

Sensors and Signal Analysis

Lecture Notes 2024

Nuno Azevedo Silva

*Faculdade de Ciências da Universidade do Porto
INESC TEC*

E-mail: nuno.a.silva@inesctec.pt

Contents

Introduction	1
1 Program	1
2 Evaluation	2
3 Course Materials	3
4 Recommended Bibliography	3
I The Perception Layer	4
1 What is a sensor?	4
2 Types of sensors	6
2.1 Active <i>vs</i> Passive	6
2.2 Analogue <i>vs</i> Digital	6
2.3 Direct <i>vs</i> Hybrid	7
2.4 External <i>vs</i> Internal	7
2.5 By operating physical principles	7
3 Characteristics of a sensor	7
3.1 Ranges and full-scales	9
3.2 Transfer function characteristics	10
3.3 Operation characteristics	11
3.4 Dynamic Characteristics	12
II Signal Analysis	14
1 Analog Signals	15
1.1 Static Response	16
1.1.1 Central Tendency	16

1.1.2	Uncertainty	17
1.1.3	Signal-to-noise Ratio	18
1.1.4	Analyzing the Transfer function and Calibration Function	19
1.2	Time-Varying Signals	21
1.2.1	Time-domain Features	21
1.2.2	Frequency-domain Features	21
1.2.3	Time-frequency Features	23
1.2.4	Statistical and correlation Features	25
1.3	Signal Pre-Processing	25
1.3.1	Sampling Rate adjustment	26
1.3.2	Signal Filtering in the spectral domain	26
1.3.3	Smoothing in time domain	26
1.3.4	Windowing and Leakage	28
2	Digital Signals	28
2.1	Typical conversion schemes	29
3	From sensors to automation: a brief note on controllers	29
4	Machine Learning and Sensors	30
5	Concluding Remarks	32
IV	Enabling Optical Sensing with Wave Interference	33
1	Wave Optics in a Nutshell	34
1.1	Maxwell's Equations and Wave Equation	35
1.2	Wave equation, Waves, and Wavefronts	36
1.3	Reflection and Refraction	40
2	Introduction to Optical Interferometry	42
2.1	Newton Rings	44
2.2	Young Double-Slit	45
3	Overview of Experimental Components	47
3.1	Sources	47
3.2	Optical components: Free space vs Fiber Optics	49

3.2.1	Beam steering	49
3.2.2	Beam division and combining	49
3.2.3	Polarization control	50
3.2.4	Lens and Lens systems	51
3.2.5	Pin-holes and slits	53
3.3	Optical Detectors	53
3.3.1	Detectors	54
3.3.2	Cameras	55
4	Concluding remarks	57
V	Optical Interferometers	58
1	Transfer Matrix Formalism	58
2	Michelson Interferometer	59
3	Mach-Zehnder Interferometer	61
3.1	Digital Off-Axis Holography	62
4	Sagnac Interferometer	64
5	Fabry-Perot Interferometer	66
6	Concluding Remarks	69
VII	Signal Detection in Interferometric Sensors	71
1	Homodyne detection	71
1.1	Active Homodyne	72
2	Heterodyne detection	74
3	Pseudo-Heterodyne detection	75
4	Concluding Remarks	76

VIII Beating The Limits of Noise	78
1 The Origin of Gravitational waves	78
1.1 How intense are these?	82
2 LIGO: The Classical Design	82
2.1 The Overall Design and Notes on Components	83
2.2 Michelson Interferometer and Dark fringe Operation	85
2.3 Fabry-Perot	86
2.4 Power Recycling	87
3 The Noise Limitations	88
3.1 The Quantum Noise Limit	90
3.2 Shot Noise Contribution	90
3.3 Radiation Pressure Noise	91
3.4 Reducing noise classically	92
4 Quantum Limit and Beyond	92
4.1 The origin of the noise	93
4.1.1 Quantization of the Electromagnetic Field	93
4.1.2 States and Operators to handle the quantized electromagnetic field	94
4.2 A Quantum Model for the Michelson Interferometer	98
4.2.1 Radiation Pressure	98
4.2.2 Shot Noise	99
4.3 Squeezed States and Reduction of noise	100
4.3.1 Surpassing the SQL using squeezing	102
5 Concluding Remarks	104
X Sensing with Quantum Optics	107
1 Generating Quantum States of Light and Single Photon Sources	108
2 Sensing with Entangled States: NOON states in Interferometers	109
2.1 NOON states	111
2.2 Enhanced Phase Sensitivity	112

3	Quantum Interference	114
3.1	Hong-Ou-Mandel Effect	114
3.2	Second-Order Correlation Function	115
3.3	Monitoring the Hong-Ou-Mandel Shift for Quantum Sensing	116
4	Quantum for metrology and the Redefinition of SI Units	118

Introduction

In this curricular unit, we aim to discuss the inner workings, characteristics, and types of Sensors and their signals, approaching them in a scientific manner. In particular, we will focus on **Optical Sensors** and specifically **interferometric sensors** to understand the potential of one of the most high-performance tools in physics these days and reflect on the challenges of the field, from classical to quantum regimes.

One of the major goals of the course is to provide a critical perspective on these topics introducing key concepts while engaging in interactive discussions and hands-on activities to see the concepts at work. By focusing on the concepts, the aim is to provide the necessary tools and solid cornerstones for a student who wishes to explore the topic of optical sensors in more detail. **The second goal**, and more general, is to provide the students close contact with techniques, materials, and skills closer to scientific research, acting as additional building blocks for their future career.

1 Program

A. Sensing and data analysis techniques and methods

Week 1-What is a sensor and how to characterize one.

Week 2 and 3-Handling output signals: static and time-varying; frequency analysis; brief introduction to data visualization and machine learning for science.

B. Classical Optical Interferometers for Measurement/Sensing

Week 4-Why Light Excels in Sensing and what is an Interferometer. Interference as the building block of interferometric sensing. Newton's Rings and Double Slit experimental demonstrations.

Week 5-Interferometers: matrix formalism and overview of Michelson, Mach-Zehnde, and Sagnac Interferometers. Hands-on with a Michelson Interferometer and Mach-Zehnder.

Week 6-Fabry-Perot interferometer and outlook and perspectives of classical interferometric sensing. Discussion of the past, present, and future of interferometers.

Week 7- Homodyne, Heterodyne detection: Pros and cons of heterodyne

detection for phase recovery in optical interferometers and characterization of their performance. A case study with pseudo-heterodyne detection in the context of optical fiber sensors.

C. Semi-Classical Optical Interferometers for Measurement/Sensing

Week 8- LIGO Case study - tackling the challenge of ultra-sensitivity. The role of noise and quantum fluctuations. The standard quantum limit.

Week 9- LIGO Case study with quantum formalism - Squeezed light: concepts and generation. Noise reduction below the standard quantum limit.

D. Quantum Sensing with Optical Interferometry

*Week 10-*Quantum light sources and detectors. Optical interferometers with entangled states and non-classical light sources. Quantum Metrology.

E. Additional topics in Optical sensors

*Week 11-*Other optical sensors: back-scattering and distributed acoustic sensing. Optical sensors exploiting resonant phenomena: surface plasmon resonance and whispering gallery modes. Spectroscopy: techniques and tools.

*Week 12-*Project tutorials and exercises.

2 Evaluation

The evaluation method includes a final exam. The students can choose to be assessed only based on the final exam, in which case the exam grade is the final grade, or presenting an optional written report can be considered for 40% of the grade. In this case, the final grade is computed as:

$$\text{Grade} = 0.6 \times \text{Exam} + 0.4 \times \text{Report}$$

It is recommended that students choose the evaluation with a project component.

3 Course Materials

1. Lecture Notes

Compilation of key ideas and contextual help for the activities. These should be utilized by the students to better prepare each week.

2. Jupyter Notebooks

To be utilized during the class to provide a hands-on approach to the topics of signal processing and analysis.

4 Recommended Bibliography

In addition to the **Lecture Notes**, which compile most of the relevant information for the curricular unit, students may utilize the following bibliographic references:

1. Robert D. Guenther; Modern optics. ISBN: 0-471-51288-5
2. Fraden, Jacob, and J. G. King. Handbook of modern sensors: physics, designs, and applications. Vol. 3. New York: Springer, 2010.
3. Reitze, David; Saulson, Peter; Grote, Hartmut (Editors); Advanced Interferometric Gravitational-wave Detectors: Essentials of Gravitational Wave Detectors. ISBN: 978-9813146075
4. Bachor, Hans-A.; Ralph, Timothy C.; A Guide to Experiments in Quantum Optics. ISBN: 978-3-527-41193-1
5. Max Born; Principles of optics. ISBN: 0-08-018018-3

The Perception Layer

Week

I

1 What is a sensor?

Understanding what is a sensor is the trailhead of this curricular unit, and sets the tone for the discussion we want to engage in during it.

| What would you call a sensor?

In a broad perspective, one could argue that a sensor would be any abstract device that reacts to something that happens in the environment. Yet, with such a broad definition you would consider almost anything as a sensor. Going to the very fundamentals of physics, and focusing on the fundamental interactions (gravity, weak force, electromagnetic force, and strong nuclear force) a simple atom would be a sensor for any process in physics. This extreme example shows that the previously mentioned definition is overly broad, and as a result, it becomes meaningless. To better define what is a sensor, a better question we can ask ourselves is to start with why, i.e.

| Why do you need sensors?

We need sensors because we want to know more about the environment so we can make decisions, build models, or construct machines that are aware of their surroundings. Therefore, being more strict, a sensor is a device that not only interacts with the environment but allows you to get information from it in order to make a **measurement**. Therefore:

Definition. Sensor: A sensor is a device that detects and measures physical properties or changes in a physical system, producing a usable output signal (typically an electric signal) in response to a *stimulus*. For example, a microphone is an acoustic sensor that measures pressure variations that reach the diaphragm associated with sound. Another example is a temperature sensor that measures the ambient temperature and outputs a signal related to that temperature.

In science and engineering, sensors are also closely related to the transduction principle, i.e., the conversion of one form of energy into another. Transducers and actuators are therefore important concepts to complete the backbone of any measurement and control system.

Section 1. What is a sensor?
Section 2. Types of sensors
Section 3. Characteristics of a sensor

Sensor is a device that provides usable output in response to a specified quantity that is measured.

Instrument Society of America

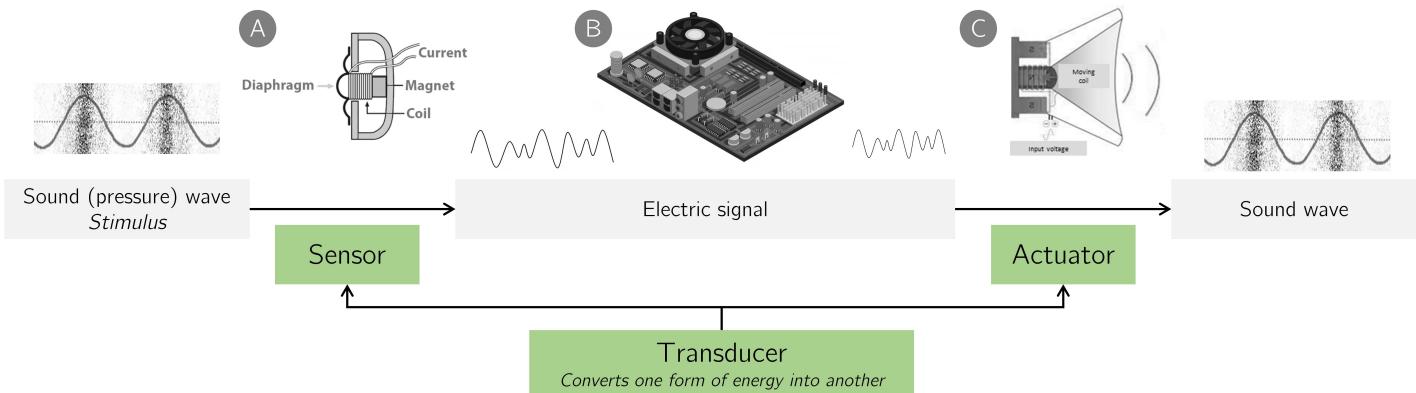


Figure 1. Sensors and actuators are both transducers. An illustrative and straightforward example is the sound components of our laptop. The sensor is the microphone, that converts sound (a signal transmitted as pressure waves) into electric signal. With this electric signal, some processing can be made to encode, decode, filter, amplify, storing, etc. The actuator in this case is the loudspeaker, which receives electrical signals and converts them into sound signals, with possibly altered properties.

Definition. Transducer: a device that converts one form of energy into another. For example, our microphone is a transducer that converts acoustic energy (sound waves) into an electrical signal. Yet, similarly, a loudspeaker can also be considered a transducer, but it is not a sensor.

Definition. Actuator: a device that produces a physical action or movement in response to a control signal. For example, an actuator may convert an electrical signal or other forms of energy into mechanical motion, as it happens in a loudspeaker. Another example is an electric motor in a robotic arm, that moves in response to signals from a control system.

At this point, our sensing machine has its backbone set but still incomplete. As we saw before, sensing is the process of getting information from the environment. Our definition of sensor utilizes two interesting concepts that we use every day in physics and engineering: **signal** and **measurement**. The signal definition is broad, but we can say that it refers to any quantity or process that conveys information, in the case of sensors, an electric signal. Measurement is the process of assigning a numerical value to a property or characteristic of an object or phenomenon (e.g. temperature, pressure, displacement). Ultimately, the sensor is only usable when we can translate a signal into a measurement.

All sensors are transducers, but not every transducer is a sensor.

Thus, measuring with a sensor has three steps:

1. the sensor is affected by the environment, outputting a response in the form of an electric signal;
2. the signal is transmitted to a processing unit;
3. the signal is processed and analyzed to convert raw information into the final measurement.

If we think about we will find this conceptual structure replicated everywhere, from man-made technology to nature's innerworkings. For example, human beings are equipped with 5 different types of sensors: eyes, which detect energy of the electromagnetic field¹; ears that respond to acoustic pressure; a tongue and nose that are produce responses to the presence of given chemicals; and skin which can detect pressure and temperature. All the signals that come out of these sensory devices then travel through our body as electric signals using nervous pathways, reaching our nervous system where they are processed².

¹ mostly in the visible range, which is usually called light

² which may lead to a response by one of the many actuators in our body - such as a muscle.

2 Types of sensors

Grouping sensors into families is a challenge due to the vast diversity and specificity of sensors. Yet, such exercise offers significant learning opportunities while it also allows to establish common language across various fields of technology and science.

2.1 Active vs Passive

Active sensors require external power from an external source to operate. This signal is often referred to as an excitation signal. For example, a strain gauge is a sensor that explores the fact that a wire resistance changes when it is stretched or compressed. Thus, it is an active sensor as it requires passing a current in the sensing head to detect a resistance variation.

In contrast, passive sensors can output an electrical signal in response to the input stimulus without any additional power. The energy utilized is obtained directly from the measurand, as it happens for example in a thermocouple device, that makes use of the Seebeck effect.

2.2 Analogue vs Digital

An **analog sensor** generates an output signal that varies continuously within a specific range, typically producing a voltage that is directly proportional to the measured parameter. Parameters such as speed, temperature, pressure, and strain - inherently analog due to their continuous nature - align well with the capabilities of analog sensors. The output from these sensors changes in a smooth, continuous manner over time. However, the use of analog sensors

may feature slower response times and often requires an external power supply for amplification (e.g. Op amp circuit) of the output signals.

On its turn, a **digital sensor** emits output signals in discrete digital form, essentially toggling between two distinct states: ON/1 and OFF/0. An illustrative example of a digital sensor is a push-button switch. Depending upon the bit depth employed to represent the measured parameter, digital sensors may feature higher accuracy compared to their analog counterparts. Besides, they are also preferable for direct integration into microcontroller-based systems. Note yet that it is still possible to integrate analog sensor outputs in similar manners by utilizing analog-to-digital converters.

2.3 Direct vs Hybrid

A sensor that incorporates more than one transducer to output an electric signal³ is called **hybrid**, in opposition to a **direct** sensor.

³ Can you give an example of a complex sensor? Hint: Chemical sensors.

2.4 External vs Internal

A sensor may be incorporated into a larger system with other sensors, transducers, processing units, and actuators. Considering the position in this system, the sensor is classified as **external** if it responds to the environment outside the system, or **internal** if it aims to monitor and measure the properties of its internal parts.

2.5 By operating physical principles

If we want to design a sensor from scratch, our better shot would be to look at physical laws and processes to act as the design unit of our sensing system. This approach entails identifying the conversion phenomena⁴ that can be derived from various physical phenomena involving interrelated variables to react to a stimulus. This diversity can also be utilized to group sensors as shown below in the table.

⁴ not exclusive

Exercise 1. Consider the sensors in your smartphone: name some of them, their function, and their type (active/passive, analog/digital, direct/hybrid, extrinsic/intrinsic).

3 Characteristics of a sensor

When we are selecting a sensor, the above qualitative classification is often secondary. Indeed, the characteristics/specifications of the sensor and how they compare with those required for the application are the most important factors. These characteristics of the sensors may be divided into **static** and **dynamic** characteristics.

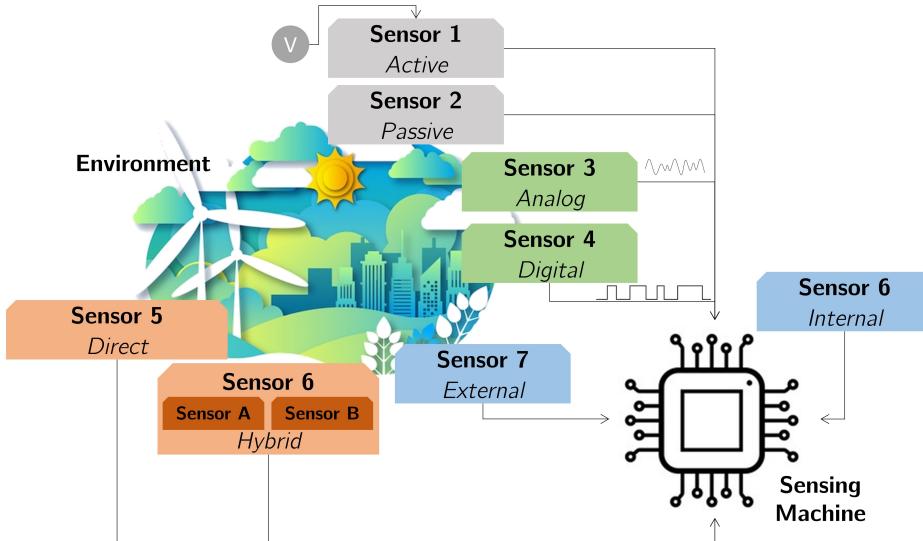


Figure 2. Wrapping up the qualitative classification of sensors into a single scheme.

Conversion Phenomena		Stimuli	
Physical	Photoelectric Photomagnetic Magnetoelastic Electromagnetic Thermo-optic etc	Electric	Charge Current Voltage Conductivity Permittivity etc
Chemical	Chemical reaction Spectroscopy Electrochemical process Physical transformation etc.	Mechanic	Position Acceleration Force Mass etc.

Table 1. Examples of conversion phenomena and stimuli.

The **static** characteristics of a sensor entail the algebraic relation between the input - e.g. a stimulus S and its output - a signal E . A characteristic curve can then be written and graphically drawn as

$$E = F(S) \quad (3.1)$$

and the function F is denominated the **transfer function** of the sensor. Measuring is, therefore, nothing more than inverting the relation, i.e. $S_m = F^{-1}(E)$ being S_m the measurement. A sensor is normally provided by the

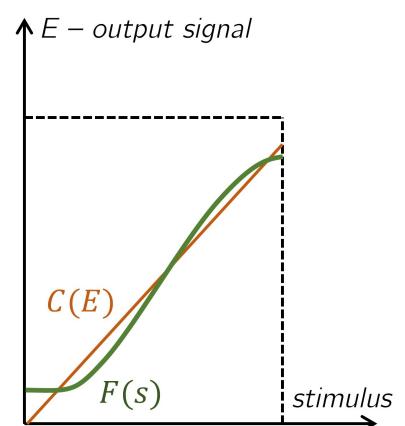


Figure 3. Transfer function and calibration.

manufacturer with a transfer function which is obtained using a calibration process. The provided mathematical model to approximate $F^{-1}(E)$ is called the **calibration curve**, which is normally a linear model $C(E) = mE + b$. Note however that the transfer function itself is often nonlinear⁵ and may even be **multidimensional**.

Dynamic characteristics on the other hand refer to the sensor response when the input changes. Usually, these are often expressed in terms of response to certain standard input signals (e.g. response to a step signal, frequency response bandwidth, possible damping, or overshooting - see figure 4).

3.1 Ranges and full-scales

A class of important static characteristics of a sensor aim to describe its range of operation. Depending on the perspective we are using or looking at the sensor, we may establish a few distinct characteristics:

Span/Full-scale Input

It indicates the difference between the minimum and maximum stimulus, s_{min} and s_{max} respectively, that produces a measurable response of the sensor, without causing unacceptably large error.

$$Span = s_{max} - s_{min}$$

Full-scale Output

For an analog output it indicates the difference between the minimum and maximum output signal generated. For a digital output, it is related to the maximum digital count that the convertor can resolve for the span.

$$FSO = E_{max} - E_{min}$$

Dynamic Range

Usually, as the above scales span across multiple orders of magnitude, the ratio of the largest measurable stimulus to the smallest measurable stimulus may be a better metric for measuring the full-scale of a sensor. This is usually referred to as **dynamic range** and is given in decibels, which is a logarithmic measure of ratios of either powers or amplitudes. By definition, for powers, it is ten times the log of the ratio of powers,

$$\text{Dynamic Range} = 10 \log_{10} \frac{P_{max}}{P_{min}} dB \quad (3.2)$$

⁵ e.g. $F(S) = Ae^{bS}$

s – stimulus

E – output signal

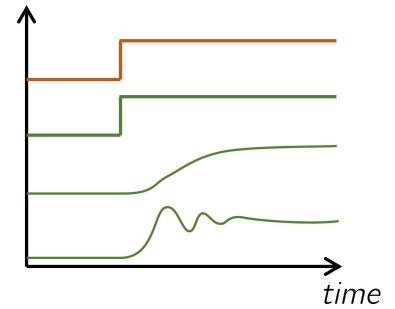


Figure 4. Various types of dynamic responses to a step stimulus.

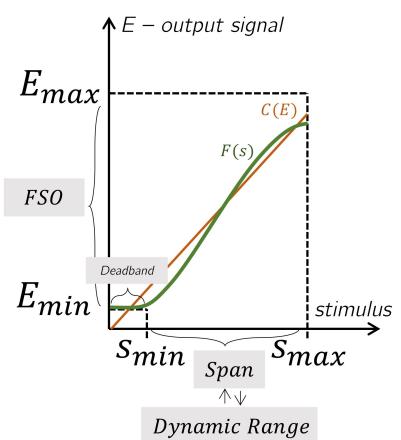


Figure 5. Range characteristics.

while for amplitudes(e.g.voltages) we have 20 times the log of the ratio of the minimum and maximum values, i.e.

$$\text{Dynamic Range} = 20 \log_{10} \frac{V_{max}}{V_{min}} \text{dB} \quad (3.3)$$

where P_{min} and V_{min} is the smallest measurable signal, typically defined as that equal to the noise level. In optics it is also common to use $P_{min} = 1\text{mW}$ with the unit being called **dBm** in that case.

Although the dynamic range or range normally applies to the span, it may also be applied to the full-scale output(it should be clearly stated in the spec sheet).

Deadband

A deadband (also known as a dead zone or a neutral zone) is a band of input values in the domain of a transfer function of a sensor for which the output value is 0 or equal to the noise (i.e. signal to noise ratio equal to 1).

3.2 Transfer function characteristics

Another set of important static characteristics of a sensor aims to describe how the sensor responds to the stimulus and how it relates to the true value.

Sensitivity

Sensitivity is a relationship between the input physical signal and the output electrical signal. It is the ratio of the change in the output of the sensor to the input value and mathematically can be expressed as

$$\text{Sensitivity} = \frac{dE}{ds} \quad (3.4)$$

Accuracy

It is the degree of exactness between actual measurement and true value. Note that the accuracy corresponds to a metric computed from the measured stimulus and not from the signal obtained itself, i.e. $\delta = s_m - s_{true}$. This means that accuracy may depend on the calibration function but also on the uncertainty of the measured signal. Besides, in typical real-world conditions, accuracy δ may vary with the stimulus as we will later discuss in this curricular unit.

Accuracy may be expressed in multiple manners, from absolute values to percentages of the full-scale input. The most common way to estimate a maximum value Δ is to calculate the calibration function and maximize

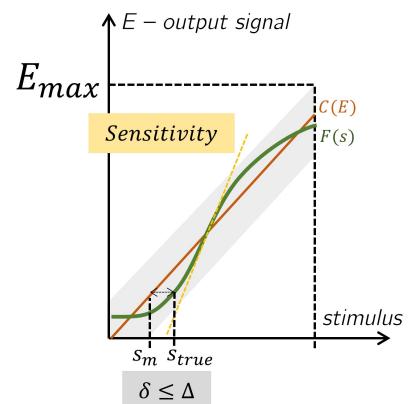


Figure 6. Sensitivity and accuracy.

$|\delta|$ for the calibration dataset, which will be in already in the units of measurement.

Linearity

When we talk about linearity we are referring to how nonlinear is the transfer function of the model. This corresponds to the deviation of the curve of actual measurement from the curve of ideal linear measurement and can be expressed in percentage of the span. Normally it is tightly connected with the dynamic range.

Saturation

Saturation is intrinsically related to the operation limits of the sensor and is identified by the lack of variation of the output signal above a threshold of the input stimuli.

Hysteresis

The hysteresis of a sensor can be defined as the maximum difference of the measurable value when approaching the point in opposite directions, i.e. first with increasing and then with decreasing the input stimuli. Hysteresis is commonly related to design, friction, or structural changes in materials, such as plastic deformation for example.

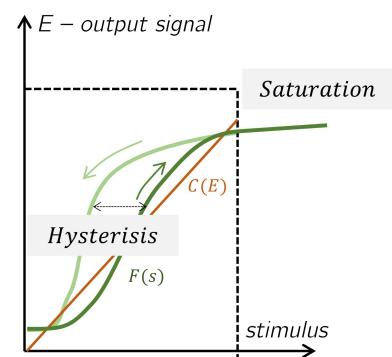


Figure 7. Hysterisis and Saturation.

3.3 Operation characteristics

Repeatability

Repeatability or reproducibility is the inability of the sensor to produce the same output for the same stimuli under similar conditions. This may be related with thermal noise or charge accumulation for example. Typically it is computed as the maximum difference of measured stimulus for two distinct runs and can be represented in the same units or in a percentage of the scale.

Stability and Drift

Over time, the characteristics of a sensor can change, which may affect the accuracy of the measurements. This may happen in short (typically minutes and hours) or long (days to years) timescales, for which the degradation in accuracy is commonly denominated stability (also short-term stability) or drift respectively.

Reliability

Reliability is closely related to the concept of stability and measures the sensor's ability to perform as expected over a period of time. Common metrics are presented in time instead of accuracy, for example, the mean-time-to-failure

$$MTTF = \frac{1}{n} \sum_i t_f^i \quad (3.5)$$

with t_f the time of failure of device i of n devices in total, starting the test at $t = 0$. It may also include parameters under extreme conditions(such as temperature, humidity, corrosion, etc.).

3.4 Dynamic Characteristics

Dynamic characteristics include a vast number of characteristics that allow one to characterize the response of a sensor to a time-varying signal. Ultimately they are related to the components constituents, principles of operation, and governing equations⁶. There are multiple parameters and strongly depend on the final application. For the sake of simplicity, we will discuss two of the most important ones: the rise time and bandwidth.

Rise Time

The rise time τ_{rise} is defined as the time it takes for the output signal waveform transitions from 10% to 90% (sometimes 20% to 80%) of the final value when the input is a step signal. It can be connected with the concept of **slew rate** in electronics, which is the resulting rate of change of the output signal, i.e. $\Delta V/\tau_{rise}$.

Bandwidth

The bandwidth of a sensor indicates how sensors respond to signals at different frequencies. It is usually measured at the -3dB point, i.e. the output signal drops $\approx 30\%$ in amplitude with the increase of frequency. Normally we only look at the upper frequency limit and assume that the sensor works well in continuous response.

Exercise 2. In the class we will distribute a sensor for each group, accompanied by a component spec sheet if possible. In groups, prepare a short presentation about your sensor, focusing on the *stimulus* to be measured, outlining the working physical principle, classifying the type of sensor qualitatively, and presenting some characteristics (from 3 to 5) of the sensor.

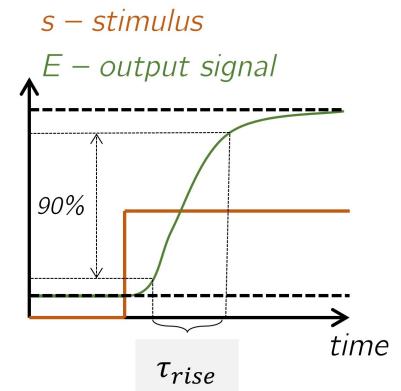


Figure 8. Rise time.

⁶ e.g. first order or second order differential equations

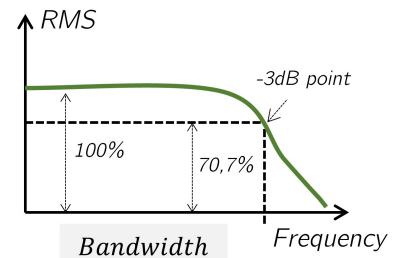


Figure 9. Bandwidth.

Concluding remarks

In this first week, we focused on understanding what is a sensor, and how to characterize one in qualitative and quantitative terms. As we have seen a sensor is not complete without a calibration function, that tries to approximate the transfer function of the sensor. This is only possible when one applies known stimuli, and analyzes and processes the output signal, extracting the necessary information from the signal that allows one to construct the mathematical model by numerical curve fitting. That is the motivation for next week.

Signal Analysis

Week
II

As we saw last week, sensors are the perception layer of our machines, interacting and actively reacting to the environment. This reaction to a given stimuli s_i translates into an output signal E_i following what we call the sensor transfer function $F(s)$. In particular, when designing a sensor for the end-user one needs to provide a calibration function $C(E_i) = s_m$ that for an output signal E_i provides a measurement s_m and which approximates the inverse of the sensor transfer function $F(s)$.

Although it may seem from the last week (and from most of the sensors around us) that the output signal E always corresponds to a voltage, that is not always true. In practice, during the design stage of a sensor⁷, E may correspond to any information extracted from the electric signal response of the sensor by using convenient processing techniques that explore its underlying physical principles. To give some practical examples:

Doppler radar: The information of object velocity is encoded in the changes of reflected wave frequency due to the movement of the object.

Surface plasmon resonance based sensor: The information is encoded in the wavelength/angle of highest absorption.

Section 1. Analog Signals
Subsection 1. Continuous
Subsection 1. Time-varying
Section 4. Digital Signals

Table 2. Contents for WEEK II

⁷ usually our work as physicists

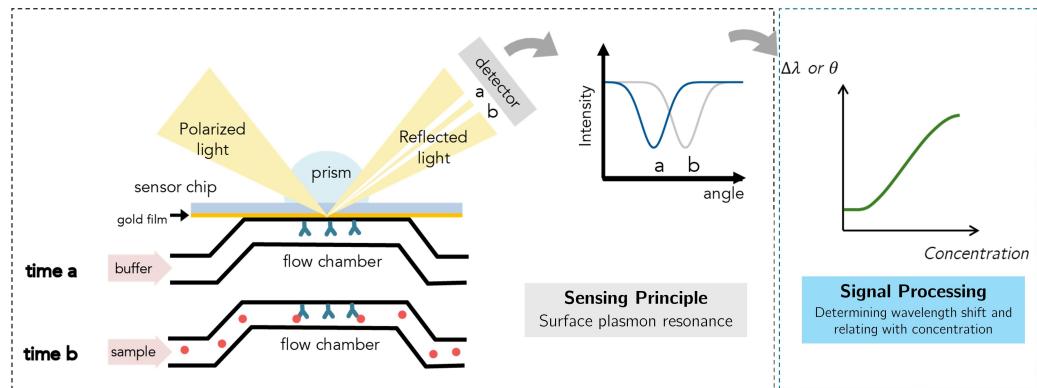


Figure 10. Sensing principle of a Surface Plasmon resonance sensor. The sensor is based on the fact that due to light confinement, plasmon coupling is extremely sensitive to the environment, with small changes in the refractive index near the surface translating into major responses. In this case, the presence of a distinct sample in the flow chamber modifies the refractive index and leads to a shift in the absorption band/angle.

