$S_d = e^{-\alpha L2,i}e^{-\beta L2,i} \cdot \begin{bmatrix} e^{\alpha \Delta Li}e^{\beta \Delta Li} & 0 \\ 0 & 1 \end{bmatrix}$$

Where $\Delta L_i = L_{2,i} - L_{1,i}$ is the difference between the upper an lower arm with $L_{2,i} > L_{1,i}$. The delay matrix contains the differential loss factor $\alpha \Delta L_i$ which is usually small and may be ignored [2].

Finally, we assume that only one input port is used, so we get the output power $P_i^{out} = |a_i^{out}|^2$ at both ports of one single MZI after splitting, propagation, and recombining:

$$\begin{bmatrix} P_{1,i}^{out} \\ P_{2,i}^{out} \end{bmatrix} = \left(\begin{bmatrix} \frac{1}{\sqrt{2}} & -i\frac{1}{\sqrt{2}} \\ -i\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot e^{-\alpha L2,i}e^{-\beta L2,i} \cdot \begin{bmatrix} e^{\beta \Delta Li} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -i\frac{1}{\sqrt{2}} \\ -i\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} a_{1,i}^{in} \\ 0 \end{bmatrix} \right)^2$$

$$(2.2)$$

2.3.2   Operating principle

The MZ-based spectrometer is obtained by assembling a number of unbalanced MZ interferometers, with an increasing optical path delay along the array, the input light is distributed equally between the MZs over a serie of Y-like couplers, the transmittance of each MZ is then measured Fig.2.7.

For a monochromatic input, different transmittance of each MZ results in a sinusoidal spatial distribution of power across the outputs of the array.
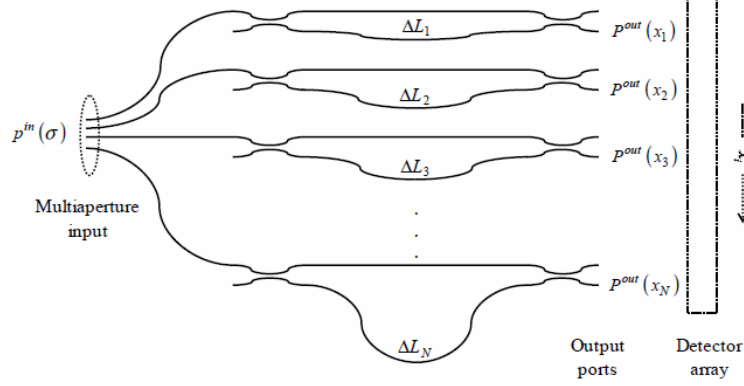
Figure 2.7: schematic view of planar waveguide SHS spectrometer formed by an array of unbalanced Mach-Zehnder interferometers. Image source: [2]

The spatially distributed power through the array's output is called an interferogram. Each point on the interferogram corresponds to the output of one single MZ port, which is the result of light interference between the two arms of a single MZ waveguide. Depending on the device characteristics, there is a certain monochromatic input for which the output interferogram is constant with zero spacial frequency through all MZs, this particular input frequency is called Littrow wavelength. So, starting from this wavelength, shifting the input frequency by a small amount(the device's resolution) results in a single fringe pattern on the interferogram. If we shift by a vector n times the resolution, we get n fringe patterns in the itnerferogram.

A polychromatic input gives a superposition of multiple periodic interferograms as shown in Fig. 2.8.

Once the interferogram is obtained, we can easily retrieve the input spectrum by performing a discrete cosine Fourier transform:

$$p^{in}(\bar{\sigma}) = \frac{\Delta x}{N} P^{in} + 2 \frac{\Delta x}{N} \sum_{i=1}^{N} F(x_i) cos 2\pi\bar{\sigma}x_i \qquad (2.3)$$

where $\Delta x = n_{eff}\Delta L_{max}$ is the maximum path delay in the array, n_{eff} is the effective index of the mode of light propagating in the waveguide and P^{in} is the input power measured at each MZI. The finite summation in Eq. (2.3) corresponds to a step-like cut-off (truncation) in the spatial interferogram distribution [2]. It is known that such truncation causes oscillatory features (ripple) in the retrieved spectrum. The ripple can be reduced by apodization. For this purpose, we apodize the interferogram by using the weighting
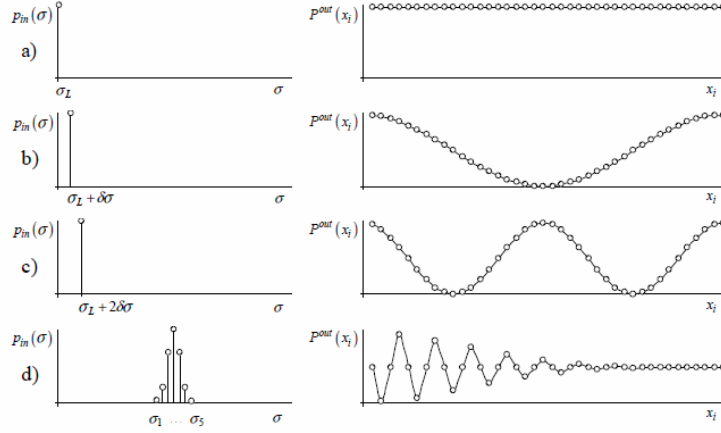
11

Figure 2.8: different figures of the output interferograms corresponding to a given monochromatic input. At Littrow wavelength, the interferogram is a a constant function, patterns appear as soon as we shift from this condition. a polychromatic input results in a superposition of sinusoidal outputs. Image source: [2]

function $W(x)$, thus obtaining the following spectrum retrieval formula :

$$p^{in}(\bar{\sigma}) = \frac{\Delta x}{N}P^{in} + 2\frac{\Delta x}{N}\sum_{i=1}^{N}W(x_i)F(x_i)cos2\pi\bar{\sigma}x_i \qquad (2.4)$$

Moreover, the Littrow condition as described earlier occurs periodically for some particular wavelengths, the periodicity is determined by the device characteristics. The OPD determines the interferogram's sampling frequency. According to Nyquist sampling condition, this frequency should equal at least twice the highest frequency of the transmission function along the interferogram. This condition implies that the free spectral range FSR of the device is half the separation between two Littrow conditions. The FSR is given by:

$$FSR = \frac{\lambda_0^2}{2n_g\Delta L_0} \qquad (2.5)$$

where λ_0 is the central operating wavelength of the spectrometer, n_g is the group index, and ΔL_0 is the incremental path delay.

Visualized in the interferogram, the FSR in wavelength domain is the separation between the wavelength giving a constant interferogram and the one resulting in a highest frequency.

Knowing the number N of the MZI constituting the device, we can writ

the device's resolution as:

$$\delta\lambda = \frac{2 \times FSR}{N} \tag{2.6}$$

The resolution, the FSR and the throughput are the key parameters to evaluate the device's performance.

We explained so far the concept of spectrometers based on Fourier Transform, and we showed how it converts the spectral domain into a spacial domain via interferences through the static increasing of the path delay. We discussed the advantage of using such configuration to measure the output power by a column of photo-detectors in one single shot. The Fourier transform of the obtained interferogram is then used to retrieve the input spectrum.

This method is simple and efficient, however, in practice, each device present some imperfections: fabrication errors, thermal fluctuations.., these effects distort the sinusoidal character of the interferogram. Consequently, the retrieval by FT is no longer the exact input peaks.

That's why modelling the device by it's calibration matrix, and using the inverse of this matrix to retrieve the spectrum was proposed. The advantage of this method is that all imperfections are embedded in the matrix, and are automatically inverted when calculating the retrieval.

The calibration matrix is constructed by measuring the transmittance function of each MZ interferometer along the array, at different input wavelengths. The concept of this method is that the interferogram of a certain input is just the product of this input by the calibration matrix. The Fig. 2.9 shows the calibration matrix in the ideal case, we see the Littrow condition perfectly aligned, with a certain periodicity. On the other hand, the Fig. 2.10, where fabrication errors are implemented, we see phase misalignment and the Littrow condition is no longer aligned. In such condition it is not possible to use the Fourier transform method for spectral retrieval. Instead, the spectrum is obtained by multiplying the interferogram by the inverse of the calibration matrix.

$$B(\lambda) = I(x_i) \times M^{-1} \tag{2.7}$$

where B is the retrieved spectrum, I is the interferogram's power amplitude as function of position and M^{-1} is the inverse of the transfer matrix. The next chapter will be dedicated to the study of the transfer matrix method, simulation of the device will be discussed step by step.
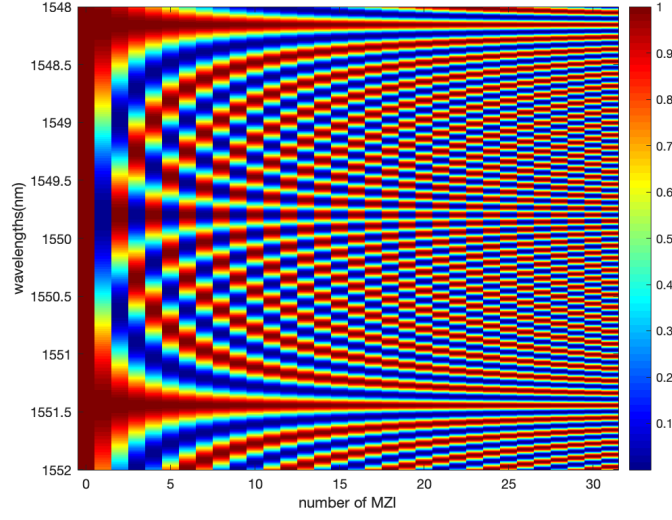
Figure 2.9: Calibration matrix of an ideal array of 32 MZIs, each point in the matrix gives the output coefficient of one MZ at a given wavelength
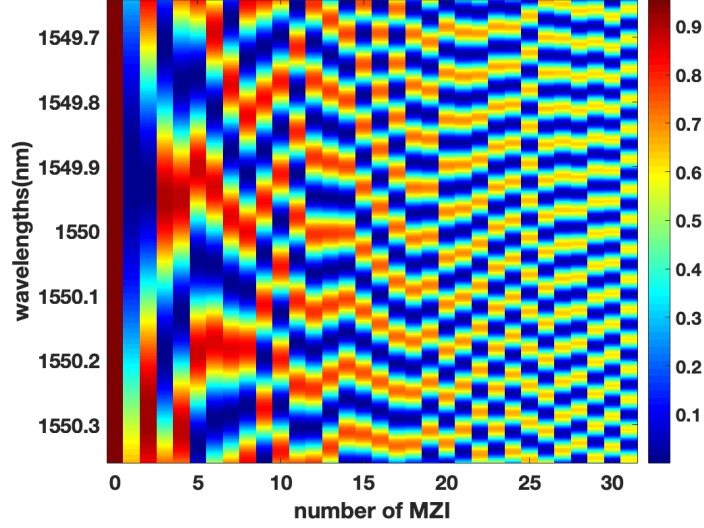

Figure 2.10: Calibration matrix with fabrication errors causing phase misalignment. 4dB/cm losses are also implemented here where the effect is more emphasized for the last MZIs

Chapter 3

Device simulation

In this chapter we simulate the spectrometer's fuctionalities using Matlab. First, the profile of the mode propagating in the waveguide is obtained by Lumerical. Then the building bloc is represented by assembling the different functional parts to obtain one MZI. The final device is obtained by assembling this building blocs in an array. The knowledge of how the light is processed in all parts of the spectrometer allows us to define the transfer matrix of the device. Different effects are then to be added to this elementary model in order to make it more 'real'. Effective index dependence on wavelength and temperature are implemented via the dispersion and the thermo-optic coefficients respectively. Propagation losses are then considered, and fluctuations in waveguide width are taken into account. Finally the white noise is implemented. The effect of all these parameters on the Free spectral range and on resolution are studied and emphasized.