Of course, most of the sensors we use every day are highly developed tools

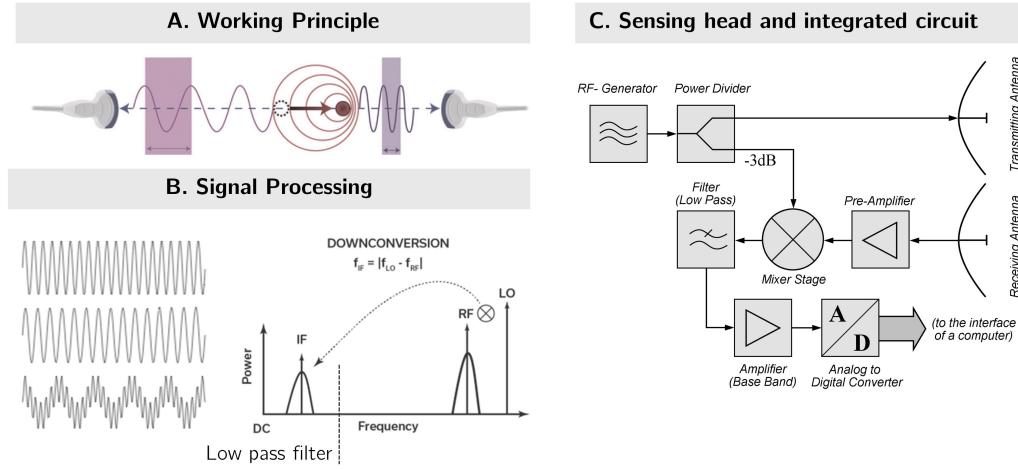


Figure 11. Doppler sensor working principle, signal processing, and circuit integration.

designed for agile integration with a processing unit, meaning that preferably all signal processing occurs in-chip (e.g. integrated circuit or even a microcontroller) to translate the output signal of the sensing head E_i into a voltage output signal E'_i for which a direct calibration function can be applied without processing. For example (see figure 11), while the information in the Doppler radar is in the frequency shift, a frequency mixer integrated circuit, followed by a low-pass filter, and a frequency-to-voltage converter can be utilized to convert the frequency shift to an output voltage signal E'_i that is only then communicated to the user. Nevertheless, note that designing a sensor like this involves prior knowledge of the signal processing that is needed to extract the information⁸ before being able to retrieve it with the converter.

For this second week, our goal is to introduce some tools of the art of processing data from sensors, with the ultimate aim of extracting the information needed to construct a transfer function and finally a calibration model. First, we focus our attention on analog signals, where the nuances of proper statistical analysis of static and dynamic signals will be explored. Then we will briefly discuss digital signals, highlighting their advantages and applications. We will describe some analog-to-digital conversion processes focusing on the most common modulation techniques that are vastly utilized in engineering these days.

⁸ e.g. mixing and filtering

1 Analog Signals

As we said before, our goal is to extract meaningful information⁹ from the output signal of the sensing head. Possible information contained in the signal largely depends on the physical process utilized as a sensing principle

⁹ Which strongly relates to the concept of feature extraction in machine learning.

and we can divide it into major families: time-constant signals and time-varying signals.

1.1 Static Response

The static response of a sensor refers to its behavior under steady-state conditions, where the quantity being measured does not change over time. Knowing the applied stimulus (normally well-known standards), one can **determine the transfer function** of the sensor and deploy a **calibration model** by curve fitting procedures.

Towards this final goal, we must start by assigning a value E_i for a given stimulus. At first glance, this may seem a simple task, apply s_i , register E_i , and compute the transfer function. However, such an approach largely overlooks experimental challenges of the real world, such as noise and variability.

Definition. **Noise** refers to the random fluctuations that can distort the output signal of a sensor. It can originate from various sources, including **external noise** (when it comes from the environment), **internal noise** (when it comes from within the sensor and electronics itself), or **quantization noise** (from the analog-to-digital conversion procedure).

Solid experimental procedures to determine the transfer function shall then involve recording multiple values of $E_i^{(j)}$ for $j \in \{0, \dots, N\}$ to a known stimulus s_i , preferably over various runs. Having a set of measures for the same stimuli s_i it is now possible to extract a good estimate E_i utilizing statistical procedures.

1.1.1 Central Tendency

The first information that we can extract is the estimated value E_i itself, which may be obtained in the form of a central tendency¹⁰. While many measures of central tendency exist, a good rule of thumb is to analyze the distribution of registered $E_i^{(j)}$ values and choose:

Mean: if the data is normally (or symmetrically) distributed and there are no significant outliers.

Median: if the data distribution is skewed (e.g. has a long tail on one side), if there are significant outliers (e.g. if anomalies are common), or if you want a more resilient measurement.

Mode: if the data is of categorical type (e.g. On/Off).

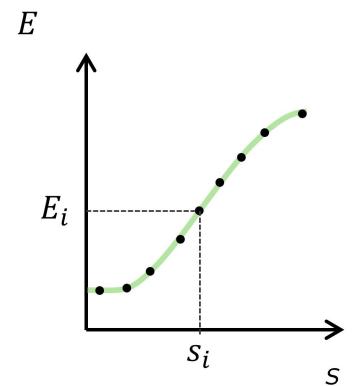


Figure 12. Experimental determination of the transfer function.

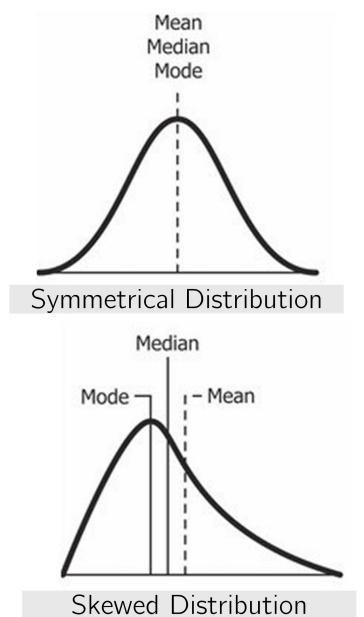


Figure 13. Central tendencies in symmetric and asymmetric distributions.

¹⁰ *tendency of quantitative data to cluster around some central value*

Less utilized statistical measures include geometrical, weighted, trimmed, and weighted means. These are utilized in very specific situations which although they may appear in a physics laboratory, usually do not appear in the context of sensors.

1.1.2 Uncertainty

Now that we have a central tendency, it would be important also to quantify the lack of exactness of the E_i considering its variation, which relates to the concept of **uncertainty**.

In science, the word **uncertainty** is not necessarily associated with a mistake but rather with the lack of exactness of a measurement. The fact is that it is impossible to measure with infinite accuracy and precision¹¹ even if we have the best equipment and conditions available.

Uncertainty encompasses all potential errors in the measurement process, which are categorized into:

- **Type A Uncertainty:** Related to statistical variation of measured data (often associated with random noise) and thus can be estimated by statistical methods.
- **Type B Uncertainty:** Related to all other sources of uncertainty in the measurement process, such as properties of the instrument and environmental conditions. These are not necessarily random and are often evaluated by non-statistical means.

There are multiple manners to estimate the uncertainty of type A, u_A being the most common:

Standard deviation is the preferred method for computing Type A uncertainty because it takes into account the variability of all measurements and provides a basis for further statistical analysis. In particular it may define a **prediction interval** for a normal distribution: e.g. the interval $[\bar{E}_i - \sigma_i, \bar{E}_i + \sigma_i]$ contains around 68% of the registered values, whereas $[\bar{E}_i - 2\sigma_i, \bar{E}_i + 2\sigma_i]$ contains around 95%.

Maximum deviation, might be used for quick assessments or when the data set is very small, but it does not provide as detailed an understanding of the data variability.

Percentiles and Interquartile Range, when the data distribution is significantly non-normal or to mitigate the presence of outliers.

For type B uncertainty u_B , one should consider all known or estimated sources of uncertainty other than those derived from statistical variation.

¹¹ Accuracy - how close a measured value is to the true value. Precision - how close the measurements of the same item are to each other;

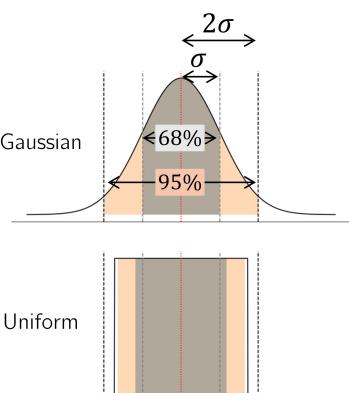


Figure 14. Standard deviation and Interquartile range.

This can include uncertainties due to instrument calibration, environmental factors, assumptions in the measurement process, and more. Each source of Type B uncertainty is quantified, usually as a standard deviation equivalent, based on the best available information.

Once you have quantified Type A and Type B uncertainties, you combine them to calculate the combined standard uncertainty. Assuming that all uncertainties are independent, the combined standard uncertainty is calculated using the square root of the sum of the squares of the individual standard uncertainties:

$$u_i = \sqrt{u_A^2 + u_B^2}. \quad (1.1)$$

Your output signal is then represented as $E_i \pm u_i^{12}$.

1.1.3 Signal-to-noise Ratio

Signal-to-noise Ratio (SNR) is another common measure used in science and engineering in the context of sensors. It calculates the ratio of signal power to noise power, **being an indicator of how much the signal may be corrupted by noise**. A higher SNR indicates a cleaner signal with less background noise, whereas a lower SNR indicates more noise in comparison to the signal.

$$SNR = \frac{P_{signal}}{P_{noise}} \quad (1.2)$$

for power or

$$SNR = \frac{A_{signal}^2}{A_{noise}^2} \quad (1.3)$$

for amplitude signals. Normally, as it covers it may be expressed in decibels (dB) being

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \text{ dB} \quad (1.4)$$

when involving powers, or

$$SNR = 20 \log_{10} \frac{A_{signal}}{A_{noise}} \text{ dB} \quad (1.5)$$

when involving amplitudes.

So you can utilize our previous estimate $E(s_i)$ obtained from central tendency as the numerator for any of these formulas. **But what should you use in the denominator?**

The most common is to utilize one of two methods:

Direct Measurement: For electronic sensors, you can, when possible, directly register the signal in a quiet (noise-free) environment. This measurement gives you the peak-to-peak noise voltage or current, from

¹² Use 2 significant figures, according to SI regulations.

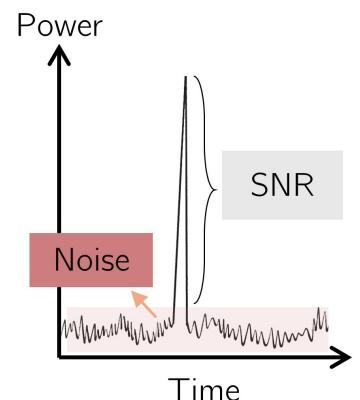


Figure 15. Signal-to-noise Ratio.

which you can calculate the RMS (root mean square) or the standard deviation for the noise signal.

Statistical Analysis: if you cannot perform direct measurement of the noise level, and you already estimated type A uncertainties, you can usually utilize the standard deviation as a measure of the sensor random noise.

To choose what to use in real-world scenarios you must analyze your situation and see what choices you have. But more importantly than that is to **to present the utilized formula**. This is usually sufficient to maintain the scientific accuracy of your works.

1.1.4 Analyzing the Transfer function and Calibration Function

With the values $E(s_i) + u_{E(s_i)}$ calculated for a given stimulus set $\{s_i\}$ we are now ready to construct the **Transfer function** for our sensor. As we saw last week, the transfer function allows us to characterize our sensor in terms of meaningful quantitative and qualitative parameters namely:

From direct graphical analysis:

- Span, Full-scale Output, Dynamic Range, and Deadband;
- Hysteresis, linearity, and saturation;

From numerical analysis:

- Sensitivity, by computing numerical derivatives;

From multiple runs:

- Repeatability, Stability (if you measure various runs over a large period of time).

Besides, having the transfer function set we are now in conditions to build the calibration curve, i.e. a mathematical model that we will use in the real world to estimate a measurement s_m if we record the value E_m at the output signal. Again, it is up to you to decide which analytical model is best. Yet it is always a good practice¹³ to prefer **linear models** when possible,

$$C(E) = mE + b. \quad (1.6)$$

If you need to utilize a nonlinear model, opt for the **lowest number of**

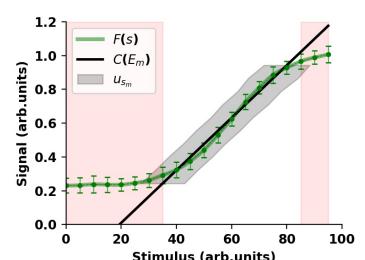


Figure 16. Experimental determination of a transfer function and calibration.

¹³ and following Occam's razor principle

fitting parameters¹⁴.

To estimate the value of parameters you need to perform what it is called in statistics and data science a numerical **regression**. A regression task provides estimates for the parameters of a given mathematical model by minimizing a loss function¹⁵.

These days you have multiple computational and statistical libraries available (in *python* you have for example *scipy curve_fit* function or the *lmfit* library) that may provide you with convenient parameter estimation and related uncertainty. Besides, most of the tools allow you to provide uncertainty for each point, performing a weighted regression which exploits the variability of your data¹⁶. Then, if you want to estimate the uncertainty on your final measurement s_m you may utilize the propagation of uncertainties (assuming uncorrelated variables)

$$u_{s_m} = \sqrt{\left(u_{E(s_m)} \frac{\partial C}{\partial E} \right)^2 + \sum_n \left(u_{a_n} \frac{\partial C}{\partial a_n} \right)^2} \quad (1.7)$$

where a_n are the parameters of the model estimated during curve fitting. Note that you do not only the uncertainty of the parameters of the fit but also of the output signal. This is expected but often disregarded as it is not trivial to compute the uncertainty $u_{E(s_m)}$ if s_m is outside the calibration set $\{s_i\}$. In that case, you may either construct an uncertainty model (e.g. spline interpolation), take the maximum uncertainty, or resort to some more advanced statistical methods.

Finally, you can compare the calibration curve and the transfer function to estimate the **accuracy** Δ of your sensor. As we saw previously, a good¹⁷ estimate is to **provide the maximum distance between the calibration curve and the transfer function**. Nevertheless, you should also consider the uncertainty previously determined as u_{s_m} . If the uncertainty is larger, it would be preferable to opt for this value as an estimate of the accuracy. Note that we are now in a condition to better estimate other parameters of our sensor such as **sensitivity**.

¹⁴ With four parameters I can fit an elephant. - John Von Neumann

¹⁵ such as the sum of squared residuals $\sum(C(E(s_i)) - s_i)^2$, as in the least squares method, but not exclusively

¹⁶ e.g. data with lower uncertainty are benefited

¹⁷ and safer!

Exercise 3. Computational Activity: Complete the activity in the Jupyter notebook for Week II that revolves around time-constant signals and constructing a transfer and calibration function.

Exercise 4. Hands-on Activity: With an Arduino, a CNY70 distance sensor, and a linear stage to precisely monitor the distance, follow the same strategy to create a transfer function and calibration function, and provide some specifications, for this simple yet real-world scenario.

1.2 Time-Varying Signals

Time-varying signals or simply time signals correspond to all the situations when the amplitude of a signal varies with time. Time signals can be separated into **deterministic** and **non-deterministic** signals. Deterministic signals follow an analytical formula or equation and can be either periodic or non-periodic. Non-deterministic are stochastic in nature and can be either stationary or non-stationary.

1.2.1 Time-domain Features

Mean and median: The average and middle value of the signal.

Standard Deviation and variance: measures the dispersion of the signal, with the variance being the square of the standard deviation.

Root Mean Square (RMS): Square root of the average of squared values. It is useful for assessing the power of the signal.

Peak-to-Peak: Difference between the maximum and minimum values in the signal.

Skewness: Measure of the asymmetry of the signal distribution.

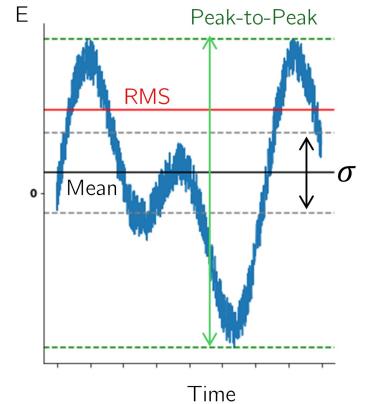


Figure 17. Some time-domain features.

1.2.2 Frequency-domain Features

Fourier transform is one of the most important and universal tools of mathematicians, engineers, and physicists. In general, it can be utilized for analyzing the frequencies contained within a signal or a function and its essence lies in its ability to decompose a complex signal into its constituent sinusoidal waves of varying frequencies, amplitudes, and phases.

Mathematically, the Fourier transform \mathcal{F} of a signal $f(t)$ is a function $F(\omega)$ given by the integral

$$\mathcal{F}[f(t)] = F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (1.8)$$

and transforms a time-domain signal into its dual frequency-domain representation. The signal can be recovered using the inverse Fourier transform

$$\mathcal{F}^{-1}[F(\omega)] = f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\omega \quad (1.9)$$

The Discrete Fourier Transform (DFT) is a variant of the Fourier Transform that is utilized for analyzing discrete signals. Its computational implementation is the Fast Fourier Transform (FFT), one of the most impactful

computer algorithms of all time¹⁸. There are also two-dimensional Fourier Transforms used for image processing and many more specialized forms.

Performing the Fourier Transform of the signal we can extract a few important features of our time signal which may contain relevant information for our sensor:

Spectral Centroid: The center of mass of the spectrum, being defined as $c = \sum_f f P(f) / \sum_f P(f)$ is the spectral centroid and $P(f)$ is either the magnitude (i.e. absolute value) of the Fourier Transform or its Power Spectral Density (see below).

Fundamental Frequency: The lowest frequency of a periodic waveform, determines its pitch.

Spectral Bandwidth: Usually the width of the band at one-half the peak maximum but in signal processing is common to define the spectral bandwidth as

$$BW = \sqrt{\frac{\sum_f (f - c)^2 P(f)}{\sum_f P(f)}} \quad (1.10)$$

where $c = \sum_f f P(f) / \sum_f P(f)$ is the spectral centroid and $P(f)$ is either the magnitude of the Fourier Transform or its Power Spectral Density (see below). This definition gives a more precise measure of the bandwidth as it accounts for the distribution of energy across the spectrum.

Spectral Entropy: Entropy of the spectral distribution, indicating its randomness. Usually computed as the Shannon entropy of the power spectral density (PSD).

At this point, we should note that in real-world applications noise may be present in the signal. In this scenario the Fourier transform and its magnitude may be insufficient to get a clear picture of the signal under analysis. In this context, the Power spectral density is a much more resilient tool:

Power Spectral Density (PSD): Measure of signal power in function of the frequency. The PSD is mathematically equivalent to the Fourier Transform of the autocorrelation function of the signal

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-2\pi i f \tau} d\tau \quad (1.11)$$

and can be considered as a normalized or averaged version of the squared magnitude of the Fourier Transform over multiple instances or realiza-

¹⁸ A nice video can be found [here](#).

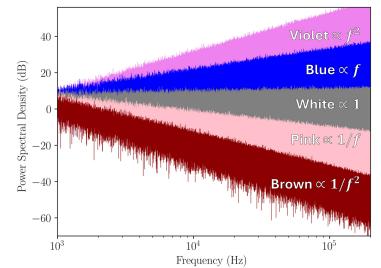


Figure 18. Analyzing the type of noise with Power Spectral Density.

tions of a process, especially in the context of random or time-varying signals. It is important to stress that **PSD is not equivalent to the magnitude of the Fourier transform**. It is rather utilized to analyze the power distribution over frequency of signals that are often considered over longer durations or as processes, being thus particularly useful when there is noise in the signal, as well as understanding noise properties and its characteristics.

As a rule of thumb, you may utilize the magnitude spectrum from the Fourier transform for identifying *what frequencies are present*. Yet, in the presence of noise, estimating the above quantities may benefit from using PSD as it answers *how is the signal's power distributed across frequencies*.

Magnitude of the Fourier Transform corresponds to the Energy, not the Power.

1.2.3 Time-frequency Features

The Fourier transform considers the signal as an infinite duration series and decomposes it into sinusoidal components. Yet, sometimes the frequency signature of the output signal may vary in time¹⁹ which can be analyzed in the form of a **spectrogram** - frequency *vs* time.

¹⁹ e.g. a Doppler sensor.

To construct a spectrogram we may use **short-time Fourier Transforms** (STFT) which is a mathematical technique used to analyze signals whose frequency components change over time. It combines the concepts of the Fourier Transform with a sliding time window to provide a two-dimensional representation of a signal, showing how its frequency content evolves. There are however two major problems of STFT, the computational load and the Heisenberg Uncertainty, which limits the simultaneous precision of measuring both time and frequency properties of a signal.

A possible workaround appears as **wavelet transforms**, mathematical tools that decompose a signal into components that are localized in both time and frequency. Wavelets are some kind of a **mathematical microscope** that you can utilize to analyze signals in detail, in particular those that have non-stationary characteristics (such as audio signals or images just to name a few examples). Wavelets are also very useful for analyzing **very large datasets** in an efficient manner.

To briefly grasp the working principles of wavelet transforms, the continuous wavelet transform of $f(t)$ may be defined by the coefficients

$$[W_\psi f](a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \psi \left(\frac{t-b}{a} \right) dt \quad (1.12)$$

with a a scale parameter and b a translation parameter. The mother wavelet ψ is a function localized in time and frequency, and thus, by varying scale a (scanning in frequency) and time with b (scanning in time), it is possible to

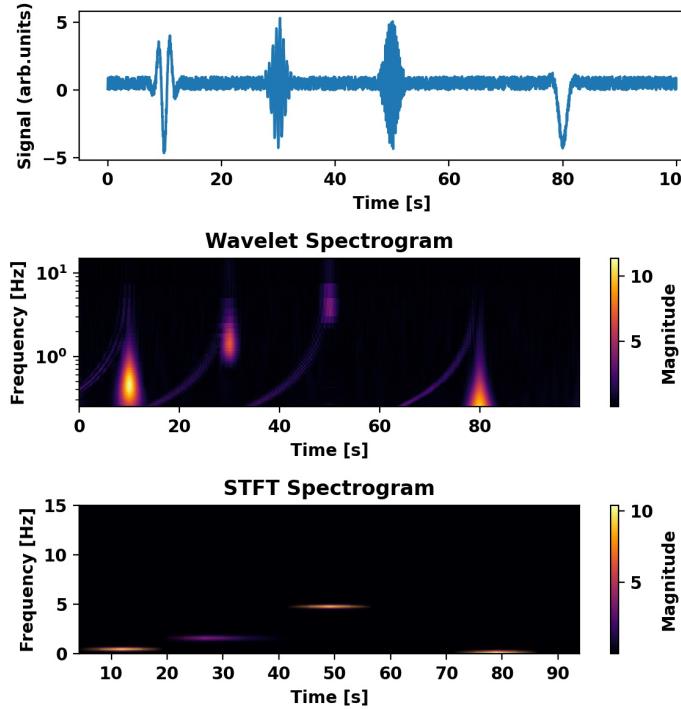


Figure 19. Spectrograms analysis with wavelets and short-time Fourier transform, highlighting the major problems of both: wavelets provide better time characterization, especially at higher frequencies, but do that at the cost of additional artifacts and non-linear scale in the frequency.

transform a signal into a localized distribution of frequencies in function of time via a **multi-resolution analysis**.

In practice, computational implementations utilize a discrete version - discrete wavelet transform - which utilizes a discretized computation of the integral to obtain the wavelet coefficients

$$c_{jk} = [W_\psi f](a = 2^{-j}, b = k2^{-j}) \quad (1.13)$$

with $a = 2^{-j}$ known as binary dilation and $b = k2^{-j}$ the binary position. This formulation is particularly useful in the case of an orthonormal wavelet, for which it is possible to recover the original signal from

$$f(t) = \sum_{j,k=-\infty}^{\infty} c_{jk} 2^{j/2} \psi(2^j t - k). \quad (1.14)$$

As k works as a linear translation in time, analyzing the variation of these coefficients at each k (e.g. centroid, bandwidth) may allow to extract of usable information in the context of sensing.

1.2.4 Statistical and correlation Features

So far we have focused on properties that are more useful in deterministic signals, whether these are periodic or non-periodic. If the signal is stochastic, however, the tools introduced previously may lack the necessary information. It is in this particular context that correlation appears.

Correlation, especially in the context of time-varying signals, is a fundamental tool in signal processing that measures the similarity or relationship between two signals as a function of the time-lag applied to one of them.

Correlation: Measures the similarity and relationship between two signals, mathematically defined

$$R_{fg}(t) = \int_{-\infty}^{\infty} f^*(t)g(t + \tau)dt \quad (1.15)$$

Auto-correlation: if $f = g$ and useful to identify periodic signals or the fundamental frequency of a signal;

Cross-correlation: if f and g are distinct, and useful for detecting if the signals are similar with a given time delay between them(similarly to pattern recognition).

Some key information that can be extracted:

- **Time delay estimation:** if two signals are similar with a lag, the time delay can be extracted as a maximum appearing in the correlation signal;
- **Signal Detection and Pattern recognition:** correlation is used to detect known signals within noisy environments, which is crucial for demodulating transmitted data in communications. Here, a known signal pattern (such as a **preamble** or **pilot** signal) is correlated with the received signal to detect the presence of data or to synchronize the receiver.
- **Noise reduction:** In periodic signals, correlating the signal with itself (autocorrelation) allows to enhance signal features while reducing noise, which does not correlate over time.

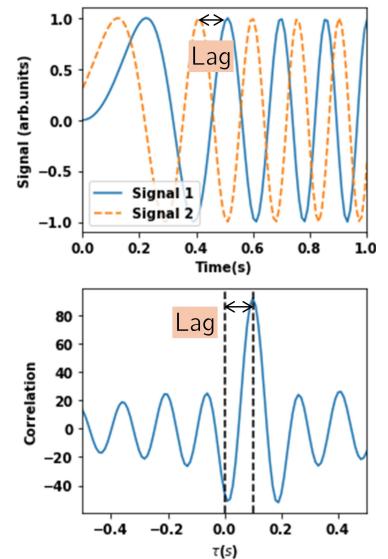


Figure 20. Determination of Lag using cross-correlation.

1.3 Signal Pre-Processing

Signal pre-processing is a critical step in the analysis and interpretation of data obtained from sensors. The goal of signal pre-processing is to improve the quality of the signal for further analysis, making it easier to extract useful information. We will focus on the most common challenges and important

Week

III

techniques, focusing on those that can be implemented both at the digital level as well as those that shall be applied at the analogic level.

1.3.1 Sampling Rate adjustment

Choosing a correct sampling frequency, i.e. frequency at which data is acquired from a sensor, is of major importance for correctly assessing time-varying data. To warrant good results, the criteria you should follow is given by the **Nyquist-Shannon theorem** that states that for recovering a signal at frequency f_0 you should sample at least at frequency $f_s \geq 2f_0$. Any value below this number will give an estimate of a frequency below the true frequency, an effect that is called **aliasing**.

An important thing you should note is that if additional frequencies are present, fake alias peaks coming from higher-frequency signals may appear in the spectrum. As this is not possible to verify prior the only way to prevent this effect is to apply spectral filtering (a low-pass, Which needs to be performed by hardware, before sampling).

1.3.2 Signal Filtering in the spectral domain

Filters are utilized to remove variations of the signal that do not contain relevant information to the sensor operation. Usually, filtering unwanted frequencies is one of the most common tasks in signal processing, in particular when we know the spectral range where the signal information is located. In this context low-pass, high-pass, band-pass, and band-reject filters are of particular importance. After sampling the data applying any kind of spectral filter in the Fourier domain is straightforward, and ideal filters can be created on-demand.

Yet, it can be inefficient in some cases, like removing aliasing effects as seen in the previous section, or if one needs real-time processing. In this context, one needs to apply filtering in the electrical domain. The Butterworth low-pass filter (see figure 22) is perhaps one of the most popular electronic filters, designed to filter out signal components above a certain frequency. Its gain function is given by

$$G(f) = \frac{1}{\sqrt{1 + (2\pi f)^{2N}}} \quad (1.16)$$

with N the order of the filter²⁰. Having higher N means sharper and better filtering, but comes at the cost of higher complexity of implementation.

1.3.3 Smoothing in time domain

Smoothing is a type of filtering that is applied in the time domain and thus may impact the whole frequency range of a signal. The most used smoothing filters include:

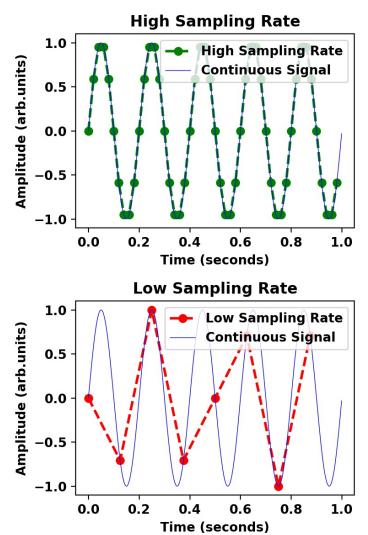


Figure 21. Aliasing.

²⁰ which translates into the number of reactive elements - inductors and capacitors - in the circuit.

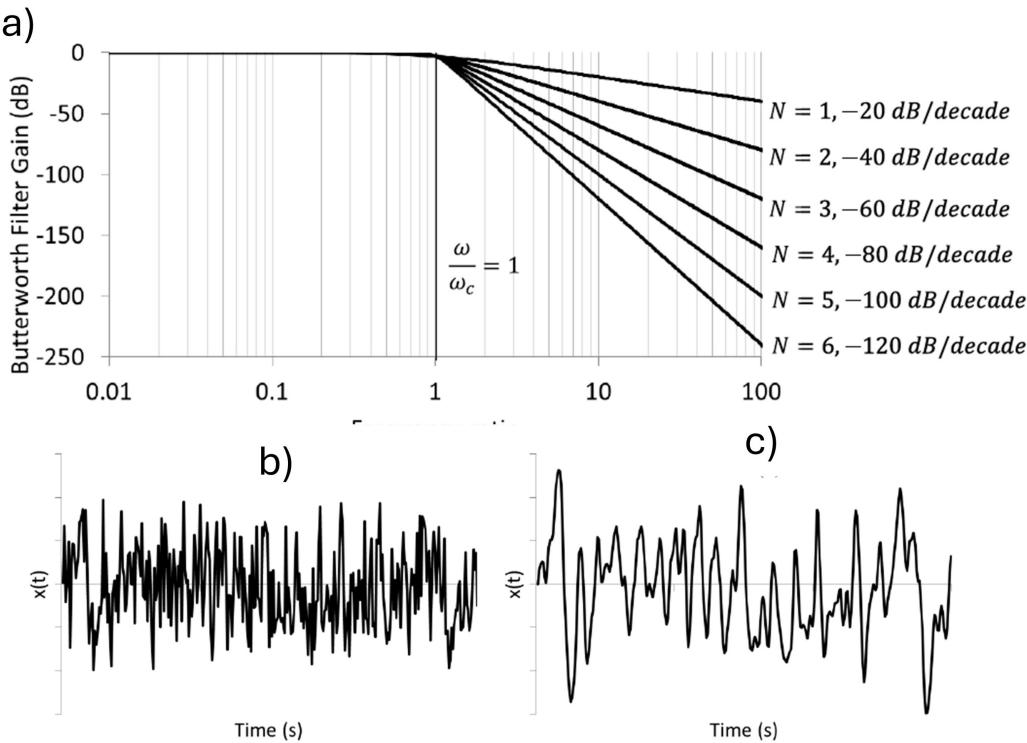


Figure 45 (a) Time waveform of a 7 Hz sine wave with added random noise. (b) The same time waveform after a Butterworth filter is applied with a cut-off frequency at 10 Hz.

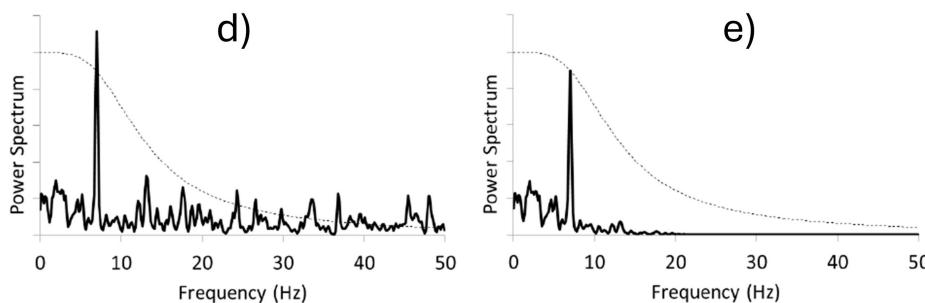


Figure 22. a) Butterworth filter gain function. b) Unfiltered and c) Filtered signal. Spectral analysis of the d) Unfiltered and e) Filtered signal.