The device showed in figure 3.1 is used in our simulation, the advantage of this design is that each MZI has its own input which leads to an increased throughput. When the signal is injected, it propagates through the interferometers, the transmittance of each one of them is collected to form the interferogram as a function of the MZI position. the interferogram is actually the product of the input signal by the transfer matrix of the device :

$$I(x_i) = B(\lambda) \times M \tag{3.1}$$

I and B are the output interferogram and the input spectrum respectively, and M is the spectrometer's calibration matrix that will be the subject of interest in this chapter. In the ideal case, the obtained interferogram is equivalent to performing a FT of the input. However, the fabrication errors should be taken in consideration, and the interferogram is then modified. In practice, we have access to the interferogram, from which we deduce the spectrum. Performing a FT of the interferogram is not a good idea because
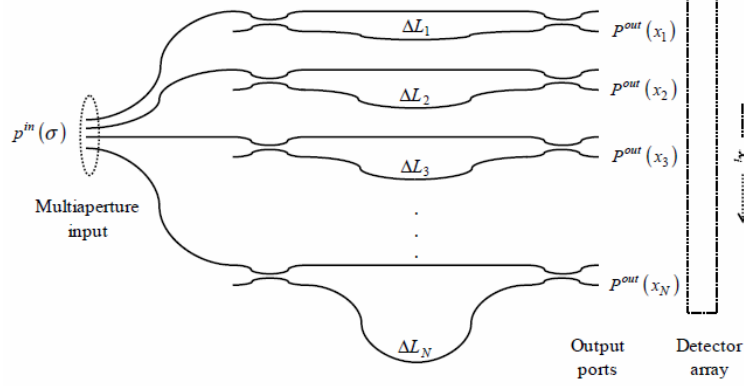
Figure 3.1: The general scheme of the spectrometer, an array of MZIs with an increasing OPD, only one port in the input and one in the output are used for each MZI. Image source: [2]

fabrication errors are not considered, the solution is then to multiply the interferogram by the inverse of the calibration matrix:

$$B(\lambda) = I(x_i) \times M^{-1} \tag{3.2}$$

3.1 Device characteristics

The spectrometer is implemented as an array of 32 MZ interferometers in silicon-on-insulator waveguide (Fig. 3.1). Most of the device characteristics that we have used in our simulation are extracted from [4]. The length of the upper arm is constant in all MZs of the array and is 500μm long, whereas the length of the lower arm increases linearly through the array. That is how we get an increasing optical path difference (OPD) with a constant increment ΔL_0=365μm, while avoiding moving mirrors as in Michelson interferometer. The last MZ has the maximum optical delay $\Delta L_{max} = 1.1315cm$.

The waveguide design with Lumerical Software gives the right effective index and the mode profile, we chose to simulate transverse magnetic TM mode, for which the perpendicular component of the electric field E_y is not concentrated in the side walls of the waveguide. This choice is made in order to minimize propagation losses, because as it will be discussed later, the waveguide width fluctuations arising from lithographic process induces losses. The Fig.3.2 shows the TM mode in a waveguide.

The spectrometer is made up of strip waveguides of silicon $450nm$ large and $260nm$ thick, the surrounding material; the cladding, is silica. The big

contrast of the refractive index between the core and the cladding allows light to be concentrated in the core. For one single MZI,when a constructive
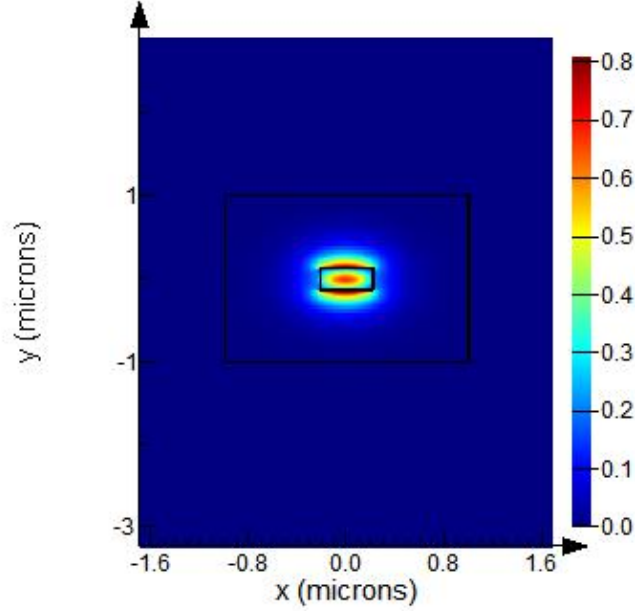

Figure 3.2: The perpendicular component E_y of the TM mode, this mode profile is suitable to minimize propagation errors. The waveguide is made of Si with $450nm$ large and $260nm$ thick surrounded by cladding of silica

interference occurs in one output port, destructive interference occurs in the other. The factor κ describes the coupling of light from one arm to another, the two coupled beams have different phases because of the difference in the optical path. We simulated the 2 output ports of one MZI with an OPD=3.285mm. The effective index of the guided mode is $n_{eff} = 2.002$, errors and propagation losses are not considered here. An input power equals to 1 is injected only in the input port1.

The Fig3.3 shows how the transmittance functions of the two output ports alternates as a function of wavelength, in such a way that the total power is always conserved and equals to 1.

Then we simulated the outputs of two ideal MZI with different optical path delays. For all the upcoming simulations, only the input port1 is used, with an input power $P_{1,i}^{in} = 1$. Similarly for the output, only one port is necessary for the retrieval, so we will focus on the output port2 $P_{2,i}^{out}$

The Fig3.4 shows the transmittance function of the two ideal MZI , we see that the transmittance frequency increases with increasing OPD, this effect
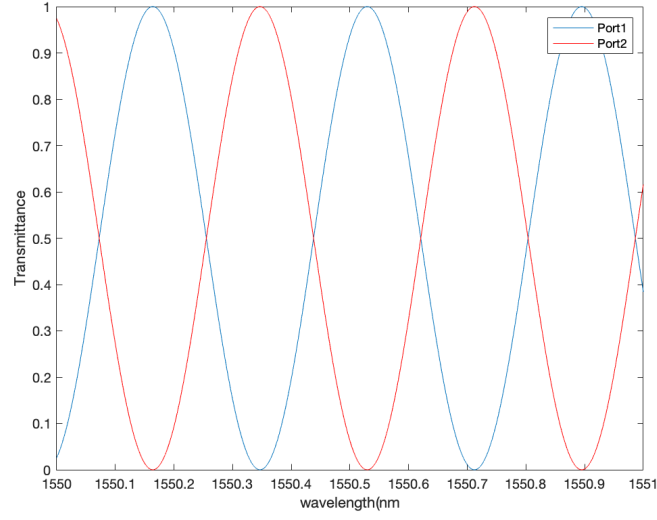
17

Figure 3.3: transmittance functions at the two output ports of one MZI with an OPD=3.285mm, the total power is conserved
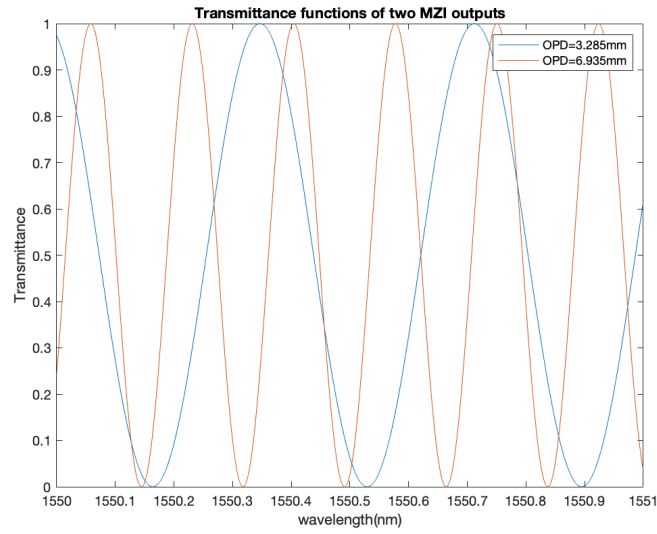

Figure 3.4: transmittance functions of two MZI, transmittance frequency is higher for bigger OPD

can be visualized clearly in the calibration matrix.

3.2 Calibration matrix

When we sweep through wavelengths, each MZI gives a vector of transmittance functions as described in the previous section. We repeat the same operation to all the set of MZIs, and we present the result as N×M dimension matrix, which is the calibration matrix for the whole system. the wavelength range is 4nm long, sampled with a step of 1pm, the matrix dimension is then (4001×32)
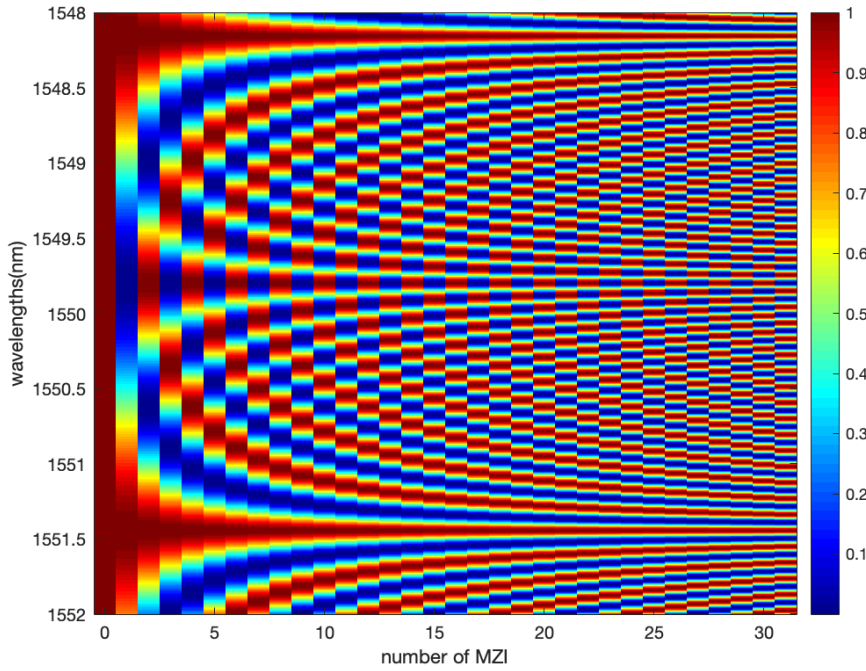

Figure 3.5: Calibration matrix of an ideal array of 32 MZIs, this matrix gives the transmission factor for each MZI at a given wavelength

Several interesting properties can be observed in this calibration matrix. First we notice two particular wavelengths λ_1 =1548.157 nm and λ_2 =1551.444 nm for which the transmittance functions are maximum and constant through all MZIs, this corresponds to two Littrow wavelengths. In this situation, the phase factor is an integer multiple of 2π.

$$\beta \Delta L_i = \frac{2\pi}{\lambda_{1,2}} n_{eff} \cdot i \cdot \Delta L_0$$

19

So for this device, the Littrow condition is observed whenever $\frac{\Delta L_0 \cdot n_{eff}}{\lambda_j}$ is an integer. As consequence of this periodicity, the free spectral range (FSR) of the device is half the separation between two Littrow conditions, in particular, we shall consider an FSR starting from Littrow wavelength:

$$FSR \in \left(\lambda_L, \lambda_L + \frac{\Delta\lambda_L}{2}\right)$$

Second, we can see in the transfer matrix that the transmittance frequency as a function of λ increases with the increasing number of MZI, the last one has the highest frequency. That's why device resolution is increased when using a large number of MZI, but also with bigger OPD increment ΔL_0, according to this formula:

$$\delta\lambda = \frac{\lambda_0^2}{n_g N \Delta L_0} \tag{3.3}$$

where λ_0 is the central operating wavelength of the spectrometer, N the number of MZIs in the spectrometer and n_g is the group index.

3.3   Interferogram calculation

For a given input, the signal propagates in parallel in all MZIs, and an interferofram is obtained at the output. To simulate this, we calculate the product of a vector containing only one none-zero wavelength (monochromatic signal in the wavelength range of the device) times the calibration matrix, the result is a vector which encompasses the transmittance function of each MZI.

The Fig. 3.6 displays the interferogram for a given wavelength. Far from the Littrow condition λ_L, the sinusoidal shape of the interferogram is obtained. The interferogram spatial frequency depends strongly on the input wavelength, it is constant with zero frequency at Littrow, then, fringe patterns start to appear as long as the input wavelength shifts from Littrow condition in both sides. The interferogram's highest frequency is reached when shifting equals the FSR, which is the half period between two Littrow wavelengths $FSR = \frac{\lambda_{L2} - \lambda_{L1}}{2}$

The Fig 3.7 shows some important properties of the interferogram. First, at Littrow wavelength ($\lambda_L = 1551.444nm$) the transmittance function is constant through all MZIs. Second, when wavelength increases, spacial fringes start to appear : one pattern at $\lambda_1 = 1551.544nm$, and two fringe patterns at $\lambda_2 = 1551.644nm$. Third, two symmetric wavelengths with respect to λ_L provide the same interferogram. In Fig 3.7 the wavelength $\lambda_3 = 1551.344nm$ and λ_1 are symmetrical, so their interferograms are identical. The same discussion is true for the wavelength $\lambda_4 = 1551.244nm$ and its symmetrical λ_2.
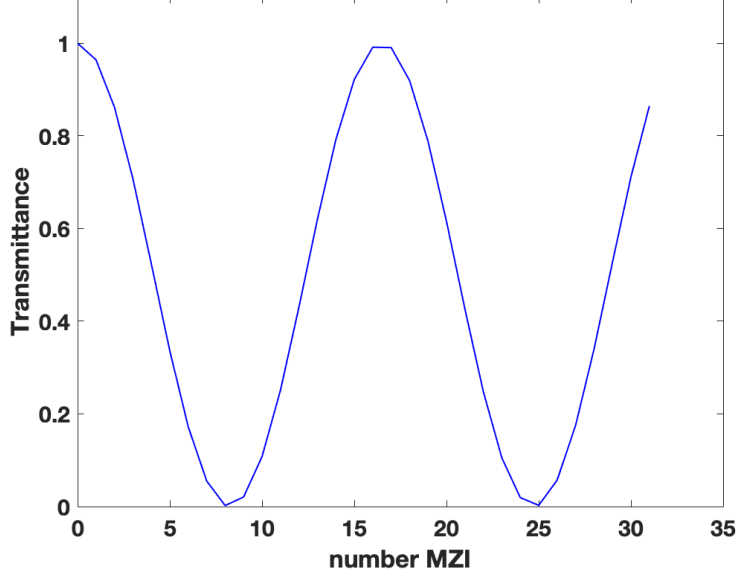
Figure 3.6: The interferogram at the output of the spectrometer for the wavelength 1551.25

As a result of this symmetry, we can restrict the study of the device only on the FSR range.

3.4   Spectral retrieval

In this section, we use the inverse of the calibration matrix to retrieve the spectrum. For the moment, the two methods; calibration matrix and Fourier transform are equivalent, but when errors and environmental conditions will be considered, the choice of using the calibration matrix becomes relevant, because the errors are inverted and eliminated when calculating the spectrum, and the exact input frequencies are retrieved back.

We have an interferogram obtained at Littrow wavelength, but in practice, the input wavelength is unknown. To retrieve it, the idea is to multiply the interferogram by the inverse of the calibration matrix. In most cases, the matrix is not square, so we use the pseudo-inverse matrix instead.

The figure 3.8a provides the calculated spectrum obtained by multiplying the interferogram at Littrow condition by the inverse of the calibration matrix. We retrieve three peaks in this wavelength range of $7nm$. The central peak corresponds to the actual input λ_L. The tow other ones at both sides
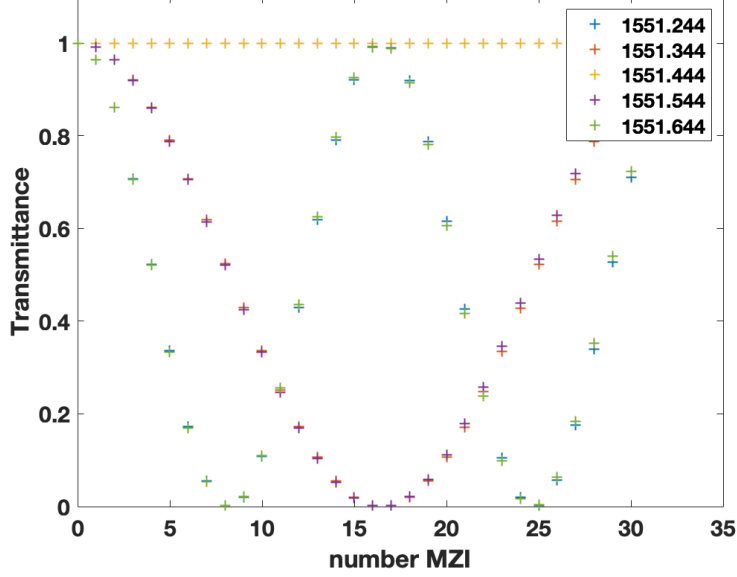
Figure 3.7: the interferogram of two wavelengths and their symmetrical ones with respect to Littrow wavelength. The interferogram of the latter is also given

correspond to annother Littrow condition each.

This periodicity can be explained as flows: The interferogram is a sinusoidal function, but it is a discrete function because of sampling, this can be seen as a comb of delta functions multiplied by sinus . The Fourier transform of that is two symmetric delta functions (FT of sinus), but repeated several times, because the FT of delta comb (sampling) is also a delta comb.

If the input wavelength was different from λ_L, the spectral retrieval would be two equidistant peaks at both sides of each Littrow wavelength, this effect can be observed in Fig.3.8b. From Fig.3.8a the measured periodicity of the Littrow conditions is $3.301nm$, from which we can deduce the free spectral range of the device: $FSR = \frac{3.301}{2} = 1.650nm$. We compare this value with the theoretical one given by :

$$FSR = \delta\lambda\frac{N}{2} = \frac{\lambda_0^2}{2n_g\Delta L_0} \tag{3.4}$$

By replacing the resolution at the central wavelength $1552nm$ by its expression from 3.3 , we get the FSR of the device : $FSR = \frac{1552^2}{2\times2.002\times365\cdot10^3} = 1.648nm$, which is in a good agreement with the measured FSR by simulation, in the limit of some approximations.
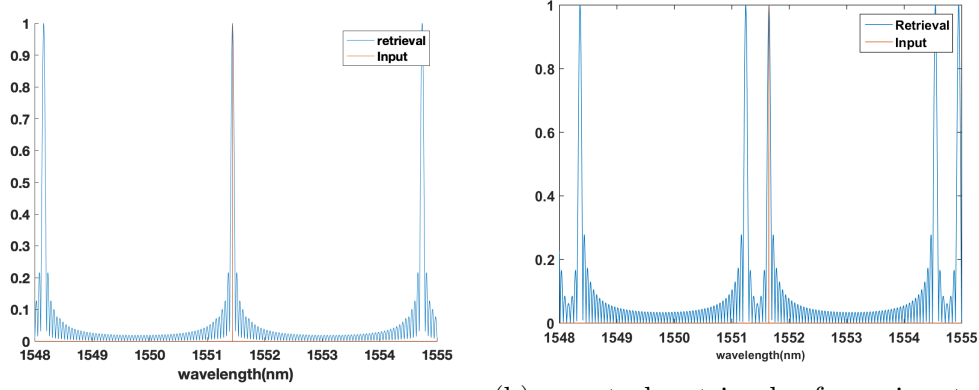
(a) spectral retrieval of an input at λ_L. Periodic repetition due to FT of delta comb (sampling)

(b) spectral retrieval of an input at λ_2. Aside of sapling effect (periodic retrieval), symmetric retrieval with respect to λ_L due to FT of sinus

Figure 3.8: (a): The spectral retrieval of an interferogram given at Littrow condition, the periodic retrieval of the Littrow wavelength is demonstrated, the monochromatic input is plot in red. The ripples are due to the square window of the interferogram for which the FT is a sinc function. (b): A monochromatic input wavelength λ_2 different from λ_L resulting in a periodic retrieved spectrum. λ_2 and its symmetric with respect to λ_L is retrieved. This symmetry is repeated around each Littrow wavelength

3.5   Resolution and FSR

We study in this section how does the number of MZI and the length of the OPD affect the resolution and the FSR. The effect of apodization which is used to eliminate ripples resulting from truncation is also studied.

We recall from the equation 3.4 that the FSR depends on the OPD and not in the number of MZI. Let's visualize that with our simulations. We inject a monochromatic input at Littrow wavelength, get the interferogram, and then use the inverse matrix method to retrieve the spectrum.

The Fig 3.9 shows the retrieved spectrum for two devices with 32 and 100 MZI, the Littrow wavelength is the same for both devices. As expected, the FSR didn't change between the two devices. However, we notice that ripples at the spectrum of 100 MZ device are reduced, that's because the truncation window is bigger and its FT is a narrowed sinc function.

We want now to see the effect of the incremental OPD length in the FSR, we set the OPD of the device at $\Delta L_0 = 400 \mu m$ in a device containing 32MZI and compare it with the last result.
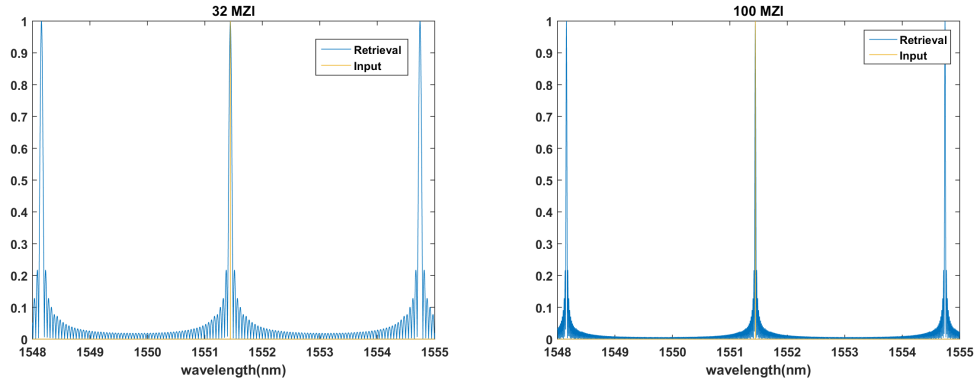
(a) spectral retrieval of an input at λ_L for a 32 MZI device. $FSR = 1.650nm$

(b) spectral retrieval of an input at λ_L for a 100 MZI device. $FSR = 1.650nm$

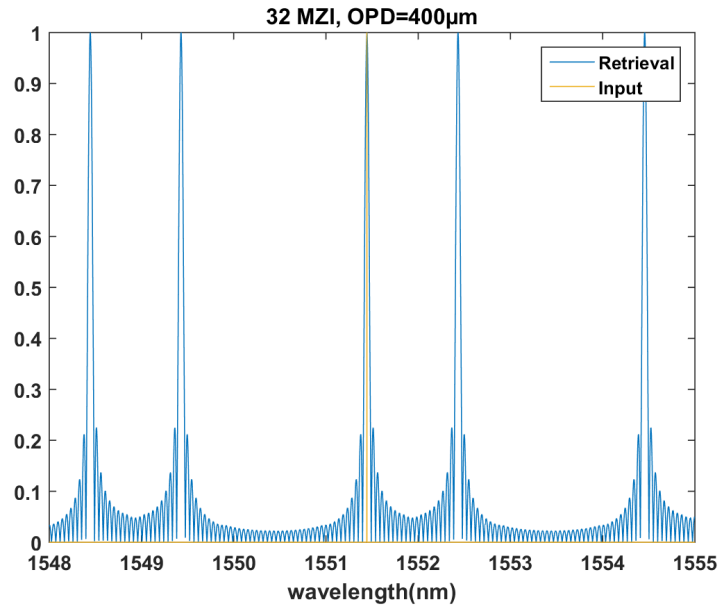Figure 3.9: The number of MZI in the device has no effect in the FSR

Figure 3.10: The FSR of a 32 MZI device with an OPD of 400 μm, $FSR = 1.5\mu m$

The first thing we notice about the retrieval of a device with an increased OPD is the shifting of the Littrow wavelength, this can be noticed in the figure 3.10 by two symmetrical retrieval of a monochromatic input with respect to the new Littrow wavelength. Compared to the device with $365\mu m$

OPD, the FSR is reduced from $1.65nm$ to $1.5nm$. The FSR is then inversely proportional to the OPD, which confirms the equation 3.4

The apodization function that we are going to use is a Gaussian function for which we control the width via the standard deviation σ, the convolution of this function with the retrieved spectrum eliminates the ripples.
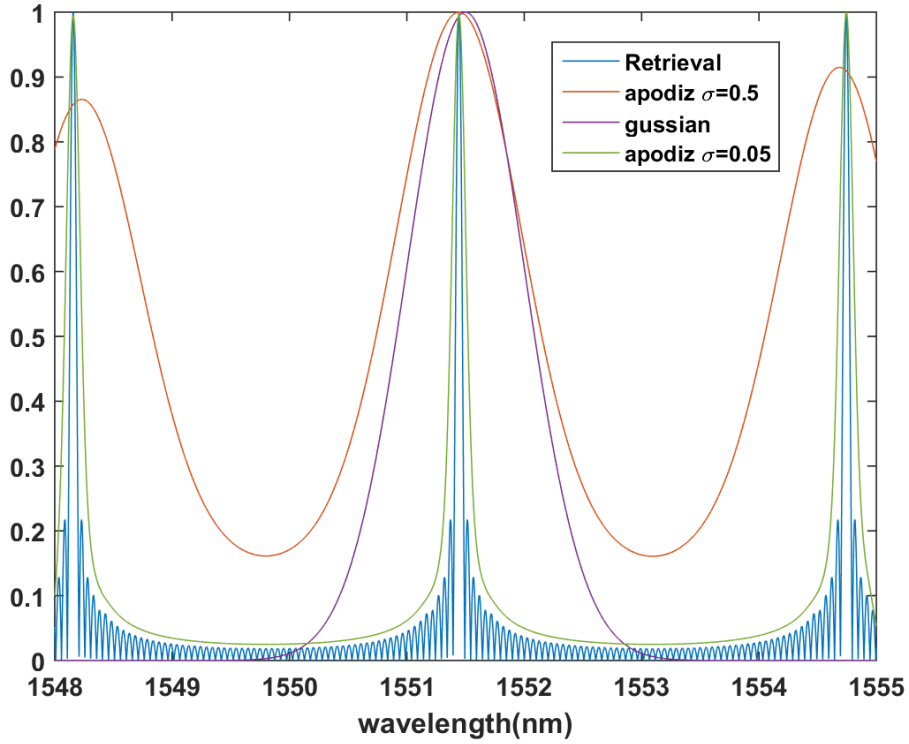

Figure 3.11: The effect of the apodization on the FSR, in Blue the actual retrieval with ripples, in green and red the retrieval with apodization with increasing Gaussian width respectively. The gaussian with $\sigma = 0.5$ is represented in purple

From the figure 3.11 we see that the width of the Gaussian with which we perform convolution should be enough to eliminate ripples, but not so big that it may over-smooth and omit some signal features.