Moving Average: Each point is replaced by an average of the neighboring data points, resulting in an effect that resembles a low-pass filter but that may attenuate the amplitude of the existing harmonics.

Savitzky-Golay: Each point is replaced by a generalized weighted average of the neighboring data points, whose coefficients (an odd number of them) are pre-computed and relate to the local polynomial fit. Compared

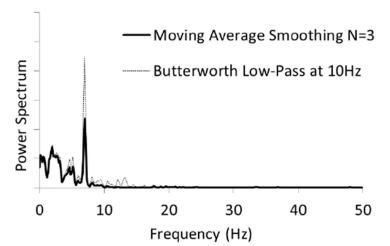


Figure 23. Moving average filter performance.

to the moving average it may preserve more of the original amplitude of the signal, but its performance in filtering noise may be lower.

1.3.4 Windowing and Leakage

Fourier analysis assumes that the signal is infinite, i.e., even if the signal is limited in time to a maximum number of samples, Fourier transform works as if that signal repeats over time. This introduces some problems when the measurement period is a non-integer multiple of the period of the signal frequency, which happens in typical conditions as perfect synchronization is not possible.

In this scenario, for the FFT algorithm the signal is no longer a pure harmonic signal and the true peak power tends to **leak** to other frequencies.

A way to mitigate leakage is to previously multiply the signal by a function - **window** - which forces the decay of the signal to zero in the limits of the acquisition period. There are multiple options for windowing, being the most utilized the Hamming, Hanning, and Blackman-Harris.

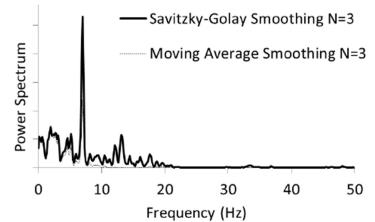


Figure 24. Savitzky-Golay filter performance.

2 Digital Signals

As we saw in the first week some sensors output analog signals while others prefer digital. This choice depends on various intertwined design preferences, from application requirements, to complexity, and even cost.

As a rule of thumb, **digital signals** may be preferable when

- **Communication:** digital signals are more resilient to degradation and thus easier to transmit over longer distances.
- **Integration:** If your sensor needs to interface with computers or digital systems, a digital output might be more straightforward to integrate.
- **Cost and Complexity of Digital Circuitry:** Digital circuitry can sometimes be cost-effective and compact, offering smart features like built-in calibration, error detection, and communication protocols. Besides, bypassing the need for an analog-to-digital converter is an important step in reducing the overall cost of the solution.

Analog signals may be preferable for the **design stage**, as well as **simpler applications, real-time applications, or high-resolution and sensitive applications**.

2.1 Typical conversion schemes

There are many analog-to-digital conversion schemes and they are not the main scope of this curricular unit. As such, we will just briefly mention two of the most used strategies: Pulse Width Modulation (PWM) and Pulse Density Modulation (PDM).

Pulse Width Modulation (PWM): PWM controls the power to devices by varying the width of the pulses in a fixed frequency signal. The average power delivered is proportional to the pulse width and thus a wider pulse increases the average voltage supplied to the load, while a narrower pulse decreases it. It is used in the control of motors or LEDs, simpler to control but re-conversion to analog may suffer higher levels of distortion.

Pulse Density Modulation (PDM): PDM represents an analog signal with the relative density of pulses in a sequence. Instead of varying the pulse width, PDM varies the number of pulses in a given time frame. A higher density of pulses represents a higher signal level, and a lower density represents a lower signal level. It is widely used in audio applications, such as digital microphones, as it offers higher resolution and is directly compatible with digital signal processing techniques. Yet is also more complex to implement and decode.

3 From sensors to automation: a brief note on controllers

Controllers are integral components of control systems used to regulate the behavior of other devices or systems. They interpret input data, often from sensors, and provide appropriate output commands to achieve desired system behavior. The objective is to maintain the system's output within a desired range, despite disturbances or changes in the environment. Among various types of controllers, Proportional-Integral-Derivative (PID) controllers, Fuzzy Logic controllers, Model Predictive Controllers (MPC), and Neural Network Controllers, just to name a few. For this curricular unit, we will focus on the most common PID controller.

A PID controller is a control loop feedback mechanism widely used in industrial control systems and a variety of other applications that require continuous monitoring and control. The working principle relies on the computation of an error value $e(t) = S - V$ as the difference between a desired setpoint S and a measured process variable V , leading to an output value

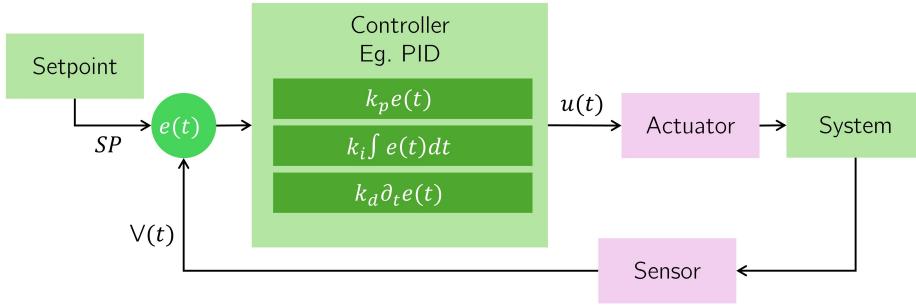


Figure 25. Conceptual role of a controller (e.g. PID) in a system monitored by a sensor.

to be passed to the actuator based on proportional, integral, and derivative terms, denoted by P , I , and D terms:

- **Proportional (P):** a proportional term to the current error value, $P = K_p e(t)$, acting as an immediate response to the deviation.
- **Integral (I):** The integral term is concerned with the accumulation of past errors, being given by $I = K_i \int_0^t e(t) dt$. It acts as an accelerator of the movement depending on the accumulated error allowing a softer response closer to the setpoint.
- **Derivative (D):** The derivative term is a prediction of future errors, based on the rate of change of the error and given as $D = K_d \partial_t e(t)$. It can provide a damping effect that reduces overshoot and improves system stability.

The output of the PID controller is thus given by $u(t) = P + I + D$. Note that the gain values shall be optimized for each system to reach the setpoint as fast as possible without significant overshoot or oscillations, and with minimal steady-state error. This can be done using methods from manual tuning, for tuning these parameters, including manual tuning, Ziegler-Nichols heuristics, or software-based optimization techniques.

4 Machine Learning and Sensors

The recent advent of machine learning has fostered significant developments in the way that we collect and analyze data gathered from sensors. In particular, machine learning²¹ offers an interesting potential for dealing with more complex scenarios in particular:

- **Highly nonlinear and multiparameter responses:** When you have multiple parameters you can have highly complex and nonlinear transfer

²¹ to be precise, numerical regression is already a type of machine learning!

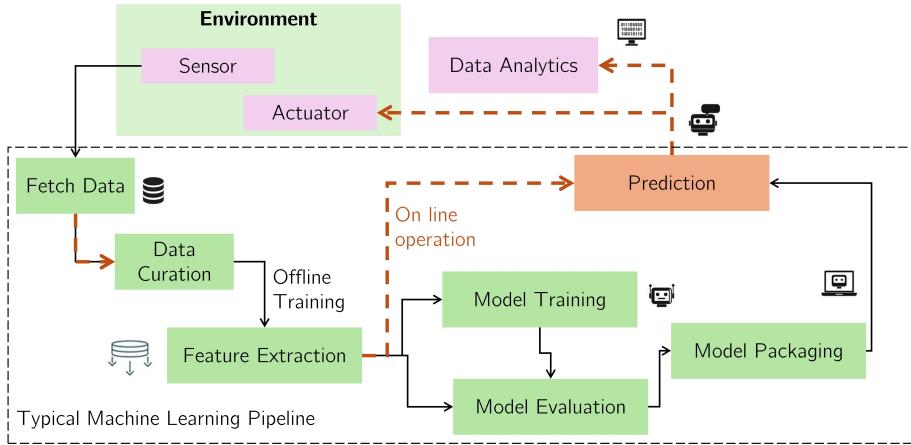


Figure 26. Typical Pipeline of Machine learning and a conceptual example of its application with a sensor.

functions. In this case, function fitting may be challenging to achieve²² and data-driven regressions such as K-Nearest Neighbors regression, Support vector machines, or multilayer neural networks may be helpful to construct reliable calibration models.

- **Sensor fusion:** Combining the information from multiple sensors is a particularly interesting topic of research. You can combine multiple measures with similar sensors (like taking a photo from distinct angles) but you can also combine measurements from very distinct sensors (e.g. combining distinct techniques, as different as sound and reflectance).
- **Data Preprocessing:** It is possible to train multiple algorithms to learn preprocessing tasks, from noise filtering to feature extraction that can be applied directly to the raw sensor data.
- **Automated Pattern Recognition:** Sometimes the information provided by sensors is extremely complex (e.g. unusual vibrations detected by installed accelerometers) to recognize complex patterns. This has many applications in industry fields such as predictive maintenance, where active maintenance strategies may be implemented to reduce downtime, predict failures, and ultimately save costs. Other interesting applications are healthcare and wearable sensors.

Before ending the discussion it is important to leave some considerations so you can learn from previous mistakes of others. Machine learning is a great tool but it is not a silver bullet and should be used only in a very informed way²³.

First, regarding models, it is always preferable to, when possible, use the standard calibration curve procedure as presented before this week. Machine

²² you would need a good estimate of a prior mathematical model

²³ Little learning is a dangerous thing.

learning virtues can often be deceiving, and being easy to train a simple regression may seem. Yet this approach may overlook some problems (e.g. outliers) that can be circumvented otherwise. If not learning models that relate to the sensing principle to black-box machine learning models.

Secondly, regarding data, take into consideration the most important saying of Machine learning: **Garbage in, Garbage out**. This saying synthesizes that the effectiveness of ML models heavily depends on the quality and quantity of the data collected by sensors. So, when possible, informed feature extraction procedures are always preferable to naive brute-force approaches. Besides, take into consideration the size of the input data that you use to train the model, which should be much larger than the number of free parameters to avoid overfitting.

Finally, consider that complex machine learning models may require the use of microprocessors or even processors. In this case, be careful to use energy-efficient algorithms and consider the possibility of performing the computing closer to the sensing head²⁴, avoiding data transmission and associated latency.

Exercise 5. Hands-on Activity: In addition to the signal analysis procedures there is a Jupyter notebook that concerns sound signal analysis and classification. There, we explore how to use a simple sensor and a machine learning system to deploy a simple classifier of spoken words.

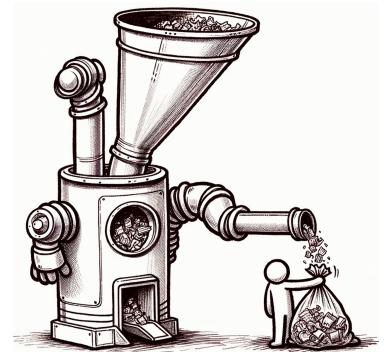


Figure 27. Garbage In, Garbage Out: perhaps the most important concept of Machine Learning.

²⁴ usually referred to as *edge computing*

5 Concluding Remarks

In this second week, we focused on signal processing, i.e., extracting good and meaningful information out of the output signal of a sensor. We covered the analysis of static signals, in particular focusing on the topic of distributions and uncertainties, allowing us to measure a transfer function from the extracted information and to build a calibration model. Then, focusing on time-varying signals, we introduced tools to extract information that can be used for the purpose. We briefly mentioned some topics on signal filtering, digital signals, and machine learning, so you can have a wide perspective of their role in modern sensors.

So far you already know:

- What is a sensor and how to characterize one;
- How to extract information from the output signal of a sensor and construct a calibration function;

and therefore we are ready to start building some of the most advanced and sensitive sensors: interferometric optical sensors.

Enabling Optical Sensing with Wave Interference

Week

IV

Light-based sensors constitute one of the most cutting-edge toolboxes of a modern-day physicist or physical engineer. Indeed, optical sensing is utilized every day in various domains and at the most diverse scales, being useful to assess either the ultra-small microscopic world or the extremely large scales of the universe and astronomy.

Sensing with light features major advantages including the most relevant:

- **High sensitivity:** utilizing principles such as interferometry or resonant light matter interaction;
- **Non-contact measurement:** Most of the sensors operate without needing to physically contact the substances they are measuring, which is particularly useful for monitoring delicate or hazardous materials;
- **Fast response time and Wide dynamic range:** Enabling real-time monitoring and control in dynamic environments.
- **Resistance to electromagnetic interference:** Unlike electronic sensors, optical sensors are generally immune to electromagnetic interference;
- **Multiplexing:** signals at distinct wavelengths do not interfere in linear optical systems, allowing the combination of multiple sensors and multiple locations in a single fiber for example;
- **Remote sensing:** Optical sensors can measure properties over long distances, making them ideal for remote sensing applications and for operation in harsh environments (e.g. space, corrosive atmospheres, high pressures, etc.);
- **Minimal footprint:** Many optical sensors, especially those based on fiber optics, are small and lightweight, thus easy to integrate into critical environments (e.g. space);

One of the ways²⁵ ways to categorize optical sensors is to divide them into three families:

Section 1. Wave Optics in a Nutshell
Section 2 Optical Interferometry
Section 3. Overview of Experimental Components

Table 3. Contents for WEEK IV

Interferometric Sensors: Sensors that exploit the interference of light waves at sub-micrometer scale to measure physical changes.

Light-Matter Interaction Sensors: Including a wide range of sensors that detect changes due to changes in the interaction of light and materials. In particular, this broad category can include photodetectors, spectroscopy-based sensors, distributed sensing based on scattering and polarization-based sensors, and plasmonic sensors.

Quantum Optical Sensors: Utilize quantum properties of light such as entanglement and quantum interference (not exactly the same as classical interference) to monitor environment changes.

In this curricular unit, we will focus essentially on Interferometric sensors. Yet, we will also present quantum optical sensors with some detail later in the semester, and briefly introduce light-matter interaction sensors with a topical presentation at the end of the semester. This week we will focus on interferometric sensors, introducing the necessary theoretical concepts, working principles, and optical elements to construct them in both free space and optical fiber configurations.

1 Wave Optics in a Nutshell

The nature of light - whether it behaves as a particle or a wave - is one of the most fascinating chapters of physics and a long discussion that spans centuries. This debate involves significant evolutions of scientific reasoning, a lot of new experimental techniques and evidence, and a very interesting history-politics-science crossover that even involves Napoleonic campaigns²⁶. Looking in retrospective, it can be said that there have been three big debates in the past about this topic, the first dating back to the 17th century and started by Isaac Newton and Christiaan Huygens.

Motivated by his successful interpretation of mechanics and forces, Newton advocated in his famous work *Opticks* (1703) that light was made up of corpuscles (small particles), explaining reflection and refraction with concepts closely related to classical mechanics. Yet, a decade before the publication of *Opticks*, Huygens²⁷ published *Traité de la Lumière*(1690), advocating that light was a wave, propagating layer by layer in the ether with each point in the wavefront acting as a spherical source - *Huygens Principle*. Now we know that Newton's ideas such as optical dispersion being related to particles of each color having distinct mass are far from being correct, while the Huygens Principle is still an interesting analogy do describe light propagation.



Figure 28. Huygens and Newton, same hairstyle but distinct sides on the wave-particle debate.

²⁶ The book "The History of Optical Interferometry and the Scientists Who Tamed Light" by David Nolte presents a very interesting historical and scientific vision on the topic.

²⁷ Who also invented the pendulum clock of our grandfather's house.

tion. But at that time Newton's immense influence in the academic world crystallized the vision of light for almost a century. The debate restarted in the 19th century with the works of Thomas Young and Fresnel. The double slit experiment by Young lectured in 1802 at the famous Royal Institution resulted in interference patterns that would only be explained if light behaved as waves. Although the double-slit sets a hallmark, Young was a polymath and a physician and never pursued a detailed mathematical formalism. This mathematical formalism was later achieved in France with Fresnel, which by the decade of 1810 was dedicated to refining the mathematical theory of Huygens, exploring light as a wave to explain more complex phenomena such as diffraction patterns around obstacles. Supported by the influent Arago, Fresnel established multiple mathematical methods based on wave behavior which consolidated the wave theory for light. A few years later, and on the other side of the English Channel, the self-taught Michael Faraday was producing remarkable experimental results in the interplay of electric and magnetic fields. In particular, his discovery of the law of induction and later of the Faraday effect²⁸ solidified a clear connection between both fields. These works inspired the work of the well-educated Faraday, unifying electric, and magnetic effects into a vector field formulation in his famous set of four partial differential equations. Besides, predicting the existence of electromagnetic waves that would propagate even in a vacuum, the classical theory for light that to the day we still use was essentially set. A third debate ignited later in the 20th century, with Planck, Einstein, and the advent of quantum mechanics. But we will come to that later in this curricular unit.

1.1 Maxwell's Equations and Wave Equation

For the next few weeks, we will focus on a classical description of light as an electromagnetic field. This picture is complete within the Maxwell equations formalism, which explains the propagation of electromagnetic waves in terms of fields \mathbf{E} , \mathbf{D} , \mathbf{B} , and \mathbf{H} . The electric displacement field \mathbf{D} and magnetic field \mathbf{B} are usually regarded as *internal fields* as they relate to the effect of fields \mathbf{E} and \mathbf{H} inside a material, being defined as

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (1.1)$$

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}, \quad (1.2)$$

where \mathbf{P} and \mathbf{M} correspond to the polarization (associated with bound charges formed in the material in the presence of an electric field) and magnetization (associated with bound currents in materials), respectively. In short, these assisting fields enclose light-matter interaction, and although a



Figure 29. Thomas Young, the polymath who shed the first light on deciphering wave optics and also the Rosetta stone.

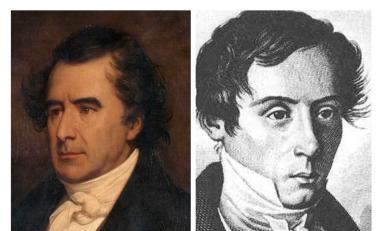


Figure 30. Arago and Fresnel, the French side.

²⁸ *in short, light polarization may be affected by a magnetic field.*

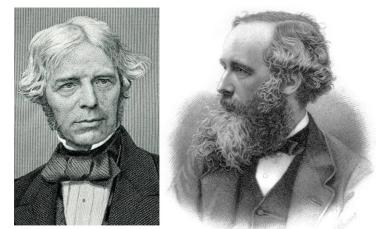


Figure 31. The experimentalist layman Faraday, and the theoretician Maxwell.

qualitative picture of both polarization and magnetization may be obtained under classical mechanics, you may require quantum mechanics to accurately predict these true observables.

The Maxwell's equations are:

$$\nabla \cdot \mathbf{D} = \rho \quad (1.3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.4)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.5)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \quad (1.6)$$

In brief, the first and second equation corresponds to the Gauss law, which relates the flux across a closed boundary with the presence of sources inside. The third corresponds to Faraday's law of induction and states that a time-varying magnetic field produces an associated electric field. Finally, the last equation corresponds to Ampere's law and states that a magnetic field can be created either by currents or time-varying electric fields.

As it can be seen, all the formalism above is sufficient to describe the propagation of waves in medium provided we know a model for the auxiliary fields and currents, i,e, the **constitutive relations**. Assuming linear relationships²⁹ we may define:

²⁹ is this always valid?

$$\mathbf{D} = \varepsilon \mathbf{E} \quad (1.7)$$

$$\mathbf{J} = \sigma \mathbf{E} \quad (1.8)$$

$$\mathbf{B} = \mu \mathbf{H} \quad (1.9)$$

where the scalars that contain an appropriate description of the light-matter interaction are the ε relative permittivity / dielectric constant, σ the conductivity, and μ the magnetic permeability.

1.2 Wave equation, Waves, and Wavefronts

The two last Maxwell's equations allow us to explain in the macroscopic regime and far from the source how light can propagate in space as a field(the far field): a time-varying electric field leads to a time-varying magnetic field, which in turn produces a time-varying electric field again and again. Assuming that:

- ε and μ are spatially uniform and time constant(generally not true);
- There are no free charges $\rho = 0$ nor currents $\mathbf{J} = 0$;

we may take the rotational of the Faraday law to get

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla \times \frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t} \nabla \times \mathbf{B} = -\varepsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad (1.10)$$

Then, by utilizing the vector calculus identity for the double rotational

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \quad (1.11)$$

and recovering the free charge assumption, i.e. $\nabla \cdot \mathbf{E} = 0$, we end up with the wave equation for the electric field. Doing the same for Ampere's law we obtain a set of wave equations:

Electromagnetic wave equations:

$$\nabla^2 \mathbf{E} - \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad (1.12)$$

$$\nabla^2 \mathbf{B} - \mu\varepsilon \frac{\partial^2 \mathbf{B}}{\partial t^2} = 0 \quad (1.13)$$

for which a wave velocity can be obtained as

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \quad (1.14)$$

which in turn defines the **index of refraction** as

$$n = \frac{c}{v} \quad (1.15)$$

where $c = 1/\sqrt{\mu_0\varepsilon_0}$ corresponds to the speed of light in vacuum.

There are multiple possible solutions for the wave equation, being the most common in the context of optics the plane wave solution, the spherical wave solution(far-field solution of a point-source, e.g. a pin-hole), and the cylindrical wave equation(far-field solution of a line-source, e.g. a slit).

The **plane wave solution** is a type of harmonic wave solution that satisfies the wave equation 1.12 and is given by the ansatz³⁰

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_0)} \quad (1.16)$$

where $\omega = 2\pi f$ corresponds to the angular frequency of the wave (f wave frequency³¹), $k = 2\pi/\lambda$ to the wavenumber (λ wavelength), and ϕ a phase at $t = 0$ and $r = 0$. Besides, the direct substitution of the ansatz leads to the relation

$$\omega = \frac{c}{n} k. \quad (1.17)$$

³⁰ we opted for the $-\omega$ convention

³¹ f from 400THz to 700THz and λ from 700nm to 400nm for visible spectrum

The term plane wave is related to the shape of the **wavefront** - points in space that oscillate in unison thus corresponding to the same phase of oscillation - which in this case corresponds to a plane. For spherical waves, the wavefront is a spherical surface, and for cylindrical waves, we have a cylindrical wavefront. Any deviations from the desired wavefront are called **aberrations**.

Some key remarks include:

Light is a transverse wave in free space: This is easily obtained from $\nabla \cdot \mathbf{E} = 0$ and means that electric and magnetic fields propagate as transverse waves, i.e. the amplitude of both fields oscillate perpendicular to the direction of propagation defined by \mathbf{k} . This transverse nature normally holds in optical media but is not always true in special cases such as conductive media (related to the presence of plasmons), in the near-field (close to sources), and as guided or evanescent waves.

\mathbf{E} and \mathbf{B} are not independent: Indeed from the plane wave solution and taking the Faraday's law it is easy to demonstrate that

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B} \quad (1.18)$$

meaning the two fields are perpendicular and $|\mathbf{E}| = \frac{c}{n} |\mathbf{B}|$. Note however this is only valid under the assumptions previously defined.

Polarization: Polarization of the wave corresponds to the direction of the electric field. Indeed, \mathbf{E} is a vector that is perpendicular to \mathbf{k} . Taking the most common choice in optics $\mathbf{k} \parallel \mathbf{e}_z$ we may have:

- **Linear polarization:** if the field always oscillates along a given direction, e.g. $\mathbf{E} \parallel \mathbf{e}_x$;
- **Circular polarization:** if the field rotates once per cycle. A circular basis can be constructed as $\mathbf{e}_{R,L} = (1/\sqrt{2})(\mathbf{e}_x \pm i\mathbf{e}_y)$ with \pm associated to right and left polarization states, respectively;
- **Elliptical polarization:** an extension of the circular polarization, but with different amplitudes for each cartesian component or phase difference $\theta \neq \pi/2$. A possible basis for this case is defined as $\mathbf{e}_{el,\pm} = (1/\sqrt{a^2 + b^2})(a\mathbf{e}_x \pm b e^{i\theta} \mathbf{e}_y)$ with a and b being two amplitudes and θ a phase difference.

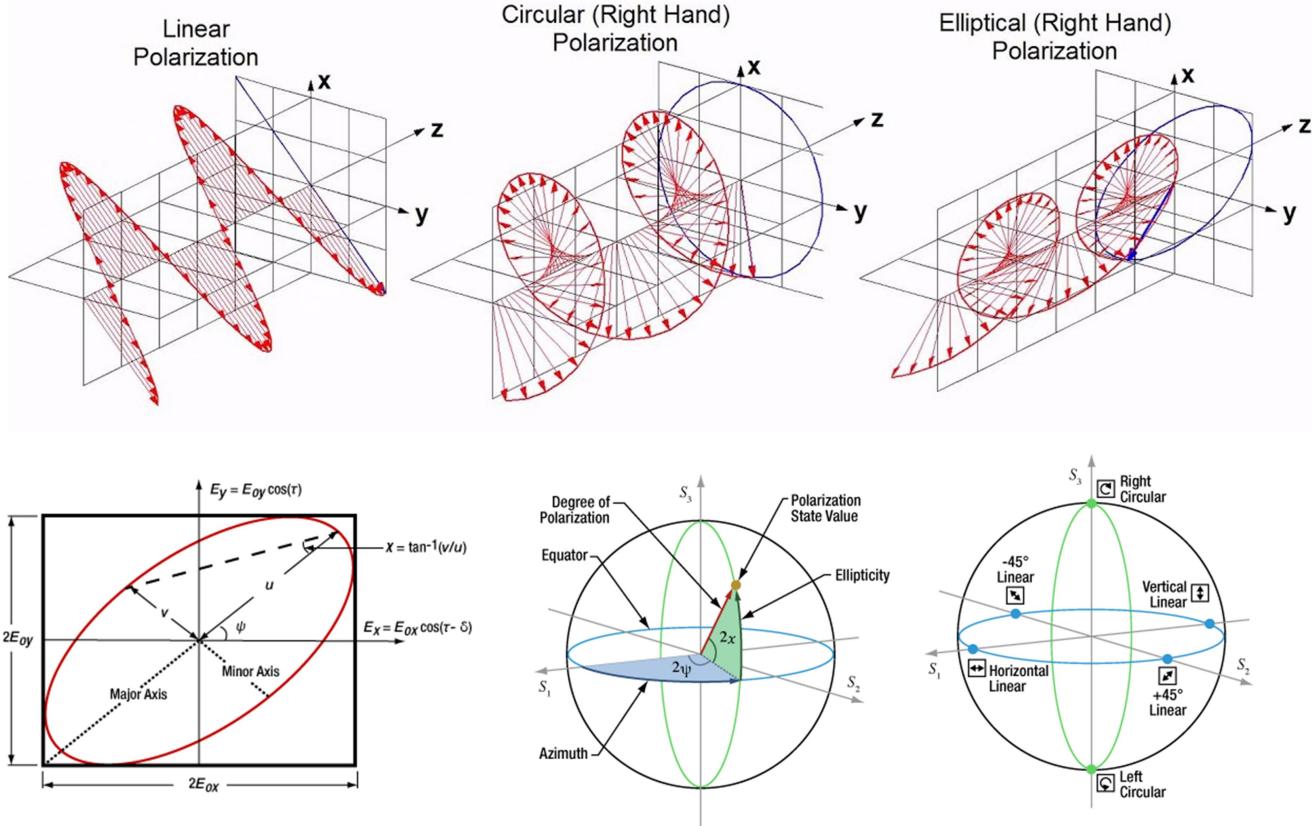


Figure 32. Visual representation of the distinct polarization states, demonstrating the absence or presence of rotation of the electric field transverse to the propagation direction. Note that in the complex field notation, this rotation along the z axis appears when you take the real value of the electric field. In the bottom row, we introduce a tool called **Poincaré Sphere** which allows us to easily identify the state of polarization(SOP) as a single vector, and possibly see some trajectories. It is a tool widely utilized in SOP sensing strategies.

Definition. Energy Density and Poynting Vector: The energy density associated with an electromagnetic wave is given by

$$U = \frac{\mathbf{D} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{H}}{2} \text{ (joules/m}^3\text{)}, \quad (1.19)$$

and the flow of the field energy is described by the **Poynting Vector**

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (1.20)$$

Exercise: Compute the energy density and Poynting vector for a plane wave in (i)vacuum and (ii)non-magnetic medium.

1.3 Reflection and Refraction

Now that we know what to expect when light propagates in homogeneous media it is important to introduce what happens when light encounters a boundary between different optical media of distinct refractive index n_1 and n_2 . For that, we establish a coordinate system as presented in Figure 33, for which we choose the condition $k_i^y = 0$. In this coordinate system, we define the *plane of incidence* as the plane containing the incident wave vector \mathbf{k}_i as well as the normal to the boundary.

First, just from D'Alembert approach to wave equation, we are expecting the waves to be solutions of the form $f(\omega t - \mathbf{k} \cdot \mathbf{r})$. This simple observation imposes that

$$\omega_i t - \mathbf{k}_i \cdot \mathbf{r}|_{z=0} = \omega_r t - \mathbf{k}_r \cdot \mathbf{r}|_{z=0} = \omega_t t - \mathbf{k}_t \cdot \mathbf{r}|_{z=0} \quad (1.21)$$

and as it should hold for any time t , it is straightforward to conclude that the frequency of the wave has to be equal at the boundary, i.e. $\omega_i = \omega_r = \omega_t$.

Now, taking the dispersion relation defined previously $k_i = n_i \omega/c$, we may take the equation 1.21 at time $t = 0$ and boundary $z = 0$ to obtain the conditions in our coordinate system

$$n_1 \sin \theta_i = n_1 \sin \theta_r = n_2 \sin \theta_t. \quad (1.22)$$

The first part of the equation,

$$\sin \theta_i = \sin \theta_r \quad (1.23)$$

gives the **Law of reflection** and states that incidence and reflectance angles are the same. The second part of the equation gives the **Law of refraction** or **Snell's law**

$$n_1 \sin \theta_i = n_2 \sin \theta_t. \quad (1.24)$$

Having the angle part solved just from a general solution of the wave equation let's look now at the amplitudes of each wave. To do this it is important to establish some boundary conditions, which will assist in solving this problem:

- When there are no surface charges, the normal components of \mathbf{D} is continuous across the boundary;
- The normal components of \mathbf{B} is also continuous across the boundary;
- The tangential components of \mathbf{E} are continuous across the boundary;

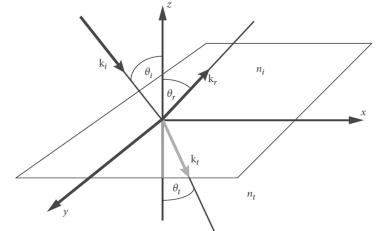


Figure 33. Coordinate system for computing the reflection and refraction angles.

- The tangential components of H are continuous across the boundary when there are no surface currents;

For a medium free of charges and currents, it is sufficient to consider the continuity of the tangential components of E and H . Note however that we need now to take into account the polarization direction. For that we will define a p-parallel polarization and s-perpendicular polarization according to the coordinate system of Figure 34. For the \mathbf{E} we can then write

$$(\mathbf{E}_i + \mathbf{E}_r - \mathbf{E}_t) \times \hat{\mathbf{n}} = 0 \quad (1.25)$$

which on our coordinate system means $\mathbf{E} \times \hat{\mathbf{n}} = E_y \hat{\mathbf{i}} - E_x \hat{\mathbf{j}}$ and taking $E_y = E_{Ni} + E_{Nr} - E_{Nt}$ and $E_x = E_{Pi} \cos \theta_i - E_{Pr} \cos \theta_i - E_{Pt} \cos \theta_t$ gives the conditions:

$$E_{Ni} + E_{Nr} = E_{Nt} \quad (1.26)$$

$$(E_{Pi} - E_{Pr}) \cos \theta_i = E_{Pt} \cos \theta_t. \quad (1.27)$$

Similarly, for the tangential continuity of \mathbf{H} , utilizing the relation of $\mathbf{H} = \frac{\sqrt{\mu\epsilon k}}{k} \times \mathbf{E}$ and considering that in our coordinate system

$$(\mathbf{k} \times \mathbf{E}) \times \hat{\mathbf{n}} = (E_z k_x - E_x k_z) \hat{\mathbf{i}} - E_y k_z \hat{\mathbf{j}} = 0 \quad (1.28)$$

we get

$$\left[\frac{1}{k_i} \sqrt{\frac{\epsilon_i}{\mu_i}} (\mathbf{k}_i \times \mathbf{E}_i + \mathbf{k}_r \times \mathbf{E}_r) - \frac{1}{k_t} \sqrt{\frac{\epsilon_t}{\mu_t}} (\mathbf{k}_t \times \mathbf{E}_t) \right] \times \hat{\mathbf{n}} = 0 \quad (1.29)$$

leading to

$$\sqrt{\frac{\epsilon_1}{\mu_1}} (E_{Pi} - E_{Pr}) = \sqrt{\frac{\epsilon_2}{\mu_2}} E_{Pt} \quad (1.30)$$

for the x component and

$$\sqrt{\frac{\epsilon_1}{\mu_1}} (E_{Ni} - E_{Nr}) \cos \theta_i = \sqrt{\frac{\epsilon_2}{\mu_2}} E_{Nt} \cos \theta_t \quad (1.31)$$

for the y component.