The next step is to look to the resolution, how is it affected by N (number of MZI) and ΔL_0. I want to emphasize here that we are only looking to this effects considering a simple model, where the effective index is independent on temperature, wavelength, fabrication errors and noise, This effects are to be added later on.

We established a script in Matlab that calculates the resolution. This script works as follows: two input wavelengths from the same FSR are injected. As an initialization, the two inputs are identical (same wavelength), the interferogram is then obtained and the retrieval is calculated, consequently only one peak is retrieved. Then one input is conserved at the same position and the other is shifted by a small step. This process is repeated by iteration, and the loop stops when the overlap at the mid-separation between the two peaks equals half the maximum of the peaks. The corresponding separation is considered as device resolution.
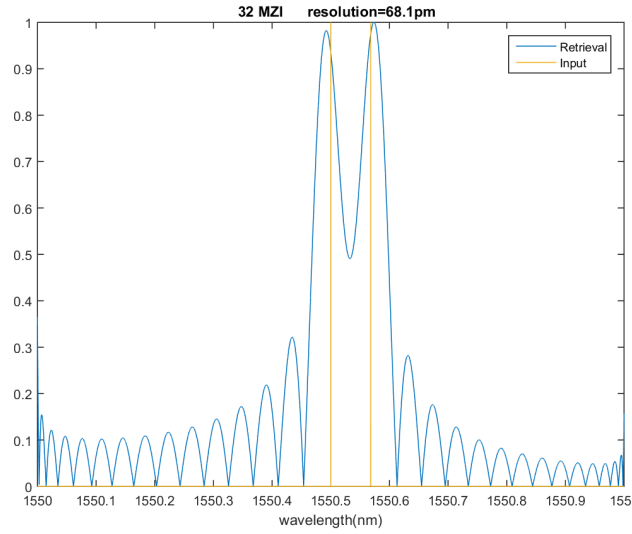

Figure 3.12: The resolution calculated for a device of 32 MZI and $OPD = 365\mu m$, the resolution is the separation between the two peaks for which the overlap is inferior to the peak's half maximum

The fig 3.12 and 3.13 show the resolution ($68.1pm$ and $24.5pm$) calculated for two devices with 32 and 100 MZI respectively, both of them having an incremental $OPD = 365\mu m$. The resolution is enhanced by increasing the number of MZI.

The effect of apodization in resolution is shown in Fig.3.14, the oversmoothing due to convolution with a wide Gaussian makes our algorithm under-estimate the resolution, it shows a resolution increased by $2.8nm$ compared to the real value.
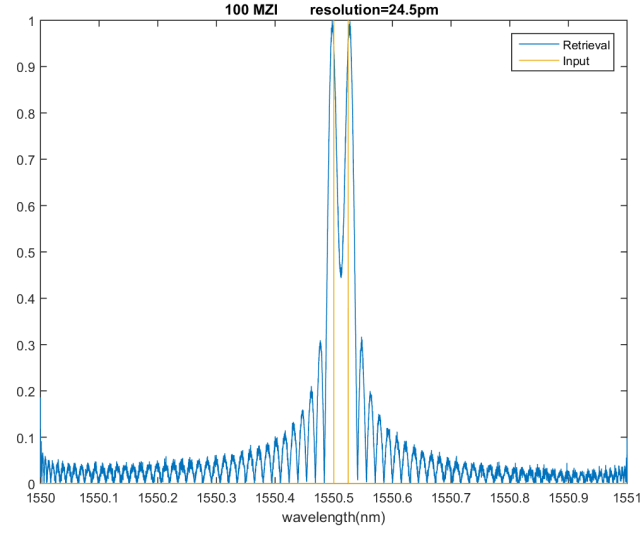
26

Figure 3.13: The resolution calculated for a device of 100 MZI and $OPD = 365\mu m$, the resolution is the separation between the two peaks for which the overlap is inferior to the peak's half maximum
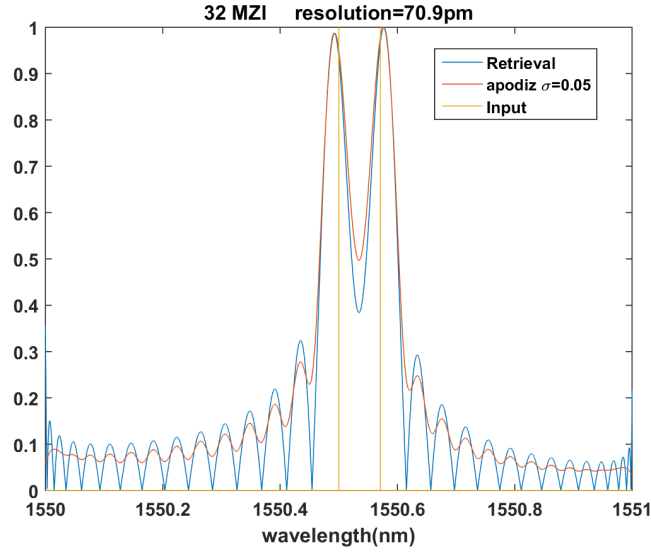

Figure 3.14: The resolution calculated for a device of 32 MZI and $OPD = 365\mu m$, in red the signal after apodization, and the estimated resolution is the separation between the two peaks

3.6   wavelength dependence

Up to now, we have considered a constant effective index of the mode propagating inside the waveguides. But this is not a rigorous assumption. The effective refractive index n_{eff} is a dimensionless number quantifying the phase delay per unit length in a waveguide, relative to the phase delay in vacuum. The effective index of the mode in the waveguide plays the same role as the refractive index n for the material: the wavenumber in the mode (or in the material) is n_{eff} (or n) times (respectively) higher than it would be in vacuum. Obviously, the effective index is not just a material property, but depends on the whole waveguide design. We study here the effect of effective index dependence on the wavelength at $300K$. The effective index depends also on temperature, this effect will be studied later.

The Si core of waveguides in our device has a width of $450nm$, and a thickness of $260nm$ ensuring monomode operation. Now that the design is fixed, we want to see how the effective index changes with the wavelength. For that we used the mode solver to calculate n_{eff} for several wavelengths by assigning the corresponding refractive index of Si at different wavelengths around $1.55\mu m$, taken from the reference [8]

$$n_{eff} = -1.6694\lambda + 4.5888 \tag{3.5}$$

we find a linear relation between n_{eff} and wavelength, from which we extract the effective dispersion coefficient $\frac{dn_{eff}}{d\lambda} = -1.6694\mu m^{-1}$

The figure 3.16a shows the resolution when the effective index depends on the wavelength. In order to compare to the last results, we simulated a device having 32 MZI and $OPD = 365\mu m$. The two peaks and their symmetrical ones with respect to Littrow wavelength are retrieved. Compared to the resolution(Fig.3.12) when effective index is constant, the resolution is reduced from $68.8pm$ to $36.5pm$.

The FSR is also reduced (Fig.3.16b)when taking the wavelength dependence into account, compared to the constant n_{eff} model (Fig.3.9a) . The effective index is reduced from $1.65nm$ to $0.717nm$. We recall that the FSR is the half distance between two Littrow wavelengths.

3.7   Fabrication errors

Another source of variation in the effective index arises from some fabrication imperfections. Actually, lithographic process is never perfect. For instance, the waveguide width varies from one waveguide to annother, but also along one waveguide. The direct consequence of this is fluctuations in the effective
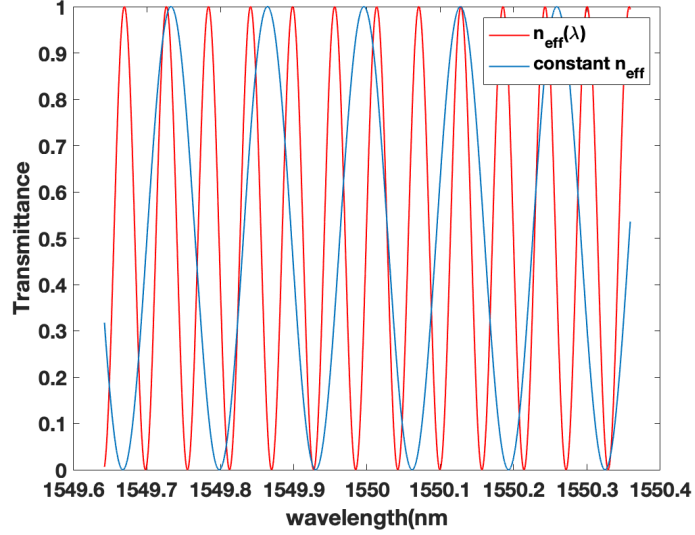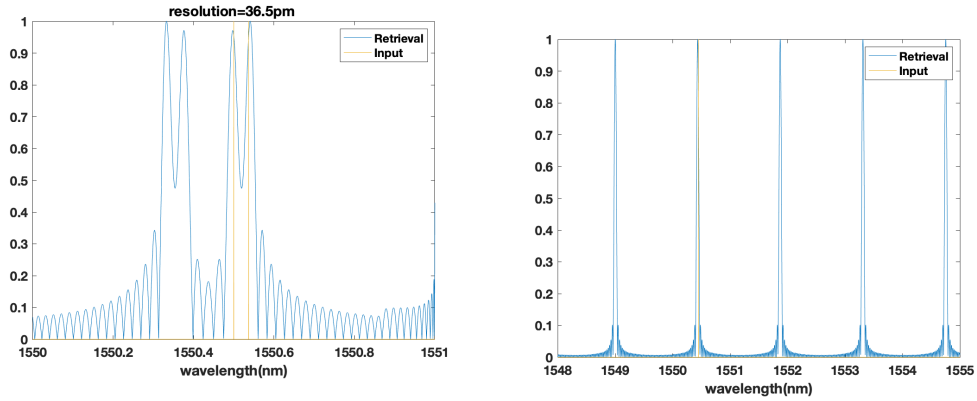
Figure 3.15: Effect of $n_{eff}(\lambda)$ on transmission function simulated at the output of one MZI. The transmittance is shifted compared to a constant n_{eff}t model

(a) resolution for 2 inputs, symmetry with respect to Littrow wavelength

(b) Free spectral range

Figure 3.16: The resolution and the FSR when the effective index depends on wavelength

index since it's sensitive to the device structure. To model this part, we found using the mode solver that the effective index changes by $\pm 2.26 \cdot 10^{-2}$ when the waveguide width varies by $\pm 20nm$ around the initial width of $450nm$. Random value belonging to the specified range is to be added to the effective

index of each MZI in order to simulate the effect of width fluctuation, the corresponding calibration matrix is shown in Fig.3.17
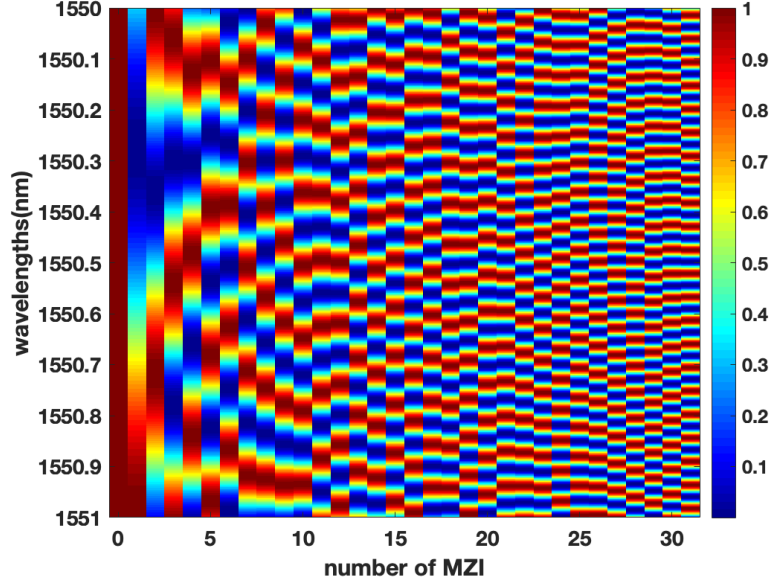

Figure 3.17: Calibration matrix with fabrication errors

Fluctuations in fabricated waveguide properties, particularly the waveguide width, produce fluctuations of the waveguide effective index, resulting in random phase errors in the MZI transmittance functions. Therefore, the phase alignment condition of the Littrow wave- number and orthogonality of the FT transformation base are not guaranteed. The phase of the transmission functions are effected, and the consequence of this is the distortion of the interferogram. The lines are misaligned in phase and a Littrow wavelength λ_L is shifted. This is a consequence of non orthogonality of the base of the cosine transform in the presence of phase errors.