S (Perpendicular) polarization: For perpendicular polarization (to the plane of incidence) we only need to consider the two equations involving E_N terms. For transmission, it can be shown that

$$t_s = \frac{E_{Nt}}{E_{Ni}} = \frac{2}{1 + \frac{\mu_1 \tan \theta_i}{\mu_2 \tan \theta_t}} \approx \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t)} \quad (1.32)$$

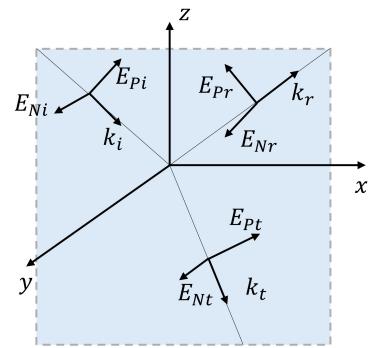


Figure 34. Coordinate system and plane of incidence.

with the approximation valid for $\mu_1 \approx \mu_2$. For the reflection coefficient we have

$$r_s = \frac{E_{Nr}}{E_{Ni}} = \frac{-\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}. \quad (1.33)$$

In particular, note that a π phase shift is observed in the reflected wave if the index of refraction satisfies $n_2 > n_1$.

P (Parallel) polarization: For parallel polarization (to the plane of incidence) we only need to consider the two equations involving E_t terms. It can be shown that

$$t_p = \frac{E_{pt}}{E_{pi}} = \frac{2 \cos \theta_i \sin \theta_t}{\cos \theta_t \sin \theta_t + \frac{\mu_1}{\mu_2} \cos \theta_i \sin \theta_i} \quad (1.34)$$

for the transmission coefficient and

$$r_p = \frac{E_{Pr}}{E_{Pi}} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \quad (1.35)$$

Note that if $\tan(\theta_i + \theta_t) = +\infty$ we may obtain no reflection for the parallel polarization. This situation is called the *Brewster's angle*.

Exercise 6. Demonstrate that a plane wave incident normally in a slab of an optical medium of length L and refractive index n acquires a phase difference term $\Delta\phi = knL$. What happens if the incidence angle is θ ?

2 Introduction to Optical Interferometry

One of the key ingredients of linear optics is the superposition principle, which warrants that the sum of two solutions of the wave equation is also a solution of the wave equation itself. While the effects of this overlap are not seen in the field directly, a square law detector³² reveals bright and dark bands called **fringes** which we could relate to the concept of **interference**.

Interference occurs when two or more coherent light waves (waves of the same frequency and constant phase difference) overlap. In particular, by definition, we say that *constructive interference* occurs where the waves are in phase (the crest of one wave meets the crest of another), leading to increased light intensity, whereas waves out of phase (the crest of one wave meets the trough of another) results in *destructive interference* with decreased intensity.

To understand the basics of wave interference we may consider the simple

³² a detector which outputs a signal proportional to the intensity.

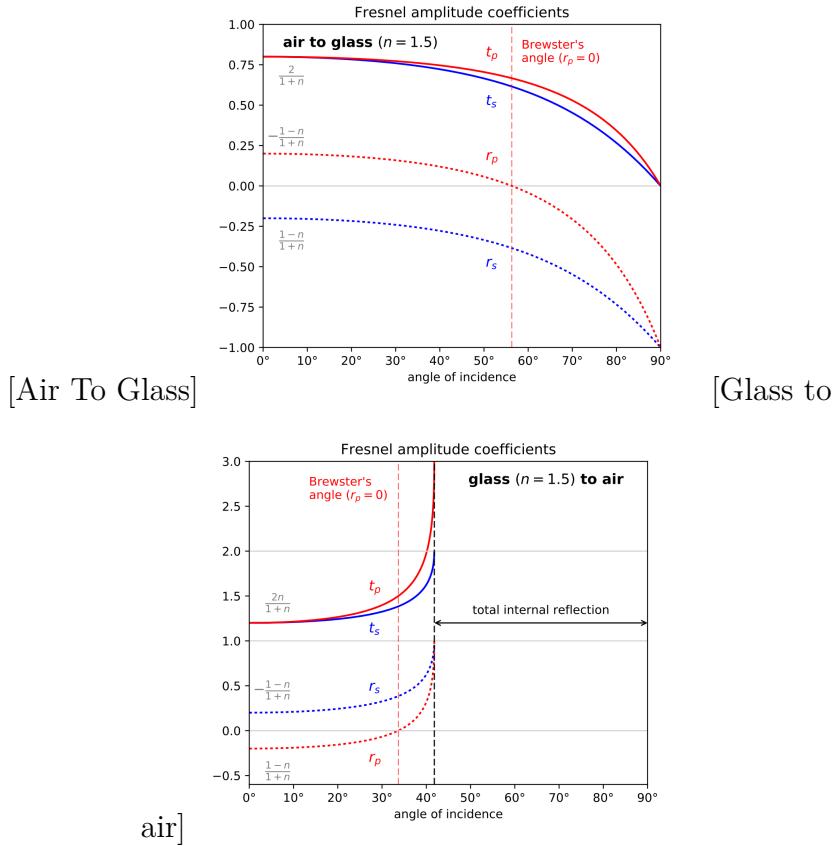


Figure 35. Fresnel amplitude coefficients. Note the minus signal corresponding to a phase shift.

case of two plane waves in the complex notation,

$$\mathbf{E} = E_1 e^{i(-\omega_1 t + \mathbf{k}_1 \cdot \mathbf{r}_1 + \phi_1)} \mathbf{e}_1 + E_2 e^{i(-\omega_2 t + \mathbf{k}_2 \cdot \mathbf{r}_2 + \phi_2)} \mathbf{e}_2 \quad (2.1)$$

where \mathbf{e}_i is a unit polarization vector for each field and \mathbf{r}_i the spatial position where the amplitude is to be added in relation to the source. The intensity of the wave can be calculated as

$$I = \frac{cn\epsilon_0}{2} |\mathbf{E}|^2 \quad (2.2)$$

$$= \frac{cn\epsilon_0}{2} (\mathbf{E}_1 \cdot \mathbf{E}_2^* + \mathbf{E}_2 \cdot \mathbf{E}_1^* + 2 \operatorname{Re}\{\mathbf{E}_1 \cdot \mathbf{E}_2^*\}) \quad (2.3)$$

$$= I_1 + I_2 + 2 \frac{cn\epsilon_0}{2} \operatorname{Re}\{E_1 E_2^* e^{i((\omega_2 - \omega_1)t + (\mathbf{k}_1 \cdot \mathbf{r}_1 - \mathbf{k}_2 \cdot \mathbf{r}_2) + \phi_1 - \phi_2)} \mathbf{e}_1 \cdot \mathbf{e}_2^*\} \quad (2.4)$$

$$= I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\Delta\omega t + (\mathbf{k}_1 \cdot \mathbf{r}_1 - \mathbf{k}_2 \cdot \mathbf{r}_2) + \Delta\phi) \operatorname{Re}\{\mathbf{e}_1 \cdot \mathbf{e}_2^*\} \quad (2.5)$$

and the interference signal is contained in the last term of the above equation, where $\Delta\omega_{21} = \omega_2 - \omega_1$ and $\Delta\phi = \phi_2 - \phi_1$. Some considerations:

- If the polarizations are collinear the interference is maximized, whereas

if perpendicular there will be no interference.

- The dependency on time is trickier than it seems at first sight. First, in real-world conditions ϕ_2 and ϕ_1 may change with time (stochastic phase jumps), meaning that for obtaining a proper interferometric signal special care must be taken concerning the coherence length. Also, for distinct time frequencies, a beating may appear and its detection will strongly depend on the frequency shift $\Delta\omega_{21}$. In the most trivial case of the same frequencies, this is not an issue.

For the next sections, we will consider the case of interference of waves with same frequency and polarization and assume that over a given time period (integration time or exposure) the phase difference $\Delta\phi = \phi_1 - \phi_2$ maintains constant, for which we obtain the temporal average $\langle \cdot \rangle$ as

$$\langle I \rangle = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\mathbf{k}_1 \cdot \mathbf{r}_1 - \mathbf{k}_2 \cdot \mathbf{r}_2 + \Delta\phi) \quad (2.6)$$

So far we disregarded the fact that the envelope function $E_i(x, y, z)$ may also contain a transverse spatial distribution. If that is the case, an explicit dependence will appear and the intensity of the interference may vary due to the variation in the intensity of each wave, which may affect visibility. This can be even more problematic in the presence of multiple wave interference or many modes as it happens in multimode waveguides.

Definition. Contrast or visibility in interference: The contrast of interference fringes may be defined as a normalized value between 0 and 1 as

$$\text{Contrast} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \quad (2.7)$$

The contrast depends on multiple factors, with the most important being the **polarization state**, the **coherence of the source**, the **amplitude** of the waves, and the **spatial modes or envelope functions**.

Finally, note that it is not problematic to add a spatial distribution to the phase difference, which ultimately establishes the principles of digital off-axis holography.

2.1 Newton Rings

Along with the thin film interference, the observation of Newton's Rings is perhaps one of the first reported evidence of light interference. Besides, and although the observation pre-dates Newton (Robert Hooke in 1665's *Micrographia*), it is quite interesting to note that Newton himself observed

this effect but never suspected that wave dynamics would be the principle underlying such phenomena.

Newton's rings refer to the pattern that emerges when a non-planar glass surface (e.g. convex lens) is placed on top of a flat glass surface. It may be observed either in transmission or reflection and although it can be observed with white light in the case of very small gaps (comparable with visible wavelengths), it is far more visible when one utilizes a monochromatic and coherent light source to illuminate the setup from above.

Let us now model the distance r for which the Newton's Rings appear. To obtain a model for this let us consider that the air gap between the lens and the flat surface varies in thickness from zero at the point of contact to a small thickness t as a function of r as in Figure 36. It is easy to show that for a lens of radius R we will have

$$R^2 = (R - t)^2 - r^2 \Leftrightarrow r^2 = 2Rt - t^2 \Leftrightarrow r \approx \sqrt{2Rt} \quad (2.8)$$

where the last approximation is valid for $R \gg t$.

For the reflection case, assuming that the observation angle and thickness t is small such that the extra path length of the transmitted wave is $\Delta L \approx 2t$, and taking into account the π phase difference that occurs due to the reflection on a medium of higher index of refraction, one obtains that constructive interference occurs for the condition

$$2t = m\lambda + \frac{\lambda}{2} \quad (2.9)$$

with $m \in \{0, 1, 2, \dots\}$. Thus maximum interference will occur at positions

$$r = \sqrt{R\lambda \left(m + \frac{1}{2} \right)} \quad (2.10)$$

for the reflection case. For the transmission case, the $\lambda/2$ factor does not exist. By counting the number of rings or by unwrapping the fringe map one can actively measure deformations on transparent media such as glass (e.g. a lens or flat surface), which is still useful for rapid testing of the quality of surfaces as they are being shaped in the optics industry.

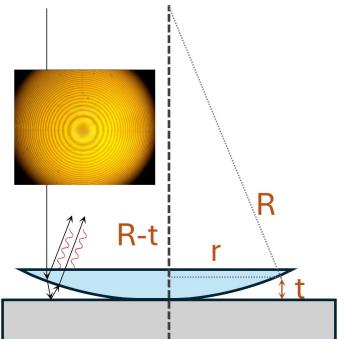


Figure 36. Newton's Rings and coordinate system.

2.2 Young Double-Slit

Although Newton's Rings and thin film interference are both manifestations of interference and wave-like behavior of light, they were not envisioned nor realized to probe that behavior at the scientific level. In this context, the double-slit experiment, conducted for the first time by Thomas Young in the early 19th century, was a **paradigm shift**. Indeed, the experiment was

designed to demonstrate the wave nature of light through the phenomenon of interference.

In general terms, the experiment is based on the principle that light, when treated as a wave, exhibits constructive and destructive interference under certain conditions. Indeed, when coherent light waves from two closely spaced slits converge on a screen, they superimpose on each other, creating an interference pattern.

To model this, let us assume that each slit acts as a source, meaning that following our framework introduced previously, we have

$$\mathbf{E}(\mathbf{r}) = E_1(\mathbf{r})e^{i(-\omega t + \mathbf{k} \cdot \mathbf{r}_1 + \phi_1)}\mathbf{e}_p + E_2(\mathbf{r})e^{i(-\omega t + \mathbf{k} \cdot \mathbf{r}_2 + \phi_2)}\mathbf{e}_p \quad (2.11)$$

where \mathbf{e}_p is the polarization vector and \mathbf{r}_1 and \mathbf{r}_2 the distances of the spatial position \mathbf{r} to the source. Assuming the relative phase $\Delta\phi = \phi_2 - \phi_1 = 0$ we get that the interference detected at a plane located at $z = L$ (see figure 37 for coordinates) is given by

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(k(r_1 - r_2)). \quad (2.12)$$

For sufficiently small angle θ (i.e. $L \gg y$) one can consider the two vectors \mathbf{r}_1 and \mathbf{r}_2 approximately parallel and the path difference becomes simply

$$r_1 - r_2 \approx d \sin \theta = \frac{dy}{L}. \quad (2.13)$$

Depending on the difference in path lengths traveled by the waves from each slit to a point on the screen, the interference can be either constructive (bright fringe) if

$$y = \frac{m\lambda L}{d} \quad (2.14)$$

for m integer or destructive (dark fringe) for

$$y = \frac{(m + 1/2)\lambda L}{d}. \quad (2.15)$$

Putting into broader perspective, the double-slit experiment underscores several key concepts in interferometry that we will utilize in the next few weeks:

- **Wavefront Division:** by dividing the wavefront and recombining it later, you can obtain interference (if within the spatial and temporal coherence parameters of your source). This property will be utilized in many different interferometers during the next few weeks;
- **Coherence of light:** contrary to Newton's ring and thin film interfer-

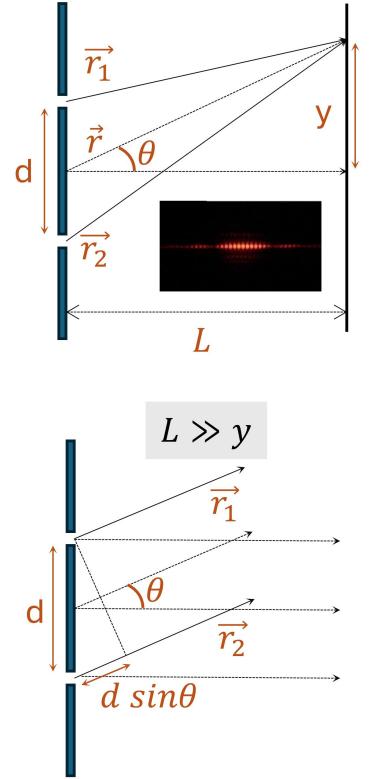


Figure 37. Young's double-slit experiment, coordinate system and approximation.

ence, double slit wave interference is only possible when using spatially coherent light sources, meaning that light sources are not the same, nor lead to the same results;

- **Path difference:** The extreme sensitivity of the interference pattern to the path difference between the two waves illustrates how interferometers can be used to detect minuscule changes in distance or optical properties of materials.

As seen, the double-slit experiment sets the cornerstone of wave-like behavior of light and interference. Understanding the principles, framework, and details behind the double-slit experiment is essential for a young student engaged in the study of interferometry and ultimately physics. Indeed, this simple yet remarkable demonstration encloses one of the most paradigm-shift ideas of physics, which about two centuries later reshaped again the landscape of physics in the form of quantum mechanics.

Exercise 7. Estimate the wavelength of the laser using the double slit experiment.

Exercise 8. What happens to the pattern if, in one of the slits, the wave acquires a phase difference, say ϕ_1 ? Can this be utilized as a sensor? What would be the form of the transfer function? What about sensitivity?

3 Overview of Experimental Components

3.1 Sources

In the context of our curricular unit and in a sensing system, a source can be considered a transducer: it converts electrical power into optical power. There are many options of sources in optics experiments. For interferometry, the key parameter to look for is **coherence**.

In broad terms, coherence measures the extent to which phase relationship is maintained across the beam in the transverse dimension - **spatial coherence** - and along the beam - **temporal coherence**. To evaluate coherence we can use simple interferometric experiments:

- Varying the distance of the slits in the double-slit experiment until visibility is affected allows us to infer the spatial coherence;
- Varying the distance of a path of a Michelson interferometer until interference degrades, establishes a **coherence length** l_c related with $\tau_c = l_c/c$ the **coherence time** of the source. This coherence time is

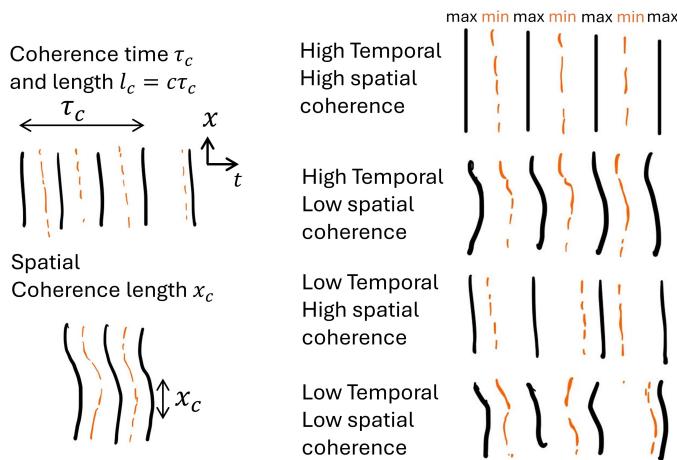


Figure 38. Spatial and temporal coherence.

related to the timescale of random variations in the phase of the laser or small wavelength variations and may be further related to the spectral width of a source by $\Delta\omega = 2\pi/\tau_c$.

Typically, sources such as common lamps and other thermal sources are not coherent, neither spatially nor temporally³³. Although white light sources and LEDs may be used as sources of small coherence length (a few wavelengths) the most typical source for optical interferometers are lasers. There are many types of lasers including gas lasers, solid state lasers, laser diodes, dye lasers, and fiber lasers, just to name a few of the most common. Typically gas lasers feature coherence lengths of the order of tenths of *cm* whereas laser diodes typically feature coherence lengths of the order of *mm*. For comparison, SLDs feature coherence lengths in tenths to hundreds of μm whereas LEDs or white light are typically in the μm range or lower.

In view of the coherence length needed, we can define two families of interference and related interferometers:

Long Coherence Interference: requires that the light source has a long coherence length, i.e. maintains its phase relationship over a long distance. This type of interference is characteristic of highly monochromatic light sources, such as lasers. Typical applications include holography (measurement of the phase and amplitude of light waves) and the most common phase-shift sensors.

Short Coherence Interference: involves light sources with short coherence lengths and exploits the fact that interference only occurs at a very small optical path difference to achieve high-resolution. In terms

³³ Note however that a pin-hole can provide spatial coherence, which combined with a narrow wavelength filter may allow some degree of temporal coherence.

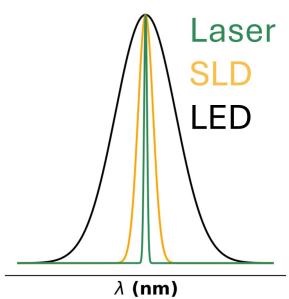


Figure 39. Shape of spectrum emitted by distinct sources.

of sources, one typically uses a laser with a small coherence length or broadband spectrum(e.g. femtosecond pulsed laser), or a broadband light source such as a superluminescent diode (SLD). Typical applications include optical coherence tomography (cross-sectional imaging of penetrable samples) and white-light interferometry (e.g. surface profiling or probing tiny cavities).

3.2 Optical components: Free space vs Fiber Optics

3.2.1 Beam steering

Beam steering in optics refers to the control of the direction of propagation of light beams, directing light to a specific location. In free space, this is usually achieved with mirrors and optomechanics, which exploit the total reflection of a coated planar surface (usually metallic coatings such as silver or aluminum for broadband applications, dielectric coating for narrowband applications). Alternatively, prisms can also be utilized for beam deflection, correction, and steering³⁴. Their functionality is based on the principles of refraction and reflection, utilizing the geometric shape of the prism and the refractive index of the material from which they are made. In terms of advantages, prisms can make setups more compact and feature better quality of reflection, yet are more costly and harder to work with.

The task in fiber optics is much simpler: having the light injected in the fiber, the optical path is defined by the fiber itself. Nevertheless, one should take special care in considering large curvatures, to avoid and losses and birefringence effects. There are multiple types of fiber, from the most common single mode³⁵ and multimode fibers, to graded-index or more complicated photonic-crystal fibers, just to name a few.

3.2.2 Beam division and combining

Beamsplitters are optical devices designed to divide an incoming light beam into two or more parts. Cube Beamsplitters are the most common and are typically made by cementing two right-angle prisms together with a special coating on the hypotenuse of one of the prisms. This coating is designed to reflect a certain percentage of the incoming light while transmitting the rest. While the most common is the 50:50 beamsplitter, other values are also available(e.g. 10:90). Another common component is a polarizing beamsplitter which utilizes birefringent materials (like calcite) or coatings to transmit/reflect light based on its polarization state. When light enters such a prism, it is split into two orthogonally polarized beams.

In fiber optics, amplitude division may be achieved using a fiber optic coupler, a device that can distribute the optical signal from one fiber to two

³⁴ e.g. porro prisms in binoculars

³⁵ The most used for interferometric applications

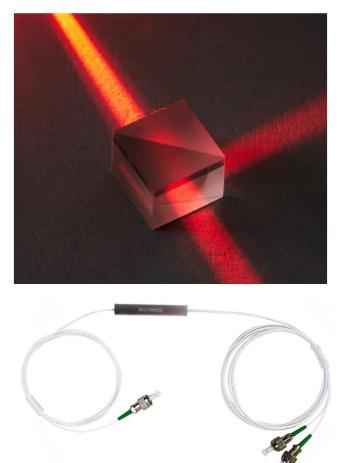


Figure 40. Free space Beamsplitter and fiber splitter.

or more fibers, or combine the optical signal from two or more fibers into a single fiber. A coupler may be either a splitter, a combiner, or an optical coupler. A basic fiber optic coupler has N input ports and M output ports. N and M typically range from 1 to 64. The number of input ports and output ports varies depending on the intended application for the coupler. Again the most common is the 2x2 50:50 (3dB) coupler, but other can exist depending on the application.

3.2.3 Polarization control

Regarding polarization control and free-space configurations, we have **polarizers** and **waveplates**.

Polarizers are optical filters that allow light waves of a certain polarization to pass through while blocking waves of other polarizations. They act as a *projector*, projecting the electric field into the linear polarization aligned with the axis of the polarizer. From this principle follows the Malus's law,

$$I = I_0 \cos^2(\theta) \quad (3.1)$$

which describes the intensity I of light after passing through a polarizer as a function of the incident light intensity I_0 and θ the angle between the initial polarization direction and the axis of the polarizer. Usually, they are utilized to convert an unpolarized light beam into a beam of linearly polarized light.

Waveplates, or retarders, are devices that alter the polarization state of light by introducing a phase shift between the orthogonal polarization components of the light wave. They are usually made from birefringent materials, which have different refractive indices for light polarized along different directions within the material. Unlike polarizers, waveplates do not project but change the relative phase polarization axis, effectively transforming the state of polarization. There are two types of waveplates:

- Half-Wave Plates ($\lambda/2$ Plates): These introduce a phase difference of π between the fast and slow axes of the waveplate, allowing the rotation of a linear polarization vector by a total amount of 2θ with θ the angle between the initial polarization and the fast axes of the waveplate.
- Quarter-Wave Plates ($\lambda/4$ Plates): Introduce a phase difference of $\pi/2$ (a quarter of a wavelength) between the fast and slow axis, allowing to convert linearly polarized light into circularly polarized light (and vice versa) if the initial polarization direction is at 45° to the waveplate axes.

Besides controlling the state of polarization, polarizers and waveplates are also frequently utilized to achieve precise control of light intensity by

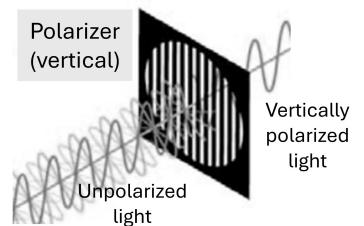


Figure 41. Action of a polarizer.

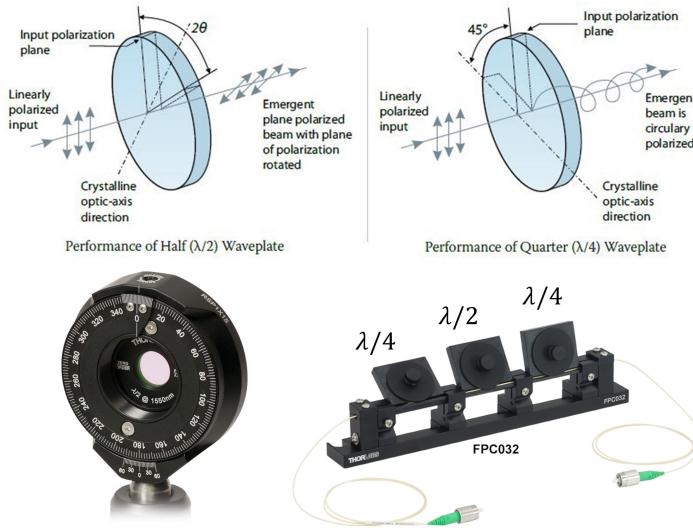


Figure 42. Polarization control in free space and fiber optics configurations.

exploiting Malus' law. For example, combining two polarizers or a waveplate and a polarizing beamsplitter³⁶ allows to control precisely the intensity of the output beam.

In the case of fiber-optic configurations polarization control can be harder to obtain. Indeed, the birefringence³⁷ of the fibers together with deformation or other responses may substantially alter the polarization state which is detrimental for interferometric purposes. Besides some free-space configurations, two common approaches allow to mitigate these effects in all-fiber configurations:

- **Manual Polarization Control:** polarization control can be achieved using stress-induced birefringence produced by wrapping the fiber around two or three spools. These paddles act as effective waveplates (usually $\lambda/4$ - $\lambda/2$ - $\lambda/4$) and thus may be utilized to alter the polarization of the transmitted light in a single-mode fiber.
- **Polarization maintaining fibers:** These types of specialty fibers allow the preservation of the polarization state of linearly polarized light over long distances.

3.2.4 Lens and Lens systems

A lens system may be as simple as a single lens or as complex as a zoom lens with 20 or more individual lenses. Concerning their surfaces, they may be plano (flat), spherical(convex or concave), or nonspherical (aspheric) and

³⁶ for higher power applications

³⁷ when the refractive index for light polarized in one direction is different from that in the orthogonal direction

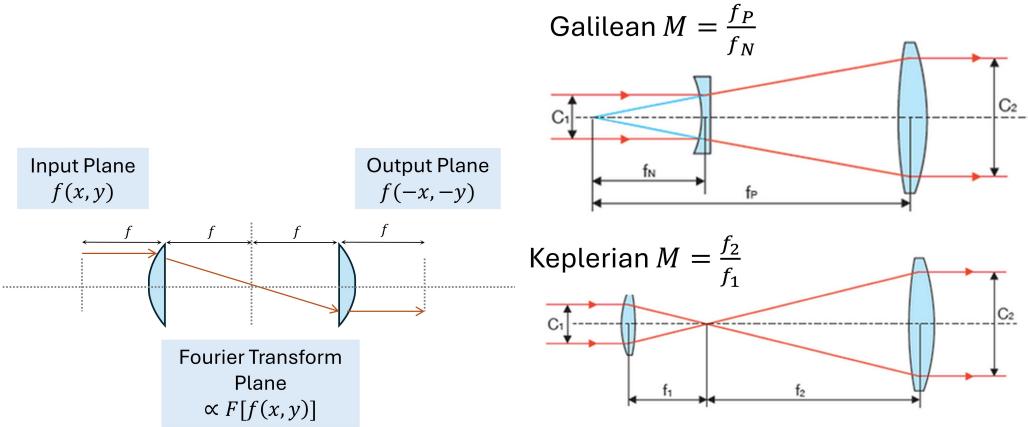


Figure 43. A 4f system, and a Kepler and Galilean beam expander/reducer. Kepler are more intuitive and easier to align, whereas Galilean are more compact, maintains the orientation of the beam and do not require beam focusing. Magnification is given by $M = \pm f_1/f_2$, with a positive sign for Galilean and negative for Keplerian configurations.

may be approximated by the thin lens equation

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o} \quad (3.2)$$

or require corrections/simulations in the case of a thicker lens. Three of the most important configurations include:

- **Single lens:** A single plano-convex lens is usually utilized for image magnification or demagnification (e.g. to match the size of the camera sensor) and to collimate a divergent beam. Besides, in the focal plane, a single lens can also be utilized to assess the Fourier transform of the field envelope;
- **Beam reducer and beam expander:** These are used for reducing or expanding collimated beams. Typical configurations include Kepler and Galilean configurations, as shown in Figure 43.
- **4f system:** The 4f system works as an optical relay utilizing two convex or plano-convex lenses of focal length f_1 and f_2 . If lens 2 is placed at a distance $f_1 + f_2$ of lens 1, it relays the amplitude and phase information of the input plane at f_1 in front of Lens 1 to an output plane located at f_2 after Lens 2. The magnification is found to be equal to $-f_2/f_1$.

Regarding fiber optics configurations, lenses are essential to couple light into a fiber or collimate it at the output. In this regard, collimators with connections (e.g. SMA or FC/PC) are the most common and practical, but

single lenses and microscope objectives are also utilized. A particularly interesting type of lens is gradient-index (GRIN) lenses, which make use of a refractive index that gradually changes across the lens diameter. This allows to focus light in a manner that closely matches the mode field diameter of the fiber and results in higher coupling efficiencies and reduced back reflections. Besides, its footprint and integration are also more practical.

3.2.5 Pin-holes and slits

Pinholes and slits serve as spatial filters, and may be utilized to perform spatial filtering, allowing to achieve a beam of desired characteristics or enhancing the resolution of an optical system³⁸. In short, a **pinhole** is a tiny circular aperture that allows light to pass through. By doing so, a nearly spherical wavefront is generated, which is ideal for high-contrast interference applications. On the other hand, a **slit** is an elongated aperture that shapes the light and may be utilized to achieve a more cylindrical wavefront. Besides, they are also fundamental for spatial filtering applications in the Fourier domain (the focal point of a lens system).

In fibers, the equivalent of a pin-hole is a single-mode optical fiber, for which the output is a pure transverse gaussian mode.

³⁸ e.g. confocal microscopy

3.3 Optical Detectors

In general terms, detectors are sensors for the electromagnetic field, that transduce the energy of a photon into an electric signal. More precisely, photon conversion into electrons is based on the absorption of the former in a semiconductor material and can involve one of three physical effects:

- **Photovoltaic(PV) effect:** the PV effect occurs in a semiconductor p-n junction, where photons with energy higher than the bandgap³⁹ are absorbed leading to the excitation of electrons from the valence to the conduction band. These diffuse towards the respective side (electrons to *n*, holes to *p*) generating a current, which establishes the operation principle of an **unbiased Photodiode**.
- **Photoconductive(PC) effect:** the PC effect explores a similar phenomenon of the PV effect but with an additional external reverse bias voltage applied to the photodiode, which increases the width of the depletion region and thus the electric field across the p-n junction. This enhances the separation of charge carriers and increases the photocurrent for a given light intensity. The photoconductive mode is used when higher sensitivity or faster response times are needed, and is the principle of operation of a **biased Photodiode**. Besides utilizing an higher bias

³⁹ Si bandgap is about 1.1eV, meaning that photons higher than 1100nm are harder to be detected.