The Fig. 3.18 shows the cosine interferogram in a perfect case, and a distorted interferogram when fabrication errors are added. The Fourier transform of the distorted interferogram is made of many frequencies, whereas the actual input is only one. In the other hand, The product of this with the inverted matrix eliminates the errors and retrieve the monochromatic input. Those are mainly the errors which prevent retrieving the exact spectrum by Fourier transform, one solution is proposed, but is requires phase correction circuits, and multiple measurement at varying temperatures.
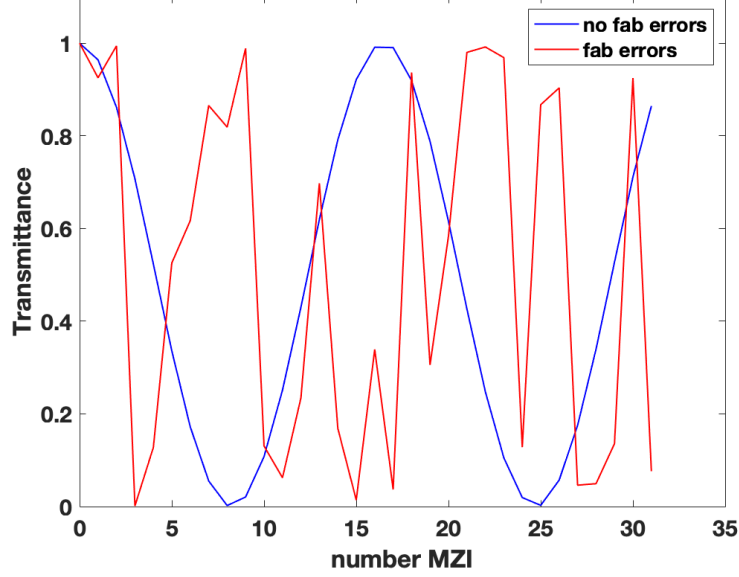
Figure 3.18: In red, the distorted interferogram compared to the theoretical cosine in blue responding to the same input when no phase errors are introduced (to visualize only the effect of imperfections, no dependence on T and λ are considered)

3.8 Effect of white noise

In order to implement white noise, Random numbers following normal distribution are added to each transmittance function on the transfer matrix. The amplitude of noise is controlled by the standard deviation σ of the normal distribution, so that 95% of the random numbers are contained in $[-2\sigma; 2\sigma]$. As a result, the interferogram is more distorted, so that retrieving the spectrum becomes more challenging.

The figure 3.19 shows the output of one MZI with and without noise. The standard deviation σ (noise amplitude) is given as a percentage rate of 1 (input power amplitude). The effect of noise is observed as fluctuations added to the cosine output. However, the noise does not change the phase and it does not introduce shifting on wavelength. Precision: In order to see only the effect of noise, no losses nor fabrication errors are taken into account here.

Let's now see the effect of noise on resolution. Since the interferogram is no longer sinusoidal when noise is added, the retrieved spectrum contains high frequency fluctuations. The Gaussian function used for apodization
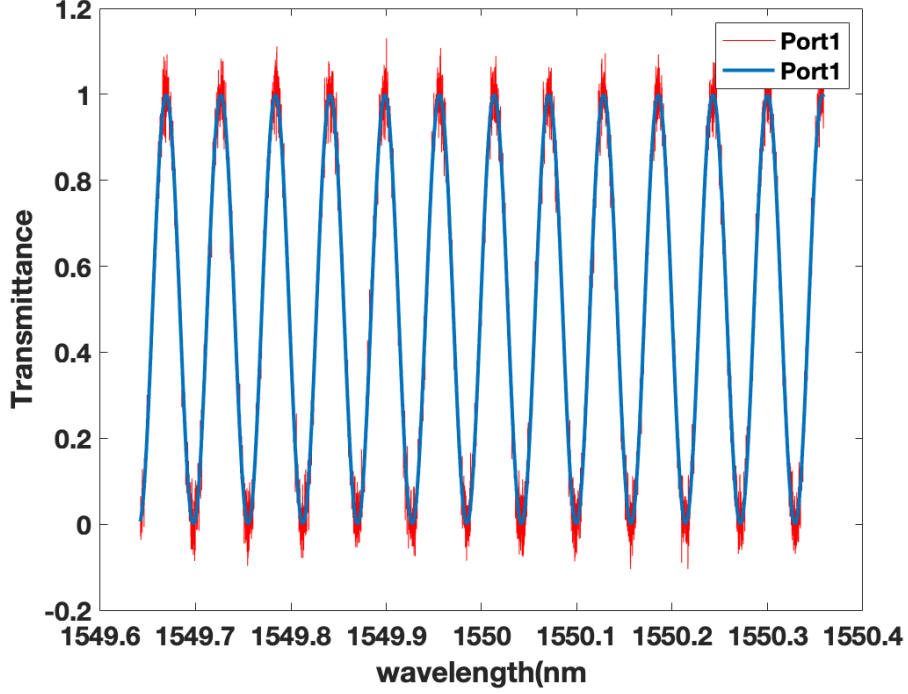
31

Figure 3.19: The output of the 26th MZI as function of the wavelength, In blue the output when no noise is added, in red 4% noise is included

eliminate ripples coming from truncation and smooth the retrieved signal. The resolution is increased from $68.1 pm$ for the same device without noise, to $44.3 pm$ when noise is added (Fig. 3.20).

3.9 Effect of Losses in presence of noise

Another aspect was neglected so far is the effect of losses. They may have different sources, but we focus on propagation losses. They become important with long distance propagation. A loss term is added to our model via the linear attenuation coefficient $\alpha(nm^{-1})$. Usually, losses are given in (dB/cm), so a simple conversion is needed:

Attenuation in dB in term of field quantity is given by:
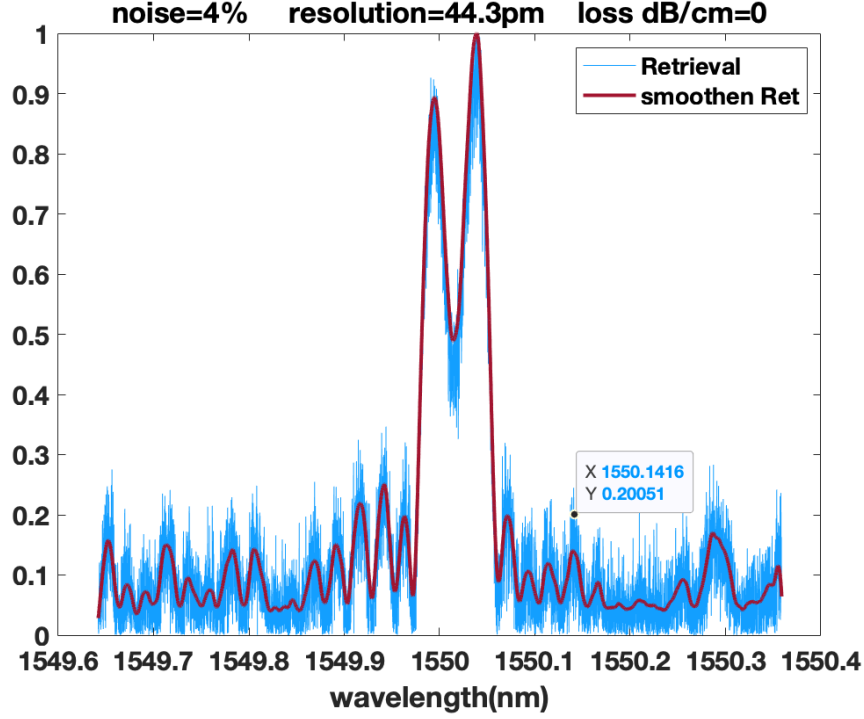
$$20 \cdot log(e^{\alpha L})$$

Figure 3.20: The noisy spectrum retrieval due to white noise in blue, the smoothen retrieval in red, obtained by a convolution with a Gaussian function

which can be written by unit of length (dB/cm) as:

$$loss = \frac{20}{ln(10)} \cdot \alpha$$

finally, the linear attenuation coefficient can be written:

$$\alpha(nm^{-1}) = \frac{ln(10)}{20} \cdot 10^{-7} * loss$$

Interferogram visibility variations are produced by uneven propagation losses in waveguides across the array. As the waveguide loss imbalance progressively increases with optical path difference between the MZI arms across the array, interferogram visibility is correspondingly reduced.

The Fig. 3.22 shows the evolution of the resolution as a function of the propagation loss and white noise, we observe that when no noise is present, the resolution doesn't depend on losses. But when noise is added, the resolution is degraded. It increases exponentially with loss. We note that for
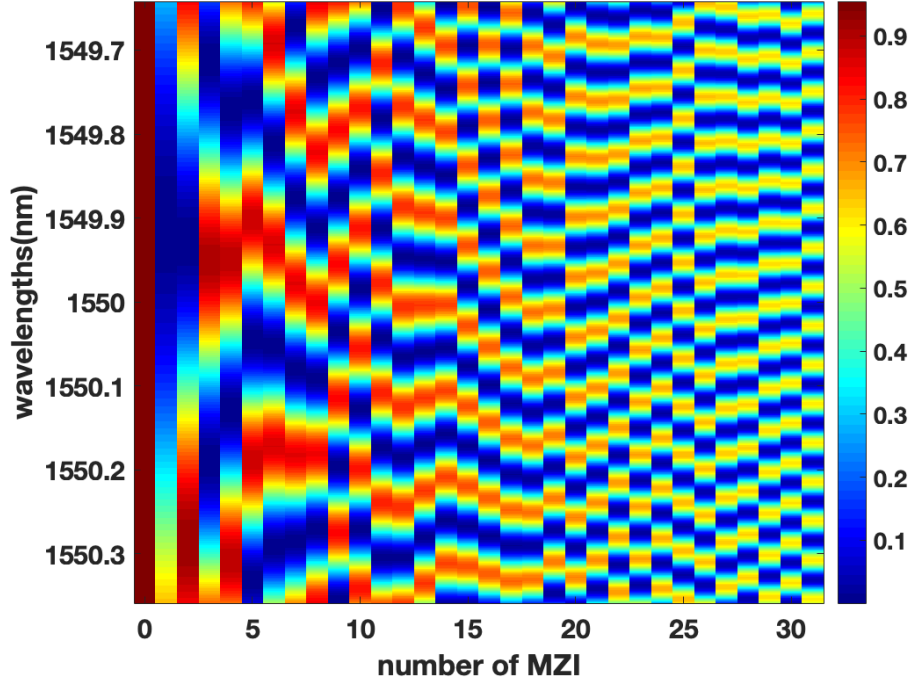
Figure 3.21: Calibration matrix with 4dB/cm losses. Two main effects causing deviations from the theoretical performance. First, waveguide insertion loss increases with the path difference of the interferometers, resulting in a reduction of the visibility of the MZI transmittance. Second, horizontal misalignment because of random phase fluctuation resulting from fabrication errors

bigger losses, the output power becomes small, so the amount of fluctuation which is to be added in order to simulate the white noise should not exceed this small output power.

3.10 Effect of temperature variation

We discussed earlier the dependence of the effective index on wavelength and we implemented that dependence in our model. In this section, we first see the dependence of n_{eff} on temperature, then we see the effect of temperature variation on the retrieval and on resolution.

Each material has its own thermo-optic coefficient, that is, the refractive index changes with temperature. Consequently, the effective index of the
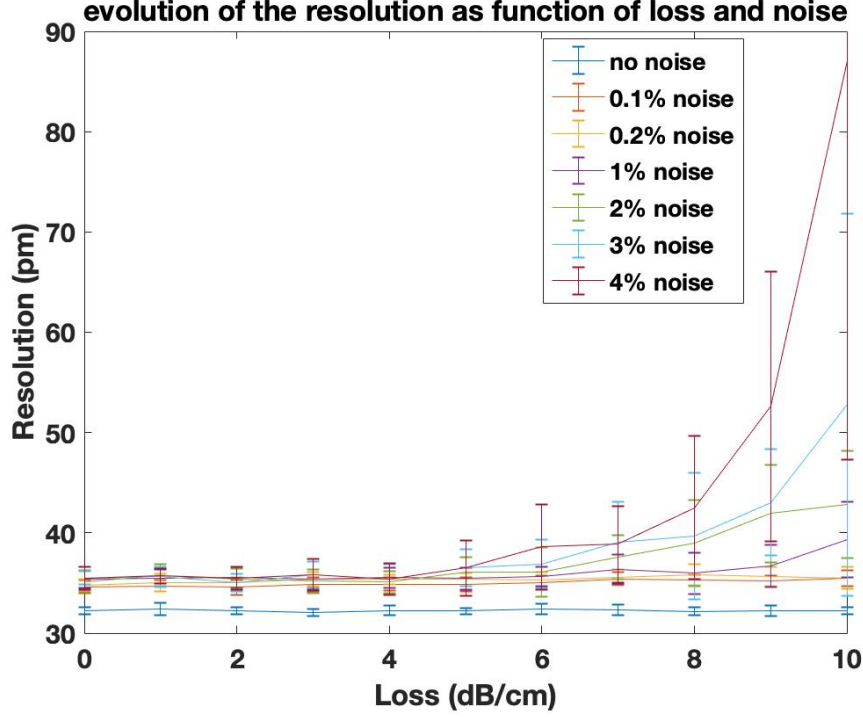
Figure 3.22: The evolution of the resolution as a function of losses, for different values of noise, the resolution was calculated for the central wavelength $1550nm$ in the FSR of the device

propagating mode changes as well. So in order to consider this dependence we used mode solver of Lumerical to calculate the effective index at several temperatures $(\lambda = 1550nm)$, by assigning the corresponding refractive index. A second order polynomial fitting of n_{eff} with respect to temperature is a good approximation:

$$n_{eff} = 1.5564 \cdot 10^{-7} \times T^2 + 7.559 \cdot 10^{-5} \times T + 1.9657$$

This study allowed us to propose the complete relation describing how the effective index varies as a function of temperature and wavelength, which will be used for the rest of our simulation:

$$n_{eff} = 1.5564 \cdot 10^{-7} \cdot T^2 + 7.559 \cdot 10^{-5} \cdot T - 1.6694 \cdot 10^{-3} \cdot \lambda + 4.55327 \quad (3.6)$$

The Fig.3.23a and Fig. 3.23b show a comparison between the obtained model and the measured values of n_{eff} as function of temperature and wavelength,

for the range of interest in temperature and wavelength.

(a) effective index variation with wavelength, model compared to actual values of n_{eff}

(b) effective index variation with temperature, model compared to actual values of n_{eff}
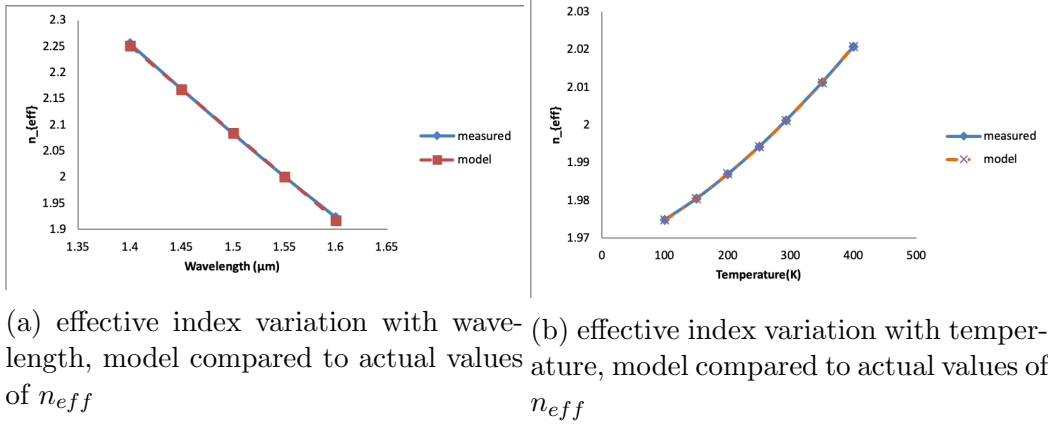
Figure 3.23: model verification, $n_{eff}(T, \lambda)$

In practice, interferogram measurement and the retrieval operation can be done in different moments, the calibration matrix is susceptible to some variations between measurement and retrieval. This effect is integrated in our script by using two different matrices, so that we can control temperature and noise, for interferogram calculation and spectrum retrieval separately.

An example at the absence of noise and losses is shown in the Fig. 3.24 where the interferogram is obtained at $300K$, and the retrieval operation is effectuated at $302K$, then at $304K$ this increasing in temperature induces shifting of the retrieval of a monochromatic input.
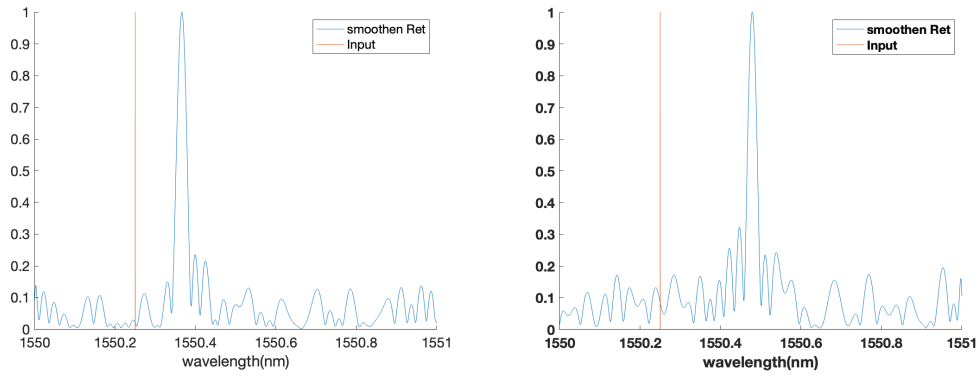
(a) variation of 2K between calibration and retrieval

(b) variation of 4K between calibration and retrieval

Figure 3.24: (a): The monochromatic input is represented in red, is is injected at 300K, the retrieval is done at 302K in (a) and 304K in (b)

Moreover, there is a linear relation linking the temperature variation to the wavelength shifting, which allows us to retrieve the exact injected wavelength by a simple shifting of the wavelength knowing the temperature difference between the moment of the interferogram measurement and the moment of spectrum calculation. The Fig. 3.25 displays the aforementioned relation.
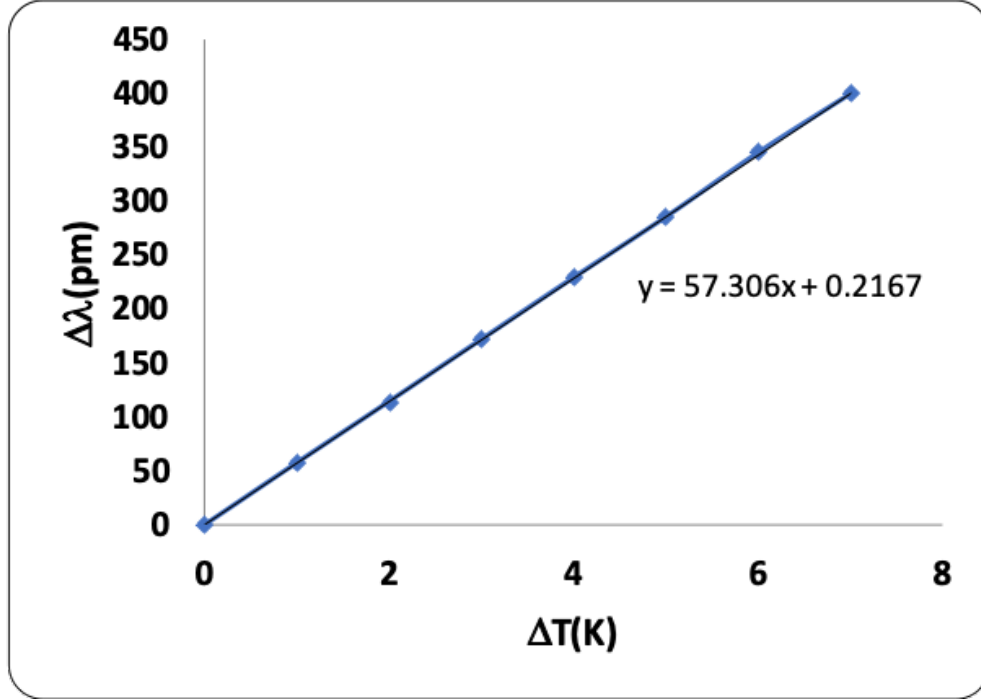

Figure 3.25: spectrum retrieval shifting as a function of temperature variation

Now that we know the effect of temperature on a single peak of the spectrum, we can see what happens for the resolution as function of temperature, for that we inject a double wavelength input, multiply it by the first matrix to get the interferogram at $300K$, then use the inverse of a second matrix to get the spectrum at a different temperature, similarly to what we have done previously, except that we make it more realistic now by adding white noise to both matrices. The noise fluctuates between -10% and 10% of the input power. We assume that the mode propagates with zero loss. The fabrication errors are kept unchanged between the measurement and calibration matrices because we are dealing with the same device. The retrieved signal should be filtered in order to smooth the signal and reduce noise.

The simulation shows that both peaks are shifted by the same value. As in the case of monochromatic spectrum, when temperature does not change,
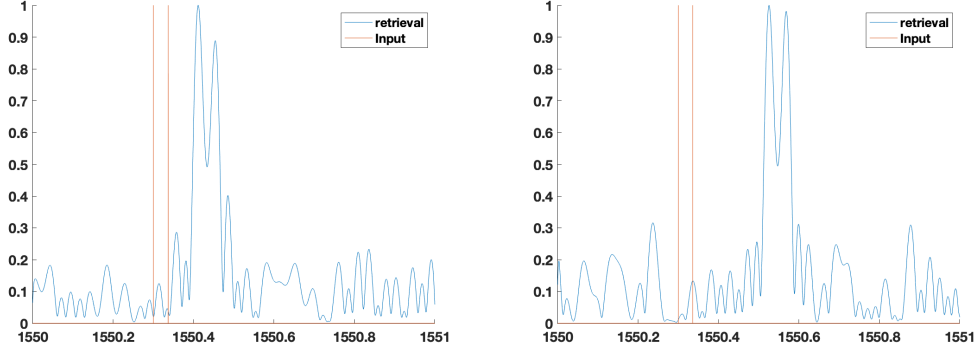
(a) variation of 2K between calibration and retrieval for two wavelengths input

(b) variation of 4K between calibration and retrieval for two wavelengths input

Figure 3.26: The two retrieved peaks shift linearly by the same amount,

we retrieve the two inputs. Otherwise, wavelength shifting of each peak is linearly dependent on temperature variation. Since the two retrieved inputs shift by the same amount, the resolution does not vary, so it's independent on temperature.

Since the shifting of the two peaks with temperature is identical and the resolution is conserved, we want to see what is the maximum temperature shifting for which a single peak shifting is smaller than resolution. we extracted that from the linear relation between temperature and wavelength (Fig.3.25). We find that wavelength shifting equals the resolution of $42pm$ corresponds to temperature shifting of $720mK$. That means that under this value, input and retrieval are considered the same because shifting is smaller than resolution.

As we can see, calibration matrix is a good tool to model the device, where we can include all device imperfections and still get the spectrum corresponding to the wright input. However, it requires a precise knowledge of all device parameters, including temperature. In the next chapter we see another model based on Machine Learning where temperature is no longer a source of restriction.

Chapter 4

Spectral prediction model

Machine Learning offers a wide range of algorithms that can predict the spectrum without specifying the exact device parameters such as temperature in our case. It is based on training, in other words, the algorithm learns how the device performs in different conditions, so that it can predict when asked in a given situation. The accuracy depends on the algorithm chosen for prediction, but they all perform generally very well. We are going to focus in this chapter in the so called k-nearest neighbors algorithm and see whether it can predict the wright input just by seeing the interferogram and without assigning the temperature. This method is an alternative of retrieving the spectrum with the inverse of calibration matrix which requires temperature knowledge.

We consider four inputs (Fig.4.1) , three absorption peaks $Pin1, Pin2, Pin3$ centred at $1549.4nm$ $1549.7nm$ and $1550nm$ respectively and one constant input with no peak $Pin4$. The four inputs are contained in the same FSR range of the device ($717pm$). Each input is called a class in our classifier algorithm. By presenting the interferogram of one of them at certain temperature, it can predict which class (input) it was.