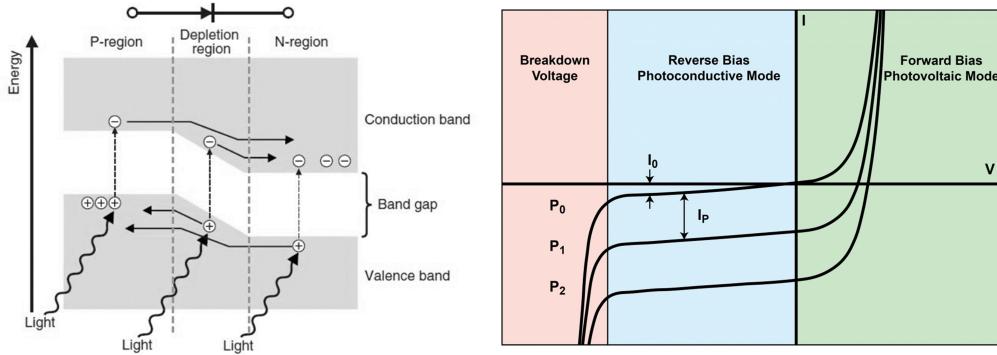


Figure 44. Semiconductor baPhotovoltaic and Photoconductive effect.

voltage, close to the breakdown voltage, one can exploit the avalanche effect to obtain extreme sensitivity down to the single photon regime⁴⁰.

- **Photoelectric(PE) effect:** the PE effect is slightly different and exploits the emission of electrons from a material when absorbing photons with energy larger than the work function⁴¹. Usually composite materials named photocathodes are utilized to convert photons into electrons that are later detected in the Anode. The photomultiplier tubes exploit these phenomena to offer extremely high sensitivities.

3.3.1 Detectors

Detectors usually refer to a single detection element that measures the intensity of the entire light beam incident upon it, working as a bucket for the whole intensity. The main advantages of single-element detectors are their high **sensitivity**, high **bandwidth** (up to GHz), high **dynamic range**, and simpler optical alignment. The main disadvantage is of course the lack of spatial information, which may be achieved with **point-wise scanning** or specific **illumination patterns**.

Depending on the working principle and physical effect discussed above, there are many solutions for detection depending on the final application. The most common are:

- **Photodiodes:** semiconductor devices that convert light into an electric current. Usually, Si detectors are utilized for applications in the visible range, whereas InGaAs detectors are more utilized for applications in the near-infrared regime.
- **Avalanche Photodiodes:** APDs are highly biased photodiodes working close to the breakdown voltage, thus being more sensitive to received signals. The single-photon avalanche diode enables it to work down to the single-photon regime.

⁴⁰ which is utilized in Avalanche Photodiodes and Single-Photon Avalanche Photodiodes.

⁴¹ the potential difference between Fermi level and vacuum levels.

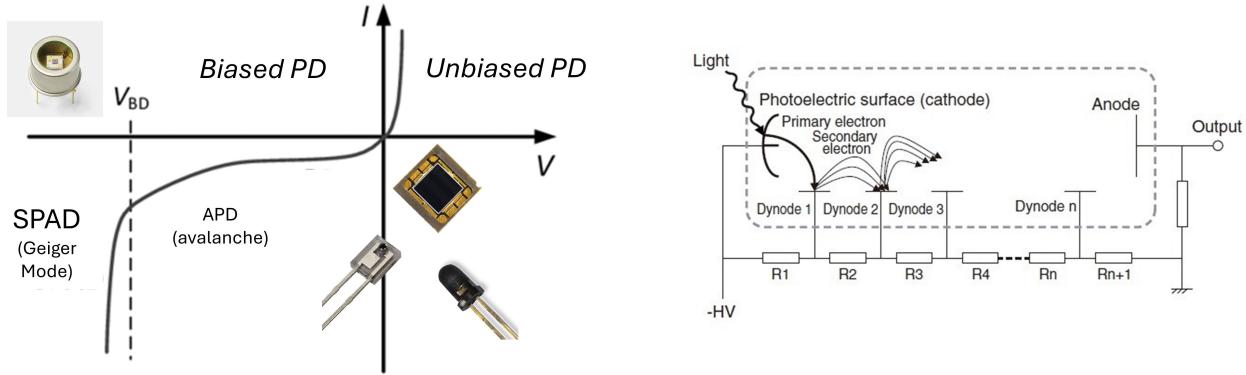


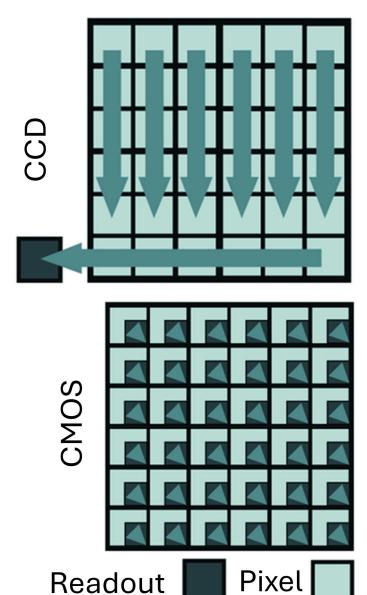
Figure 45. (Left) Various types of photodiodes. (Right) Photoelectric effect as the underlying principle of a photomultiplier.

- **Photomultiplier tubes:** Photomultipliers utilize the PE effect to detect a small signal with high efficiency, also enabling the detection of single photons.
- **Thermopiles and Pyroelectric Detectors:** Convert infrared radiation into a measurable electrical signal, usually utilized for thermal-related applications in the LWIR range.

3.3.2 Cameras

Camera sensors usually consist of an array (1D or 2D) of detection elements (usually photodiodes), each capable of independently measuring the intensity of light incident upon it. Compared with detectors (single-pixel detectors), camera sensors offer detailed **spatial information**⁴². Yet, they typically feature lower temporal response bandwidths and sensitivity. Depending on the final application, design, working principles, and specifications, camera sensors may offer unique capabilities for imaging applications:

⁴² which may be helpful for spatially varying interference patterns



- **CCD (Charge-Coupled Device) Cameras:** In a CCD, each semiconductor pixel accumulates charge during the exposure time. Then, each charge is transferred from pixel to pixel in the chip before being read converted into a voltage, amplified, and digitized at a single exit. The sequential nature of the charge transfer in CCDs is both a strength and a weakness: it ensures high-quality images with low noise but can lead to slower readout speeds and higher costs compared to CMOS sensors. Usually utilized when image quality is critical, such as science and astrophotography.

Figure 46. CCD and CMOS detectors.

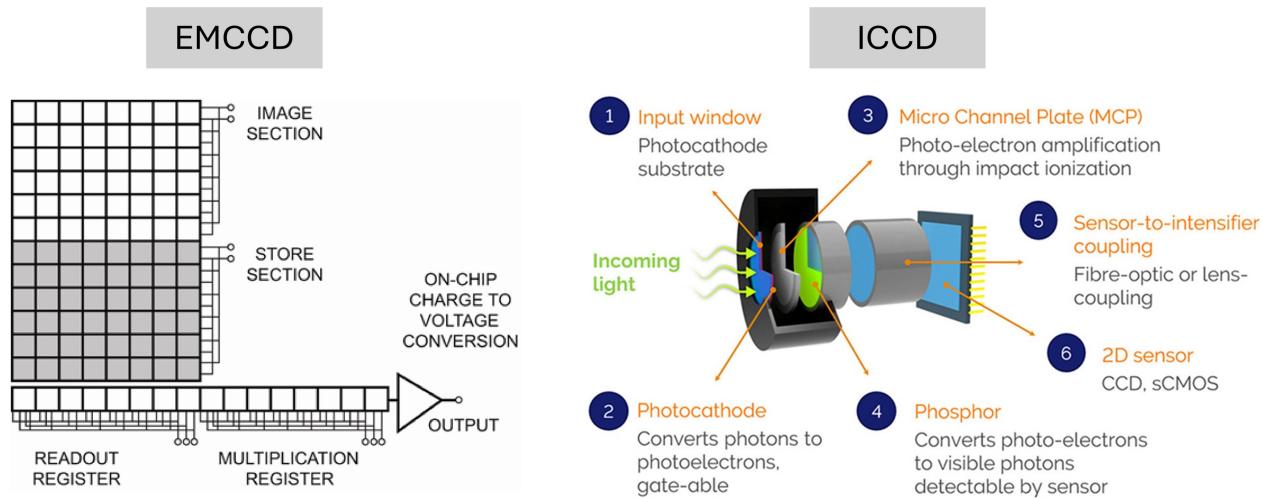


Figure 47. EMCCD and ICCD working principles towards ultra-sensitivity.

- **CMOS (Complementary Metal-Oxide-Semiconductor) Cameras:**

In a CMOS, each pixel has its own charge-to-voltage conversion mechanism, with an amplifier. This parallel process and design means faster data acquisition, readout speeds, lower power consumption, and lower manufacturing costs than CCD cameras. While the performance is lower than a CCD, its lower cost means it is usually the option to consumer electronics, security, and automotive.

- **EMCCD (Electron Multiplying CCD) Cameras:** Incorporate on-chip amplification to achieve high sensitivity, capable of single-photon detection. EMCCD cameras are used in low-light conditions, such as fluorescence microscopy and astronomical imaging, where detecting faint signals is critical.

- **ICCD (Intensified CCD) Cameras:** Combine CCD technology with a photocathode and a microchannel plate to intensify the image before it hits the CCD sensor, allowing for extremely high sensitivity and the ability to capture images in very low-light conditions. ICCDs are used in scientific research, surveillance, and night vision.

- **SPAD Cameras:** Utilize SPAD detectors in an array to capture images based on the detection of single photons, offering unparalleled sensitivity and precision in photon counting. SPAD cameras are emerging in fields requiring high temporal resolution and sensitivity, such as quantum imaging and LiDAR.

4 Concluding remarks

This week we introduced the basic concepts of wave optics, which enable us to explain wave interferometry. Under the mathematical formalism introduced, we are now capable of describing most of the typical interferometer designs, exploring their response to external stimuli and transforming them into interferometric optical sensors. Besides, we also introduced and discussed in general some of the most common components involved in optical interferometric sensors, both for free-space and fiber-optic configurations. With this knowledge, you are now ready for that hands-on approach to interferometric sensors of the next week.

Optical Interferometers

As we already started to unveil last week, interferometry is an extremely powerful toolset for engineering and modern physics. In the next three weeks you will see how the same tools and concepts have been used for over a century to support fundamental and theoretical research - from the foundations of electromagnetism to general relativity - as well as support some of the most sensible and robust state-of-the-art optical sensors - from very precise accelerometers to robust magnetic field sensing. Besides, we will try to capture the advantages and disadvantages of various designs and their potential applications, while also studying the challenges involved in detection and ways to circumvent them, from classical to quantum theory.

The essence of interferometry resides in the precise measurement of the phase difference between waves as they combine. Such phase differences occur due to variations on characteristic length scales below light wavelength, acting like a **microscopic ruler** and thus offering humans a window into the mesoscopic realm.

For this fifth week, we delve into three well-known designs of interferometers - **Michelson**, **Mach-Zehnder**, and **Sagnac** interferometers - exploring their historical framing, mathematical framework, and some applications. Then, connecting to the common language we established for the sensors and through some hands-on activities (experimental and computational), we will explore the potential of interferometry for optical sensing.

But before entering on the description of the interferometers themselves, let us introduce a handy matrix formalism to deal with interferometers when these involve beam splitters and other optical elements.

1 Transfer Matrix Formalism

While it is easy to explain wave interference with just two waves, analyzing optical systems involving components such as beam splitters and forward and backward propagating waves traveling the same paths can become rather complicated if one wants to keep all the mathematical details. In this case, the transfer matrix formalism comes as a very handy tool, simplifying the approach. In short, assuming a 50:50 beam splitter and the ports nomenclature of Figure 48, we have the resulting fields⁴³

Week

V

Section 2. Michelson Interferometer

Section 3. Mach-Zehnder Interferometer

Section 4. Sagnac Interferometer

Table 4. Contents for WEEK V

⁴³ we assume the Loudon symmetric beam splitter

$$\begin{aligned} \begin{bmatrix} \mathbf{E}_3 \\ \mathbf{E}_4 \end{bmatrix} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} \\ &= \overline{\mathbf{PS}} \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} \end{aligned} \quad (1.1)$$

In similar manner, one case use a matrix of the form

$$\overline{\mathbf{PS}} = \begin{bmatrix} e^{i\Delta\phi} & 0 \\ 0 & 1 \end{bmatrix} \quad (1.2)$$

to account for a phase shift $\Delta\phi$ in one of the paths. For example, the propagation over a distance L may be taken into account as a phase term kL (or $-kL$ for $+\omega$ convention). Furthermore, this phase shift may also be a spatial distribution of phase in the transverse directions, say $\Delta\phi(x, y)$. Finally, a mirror can be taken into account as a phase shift term, usually of π .

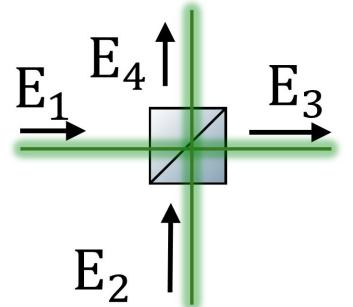


Figure 48. Beamsplitter port nomenclature for the transfer matrix formalism.

2 Michelson Interferometer

Born in 1852, Albert Michelson is one of the fathers of the North American Physics community, being the first one to be awarded the Nobel Prize in 1907 "for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid".

But interestingly, the success of Michelson was mostly a product of his drive to probe a wrong hypothesis⁴⁴: the presence of the luminiferous aether, a medium utilized for light waves to propagate through space. Indeed, Michelson designed an ultraprecise interferometer - **Michelson Interferometer** - and mounted their apparatus on a large stone block floating in a pool of mercury to isolate vibrations and to allow easy rotation of the experiment⁴⁵. At that time, aether was thought to be a fixed, invisible medium filling all space, and as Earth was believed to move through this aether, Michelson and Morley were expecting a variation of the signal of the interferometer depending on the orientation of the experiment. With unbeatable precision for that time, the negative outcome of the Michelson-Morley experiment(1887) was strong evidence of the inexistence of aether. Despite the "failure" to detect aether⁴⁶, the Michelson interferometer became an essential tool in the scientific community due to its unparalleled precision. The design of the Michelson interferometer (Figure 49) starts with a coherent light source, such as a laser, that emits a beam steered towards a beamsplitter. This beamsplitter (say 50:50) divides the incoming light into two paths - call it reference path 4 and measurement path 3. The light is then reflected by a mirror in each path and recombined at the beamsplitter. For the present case, we consider that

⁴⁴ "The only real mistake is the one from which we learn nothing"

⁴⁵ a very rudimentary optical table

⁴⁶ and which apparently contradicted the Fizeau moving media experiment, a discrepancy solved later by Einstein's special relativity

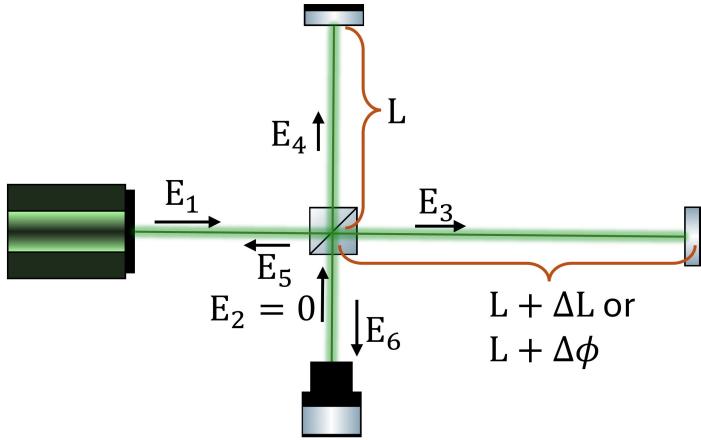


Figure 49. Configuration of the Michelson Interferometer.

there is an additional path length of t in the measurement arm, which going back and forth amounts to a total of path length $2\Delta L$, and corresponding phase shift $2\Delta\phi = 2k\Delta L$. Assuming that the polarization stays the same, it is easy to get under the transfer matrix formalism (discarding the effect of the mirrors and the phase term of propagation along the same path)

$$\begin{bmatrix} \mathbf{E}_5 \\ \mathbf{E}_6 \end{bmatrix} = \overline{BS} \cdot \overline{PS} \cdot \overline{PS} \cdot \overline{BS} \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} \\ = \frac{1}{2} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\Delta\phi} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{i\Delta\phi} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix}. \quad (2.1)$$

For the simple case of input state only in one port, i.e. $\mathbf{E}_2 = 0$, it is straightforward to get

$$\begin{bmatrix} \mathbf{E}_5 \\ \mathbf{E}_6 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (e^{2i\Delta\phi} - 1) \mathbf{E}_1 \\ i(e^{2i\Delta\phi} + 1) \mathbf{E}_1 \end{bmatrix}. \quad (2.2)$$

which taking the intensities at each port gives

$$\begin{aligned} I_5 &= \frac{I_1}{2}(1 - \cos(2\Delta\phi)) = \frac{I_1}{2}(1 - \cos\left(\frac{4\pi\Delta L}{\lambda}\right)) \\ I_6 &= \frac{I_1}{2}(1 + \cos(2\Delta\phi)) = \frac{I_1}{2}(1 + \cos\left(\frac{4\pi\Delta L}{\lambda}\right)). \end{aligned} \quad (2.3)$$

It is thus trivial to observe that if $\Delta L = 0$ there will be constructive interference at port 6 and destructive interference at path 5. Besides, the signal recorded at 6 will be sensitive to variations of ΔL and therefore may be suitable for multiple sensing applications. This way, the Michelson interferometer has a plethora of applications ranging from fundamental science to

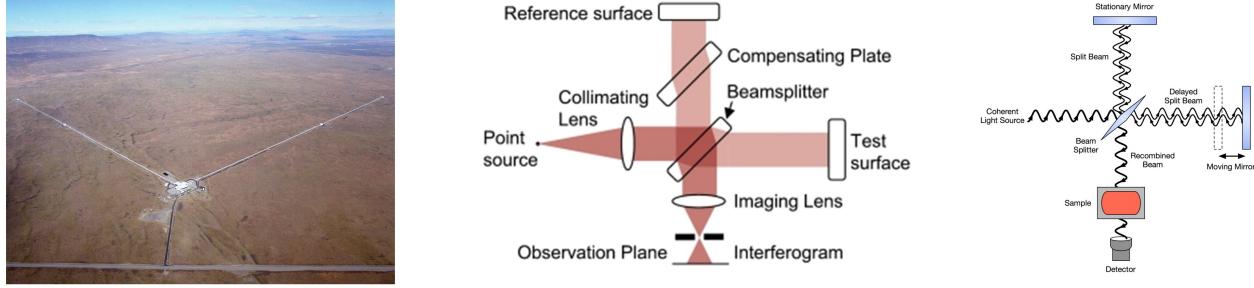


Figure 50. Applications of the Michelson interferometer: LIGO, Twyman-Green, and FTIR.

practical technology. Notable applications include precise metrology such as gravitational wave detection in LIGO, but also optical components testing (under the Twyman-Green configuration) and spectroscopy (being one of the working principles of Fourier Transform Infrared (FTIR) spectroscopy).

3 Mach-Zehnder Interferometer

With a similar working principle to the Twyman-Green configuration, the Mach-Zehnder interferometer (1891) also uses of an expanded beam of collimated light but simplifies the interpretation of the observed fringes by passing light through the sample only once. Another advantage is the access to an additional port, which comes in phase opposition to the other and may be utilized for noise mitigation strategies or to enable other types of signal analysis and interrogation as we will explore next week.

The Mach-Zehnder interferometer utilizes two beam splitters together with two mirrors to divide the beam into 2 paths before recombining the beams. Following the configuration of Figure 51, we get a very similar formula to that obtained for the Michelson interferometer,

$$\begin{aligned} I_5 &= \frac{I_1}{2}(1 - \cos(\Delta\phi)) \\ I_6 &= \frac{I_1}{2}(1 + \cos(\Delta\phi)) \end{aligned} \quad (3.1)$$

and thus the signal at the output ports of the beamsplitter will be sensitive to any phase variation induced in the measurement arm 3.

A very common way to use the Mach-Zehnder is to add a small displacement on the lateral separation of both beams, together with a small angle that can be controlled by the reference mirror (see Figure 52). If needed, one can also expand the reference arm. Considering a small angle θ in the $x0z$ plane,

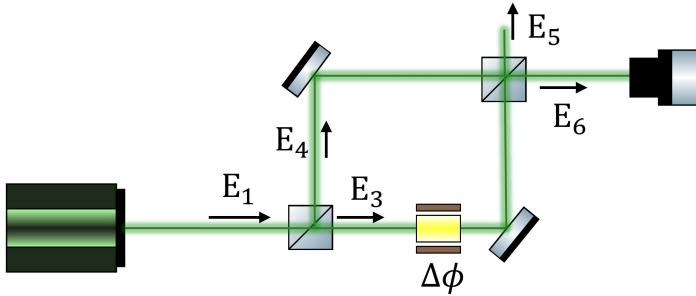


Figure 51. Configuration of the Mach-Zehnder Interferometer.

this will lead to a phase difference on the reference path $\Delta\phi_{ref} = (k \sin \theta)x$

$$\begin{aligned} I_5 &= \frac{I_1}{2} [1 - \cos(\Delta\phi + (k \sin \theta)x)] \\ I_6 &= \frac{I_1}{2} [1 + \cos(\Delta\phi + (k \sin \theta)x)] \end{aligned} \quad (3.2)$$

This way, a fringe pattern varying across the transverse plane may be generated (see Figure 52), which can be used to observe wavefront deformations and relate with sample properties(e.g. spatial distribution of the refractive index).

The versatility of the Mach-Zehnder Interferometer has led to its widespread use across various fields, including metrology for gas flows and plasmas, and biological imaging (e.g. Optical coherence tomography). In addition, Mach-Zehnder is also widely utilized as a modulator for telecommunication purposes, utilizing phase shifts in one arm to generate amplitude modulation at the output ports. Finally, the Mach-Zehnder also played a crucial role in foundational experiments in quantum mechanics, including tests of quantum entanglement and the principle of superposition, as we will later explore in this curricular unit.

3.1 Digital Off-Axis Holography

Digital off-axis holography concerns a technique of using interference and digital signal processing to reconstruct the complex electric field and thus access the spatial phase distribution on the measurement arm. In particular, the Mach-Zehnder interferometer serves as an excellent platform for implementing digital off-axis holography due to the possibility of controlling fringe spacing and shape.

Unlike traditional in-line holography, off-axis holography utilizes a deliberate angle between the object and reference beams as they are recombined. This angle separation (that controls fringe spacing and shape) ensures the spatial separation of the hologram zeroth-order and conjugate images in the

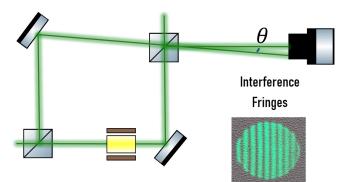


Figure 52. Adding an angle between the beams for the interference fringes.

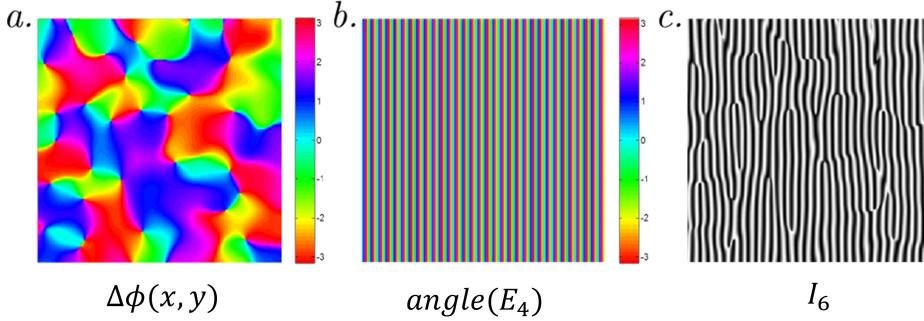


Figure 53. Filtering Process of the Digital Off-Axis holography.

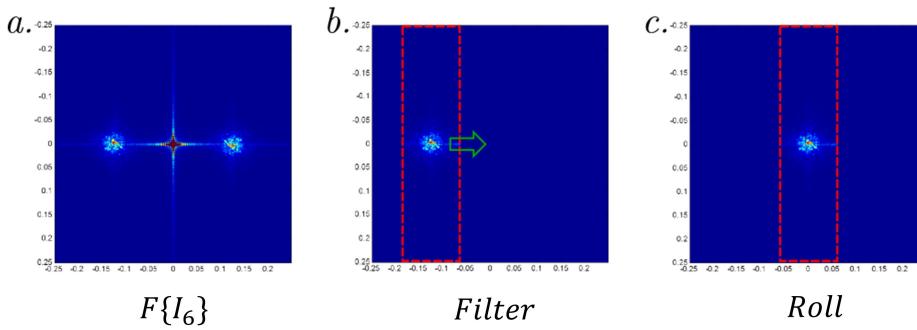


Figure 54. Filtering Process of the Digital Off-Axis holography.

frequency domain, facilitating easier and more efficient hologram reconstruction. Considering the spatial phase distribution we want to measure in path 3 to be described by $\Delta\phi(x, y)$ you can recover the previously derived formula for the path 6 and generalize it for the following assumptions:

- Distinct intensities $I_3(x, y)$ and $I_4 = \text{const.}$;
- Reference and measurement wavevectors define two angles θ_1 and θ_2 in the $x0z$ and $y0z$ planes, for which we also define $k_x = k \sin \theta_1$ and $k_y = k \cos \theta_2$.

The signal at port 6 becomes then

$$I_6 = I_3(x, y) + I_4 + \sqrt{I_3(x, y)I_4} \cos(\Delta\phi + k_x x + k_y y). \quad (3.3)$$

The holographic image - i.e. the complex field $E_3 = \sqrt{I_3}e^{i\Delta\phi(x,y)}$ - can then be reconstructed by digital signal processing in the following manner:

- First, blocking the reference path, you record with a camera the profile $I_3(x, y)$;
- Secondly, with the reference path on, you record with a camera the interferogram profile $I_6(x, y)$;

- Transforming I_6 to Fourier domain using a numerical 2D FFT, you will encounter:
 - A spot close to the center $(k_x, k_y) = 0$ corresponding to the zeroth order and to the Fourier transform of the intensity $F\{I_3 + I_4\}$;
 - Two spots localized in two of the quadrants: one will correspond to the term $E_3 E_4^*$ (i.e. centered in $(-k_x, -k_y)$ in the Fourier domain, denominated -1 order) whereas the other will correspond to $E_3^* E_4$ (i.e. centered in (k_x, k_y) in the Fourier domain, denominated +1 order);
- Localize the -1 order point and optimize the tilt of the mirror to control the center point in order to prevent overlap of order 0 and order -1;
- Apply a filter in the Fourier domain, keeping only the terms related to that order - for example applying a circular window - which leaves you with the term $F\{\sqrt{I_3 I_4} e^{\Delta\phi(x,y) - i(k_x x + k_y y)}\}$;
- Choosing the correct center point, execute a roll necessary to recenter the transform at the origin, giving you $F\{\sqrt{I_3 I_4} e^{\Delta\phi(x,y) - i(k_x x + k_y y)} e^{+i(k_x x + k_y y)}\}$, where the second term is added due to the shift in the Fourier domain and Fourier transform properties;
- Apply the inverse 2D Fourier transform to obtain the complex field, $F^{-1}\{F\{\sqrt{I_3 I_4} e^{\Delta\phi(x,y)}\}\} = \sqrt{I_3 I_4} e^{\Delta\phi(x,y)}$;

This technique allows measuring the complex field with just a single shot and is widely used in advanced imaging solutions. For example, it has found applications in biological imaging, microfluidics, and materials science, where detailed three-dimensional reconstructions and precise measurements of optical path length changes are invaluable. Another common application is imaging through complex media, where it enables one to see through diffusive media or multimode fibers, for example.

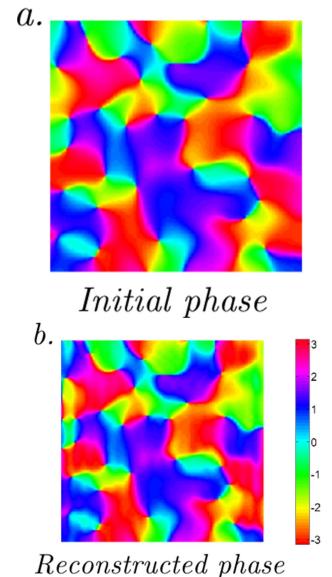


Figure 55. Result of the Off-axis holography.

4 Sagnac Interferometer

Another interesting interferometer design is the Sagnac interferometer, developed by French physicist Georges Sagnac in 1913. As the Michelson interferometer, this instrument was also designed to probe the existence of the luminiferous aether. Aether was not found but instead, it was discovered a rotation-induced phase shift of light, a phenomenon now known as the Sagnac effect.

In the Sagnac interferometer, as shown in Figure 56, the two beams traverse the same closed path in opposite directions. Its unique configuration allows to detect phase shifts caused by non-reciprocal effects⁴⁷ such as rotational motion. The closed path feature also offers several advantages:

- **Easier to align:** utilizing the same path makes it very easy to align;
- **Robustness to External Vibrations:** featuring symmetry in both arms, it is less susceptible to external vibrations and thermal fluctuations;
- **Simplicity and Versatility:** involves just a few components and does not require special care about the coherence length of the laser.

Not surprisingly the mathematical formulation is the same as the last two examples. Regarding the phase difference, we will consider the simplified case of the ring interferometer⁴⁸. Considering the rotation motion of angular velocity ω we will have that the time taken by field E_3 to travel the distance of the interferometer is

$$t_3 = \frac{2\pi R + \Delta L}{c} = \frac{2\pi R + R\omega t_3}{c} \quad (4.1)$$

where $\Delta L = R\omega t_3$. Conversely, field E_4 will travel less distance, leading to

$$t_4 = \frac{2\pi R - R\omega t_4}{c}. \quad (4.2)$$

The total time difference becomes

$$\Delta t = t_3 - t_4 = \frac{4\pi R^2 \omega}{c^2 - R^2 \omega^2} \quad (4.3)$$

which for $R\omega = v \ll c$ translates into a phase difference

$$\Delta\phi = \frac{c\Delta t}{\lambda} = \frac{4\pi A\omega}{\lambda c} \quad (4.4)$$

where A stands for the area enclosed by the light path, λ the wavelength of the light, and c the speed of light vacuum.

In the particular case of the Sagnac interferometer, a fiber-optic configuration is particularly interesting as it allows for a significant increase in the effective path length by utilizing a coiled version, multiplying the phase length by a factor of N with N being the number of fiber optic coils utilized.

Such ultra-high sensitivity combined with the robustness and compactness of the system, enable one of the most successful applications of the Sagnac interferometer as an optical sensor: the high-precision fiber optic gyroscopes. In

⁴⁷ i.e. distinct effects for the distinct directions

⁴⁸ although a general formulation can be achieved under relativistic theory and infinitesimal path arguments

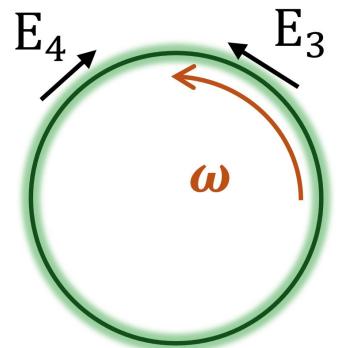


Figure 56. Sagnac configuration.

addition, such gyroscopes also feature immunity to electromagnetic interference, making them ideal for use in harsh environments or critical applications environments such as satellites, submarines, and autonomous vehicles.

Week

VI

5 Fabry-Perot Interferometer

So far we explored interferometric configurations involving a simple two-wave interference process. However, besides the possible mathematical challenge that may come with it, there is no limitation on extending the interferometer principle to a situation with more than two waves. The Fabry-Perot Interferometer (FPI) is an example of a device in the field of optical sensing that exploits the principles of multiple wave interference to measure, amongst others, the spectral properties of light. Its design consists fundamentally of two parallel, highly reflective mirrors, creating a resonant cavity that enhances certain wavelengths of light through constructive interference.

Our starting point is the analysis of the combination of multiple wave reflections occurring on a thin dielectric film. Looking at Figure 57, our intuition on wave interference leads us to think that the light waves reflecting back and forth between the boundaries of the dielectric layer may result in a resonant cavity. For the sake of simplicity, we will consider that we have the effective reflection and transmission coefficients t_0 , t'_0 , and r_1 , which depending on the polarization can change according to Fresnel equations.

Looking at the first transmitted wave we have

$$E_{t_0} = E_0 t_0 t'_0. \quad (5.1)$$

It is easy to generalize the N-th transmitted wave

$$E_{t_N} = E_0 r^{2N} t_0 t'_0 e^{iN\delta} \quad (5.2)$$

with

$$\delta = \frac{4\pi n d \cos(\theta)}{\lambda} \quad (5.3)$$

with n the refractive index inside, meaning that the total transmitted field will be given by a geometric series

$$E_t = E_0 \sum_{n=0}^{\infty} r^{2N} t_0 t'_0 = E_0 t_0 t'_0 \frac{1}{1 - r^2 e^{i\delta}}. \quad (5.4)$$

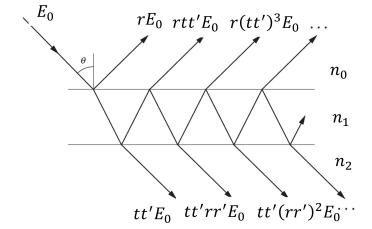


Figure 57. Multiple reflections and total transmitted intensity.

The most usual case is unpolarized light.

The transmitted intensity can be derived from the electric field as

$$I_t = I_0 T^2 \frac{1 - R^2}{1 + R^2 - 2R\cos(\delta)} \quad (5.5)$$

where we consider $T = |t_0|^2 = |t'_0|^2$ and $R = |r|^2$. Using the definition of **coefficient of finesse** $F = 4R/(1 - R^2)$ it can be further simplified to

$$I_t = I_0 T^2 \frac{1}{1 + F \sin^2(\delta/2)}. \quad (5.6)$$

By plotting the transmitted intensity over the total intensity in function of the phase difference it is easy to understand what happens in a Fabry-Perot interferometer and the actual role of the finesse coefficient. By analyzing Figure X it is possible to see that higher finesse coefficient relates to higher visibility whereas a lower finesse coefficient leads to lower (almost sinusoidal-like) visibility.

Recovering the phase difference formula, it is possible to demonstrate that the maximum will occur at

$$\delta/2 = m\pi \rightarrow \cos(\theta) = \frac{m\lambda}{2nd} \quad (5.7)$$

meaning that:

- For normal incidence $\theta = 0$, the FPI is sensitive to variations of thickness and wavelength;
- With a lens at the output (incident angles are transformed into distances in the plane) you will obtain different peaks depending on the wavelength of your source.

Motivated by the last observation we can also explore what happens in an FPI as a function of the wavelength (see Figure 58). In this case, two figures of merit can be helpful to describe the experimental settings and the characteristics of a sensor: the free spectral range and the Finesse.

Free Spectral Range: The Free Spectral Range (FSR) corresponds to the separation in frequency between successive transmission peaks. In frequency it is given by:

$$FSR_f = \frac{c}{2n_1 d \cos(\theta)} \quad (5.8)$$

where c is the speed of light, and d is the mirror separation. In wavelength, using $f = c/\lambda$ which leads to $df = -c/\lambda^2 d\lambda$ this can be expressed

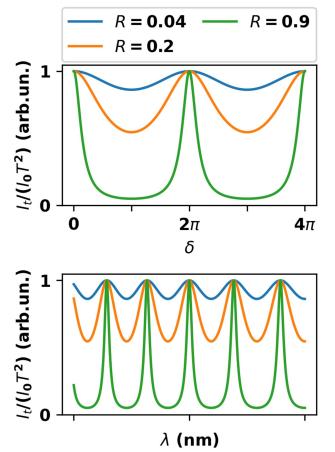


Figure 58. Transmitted intensity in function of path distance and wavelength.

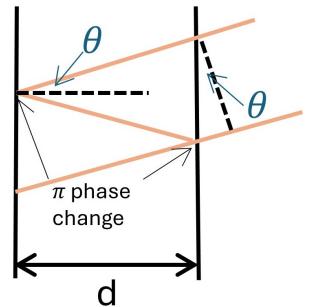


Figure 59. Path distance.

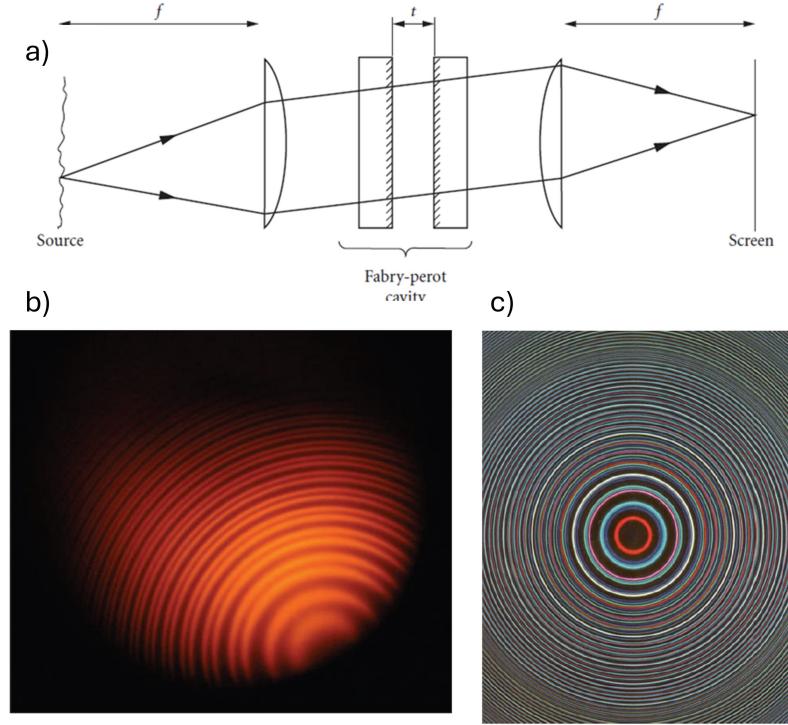


Figure 60. Spectrometry using a Fabry-Perot. Different wavelengths will have distinct resonance angles, ultimately leading to circle structures when passing the setup (a). (b) Sodium lamp doublet and (b) white light spectrum.

as

$$FSR_\lambda = \frac{\lambda_0^2}{2n_1 d \cos(\theta)} \quad (5.9)$$

Finesse: The finesse \mathcal{F} (not to be confounded with the coefficient of finesse) measures the spectral resolution of the Fabry-Perot and is defined as the ratio of the FSR to the full width at half maximum (FWHM) of the transmission peaks:

$$\mathcal{F} = \frac{FSR}{FWHM} = \frac{\pi}{2 \arcsin(1/F)} \approx \frac{\pi \sqrt{F}}{2} \quad (5.10)$$

where the last approximation is valid for high-reflective mirrors $R \approx 1$.

Overall, Fabry-Perot interferometers offer several distinct advantages that make them particularly valuable for optical sensing applications:

- **Versatility, compactness, and simplicity of design:** FPIs operate from UV to IR wavelength ranges, needing only a few optical components and can be deployed either in free space or fiber configurations;
- **Sensitivity and High Spectral Resolution:** FPIs are highly sensi-

tive to changes in wavelength, thus useful to monitor spectral changes (e.g. environmental monitoring applications);

- **Tunability and Multiplexing capabilities:** The resonant cavity of an FPI can be easily tuned by adjusting the distance between the mirrors, which changes the resonance conditions of the cavity. This tunability allows for selective filtering and analysis of specific wavelengths;
- **Stability and Reliability:** FPIs are known for their stability and reliability over time and can be easily integrated with other optical elements and electronic detection systems.

Finally, in terms of applications, Fabry-Perot interferometers are a popular choice for wavelength-related or wavelength-based analysis, from **environmental monitoring** to **fiber-based microphones**, and from **telecommunications** to **laser spectrometry** (i.e. measuring the spectrum of a laser).

6 Concluding Remarks

Throughout this chapter, we have seen how different interferometric designs present opportunities and unique advantages for sensing applications that span a wide range of scientific and technological fields. Although we covered the most common ones - **Michelson**, **Mach-Zehnder**, Sagnac, and **Fabry-Perot** interferometers - there many other configurations (see Figure ?? for classical and historical examples). Yet, the same principles you learned in this chapter hold in every other configuration, meaning that in the classical picture, you have all the tools that you need to analyze any configuration, from free space to fiber-based and even integrated optics.

To conclude, interferometers translate variations of the environment into phase differences that ultimately lead to variations in the intensity measured at the output. Now, it is time to discuss how to measure this intensity.

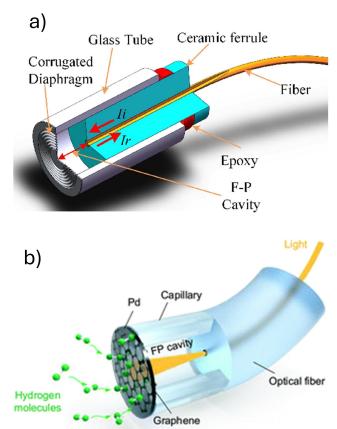


Figure 61. Example applications of the Fabry-Perot interferometer in Fiber Sensing solutions for (a) a microphone and (b) biochemical sensing.

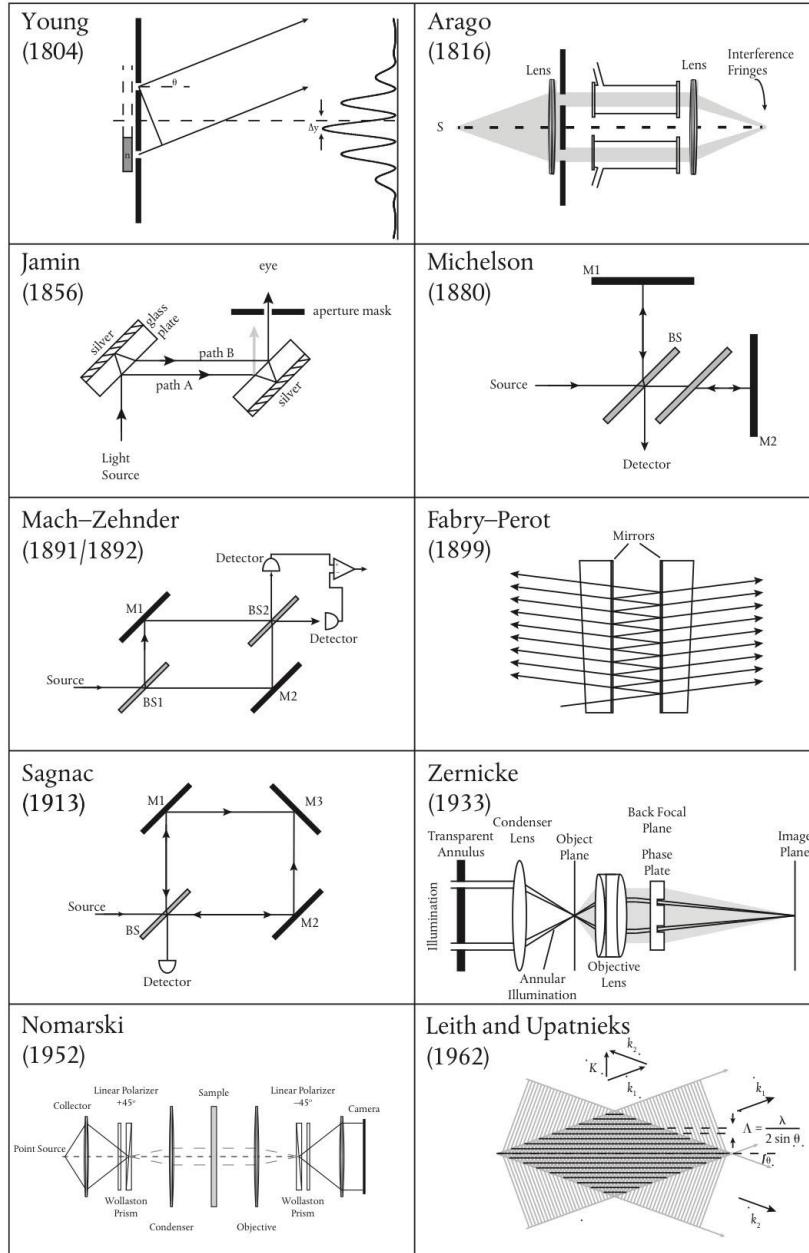


Figure 62. Multiple configurations of interferometers over the last two centuries.

Signal Detection in Interferometric Sensors

Week

VII

So far we discussed interferometers from the perspective of working principles and took for granted the detection part: we consider the ideal case where we use a simple square-law detector and that we are able to obtain an intensity related to the phase and ultimately to our target quantity. Yet, in real-world conditions, detection can be more critical than the interferometer design itself, as assuring the detection of a good signal with meaningful information is a step of utmost importance. This requires exploring possible problems and challenges of the configuration used, characteristics of the equipment, and also noise.

In general terms, when we think about signal detection in the context of optical interferometric sensors we are referring to the process of translating the phase information into a usable electrical signal. Referring to ideal characteristics of our detection scheme we want to have an **accurate and stable output signal**, a **large Dynamic range**, and possibly a **large signal-to-noise ratio**.

Detection may be divided into **passive** or **active** schemes, depending if they require feedback of some information and signal back to the components of the interferometer, including to the sensor, to the laser source, or to other optical element. Currently, there are many detection schemes depending on the application and performance required. For the sake of simplicity, we will focus on three specific configurations which we will discuss in detail during this class: *homodyne*, *heterodyne*, and *pseudo-heterodyne* detection.

1 Homodyne detection

As we saw in the last few weeks, a standard two wave interferometer usually leads to intensities at the output ports given by the formula

$$I = \frac{1}{2} (I_0 + I_0 \cos(\Delta\phi)) \quad (1.1)$$

where $\Delta\phi$ is the phase difference of the two optical beams that ultimately is related to what we want to measure. For the present chapter we will consider

- Section 1. Homodyne Detection
- Section 2. Heterodyne Detection
- Section 3. Pseudo-heterodyne detection

Table 5. Contents for WEEK VII

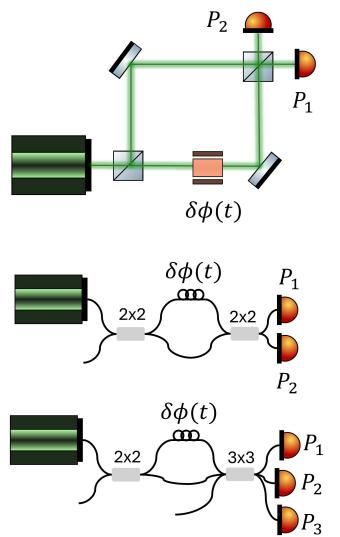


Figure 63. Passive Homodyne detection.

this phase difference to be composed as

$$\Delta\phi = \phi_0 + \delta\phi(t), \quad (1.2)$$

i.e. a static or quasi-static phase difference ϕ_0 related to the configuration of the interferometer plus a time-varying signal $\delta\phi(t)$ which corresponds to the signal we want to detect. Substituted into equation 1.1 it gives

$$I = \frac{I_0}{2} + \frac{I_0}{2} \cos(\phi_0 + \delta\phi(t)), \quad (1.3)$$

being straightforward to conclude that inverting the relation between measured I to obtain $\delta\phi(t)$ is highly nonlinear and non-trivial depending on factors such as the amplitude of $\delta\phi(t)$ and initial interferometer configuration. For small signal perturbations $\delta\phi(t) \ll 1$ it is possible to expand equation 1.3 to the first order terms as

$$I = \frac{I_0}{2} + \frac{I_0}{2} \cos(\phi_0) - \sin(\phi_0)\phi(t) + O(\delta\phi^2). \quad (1.4)$$

Looking at this formula two conclusions emerge, eventually related to the problems of homodyne detection:

- **Limited Dynamic Range:** The formula is only valid for small perturbations.
- **Varying Sensitivity:** depending on the phase ϕ_0 one can have maximum sensitivity (for $\phi_0 = (2m + 1)\pi/2$) or even reach zero sensitivity ($\phi_0 = m\pi$). As this phase difference usually depends on the initial optical path difference, say OPL_0 but also on the wavelength itself, i.e. $\phi_0 = 2\pi OPL_0/\lambda$, the sensitivity may be strongly affected by **mechanical or laser variations and drifts** over time.

Combining the signals of two or more output ports and multiple wavelengths, one can work out numerical methodologies to circumvent the varying sensitivity problem(see examples in Figure 64). This can be done without dramatically changing the configuration nor requiring active elements operating on the system, yet their performance is typically less stable compared to active homodyne configurations.

1.1 Active Homodyne

To circumvent the instability of **passive homodyne** detection in terms of sensitivity one can employ **active** configurations by adding optical components for active phase tracking. The idea is to incorporate a component on the reference arm of the interferometer (e.g. a fiber stretcher or a piezo-electric

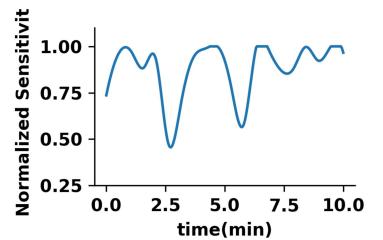


Figure 64. Typical temporal drift of the sensitivity in Passive Homodyne.

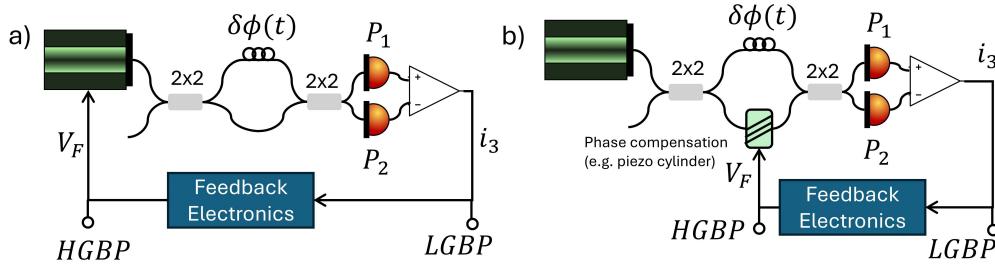


Figure 65. Active Homodyne Configurations.

controlled mirror) to maintain the interferometer on the point of maximum sensitivity, defined as **quadrature**,

$$\phi_0 = \frac{(2m+1)\pi}{2} \quad (1.5)$$

where m is an arbitrary integer number. In this case, it is typical to subtract the two ports of the interferometer⁴⁹ to get I_3 without a DC component as

⁴⁹ if possible

$$I_3 = I_0 \cos(\phi_0 + \delta\phi(t) - \Delta\phi) \quad (1.6)$$

where ϕ_{act} is the phase introduced by the active element. Considering the schematic of Figure ??, the feedback signal following to the active element is just $V = G \int I_3(t') dt'$ where G is a gain parameter, being straightforward to demonstrate that in the case of ϕ_{act} correcting to the quadrature point, the feedback loop follows the differential equation

$$\frac{dV}{dt} \approx I_0 G (\phi_0 + \delta\phi(t) - \phi_{act} - \pi/2) \quad (1.7)$$

or considering the introduced phase change to be proportional to an applied voltage as $\phi_{act} = kV$,

$$\frac{dV}{dt} + I_0 G k V = I_0 G (\phi_0 + \delta\phi(t) - \pi/2). \quad (1.8)$$

In this form, the term $I_0 G k$ is usually called the gain-bandwidth product of the feedback circuit and establishes two regimes for operation:

- For frequencies of $\delta\phi(t)$ much lower than the gain-bandwidth product, the term $\delta\phi(t)$ may be extracted by *reading and filtering the signal information at the feedback voltage* as it is only given by

$$kV = \phi_0 + \delta\phi(t) - \pi/2. \quad (1.9)$$

- For frequencies of $\delta\phi(t)$ larger than the gain-bandwidth product, the term $\delta\phi(t)$ needs to be extracted from the signal of I_3 and the correction only accounts for the drift term ϕ_0 .

This detection scheme is easy to implement, provides a highly linear operation, and the noise introduced in the sensor is usually negligible. Yet, there are still drawbacks:

- **Limited Dynamic Range:** The amount of correction that you can get with a piezoelectric element is usually small, meaning that the dynamic range is still limited;
- **Piezoelectric Devices:** One of the disadvantages of piezoelectric devices is that they usually require high voltages (up to thousands of Volts) which can be detrimental for the versatility of the final solution (e.g. electric sensors).

A possible configuration to bypass the second problem is to use act directly on the laser, using wavelength tuning. In particular, in the case of diode lasers, it is possible to modulate (in a typically small range) the emission frequency of the laser by $\Delta\nu = k_d \Delta i_L$ via the provided current change Δi_L . In this case, the phase difference generated will be $\phi_{act} = \frac{2\pi OPL_0}{c} \Delta\nu$ which can easily be worked out and solved for Δi for achieving the condition $\phi_0 - \phi_{act} = \frac{(2m+1)\pi}{2}$.

2 Heterodyne detection

In very generic terms heterodyne detection involves mixing a signal of interest with a reference signal, called the **local oscillator**, which has a slightly different oscillation frequency. Under this mixing process, new frequencies at the sum and difference of the original frequencies appear at the detection level⁵⁰. For example, going back to the Mach-Zehnder interferometer configuration setup, we can add an optical component - called an acoustic-optic modulator or Bragg Cell (see Figure 67) - which is the first order of transmission provides a frequency shift to the original wave frequency

$$\omega_{LO} = \omega - \omega_{IF}. \quad (2.1)$$

In this case, it is straightforward to recover the interferometer formula as dependent on this frequency difference:

$$I_6(t) = \frac{I_s + I_{LO}}{2} - 2\sqrt{I_s I_{LO}} \cos(\omega_{IF}t + \phi_0 - \phi_{LO} + \delta\phi(t)) \quad (2.2)$$

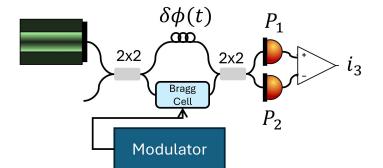


Figure 66. Heterodyne Detection.

⁵⁰ These frequencies only appear due to the square-law detection, and are not related to the oscillatory field itself.

whose DC term can be eliminated in differential detection if we have access to a second port of the interferometer.

Typically, the frequency of interest in an interferometric setup is the one that is shifted to a more manageable frequency range for easier signal processing. In this context, the objective is to convert a high-frequency signal into a lower frequency that retains the amplitude and phase information of the original signal and thus the term to be considered is the difference term, i.e. with frequency ω_{IF} . The demodulation of this signal can be experimentally done using a **Lock-in amplifier** or a **Phase-tracking circuit**.

The major advantages of heterodyne detection are:

- **Sensitivity and Selectivity:** Multiplied by the local oscillator amplitude, heterodyne detection allows higher sensitivity to small changes in phase even if the amplitude of the signal is very faint;
- **Robustness to noise and frequency translation:** While typical noise spans over a wide frequency range, the possibility of measuring signals at very specific frequencies makes it ideal to bypass the challenges of noisy environments;
- **Large Dynamic Range:** Conceptually, if interrogated properly, it may feature an infinite dynamic range.

In this line, one may be inclined to conclude that if possible, heterodyne detection is preferable. Yet, in practice, there are however also a few challenges:

- **Requirement of additional components:** in particular, a rather bulky acoustic-optic modulator, which introduces more complexity and has a large footprint incompatible with miniaturization and may increase the cost;
- **Complexity and Stability of frequency:** If not on the laser side (e.g. two frequency lasers), the setup for heterodyne detection is typically more complex than that for homodyne detection. It requires precise control of the local oscillator frequency and phase.

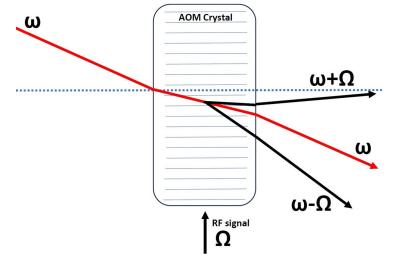


Figure 67. Bragg Cell working principle.

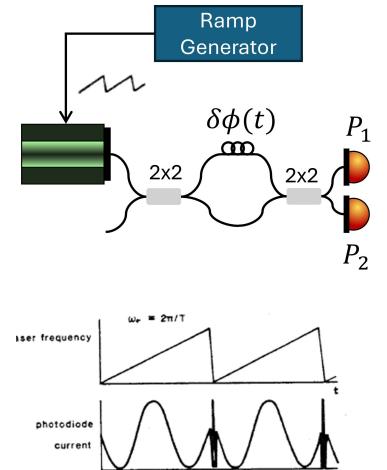


Figure 68. Typical Pseudo-Heterodyne Detection.

3 Pseudo-Heterodyne detection

As seen, heterodyne detection features interesting mathematical advantages, yet true heterodyne detection is highly limited in the experimental world by practical applications. The pseudo-heterodyne detection aims to circumvent these problems by modulating the laser source.

Considering the case of Figure 68, consisting of an interferometer with unbalanced arms with path difference L , it is trivial to demonstrate that a shift in the laser emission frequency $\Delta\nu$ leads to a phase shift

$$\phi_{act} = \frac{2\pi Ln}{c} \Delta\nu. \quad (3.1)$$

Again, in diode lasers, this can be achieved using a current modulation. It is trivial to demonstrate that if you modulate the laser with a ramp, it will act as an effective modulation frequency $\omega_m = \frac{2\pi Ln}{c} \frac{di}{dt} K_d(\omega)$ (see Figure 69). In this way, you can recover the same results that you would have if you measured it with true heterodyne.

4 Concluding Remarks

During this chapter, we explored the intricacies of signal detection in interferometric sensors, focusing on the comparative advantages and challenges of homodyne and heterodyne detection methods. We have shown that while the most classical homodyne detection seems straightforward and cost-effective, it suffers from several practical limitations for supporting real-world sensing devices. If limitations for the final application, these drawbacks can be mitigated to some extent by using active homodyne, heterodyne, and pseudo-heterodyne detection schemes. Yet, the complexity of the setup and the need for additional components can also pose challenges, meaning that the ultimate choice between homodyne and heterodyne detection methods depends on the specific requirements of the application, as synthesized in the final Table 6.

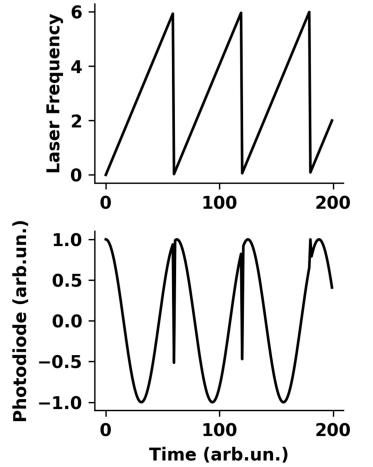


Figure 69. Ramp Modulation and resulting signal.

	Active homodyne	Passive homodyne	True heterodyne	Pseudo-heterodyne
Active compensation or bias elements required	Yes ¹	No	Yes	No
Special optical elements required	No	Yes	Yes	No
Complexity of electronics	Low	Ave.	Ave.	Ave.
Linearity	V. good	Fair	Good	Fair
Phase tracking range	Limited	Infinite	Infinite	Practically unlimited
Unbalanced interferometer and laser emission frequency modulation required	No	No	No	Yes
Suitable for remote passive sensor arrays	No	Yes	No	Yes

Table 6. Comparison of different interrogation schemes of optical interferometers, with an overview of advantages and disadvantages.

Beating The Limits of Noise

Week

VIII

The Laser Interferometer Gravitational-Wave Observatory (LIGO) stands as one of the most impressive examples of optical interferometric sensing. As we will see, LIGO builds upon a very simple Michelson interferometer design, enhanced with a few additional components, tweaks, and ideas in the realm of optical interferometric sensing that we already discussed in our course. These enhancements are designed to match the capabilities of the equipment to the requirements of the problem, being able to detect minuscule ripples in spacetime caused by gravitational waves, with arm lengths extending four kilometers to maximize sensitivity.

For the next two weeks, and leveraging such extreme operation conditions and required sensitivities, we will focus on the LIGO as a case study to discuss the **role of noise when extreme precision is required**, showing that even in ideal conditions, the quantum nature of physics adds a certain noise limit - the **standard quantum limit**. We will discuss the origin and role of quantum noise in LIGO interferometric sensors, and how we can tame it when we need precise measurements utilizing a quantum-related tool, the vacuum **squeezed states**.

1 The Origin of Gravitational waves

In very generic terms, gravitational waves are disturbances in the fabric of spacetime caused by movements of masses in the universe. Predicted by Albert Einstein in 1916 as a consequence of his General Theory of Relativity, gravitational waves carry information about their origins as well as about the nature of gravity itself, providing insights into regions of the cosmos that are otherwise hidden from view. In particular, they may be helpful to:

- Detect astrophysical events or entities that due to their physical nature are impossible to detect in the electromagnetic signatures, such as the case of black holes and their dynamics;
- Reveal information about the internal structure of astrophysics bodies such as neutron stars and events such as supernovae. Indeed, traveling unimpeded through internal layers, they provide unique insight into physics below the surface level of an event.

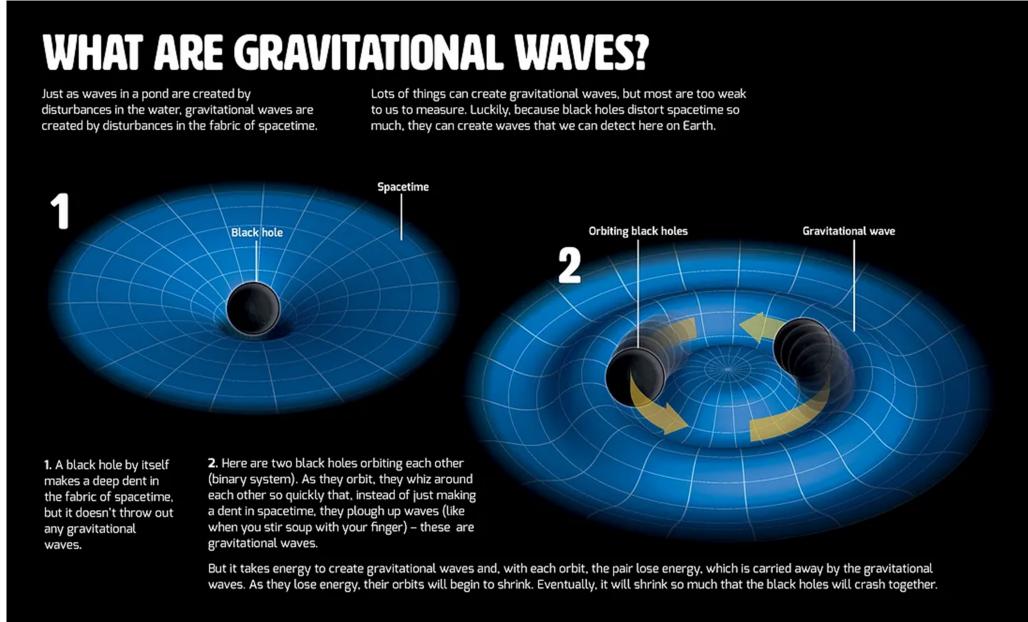


Figure 70. Illustration of the underlying physics of gravitational waves.

Figure 71. Visual Guide to Gravitational Waves.

- Correlate distinct astronomical phenomena, with particular relevance to explain intense and localized phenomena in the electromagnetic spectrum (e.g. gamma-ray bursts and merging of binary neutron stars).