To prepare our training data, we simulate the output of each MZI as function of temperature when each of the four inputs is injected. The operation is done for a range of temperatures between $300K$ and $305K$ with a step of $0.1K$. we use the device with $\Delta L_0 = 365\mu m$ and 32 MZI and noise is set at 2%. As a result, we get for each of the four inputs an $N \times D$ dimension matrix of interferograms at N different temperatures and D number of MZI.

The output of each MZI as function of temperature for each input has a periodic form, so we use a fitting with Fourier series:

$$f(T) = a_0 + a_1 cos(\omega T) + b_1 sin(\omega T) \tag{4.1}$$

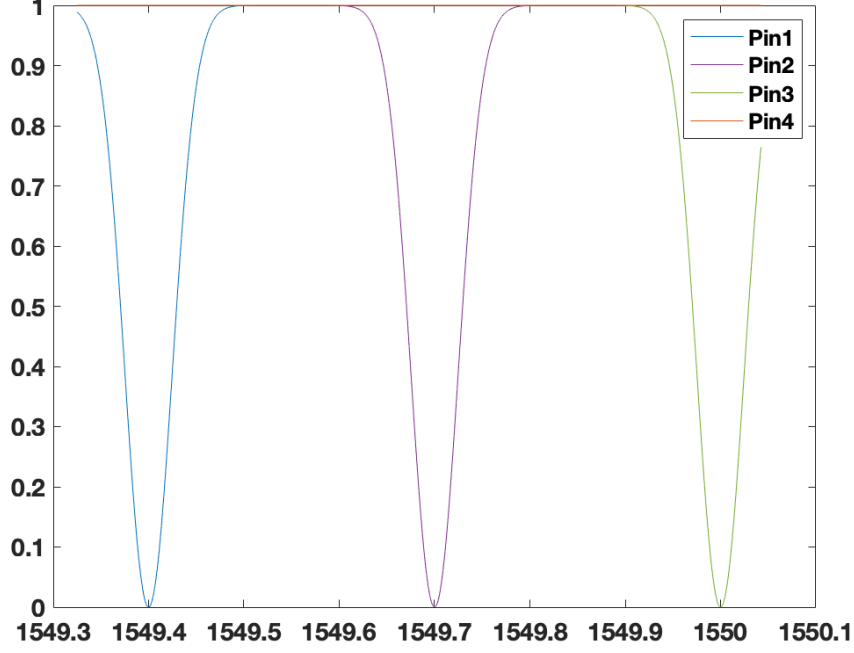The fitted data will be considered as training data, as an example, fitting

Figure 4.1: The four inputs of interest

coefficients for the 20th MZI having as input one of the four classes are given in Tab.4:

| fitting coefficients $(\times 10^3)$				
Input	a_0	a_1	b_1	ω
Input1	3.2757	-0.0568	0.1148	0.0048
Input2	3.2744	-0.1293	0.0247	0.0048
Input3	3.2889	-0.1077	-0.0726	0.0048
Input4	3.5883	-0.1020	0.0637	0.0048

The fitted output matrices are concatenated vertically to create the training data, the dimension of the resulting matrix is 204×32 , each line is an interferogram of one of the four classes at given temperature. We introduce the notion of feature and label, the feature is any input variable used in making predictions, in our case, it is each point in the interferogram at certain temperature. A label is the "answer" or "result" portion of an example [20]. Each class has its own label, so we have four labels in our example. We can
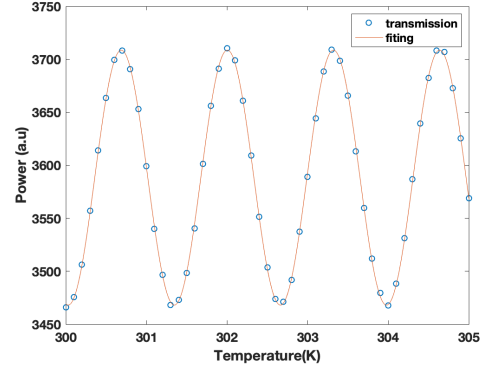
40
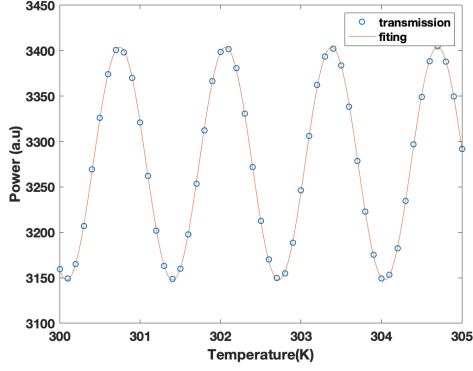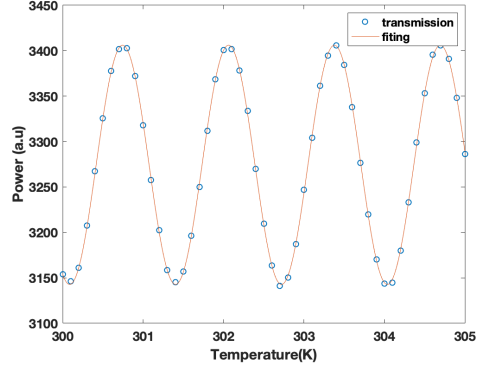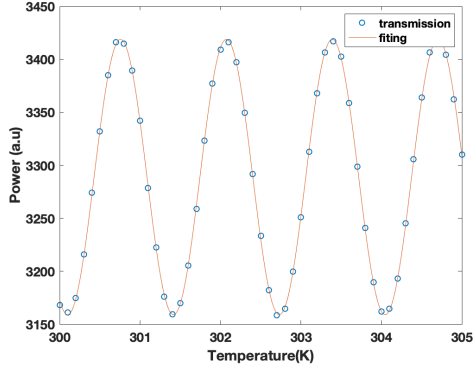
Figure 4.2: output transmissions of the 20th MZI as function of temperature for each of the injected inputs. Fitting with with Fourier series

test a prediction algorithm as follows: training data with their labels are given to the algorithm beforehand, when asking for prediction, we present features that are never seen by the model, they will be proceeded depending on the used algorithm, then it gives the predicted labels that we compare to the real ones (testing labels) to get the accuracy of the model.

K-Nearest Neighbors algorithm (or KNN) algorithm is one of the simplest classification algorithm and it is one of the most used learning algorithms [18]. KNN is a non-parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point.

A non-parametric technique means that it does not make any assumptions on the underlying data distribution. In other words, the model structure is

determined from the data. Therefore, KNN is one of the best classifiers when there is little or no prior knowledge about the distribution data.

KNN is also a lazy algorithm (as opposed to an eager algorithm). That means that it does not use the training data points to do any generalization. In other words, there is no explicit training phase [19]. Lack of generalization means that KNN keeps all the training data. To be more exact, all the training data are needed during the testing phase. KNN Algorithm is based on feature similarity: How closely out-of-sample features resemble our training set determines how we classify a given data point.
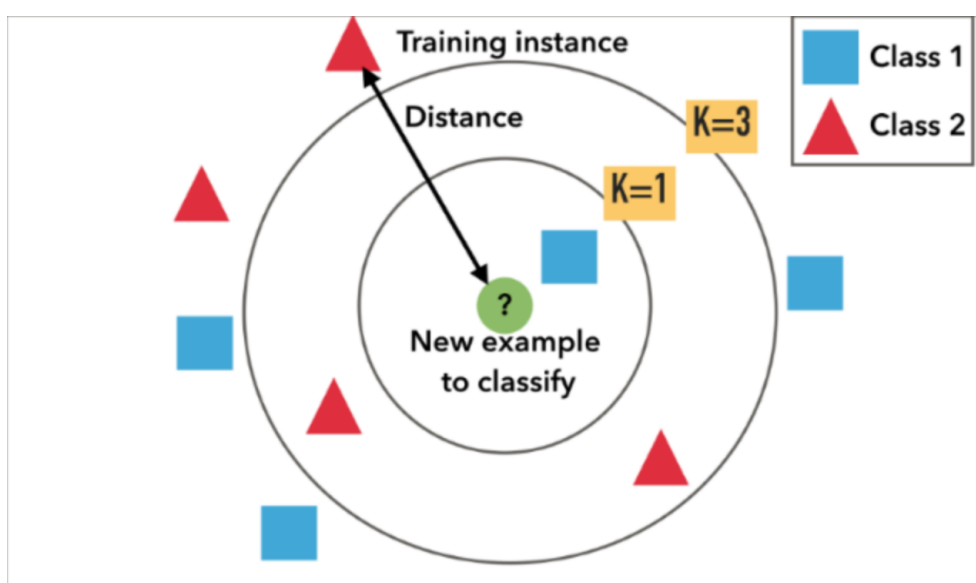

Figure 4.3: classification depends o the number of neighbors which are considered. Class 1 is predicted for k=1, class 2 is predicted for k=3. Image source: [18]

the Fig. 4.3 shows an example of classification, the test sample should be classified either to the first class of blue squares or to the second class of red triangles. If k = 1 (inner circle) it is assigned to the first class because there is only one square and no triangle. If k=3 (outer circle), it is assigned to second class because there are are 2 triangles and only 1 square inside.

KNN can be used for classification — the output is a class membership (predicts a class — a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It can also be used for regression — output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

The KNN advantage resides in its simplicity to be implemented and it gives relatively high accuracy and it can be used for classification or regression. The drawback, in the other hand is that it is computationally expensive and relatively slow because it has to run over all training data for each prediction, it is also sensitive to irrelevant features[18].

We have a function that takes as input the number k of the nearest neighbors that should be considered, training and testing matrices (should have the same number of features, ie: 32 corresponding to the output of each MZI), finally training and testing labels (a column vector each containing integers from 1 to 4 to label the interferogram of each input class)

The idea is to keep all training samples in hand and when we receive a new data point (represented as a vector), the classifier measures the euclidean distance between data point and all training data it has [19]. We have to provide the algorithm with the labels of each training data point. Then, the classifier will find the nearest k data points in the training samples and assume the label of the given data point is the major label of the nearest k training data points.

The function returns three outputs, first the predicted labels based on the k nearest data points of the training data, and the label will be estimated via majority voting of the labels of these nearest points. Second, the index of the nearest training data point for each training sample. Third, The accuracy (when testing) which is the ratio between the number of the correct predicted labels over the length of the testing labels.

We used the four inputs to test our algorithm, training data are fixed and they are obtained with fitting at 2% noise as explained before. We present an input with some noise rate to the classifier and see the rate of correct predictions it does, and eventually, what class it predicts if it does a wrong prediction.

The results were astonishing with 100% correct predictions of the four inputs for added noise up to 100% of the input power. An explanation of this perfect prediction may be because the input peaks are far from each other so that the algorithm can easily distinguish between them.

That leads us a step further in our study, where we consider peaks with the same width and which are separated by the width at half maximum (WHM). The inputs this time are $Pin1, Pin5$, $Pin6$ and $Pin7$ centred at $1549.4nm$, $1549.459nm$, $1549.517nm$ and $1549.576nm$ respectively, and the fifth input is the same constant $Pin4$ of the last part.
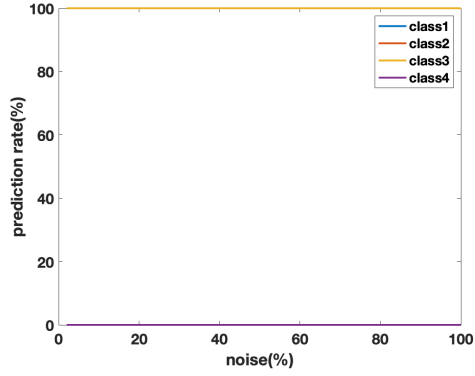
New training data were created by fitting the outputs of MZI as function of temperature for each input class, the data were then organized in a 255×32 matrix. Each line corresponding to an interferogram at a given temperature, for each of the five inputs. Training data were generated with 2%of noise.
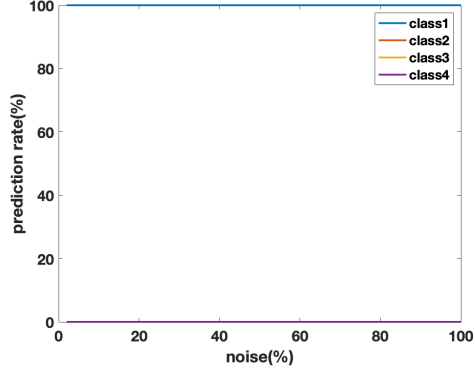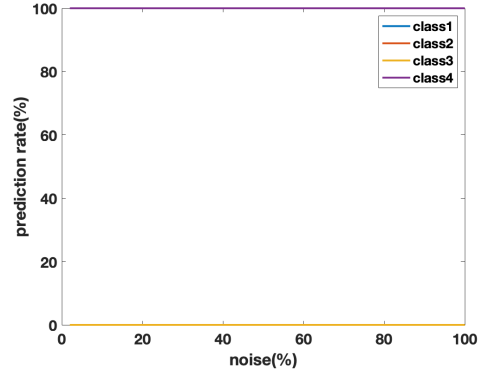
(a) **Input1**
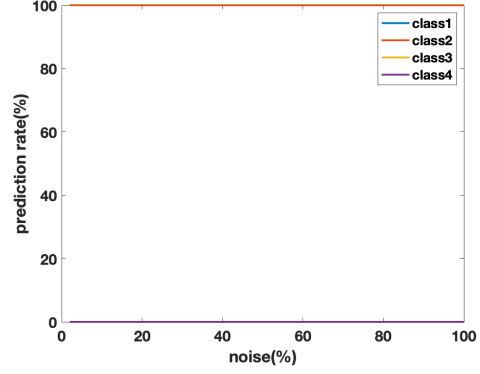
(b) **Input2**

(c) **Input3**

(d) **Input4**

Figure 4.4: Predicted classes for each of the four injected inputs, The algorithm is always predicting the wright class.

Testing data were then generated by collecting the interferograms at different temperatures for each input class at different noise levels, each testing data is a 51×32 matrix.

The Fig.4.6 shows the euclidean distances between testing data for class1 and training data allowing to find the closest neighbors, It's normal that the distance increases when testing noise increases. We note that the distance between two interferograms of two close classes in training data is about 650. The corresponding noise for this value in Fig.4.6 is 120% for which the algorithm actually starts to make errors in classification by predicting not the correct one, but the nearest class next to the correct one instead.

The Fig.4.7 shows that we can still retrieve a good classification accuracy up to 100% noise. Input 5 is equidistant from Input1 and Input6, this can be
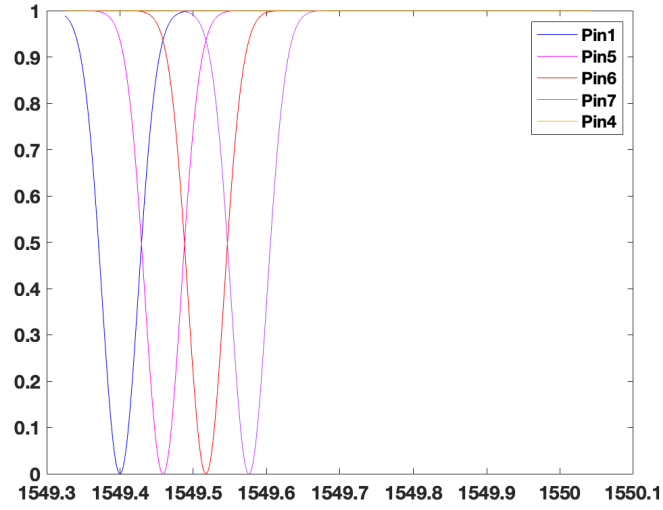
Figure 4.5: Four input peaks separated by WHM and a constant input with no peak
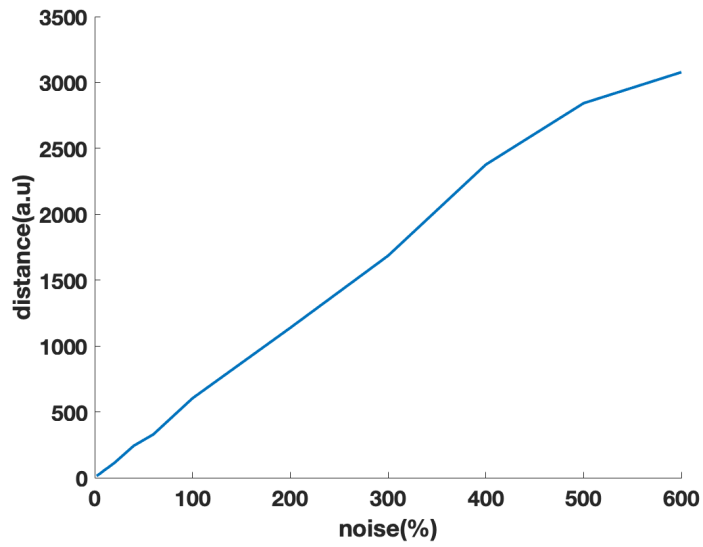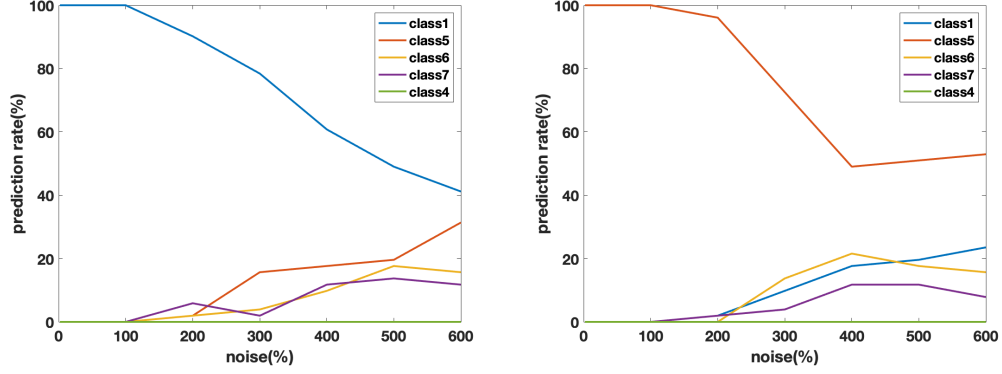

Figure 4.6: Closest distances between interferograms in testing data (class1) and the corresponding ones in training data, we see that the distance increases with increasing noise

(a) Prediction as function of increasing testing noise for input class1

(b) Prediction as function of increasing testing noise for input class5

Figure 4.7: Predicted classes for two different inputs, the accuracy decreases with increasing noise

seen in the predicted classification (4.7). Indeed, when noise increases, prediction accuracy decreases and the peaks which are mostly predicted instead of the correct one are the closest ones.

This chapter was an introduction in how we can use prediction models to FT spectroscopy. The proposed model is K-nearest neighbors which gives high prediction accuracy as discussed earlier, we were able to retrieve the correct injected peak when absorption peaks are separated by $tenth(1/10)$ the FSR. The high accuracy was demonstrated for noise up to 100% of the injected input power. More investigation can be done such as the effect of coupling efficiency between the integrated waveguide in the chip and the optical fibers. The accuracy is highly effected when the coupling coefficient are under 90% next improvements of the model can go in this perspective.

Chapter 5

Conclusion and perspectives

In this report We saw how The Fourier transform spectrometer works, we discussed the different configurations in bulk optics and in integrated optics. We emphasized the advantages it brings among other spectrometer: resolution, FSR and throughput. the configuration of interest was an array of Mach-Zehnder interferometers.

Fourier transform which is required in this kind of spectrometer is a good tool to retrieve the spectrum by measuring the interferogram output. However, devices actually present fabrication errors which distort the interferogram, so that the retrieval can't be gotten with FT.

A solution of this concern is proposed in literature by using the calibration matrix of the device. The good thing about this method is that all the device imperfection are embedded in the matrix. Since the spectral retrieval is the product between the interferogram and the inverse of the calibration matrix, device imperfection are inverted and the spectrum is retrieved correctly. We did a large simulation of this method and studied the effect of noise, propagation loss, temperature.. in the resolution and FSR.

This method is quickly limited because we found that temperature variation between the moment of collecting the interferograms and the moment of spectral retrieving induces shifing in the spectrum. So temperature knowledge is requirement which adds another degree of freedom to our device.

The solution we propose here is to take advantage of Machine Learning models which are largely used in many domains. These models are based on training and prediction of the spectrum without knowing all the device's parameters.

The goal was to implement use a specific family of Machine learning (ML) called "Deep Learning" via artificial neural networks, but by lack of time and the short period of the internship (four months to understand the physics behind the system, study the calibration matrix model and implement the

ML algorithm) we ended up using another algorithm more simple and one of the first algorithm to think of in Machine learning: K-nearest neighbors. we studied this classifier and see how it can be used to get a high accuracy prediction.

References

[1] Kirk, J., Fridley, G., Chamberlain, J., Christensen, E., Hochberg, M. and Ratner, D. (2011). Multiplexed inkjet functionalization of silicon photonic biosensors. Lab on a Chip, 11(7), p.1372.

[2] Florjanczyk, M., Cheben, P., Janz, S., Scott, A., Solheim, B. and Xu, D. (2007). Multiaperture planar waveguide spectrometer formed by arrayed Mach-Zehnder interferometers. Optics Express, 15(26), p.18176.

[3] Florjańczyk, M., Alonso-Ramos, C., Bock, P., Bogdanov, A., Cheben, P., Molina-Fernández, Í., Janz, S., Lamontagne, B., Ortega-Moñux, A., Scott, A., Sinclair, K., Solheim, B. and Xu, D. (2012). Development of a Fourier-transform waveguide spectrometer for space applications. Optical and Quantum Electronics, 44(12-13), pp.549-556.

[4] Velasco, A et al. (2013). High-resolution Fourier-transform spectrometer chip with microphotonic silicon spiral waveguides. Optics Letters, 38(5), p.706.

[5] Podmore, H., Scott, A., Cheben, P., Sioris, C., Cameron, P., Schmid, J., Lohmann, A., Corriveau, Z. and Lee, R. (2017). Athermal planar-waveguide Fourier-transform spectrometer for methane detection. Optics Express, 25(26), p.33018.

[6] Herrero-Bermello, A., Velasco, A., Podmore, H., Cheben, P., Schmid, J., Janz, S., Calvo, M., Xu, D., Scott, A. and Corredera, P. (2017). Temperature dependence mitigation in stationary Fourier-transform on-chip spectrometers. Optics Letters, 42(11), p.2239.

[7] Bradley J. Frey, Douglas B. Leviton, and Timothy J. Madison "Temperature-dependent refractive index of silicon and germanium", Proc. SPIE 6273, Optomechanical Technologies for Astronomy, 62732J (6 July 2006)

[8] Refractiveindex.info. (2019). RefractiveIndex.INFO - Refractive index database. [online] Available at: https://refractiveindex.info/ [Accessed 25 Jun. 2019].

[9] En.wikipedia.org. (2019). Fourier-transform infrared spectroscopy. [online] Available at: https://en.wikipedia.org/wiki/Fourier-transfor-infrared-spectroscopy [Accessed 25 Jun. 2019].

[10] LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. Nature, 521(7553), pp.436-444.

[11] Fang, Z. and Zhao, C. (2012). Recent Progress in Silicon Photonics: A Review. ISRN Optics, 2012, pp.1-27.

[12] Wanda.fiu.edu. (2019). Michelson Interferometer — Modern Lab Experiments documentation. [online] Available at: http://wanda.fiu.edu/teaching/courses/Modern-lab-manual/michelson.html [Accessed 26 Jun. 2019].

[13] Cheben, P., Powell, I., Janz, S. and Xu, D. (2005). Wavelength-dispersive device based on a Fourier-transform Michelson-type arrayed waveguide grating. Optics Letters, 30(14), p.1824.

[14] Kumar, A., Kumar, S. and Raghuwanshi, S. (2013). Analysis of various optical filtering techniques. 2013 International Conference on Microwave and Photonics (ICMAP).

[15] En.wikipedia.org. (2019). Diffraction grating. [online] Available at: https://en.wikipedia.org/wiki/Diffraction-grating [Accessed 9 Jul. 2019].

[16] Wim Bogaerts, Shibnath Pathak, Alfonso Ruocco, and Sarvagya Dwivedi "Silicon photonics non-resonant wavelength filters: comparison between AWGs, echelle gratings, and cascaded Mach-Zehnder filters", Proc. SPIE 9365, Integrated Optics: Devices, Materials, and Technologies XIX, 93650H (2 April 2015).

[17] Nirmal, K., Rengaswamy, S., Sriram, S., Murthy, J., Ambily, S. and Safonova, M. (2018). Design and modeling of a tunable spatial heterodyne spectrometer for emission line studies. Journal of Astronomical Telescopes, Instruments, and Systems, 4(02), p.1.

[18] Medium. (2019). A Quick Introduction to K-Nearest Neighbors Algorithm. [online] Available at: https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7 [Accessed 16 Jul. 2019].

[19] Mahmoud Afifi (2019). kNN classifier (https://www.mathworks.com/matlabcentral/fileexchange/63621-knn-classifier), MATLAB Central File Exchange. Retrieved July 16, 2019.

[20] Google Developers. (2019). Machine Learning Glossary — Google Developers. [online] Available at: https://developers.google.com/machine-learning/glossary/ [Accessed 16 Jul. 2019].