The innerworkings of gravitational waves stem directly from Einstein's field equations of general relativity(GR), which describe how matter and energy influence the curvature of spacetime⁵¹. In very generic terms, GR does not allow the instantaneous (which would violate causality) Newton-gravity interaction but instead introduces a formalism where massive objects distort spacetime which in turn influences the dynamics of bodies. Besides, as this distortion is not instantaneous, it propagates outward at the speed of light in the form of *gravitational waves* when the mass undergoes acceleration, particularly asymmetric acceleration such as in binary orbits. Adding to solving the causality, GR explains also gravitational redshifts, gravitational lensing, and the anomalous precession of the orbit of Mercury around the Sun, being one of the most successful physical theories of the 20th century. A complete derivation of the formalism to describe gravitational waves would require an entire course of GR which is not the intent of this chapter⁵². But we can start by introducing the Einstein equation as the foundation of general relativity, describing how matter and energy influence the curvature of spacetime as

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \quad (1.1)$$

⁵¹ Matter tells spacetime how to curve, and spacetime tells matter how to move. - John Wheeler

⁵² A good reference to start such endeavor would be "A first course in general relativity" by Schultz

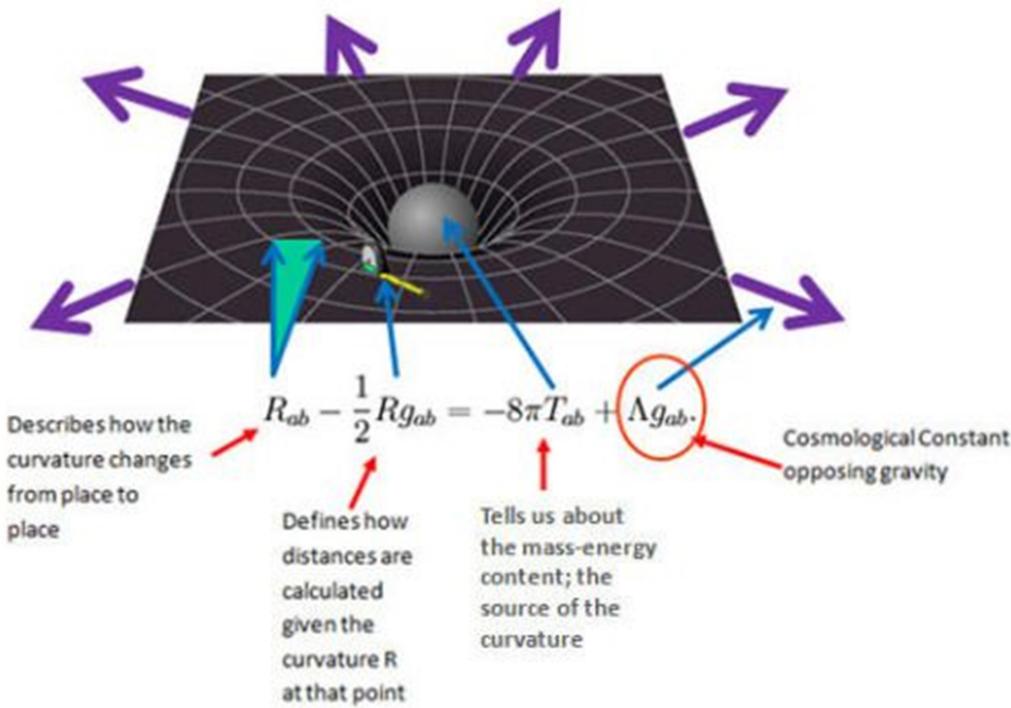


Figure 72. Illustration of the underlying physics of Einstein Equation

with $G_{\mu\nu}$ the Einstein tensor, Λ the cosmological constant, $g_{\mu\nu}$ the metric tensor, G the gravitational constant, c the speed of light and $T_{\mu\nu}$ the stress-energy tensor. In short, the Einstein tensor can be constructed mathematically from the metric as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \quad (1.2)$$

where the scalar curvature R is the trace of the Ricci tensor $R_{\mu\nu}$, i.e.

$$R = g^{\mu\nu}R_{\mu\nu} \quad (1.3)$$

which in turn is obtained from the Riemann tensor as

$$R_{\mu\nu} = R_{\mu\lambda\nu}^\lambda. \quad (1.4)$$

The Riemann curvature tensor is obtained from

$$R_{\sigma\mu\nu}^\rho = \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\mu\sigma}^\lambda \quad (1.5)$$

where the Christoffel symbols of the second kind are given by

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \quad (1.6)$$

On its side, the energy-momentum tensor represents the distribution and

Figure 73. Polarization of gravitational waves.

flow of energy and momentum in spacetime, with derivation depending on the context. But overall, the Einstein equation relates how matter and energy influence the metric of spacetime, affecting the equations of motion of matter in this way. Straightforwardly, any changes of matter and energy propagate as a perturbation of the metric as **gravitational waves**.

For simplicity, we usually assume that the spacetime is only slightly perturbed from a flat condition. In such a case, the metric tensor, which describes spacetime curvature, can be expressed as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \quad (1.7)$$

where $\eta_{\mu\nu}$ represents the Minkowski (flat spacetime) metric and $h_{\mu\nu}$ represents a small perturbation due to gravitational waves. Using the linearized version of Einstein's equations under the assumption of weak fields, and for a particular gauge choice, the wave equation for the perturbations can be derived as

$$\left(\nabla^2 - \frac{1}{c^2} \partial_{t^2}\right) h_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.8)$$

which can be simplified in vacuum using $T_{\mu\nu} = 0$. A common choice for $h_{\mu\nu}$ in this case is to warrant that it is transverse and traceless, meaning that for a wave propagating in the z direction, it can be written as

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (1.9)$$

where h_+ and h_\times can be understood as the amplitudes of the gravitational waves in two orthogonal polarizations (plus and cross), that will depend on the orientation of the source dynamics relative to the detector but that:

- Plus: stretches and compresses space along the x and y axes, respectively;
- Cross: stretches and compresses space but at 45-degree angles to the plus polarization.

In this picture, gravitational waves traveling through Earth cause tiny oscillations in spacetime, alternately stretching and compressing distances but by an amount so small that detecting them is a significant technological challenge. As an influence in the metric, gravitational waves produce a **strain** in

Figure 74. The ecosystem of GW detection in terms of frequency.

spacetime, meaning that its effect will only be felt when measuring the distance between two points in space. Roughly speaking, if completely oriented with the polarization of the wave, the deformation from the initial distance L will be equal to

$$\Delta L = |h_{\mu\nu}|L \quad (1.10)$$

1.1 How intense are these?

The gravitational wave equation does not seem friendly itself and in practice is extremely hard to tackle in an analytical manner. A popular method is to utilize a multipolar expansion, which leads to a quadrupole moment contribution that can be used to derive some figures of merit for the case of two-point masses in circular orbit about a common center of mass and that describes the binary merger of neutron stars and black holes, two of the most intense gravitational phenomena in the universe⁵³.

Using some approximations that we will not describe in detail but that can be found elsewhere⁵⁴ it is possible to put a number in the fluctuations in spacetime on the order of

$$|h_{\mu\nu}| \approx 10^{-21} \quad (1.11)$$

and frequencies up to

$$f_{gw} \approx 800 Hz. \quad (1.12)$$

meaning that ultimately we need to measure minuscule strains at audio frequencies, where environmental noise is abundant.

⁵³ and thus, the more likely to be "observed" in LIGO

⁵⁴ e.g. "Squeezed Vacuum Injection in Advanced LIGO: Enhancing Gravitational-Wave Detection Using Quantum States of Light" by Maggie Tse

2 LIGO: The Classical Design

The Laser Interferometer Gravitational-Wave Observatory (LIGO) was conceived with the goal of directly detecting gravitational waves. It employs optical interferometric concepts and is perhaps the most famous and versatile type of gravitational wave detector (others are Pulsar Timing Arrays - measuring the time variations of pulsar systems - and the future LISA, a space-based interferometer). The conceptual groundwork for LIGO began in earnest in the 1970s, led by pioneers like Rainer Weiss, Kip Thorne, and Ronald Drever, who proposed using laser interferometry to detect gravitational waves. The principle behind the detection method involves measuring the minute changes in distance between two test masses positioned kilometers apart, changes expected to be induced by a passing gravitational wave.

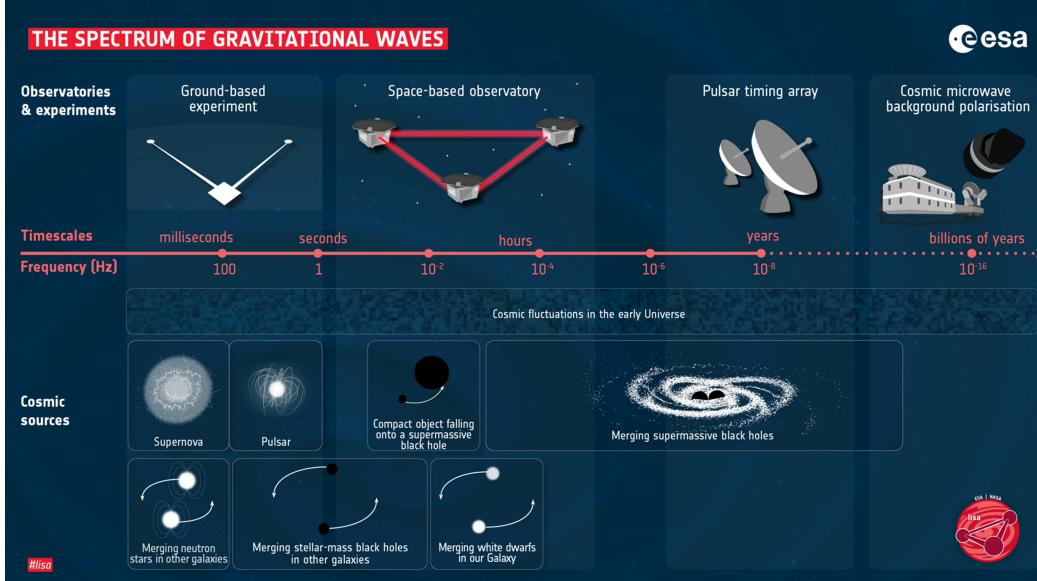


Figure 75. Timescales of gravitational waves from distinct cosmological events.

In 1992, the National Science Foundation (NSF) approved funding for LIGO construction, and by 2002, the observatories at Hanford, Washington, and Livingston, Louisiana, were operational. The first direct detection of gravitational waves by LIGO in September 2015 marked a seminal moment in physics and astronomy. This detection (event GW150914) was sourced to the merger of two black holes approximately 1.3 billion light-years away and confirmed the robustness of general relativity in the strong-field regime. Since then, LIGO, in conjunction with Virgo and other observatories, has detected several more gravitational wave events, each adding to our understanding of the universe.

2.1 The Overall Design and Notes on Components

At its heart, LIGO employs a Michelson interferometer, a design that we already discussed and that has multiple connections with astronomy and cosmology in different epochs of time. The design is simple in concept but extraordinarily complex in execution at LIGO. It consists of two perpendicular arms, each extending four kilometers in length, within which laser light is split into two beams that travel back and forth along the arms. In particular, let us introduce some notes on its components:

Laser: The primary light source in LIGO is a neodymium-doped yttrium aluminum garnet (Nd:YAG) laser capable of emitting a highly stable and coherent beam of light at 1064 nanometers. This laser type is chosen for its high power stability and minimal phase noise, crucial for the precise

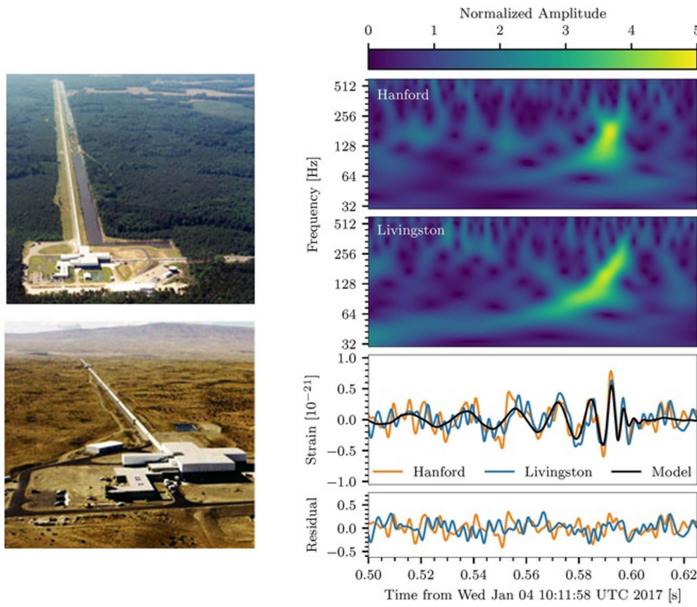


Figure 76. LIGO facilities in Washington and Livingstone and first event detection in 2015.

measurements required in gravitational wave detection. Typically, the laser output in LIGO operates at a power of about 200 watts.

Test Masses and Mirrors: At the heart of LIGO's detection capability are its test masses—four ultra-precise mirrors, two at each site, which serve as the endpoints of the interferometer's arms. These mirrors are made from fused silica, chosen for its low thermal expansion properties and excellent optical quality. Each mirror is meticulously polished to achieve the required flatness and is coated with multiple layers of reflective material to maximize the reflection of laser light. The mirrors also function as test masses due to their substantial mass (about 40 kilograms), helping to isolate them from external vibrations and noise.

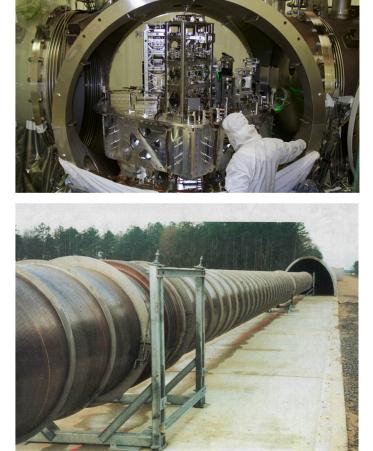
Vacuum System; LIGO's optical components are housed within an immense vacuum chamber, one of the largest and most pristine in the world, boasting a volume of 10,000 cubic meters (353,000 cubic feet) and maintaining a pressure of 10^{-9} Torr, or one-trillionth of an atmospheric pressure. This level of vacuum is surpassed only by that of the Large Hadron Collider in Switzerland. The primary purpose of such an extreme vacuum is to eliminate air molecules and microscopic debris, preventing them from i) affecting measurements and ii) settling on LIGO's mirrors or other optical elements. If contaminants were to come into contact with



Figure 77. Mirrors of LIGO.

the laser beam, they could vaporize and potentially cause irreparable damage to the equipment (each test-mass mirror is valued at approximately 2 million dollars).

Achieving the requisite vacuum conditions within LIGO's chambers involved a lengthy process: it took 1,100 hours (about 40 days) of continuous pumping to reach the optimal operational pressure. During this phase, turbo-pump (jet) vacuums initially removed the majority of the air. Concurrently, the vacuum tubes were heated to temperatures between 150-170 degrees Celsius for 30 days to eliminate any residual gases and moisture embedded in the steel, ensuring the integrity and functionality of the vacuum environment.



2.2 Michelson Interferometer and Dark fringe Operation

In LIGO, the Michelson interferometer is used to measure the difference in path length between its two arms. When a gravitational wave passes through the interferometer, it changes the relative length of the arms very slightly due to the stretching and squeezing of spacetime, which alters the interference pattern observed at the detector due to the introduction of a phase difference. As we saw in previous weeks, the output at the anti-symmetric port for a Michelson will be given

$$I_{PD} = \frac{I_0}{2}(1 + \cos(2\Delta\phi + \phi_0)) \quad (2.1)$$

where ϕ_0 encloses the initial condition for operation whereas $\Delta\phi$ corresponds to the difference in the optical path due to the strain in the spacetime. When completely aligned with the polarization of the gravitational wave we have

$$\Delta\phi = \frac{2\pi h(t)L}{\lambda} \quad (2.2)$$

valid for time-varying strain $h(t)$ with a period much longer than the total round-trip of light in each arm.

In LIGO, and contrary to the cases we previously described in the last weeks, the point of operation is close to the **Dark Fringe** instead of the quadrature point.

Indeed, this point comes with non-trivial advantages:

- **Sensitivity to noise:** while at the quadrature point the sensitivity to changes is maximized, the sensitivity to noise is also maximum.
- **Visibility:** operating near the dark fringe, LIGO ensures that the baseline signal is minimal, thereby maximizing the contrast and visibility of

Figure 78. Every part of LIGO is in a vacuum - even the 4km arms.

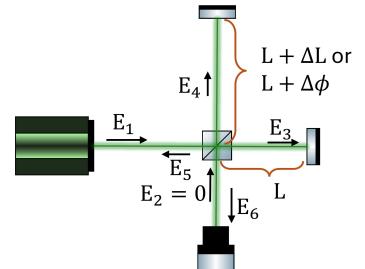


Figure 79. Simplified Scheme of the LIGO Michelson interferometer

any signals produced by gravitational waves.

- **Controllability:** Operating near the dark fringe provides a stable and well-defined reference point for calibration and control systems. It simplifies the process of adjusting the interferometer's settings to maintain optimal sensitivity and ensures that the system's response to gravitational waves can be accurately modeled and predicted. The control systems can maintain the operational point with greater precision when starting from a state of minimal light intensity, where deviations are more easily detectable and correctable.

In dark fringe operation, the interferometer is tuned such that in the absence of a gravitational wave, the path length difference causes the beams to recombine destructively, resulting in minimal or no light intensity at the detector (a dark fringe). Experimentally we choose $\phi_0 = (2m + 1)\pi$ with m an integer corresponding to a case where (utilizing $\lambda = 2\pi c/\omega$) the power at the detector is given by

$$P_{PD} = P \sin^2 \left(2\pi \frac{\Delta L}{\lambda} \right) \approx P \frac{4\omega^2}{c^2} (\Delta L)^2 \quad (2.3)$$

with P being the laser power entering the interferometer.

2.3 Fabry-Perot

The arms of LIGO are approximately 4km long, meaning that the total length change for the numbers introduced before is around

$$\Delta L \approx |h|L \approx 10^{-18} m \quad (2.4)$$

which is still 1000 times smaller than the size of the proton. One can substitute directly into the expression for the power of the photodetector and obtain an extremely small value of $P_{PD} \approx 10^{-22} W$. Taking into consideration that the energy for a single photon at 1064nm is around $1.8 \times 10^{-19} J$ this means that we will be looking at a single photon arriving at each 3 minutes. In practice, this means that while 4-km-long arms already seem enormous, they would still be too short to enable the detection of gravitational waves. Indeed, a good target to warrant detection will be above the femtowatt range (around thousands of photons per second, around 10 per millisecond, allowing to reach the 1000Hz frequency range).

A way to provide additional path length is to alter the Michelson interferometer to include Fabry-Perot cavities, just like in the Fabry-Perot interferometer we described in previous weeks. In this case, an additional mirror is placed in each arm near the beamsplitter and 4 km from the mirror at the end

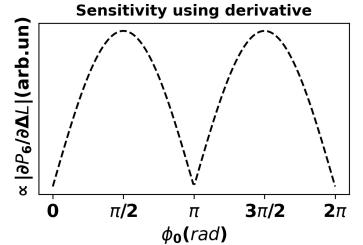


Figure 80. In LIGO the operation point is the dark fringe $\phi_0 = (2m + 1)\pi$ and not the quadrature point $\phi_0 = (2m + 1)\pi/2$.

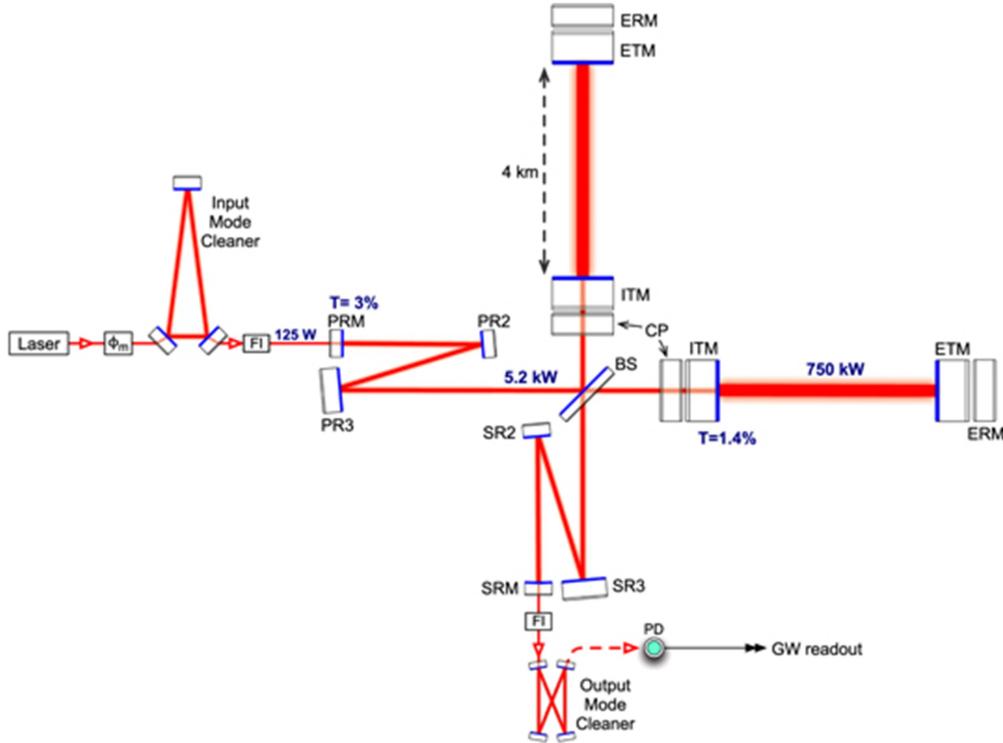


Figure 81. Closer to LIGO actual setup, adding a Fabry-Perot (ITM-ETM) in each path to increase effective path distance, together with the power recycling mirrors (PR).

of that arm. In practice, after entering the instrument via the beam splitter, it can be shown that a Fabry-Perot cavity in such configuration would lead to an increase of the effective path length from L to

$$L_{eff} = \mathcal{F}/\pi \quad (2.5)$$

where \mathcal{F} is the Finesse of the Fabry-Perot cavity.

In the existing configuration $\mathcal{F} \approx 300$ which increases the effective distance traveled by each laser from 4 km to about 1200 km, increasing the signal at the photodetector by two orders of magnitude. Still, it is not yet sufficient to reach the femtowatt or above level we previously introduced as a ballpark figure for gravitational wave detection.

2.4 Power Recycling

Looking at the expression of the power reaching the photodetector one can see that another strategy to increase the signal is to increase the power of the laser itself. While this is somehow difficult to do on the side of the laser itself, the operation in the dark fringe mode allows to employ power recycling strategies, increasing the effective power of the laser light within the

interferometer without increasing the actual laser output. This is achieved through the use of a power recycling mirror placed strategically in the LIGO setup.

The power recycling mirror (PRM) is located at the symmetric port of the interferometer (the port of the laser) and in short it exploits the fact that most of the power exits the interferometer from the original input port. Assuring that:

- the mirror is positioned in such a way that it contributes constructively to the input of the laser;
- the transmissivity of the PRM matches the level of optical loss inside the Michelson, optimizing the coupling of light into the Michelson in the form of impedance matching between the input and the interferometer

allows to utilize the optical power that does not dissipate nor exit the interferometer from the anti-symmetric port to be recirculated rather than re-entering the laser. Indeed, it is possible to demonstrate that this boost corresponds to an effective laser power of approximately 750kW, meaning that a signal in the range of $10^{-15}W - 10^{-13}W$ will be available for detection at the anti-symmetric port.

3 The Noise Limitations

At this point, and classically, we are ready to detect gravitational waves with this LIGO system. Yet, being such small variations, understanding and mitigating noise - in particular in the expected frequency bands - is of fundamental importance. We can divide the noise into two large families: **displacement noise** and the **phase noise**.

Displacement noise: Displacement noise is associated with the unwanted motion of test masses. Dividing in the major sources we have:

- **Ground vibration (including seismic):** occurring at low frequencies (below 1Hz) and can be mitigated by a combination of passive isolation through pendulum systems and active feedback control. Besides, being geographically isolated also contributes to lowering this noise.
- **Thermal noise:** coming from the optical elements - mostly from the optical coatings than from the fused silica itself - and the suspensions of the test masses - fused silica fibers that hang the test masses. The thermal noise contribution is typically small, but future state-of-the-art research is set to target these limits with research

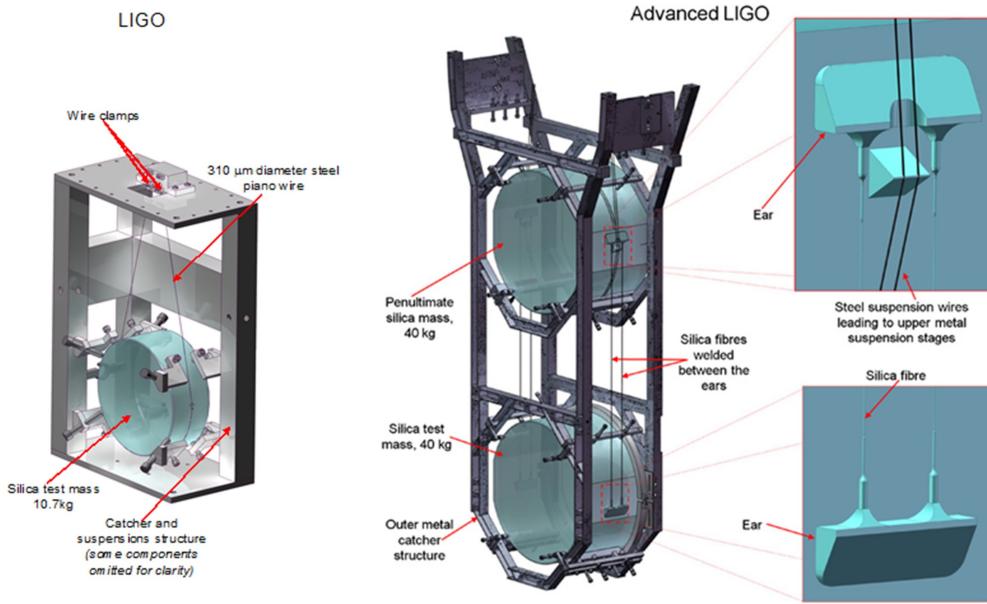


Figure 82. Mitigating ground vibration via freely moving mirrors suspended by thin fibers and differences between LIGO and Advanced LIGO configurations. Starting in 2010, the major reduction in noise of the Advanced LIGO redesign allowed us to observe in 2015 the first event.

into new materials and the choice of geometry for the suspension fibers.

Phase noise: Associates to the unwanted phase shifts and its primary sources are:

- **Residual gas:** small contribution due to the presence of residual atoms in the beam path, but very small due to the vacuum configuration;
- **Scattered light:** arises from imperfections in optics, which can scatter light out of the main beam. This light can then acquire phase shifts by reflecting off vibrating components such as the beam tube and other hardware inside the vacuum chambers, before recombining with the main beam and reaching the readout. Typically dominates below 50Hz.
- **Quantum Fluctuations:** statistical fluctuations in photon arrival times at the readout which can be interpreted as phase fluctuations.

Thus being said, engineering advances play an important role in reducing the noise in LIGO, allowing to mitigate most in classical manners. Yet, the

quantum noise represents a fundamental barrier that even if all the other sources of noise are mitigated in classical manners, there would still be a noise level at the detection - the **quantum noise limit**.

3.1 The Quantum Noise Limit

To understand where the quantum noise terms come from we will need to delve deeper into the fundamental physics of the electromagnetic field. From a quantum perspective, we know that photons are in practice discrete quanta of energy that can be measured by photodetectors. This means that there is an intrinsic **discretization** which ultimately may lead to two distinct effects:

- **Shot noise:** arises from statistical fluctuations in the arrival time of photons at the interferometer output, which can resemble a phase shift.
- **Radiation pressure noise:** with such large effective powers, light in each cavity transfers a non-negligible amount of momentum to the suspended mirrors. Even with the 40 kg mirrors, quantum fluctuations in the field amplitude will translate into fluctuations in the mirror momentum. Varying the mirror position will alter the interferometric conditions and the phase of the light, manifesting as a quantum radiation pressure noise at the exit photodetector.

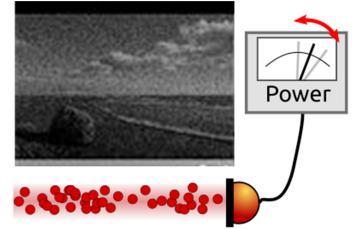


Figure 83. Fluctuations on the arrival time of photons lead to shot noise. This is also observed in low lighting photography.

3.2 Shot Noise Contribution

Shot noise is a quantum-measurement noise associated with the discrete nature of photons arriving on the detection photodiode. Considering the most simple Michelson interferometer we have

$$P_{PD} = P \sin^2 \left(2\pi \frac{\Delta L}{\lambda} \right) \approx P \frac{\omega^2}{c^2} (\Delta L)^2 \quad (3.1)$$

for which the spectral density of the power is given by

$$S_{PD}^{SN} = 2\hbar\omega P_{PD}. \quad (3.2)$$

To transform this to the noise in ΔL measurement we can utilize the fact that for small ΔL we may expand obtain $P_{PD} \approx (\partial P_{PD}/\partial \Delta L)\Delta L$ or in the converse form $\Delta L \approx P_{PD}(\partial P_{PD}/\partial \Delta L)^{-1}$ to obtain

$$S_{\Delta L}^{SN} = \frac{2\hbar\omega P_{PD}}{(\partial P/\partial \Delta L)^2} = \frac{2\hbar\omega P_{PD}}{4\omega^2/c^2 P P_{PD}} = \frac{\hbar c^2}{2\omega P} \quad (3.3)$$

Power spectral density is given by

$$S_{xx}(f) = \lim_{T \rightarrow \infty} \frac{1}{T} |\hat{x}_T(f)|^2$$

with $\hat{x}_T(f)$ the Fourier transform of signal x

Finally one can get the shot noise limited length difference noise from the square root of this power:

$$u(\Delta L_{SN}) = \sqrt{S_{\Delta L}^{SN}} = c \sqrt{\frac{\hbar}{2\omega P}} \quad (3.4)$$

which is given in units of $m\sqrt{Hz}^{-1}$ ⁵⁵. From this formula, we can right away conclude a few things:

- The shot noise is independent of the frequency of the oscillating strain, being thus of the white noise type;
- The shot noise is independent of the power reaching the photodetector;
- The shot noise can be reduced by increasing the power inside the interferometer.

⁵⁵ In practice, this gives the error in the frequency space, which is convenient for this case.

Thus, we can conclude that operating near a dark fringe with a low output power is indeed the ideal condition to reduce shot noise.

Note that in the case of the Michelson interferometer with the additional Fabry-Perot in each arm, there is an additional multiplicative term in the shot noise as

$$u(\Delta L_{SN}^{FP}) = u(L_{SN})(1 + i2\Omega\tau_s) \quad (3.5)$$

where τ_s corresponds to the storage time of the arm cavities. This adds a dependency on the gravitational wave frequency that for the sake of simplicity we will ignore in the following sections.

3.3 Radiation Pressure Noise

Radiation pressure noise in LIGO results from the momentum transfer of photons to the mirrors. As photons reflect off the mirrors, they exert a force due to their momentum, causing tiny fluctuations in the position of the mirrors. These fluctuations contribute to noise in the measurement of gravitational waves, particularly at lower frequencies.

The mirror can be modeled as a harmonic oscillator of angular frequency Ω and mass m with the equation of motion given by

$$m\ddot{x} + m\Omega^2x = F \quad (3.6)$$

which solved in the Fourier space leads to

$$\hat{x}(\Omega) = \frac{2P}{2m\Omega^2c} \quad (3.7)$$

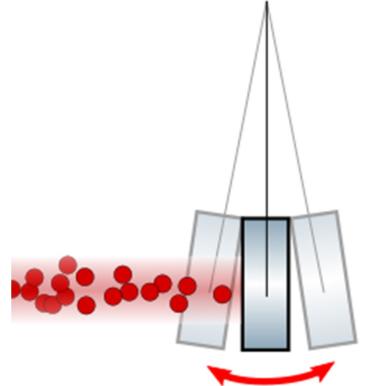


Figure 84. Fluctuations on the arrival time of photons in the mirrors may also lead to radiation pressure noise.

where we substituted the force $F = 2P/c$ for the photons reaching the mirror. The power spectral density for x and consequently ΔL can be obtained from the definition and the spectral density for the power $S_P = 2\hbar\omega P$, obtaining the differential arm length due to radiation pressure noise is then given by

$$S_{\Delta L}^{RPN} = \frac{2\hbar\omega P}{m^2\Omega^4 c^2} \quad (3.8)$$

which is associated to a limit

$$u(\Delta L_{RPN}) = \sqrt{S_{\Delta L}^{RPN}} = \frac{1}{m\Omega^2 c} \sqrt{2\hbar\omega P}. \quad (3.9)$$

Analyzing the equation we see that:

- the noise is frequency dependent, and decreases with frequency;
- increasing the circulating power adds more noise, but increasing the mirror masses can be used to arbitrarily reduce the impact of quantum noise.

3.4 Reducing noise classically

From the noise derived in the previous sections, it appears that one can increase the circulating power in the interferometer arbitrarily to reduce shot noise, while simultaneously increasing the mirror mass to compensate for the power increase, and thus still reduce radiation pressure noise. However, in practice, this would lead to encountering other limits for the usual behavior. First, the thermal deformation of the mirrors due to light absorption would create thermal lensing requiring further compensation. Then, the mirror mass is limited by the ability to produce large polished mirrors, and the ability to suspend them. In this case, solving quantum problems with classical techniques can only get us so far, once we reach the limits of current technology.

4 Quantum Limit and Beyond

From the equations introduced before it is possible to compute a lower boundary for the sum $u(\Delta L_{SN}) + u(\Delta L_{RPN})$ by carefully choosing P and assuming **no correlations between shot noise and radiation pressure noise**. Indeed, by computing the minimum in relation to P and substituting this value, it is straightforward to obtain the minimum

$$u(\Delta L_{SQL}) = \sqrt{S_{\Delta L}^{SQL}(\Omega)} = \frac{1}{\Omega} \sqrt{\frac{2\hbar}{m}} \quad (4.1)$$

each photon carries $p = h/\lambda = \hbar\omega/c$ and is reflected

which is denoted the **standard quantum limit**, represented Figure 85. In essence, the SQL is a manifestation of the Heisenberg uncertainty principle, and represents the best quantum noise performance achievable in an interferometer without the use of nonclassical techniques. In practice, from the perspective of gravitational-wave astrophysics, this limits the higher measurement precision in an interferometer and the detection of fainter signals from smaller and more distant objects. So, can we go beyond it?

4.1 The origin of the noise

To really understand where the noise comes from, we will need to enter into a full quantum picture. We will not enter into detail on the quantum theory itself and try to get only the necessary tools along the way. Yet keep in mind that the end goal is to understand two key concepts: the ground state fluctuations of the electromagnetic field, and what happens in a quantum version of a beamsplitter.

4.1.1 Quantization of the Electromagnetic Field

First, we will start with the quantization of the electromagnetic field. This process involves promoting the classical electromagnetic fields to quantum operators and expressing them in terms of creation and annihilation operators and starts from the vector potential.

The classical electromagnetic field can be described by the vector potential $\mathbf{A}(\mathbf{r}, t)$ and the scalar potential $\phi(\mathbf{r}, t)$. In the Coulomb gauge, where $\nabla \cdot \mathbf{A} = 0$ and $\phi = 0$, the electric and magnetic fields are given by:

$$\mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{A}(\mathbf{r}, t)}{\partial t},$$

$$\mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t).$$

The vector potential $\mathbf{A}(\mathbf{r}, t)$ can be expanded as a sum over modes characterized by the wave vector \mathbf{k} and polarization index λ :

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k}, \lambda} \left(\mathcal{A}_{\mathbf{k}, \lambda} \mathbf{e}_{\mathbf{k}, \lambda} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}} t)} + \mathcal{A}_{\mathbf{k}, \lambda}^* \mathbf{e}_{\mathbf{k}, \lambda}^* e^{-i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}} t)} \right),$$

where $\mathcal{A}_{\mathbf{k}, \lambda}$ are the complex amplitudes of the vector potential, $\mathbf{e}_{\mathbf{k}, \lambda}$ are the polarization vectors, and $\omega_{\mathbf{k}} = c|\mathbf{k}|$ is the angular frequency of the mode.

Now, to quantize the electromagnetic field, we simply promote the classical field amplitudes to quantum operators. The vector potential is then

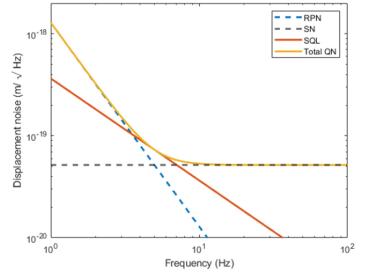


Figure 85. Shot Noise, Radiation Pressure, and Standard Quantum Limit, computed for a displacement noise spectral density for a Michelson interferometer with mirrors 10kgs mass suspended as pendulums and a total light power of 1MW.

expressed in terms of the creation and annihilation operators $\hat{a}_{\mathbf{k},\lambda}$ and $\hat{a}_{\mathbf{k},\lambda}^\dagger$:

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar}{2\epsilon_0\omega_{\mathbf{k}}V}} (\hat{a}_{\mathbf{k},\lambda} \mathbf{e}_{\mathbf{k},\lambda} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)} + \hat{a}_{\mathbf{k},\lambda}^\dagger \mathbf{e}_{\mathbf{k},\lambda}^* e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)}),$$

where $\hat{a}_{\mathbf{k},\lambda}$ and $\hat{a}_{\mathbf{k},\lambda}^\dagger$ are the annihilation and creation operators for the mode (\mathbf{k}, λ) , V is the quantization volume, and ϵ_0 is the permittivity of free space.

Using the quantized vector potential, the electric and magnetic fields can be written as:

$$\mathbf{E}(\mathbf{r}, t) = i \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0 V}} (\hat{a}_{\mathbf{k},\lambda} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)} - \hat{a}_{\mathbf{k},\lambda}^\dagger e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)}) \mathbf{e}_{\mathbf{k},\lambda},$$

$$\mathbf{B}(\mathbf{r}, t) = \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\mu_0 V}} (\hat{a}_{\mathbf{k},\lambda} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)} + \hat{a}_{\mathbf{k},\lambda}^\dagger e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)}) (\hat{\mathbf{k}} \times \mathbf{e}_{\mathbf{k},\lambda}),$$

where μ_0 is the permeability of free space.

Finally, the Hamiltonian of the quantized electromagnetic field can be derived from the energy density of the electric and magnetic fields. It is given by:

$$\hat{H} = \frac{1}{2} \int \left(\epsilon_0 |\mathbf{E}|^2(\mathbf{r}, t) + \frac{1}{\mu_0} |\mathbf{B}|^2(\mathbf{r}, t) \right) d^3 r.$$

Substituting the expressions for $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$, and integrating over the volume, we obtain:

$$\hat{H} = \sum_{\mathbf{k},\lambda} \hbar\omega_{\mathbf{k}} \left(\hat{a}_{\mathbf{k},\lambda}^\dagger \hat{a}_{\mathbf{k},\lambda} + \frac{1}{2} \right).$$

This is the Hamiltonian in terms of the creation and annihilation operators, showing the quantized nature of the electromagnetic field.

4.1.2 States and Operators to handle the quantized electromagnetic field

In practice, after the quantization of the electromagnetic field, we end up working in terms of states and operators. There will be two types of states - the Fock states and Coherent states - which are closely related to the action of the creation and annihilation operators.

Properties of the Creation and Annihilation operators: As introduced in the quantization step the Creation and Annihilation operators feature very interesting properties that we must first introduce.

Action on Fock States: These operators are defined by their action

on the Fock basis (or number basis), given by

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle, \quad (4.2)$$

and

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \quad (4.3)$$

Canonical commutation relations: The creation and annihilation operators satisfy the following canonical commutation relations:

$$[\hat{a}, \hat{a}^\dagger] = 1, \quad (4.4)$$

and

$$[\hat{a}, \hat{a}] = [\hat{a}^\dagger, \hat{a}^\dagger] = 0. \quad (4.5)$$

Vacuum State: In the Fock basis the vacuum state is given by $|0\rangle$ and is annihilated by \hat{a}

$$\hat{a}|0\rangle = 0 \quad (4.6)$$

while \hat{a}^\dagger generates the first excited state

$$\hat{a}^\dagger|0\rangle = |1\rangle. \quad (4.7)$$

Number Operator: The number operator is just

$$\hat{n} = \hat{a}^\dagger\hat{a}. \quad (4.8)$$

and it gives the number of photons in the state

$$\hat{n}|n\rangle = n|n\rangle. \quad (4.9)$$

Beamsplitters and interferometers: To handle the same matrix formalism of the classical field applies, e.g. for a beamsplitter

$$\begin{bmatrix} \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} = \overline{BS} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} \quad (4.10)$$

which is very convenient to keep track of all the contributions of the fields at the output ports.

Hamiltonian: Using these operators it is possible to write the Hamiltonian for the electromagnetic field as

$$\hat{H} = \hbar\omega \left(\hat{a}^\dagger\hat{a} + \frac{1}{2} \right). \quad (4.11)$$

which means that the vacuum state has an energy contribution!

Exercise 9. Using the action of the operators on the Fock basis, demonstrate the commutation relations.

In addition to the Fock states (or number states), coherent states are particularly relevant when handling real experimental conditions. Indeed, these approximate quite well the output of a laser beam.

Coherent states: these states are particularly important in the context of quantum optics as they can be used to describe laser light. A coherent state $|\alpha\rangle$ may be constructed in the Fock basis as

$$|\alpha\rangle = \exp\left(\frac{-|\alpha|^2}{2}\right) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \quad (4.12)$$

and it can be demonstrated that it is an eigenstate of the annihilation operator

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle, \quad (4.13)$$

and can be constructed from the vacuum state by applying the displacement operator

$$|\alpha\rangle = D(\alpha)|0\rangle, \quad (4.14)$$

given by

$$D(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a}). \quad (4.15)$$

Note that for such a state, both the mean number of photons $\langle\alpha|\hat{n}|\alpha\rangle$ and variance $\langle\alpha|\hat{n}^2|\alpha\rangle - \langle\alpha|\hat{n}|\alpha\rangle^2$ equals to $|\alpha|^2$.

Another interesting fact is that the probability of finding a state of n photons in the coherent state follows a Poisson distribution given by

$$P(n) = |\langle n|\alpha\rangle|^2 = \frac{|\alpha|^{2n} e^{-|\alpha|^2}}{n!} \quad (4.16)$$

Exercise 10. Demonstrate that the coherent state is an eigenstate of the annihilation operator.

Exercise 11. Demonstrate that the $P(n) = |\langle n|\alpha\rangle|^2 = \frac{|\alpha|^{2n} e^{-|\alpha|^2}}{n!}$.

Finally, a very interesting tool to deal with some calculations that will follow are the quadrature operators.

Quadrature operators: The quadrature operators are two operators that can be defined in terms of the creation and annihilation operators as

$$\hat{X} = \frac{\hat{a} + \hat{a}^\dagger}{2}; \hat{Y} = \frac{(\hat{a} - \hat{a}^\dagger)}{2i} \quad (4.17)$$

There are a few advantages on using these operators. First, the Hamiltonian becomes

$$\hat{H} = \hbar\omega(\hat{n} + 1/2) = \hbar\omega(\hat{X}^2 + \hat{Y}^2). \quad (4.18)$$

which means that the zero-point energy is now connected with the quadrature variances of the vacuum state. Besides, the two quadrature operators do not commute,

$$[\hat{X}, \hat{Y}] = \frac{i}{2} \quad (4.19)$$

which leads to a Heisenberg uncertainty relation for the quadrature variances defined as

$$(\Delta\hat{X})^2(\Delta\hat{Y})^2 \geq \frac{1}{16} \quad (4.20)$$

Applying the vacuum state is trivial to demonstrate that it is an example of a minimum uncertainty state with equal uncertainties in each quadrature, i.e.

$$(\Delta\hat{X})^2 = (\Delta\hat{Y})^2 = \frac{1}{4} \quad (4.21)$$

which saturates the Heisenberg uncertainty principle.

Finally, you can also define the quadrature operators in the generalized form

$$\hat{X}(\theta) = \frac{\hat{a}e^{i\theta} + \hat{a}^\dagger e^{-i\theta}}{2}; \hat{Y}(\theta) = \frac{(\hat{a}e^{i\theta} - \hat{a}^\dagger e^{-i\theta})}{2i} \quad (4.22)$$

Heisenberg uncertainty relation: If $\hat{C} = [\hat{A}, \hat{B}]$ then $(\Delta A)^2(\Delta B)^2 \geq \frac{1}{4}|\langle C \rangle|^2$.

Exercise 12. Compute the variance of the quadrature operators for the vacuum state and show that this state minimizes the Heisenberg uncertainty principle.

Exercise 13. Prove the Hamiltonian form for the quadrature operators and establish the connection between the zero point energy of the vacuum state and the variance of the quadratures.

4.2 A Quantum Model for the Michelson Interferometer

Having introduced the necessary tools we will now describe the Michelson interferometer in the quantum picture, utilizing for this the matrix formalism. Using the notation introduced in Figure 86 we have that annihilation operators after the first beam splitter are given by

$$\begin{aligned} \begin{bmatrix} \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} &= \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\frac{2\pi L}{\lambda}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\phi/2}(\hat{a}_1 + i\hat{a}_2) \\ i\hat{a}_1 + \hat{a}_2 \end{bmatrix} \end{aligned} \quad (4.23)$$

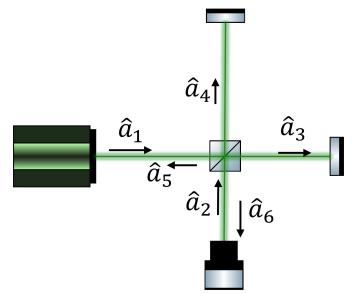


Figure 86. Michelson interferometer in operator picture.

and at the output ports

$$\begin{aligned} \begin{bmatrix} \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\frac{2\pi L}{\lambda}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{i\frac{2\pi L}{\lambda}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} (e^{i\phi} - 1)\hat{a}_1 + i(e^{i\phi} + 1)\hat{a}_2 \\ i(e^{i\phi} + 1)\hat{a}_1 + (1 - e^{i\phi})\hat{a}_2 \end{bmatrix} \end{aligned} \quad (4.24)$$

. In this form we note that the beam splitter mix the states in uneven form, meaning that results both at the mirrors and at the output ports will be different from the quantum perspective levels. Indeed, the presence of fluctuations of the ground state will require to keep track of the \hat{a}_2 terms even when the state entering port 2 is the vacuum state $|0\rangle_2$. This is different from the classical case where we drop the $E_2 = 0$ terms and will lead to very distinct results.

4.2.1 Radiation Pressure

The radiation pressure is connected to the fluctuation of the force acting on the mirrors. In this way, we can say that the force difference will be connected to the fluctuation of the number of photons reaching the two mirrors, which means that the radiation pressure will be proportional to $\hat{n}_4 - \hat{n}_3$ which leads to

$$\begin{aligned} P_{rad} \propto \hat{n}_4 - \hat{n}_3 &= \hat{a}_4^\dagger \hat{a}_4 - \hat{a}_3^\dagger \hat{a}_3 \\ &= \frac{i}{2} (\hat{a}_2 \hat{a}_1^\dagger - \hat{a}_2^\dagger \hat{a}_1) \end{aligned} \quad (4.25)$$

Using a coherent state $|\beta\rangle_1$ with $\beta = |\beta|e^{i\theta}$ we have

$$\hat{P}_{rad} \propto |\beta| \hat{Y}_2(\pi/2 - \theta) \quad (4.26)$$

meaning that the fluctuations of the radiation pressure are connected with the fluctuations of the quadrature operator \hat{Y} , and thus, with the zero point energy of the electromagnetic field. Radiation pressure can then be understood as an effect that **occurs due to the beating of the coherent field and the vacuum entering port 2**. Besides, it is proportional to the variance $Var(\hat{Y}_2(\pi/2 - \theta))$, which for the vacuum state $|0\rangle_2$ is $1/4$, giving

$$u(\Delta L_{RPN}) \propto \sqrt{Var(\hat{P}_{rad})} = |\beta| \sqrt{Var(\hat{Y}_2(\pi/2 - \theta))} = \frac{1}{2}|\beta| = \frac{1}{2}\sqrt{\frac{P}{\hbar\omega}} \quad (4.27)$$

Where we utilize the definition

4.2.2 Shot Noise

For computing the shot noise we first compute the power at the output port 6, proportional to the mean value of the number operator \hat{n}_6 , i.e.

$$\begin{aligned} \hat{P}_6 \propto \hat{a}_6^\dagger \hat{a}_6 &= \frac{1}{4}(\gamma^* \hat{a}_1^\dagger + \sigma^* \hat{a}_2^\dagger)(\gamma \hat{a}_1 + \sigma \hat{a}_2) \\ &= \frac{1}{4} \left(|\gamma|^2 \hat{a}_1^\dagger \hat{a}_1 + |\sigma|^2 \hat{a}_2^\dagger \hat{a}_2 + \gamma \sigma^* \hat{a}_2^\dagger \hat{a}_1 + \gamma^* \sigma \hat{a}_1^\dagger \hat{a}_2 \right), \end{aligned} \quad (4.28)$$

where we defined $\gamma = e^{i\phi} + 1$ and $\sigma = 1 - e^{i\phi}$. Taking the mean for the state $|\psi\rangle = |\beta\rangle_1 |0\rangle_2$ we get

$$\bar{P}_6 \propto \langle \psi | \hat{n}_6 | \psi \rangle = \frac{1}{4} |\beta|^2 |\gamma|^2. \quad (4.29)$$

Now to compute the noise, one should look at the variance of the operator \hat{P}_6 looking at its definition it is trivial to demonstrate that for state $|\psi\rangle = |\beta\rangle_1 |0\rangle_2$ we have

$$\begin{aligned} Var(\hat{P}_6) \propto Var(\hat{n}_6) &= \frac{1}{16} \left(|\gamma|^4 Var(\hat{n}_1) + |\sigma|^4 Var(\hat{n}_2) + 4|\sigma|^2 |\gamma|^2 |\beta|^2 Var(\hat{X}_2) \right) \\ &= \frac{1}{16} \left(|\gamma|^4 |\beta|^2 + 4|\gamma|^2 |\sigma|^2 |\beta|^2 Var(\hat{X}_2) \right) \end{aligned} \quad (4.30)$$

where we utilize the fact that the variance of the number operator for a Fock state is 0 and for a coherent state is $|\beta|^2$.

Before ending the discussion it is important to make three remarks:

- First, we see again a relation between noise and fluctuations of vacuum, now in the form of the other quadrature operator \hat{X}_2 . This mathematical formalism was first proposed by Caves in 1981, with the conclusion that the quantum noise at the readout is also related to the quantum vacuum state entering port 2.

- It is trivial to demonstrate that

$$\begin{aligned} |\gamma|^2 &= 2(1 + \cos(\phi)) \\ |\sigma|^2 &= 2(1 - \cos(\phi)) \end{aligned} \quad (4.31)$$

and considering the case of vacuum state, i.e. $\text{Var}(\hat{X}_2) = 1/4$ this means that

$$\text{Var}(\hat{P}_6) = \frac{|\beta|^2}{4}(1 + \cos(\phi)). \quad (4.32)$$

If we compute the signal-to-noise ratio for a small gravitational wave producing ΔL and considering $\phi = \frac{4\pi\Delta L}{\lambda} + \phi_0$, we can take the signal as a small perturbation to the power of the port 6

$$\Delta \bar{P}_6 = \frac{\partial \bar{P}_6}{\partial \Delta L} \Delta L \propto \frac{2|\beta|^2 \pi}{\lambda} \sin(\phi_0) \Delta L \quad (4.33)$$

and the noise as the root square of the variance, meaning that

$$\text{SNR} \propto \frac{\Delta \bar{P}_6}{\sqrt{\text{Var}(\hat{P}_6)}} = \frac{4|\beta|^2 \pi \sin(\phi_0) \Delta L}{\lambda |\beta| \sqrt{1 + \cos(\phi_0)}} = \frac{2|\beta| \pi \sin(\phi_0)}{\lambda \cos(\phi_0/2)} \Delta L. \quad (4.34)$$

By plotting the SNR (see Figure 87) it can be seen that it attains its maximum at $\phi_0 = (2m + 1)\pi$, i.e. **at the dark fringe**, proving our empirical statement introduced before in this chapter.

- Finally, by following the same methodology as before, we can compute the maximum uncertainty related with the shot noise as

$$u(\Delta L_{SN}) = \frac{\sqrt{\text{Var}(\hat{P}_6)}}{\frac{\partial \bar{P}_6}{\partial \Delta L}} = \frac{\lambda}{2\pi|\beta|} = c\sqrt{\frac{\hbar}{2\omega P}} \quad (4.35)$$

4.3 Squeezed States and Reduction of noise

Recovering the uncertainty relation for the quadrature operators we see that we have a fundamental limit. Yet, it still presents some degree of versatility: we can for example increase the variance in the direction of \hat{X} while decreasing in the direction of \hat{Y} . Taking into consideration our picture for noises, this would mean to reduce the shot-noise at the expense of increasing the radiation pressure noise, which depending on the zone we are working can lead to higher sensitivities. This is the concept behind **vacuum squeezing**.

In the context of quantum optics, squeezing involves manipulating the quantum state of light to reduce noise in one observable at the expense of increased noise in its conjugate variable, respecting the Heisenberg Uncertainty Principle. The process is typically realized by using a nonlinear op-

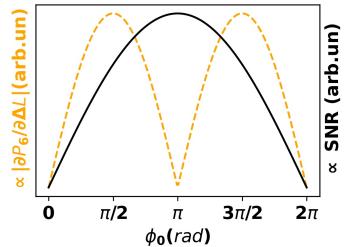


Figure 87. In LIGO the operation point is the dark fringe and not the quadrature point and that is due to the fact that the SNR for the shot noise contribution has a minimum in that region.

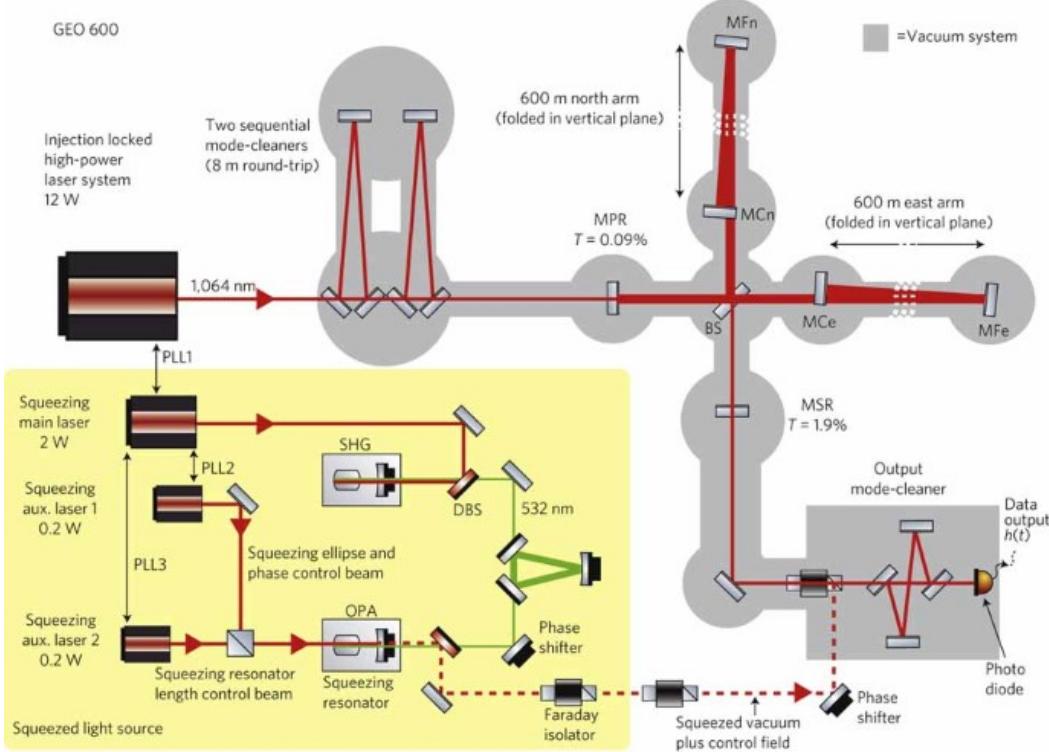


Figure 88. Squeezing module in LIGO. First, the coherent state from the laser is doubled in frequency using second harmonic generation in a first nonlinear crystal (green path), which is retrieved back again to the original frequency using four-wave mixing in the second crystal. Note that a squeezed state is not a vacuum state, but a state that is created from vacuum for which there are photon pairs with some degree of correlations.

tical medium to perform parametric down-conversion (e.g. pumping a BBO crystal with double the frequency, creating pairs of photons at half of the frequency) or four-wave mixing(see Figure 88). For example, a squeezing interaction can arise from the Hamiltonian written as

$$\hat{H}_s = i\hbar\chi(\hat{a}^\dagger \hat{b} - \hat{a}^2 \hat{b}^\dagger) \quad (4.36)$$

where \hat{a} and \hat{a}^\dagger are operators for the field that exits the crystal at frequency $\omega_2 = 2\omega_1$, whereas \hat{b} and \hat{b}^\dagger are operators for the field entering the crystal at frequency ω_1 . Also, χ is a parameter related to the nonlinear properties of the medium and intensity of the pump beam, controlling **the strength of the squeezing interaction**.

It can be noted that the form of the Hamiltonian means that one photon at frequency ω_1 is annihilated/created leading to creation/annihilation of two photons at frequency ω_2 . Beside, if we consider that we have an input state

$$|\psi_{in}\rangle = |\xi\rangle_1 |0\rangle_2 \quad (4.37)$$

meaning that at the entrance of the crystal, we have a coherent state (i.e. a laser beam) on the field oscillating at ω_1 and a vacuum state on the field oscillating at ω_2 , we can get a reduced Hamiltonian in the state space of ω_2 given by

$$\hat{H}_s = i\hbar\chi(\xi\hat{a}^{\dagger 2} - \xi^*\hat{a}^2) \quad (4.38)$$

Taking the Schrödinger equation

$$i\hbar\partial_t|\psi\rangle_2 = \hat{H}_s|\psi_{in}\rangle_2 \quad (4.39)$$

and the formal

$$|\psi_{out}\rangle_2 = \exp\left(\frac{\hat{H}_s t}{i\hbar}\right)|\psi_{in}\rangle_2 \quad (4.40)$$

we get

$$|\psi_{out}\rangle_2 = \exp\left(\chi t(\xi\hat{a}^{\dagger 2} - \xi^*\hat{a}^2)\right)|\psi_{in}\rangle_2 = \hat{S}(r)|\psi_{in}\rangle_2 \quad (4.41)$$

where the **squeezing operator** can be

$$\hat{S}(r) = \exp\left(\frac{1}{2}(r\hat{a}^{\dagger 2} - r^*\hat{a}^2)\right) \quad (4.42)$$

with $r = 2\chi|\xi|t$.

4.3.1 Surpassing the SQL using squeezing

In the context of our case study, we can proceed with our analysis by applying the squeezing operator to the vacuum state, i.e.

$$|r\rangle = \hat{S}(r)|0\rangle_2 \quad (4.43)$$

where $|r\rangle$ is now the squeezed vacuum state. For the sake of simplicity, we will consider r as a real value from now on.

It can be shown that the squeezing operation transforms the annihilation and creation operators as

$$\hat{S}^\dagger(r)\hat{a}\hat{S}(r) = \hat{a}\cosh(r) - \hat{a}^\dagger\sinh(r) \quad (4.44)$$

$$\hat{S}^\dagger(r)\hat{a}^\dagger\hat{S}(r) = \hat{a}^\dagger\cosh(r) - \hat{a}\sinh(r) \quad (4.45)$$

leading to the transformation of the quadrature operators

$$\hat{X}' = \hat{S}^\dagger(r)\hat{X}\hat{S}(r) = e^{-r}\hat{X} \quad (4.46)$$

and

$$\hat{Y}' = \hat{S}^\dagger(r)\hat{Y}\hat{S}(r) = e^r\hat{Y} \quad (4.47)$$

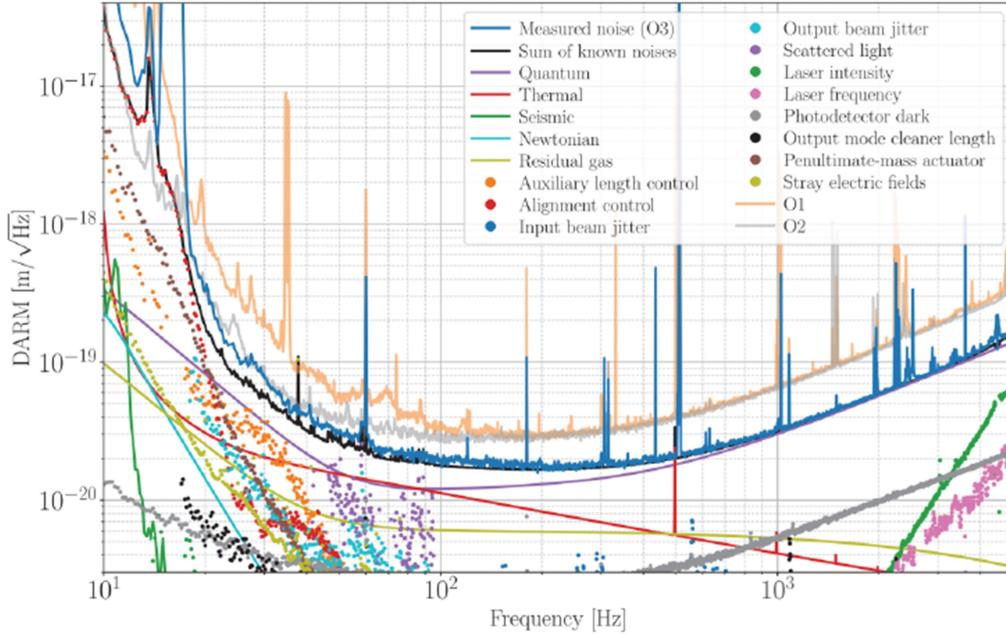


Figure 89. Squeezing noise reduction may be observed in the experimentally calculated noise for LIGO, in particular from O1 and O2 observation seasons to O3.

It is then straightforward to compute the squeezed variances

$$\text{Var}(\hat{X}') = e^{-2r} \text{Var}(\hat{X}) = \frac{1}{2} e^{-2r}, \quad (4.48)$$

and

$$\text{Var}(\hat{Y}') = e^{2r} \text{Var}(\hat{Y}) = \frac{1}{2} e^{2r}. \quad (4.49)$$

This means that the squeezing Hamiltonian generates correlations that effectively squeeze the quantum state in the phase space, reducing uncertainty in one of the directions at the expense of the other.

Putting this result against the work we have done so far, **this demonstrates that a squeezing process allows to enhance the precision of the measurement of gravitational waves, surpassing the standard quantum limit baseline introduced before.** Note however that this is done within the hard boundaries defined by the Heisenberg limit and in a broader perspective means that the use of **squeezing may help improve the precision of sensors beyond the limits imposed by quantum fluctuations.**

Exercise 14. Using the modified Baker-Campbell-Hausdorff formula

$$e^{\hat{A}} \hat{B} e^{-\hat{A}} = \hat{B} + \frac{1}{2} [\hat{A}, \hat{B}] + \frac{1}{2!} [\hat{A}, [\hat{A}, \hat{B}]] + \dots \quad (4.50)$$

prove that for a real-valued r , the annihilation operator is transformed by the squeezing process as

$$\hat{S}^\dagger(\xi) \hat{a} \hat{S}(\xi) = \hat{a} \cosh(r) - \hat{a}^\dagger \sinh(r). \quad (4.51)$$

Optical state Phase Space: A common and intuitive representation of an optical state comes in the form of the phase space defined using the quadrature operators. The precise definition involves Wigner function formalism but in simple terms, the idea is to represent each state as a probability density function centered in the mean value of quadrature operators \hat{X} and \hat{Y} and with widths that vary depending on the variances $Var(\hat{X})$ and $Var(\hat{Y})$, which better illustrates the concept of state squeezing.

Note the connection of these with the position and momentum operators of a harmonic oscillator

5 Concluding Remarks

LIGO has marked a revolutionary milestone in the field of physics, enabling the ground-breaking detection of gravitational waves that not only confirmed key predictions of Einstein's General Theory of Relativity but also pave for an entirely new way of observing the universe. Yet, achieving this breakthrough required the development of innovative engineering solutions and state-of-the-art technologies.

Indeed, the extreme sensitivity needed to detect gravitational waves first required enhancements at the engineering and technological levels. However, this quest for ultra-sensitivity also brought to light the fundamental barriers imposed by the quantum nature of light. With the increase of LIGO sensitivity, it became evident that classical sources of noise could be mitigated only up to a point: the **standard quantum limit**, related to quantum fluctuations and influencing both the power detected - **shot noise** - but also the optical force on the mirrors - **radiation pressure**.

In this chapter, we explored these problems in depth, overviewing LIGO construction and analyzing in detail the noise origins and mitigation strategies. In particular, in the last part, by modeling the interferometer within a quantum framework, we identified that the noise results from the interaction between the coherent laser light and the vacuum state of the electromagnetic field entering the unused port of the beamsplitter. The interference of vacuum fluctuations and the main laser leads to uncertainties in the phase and

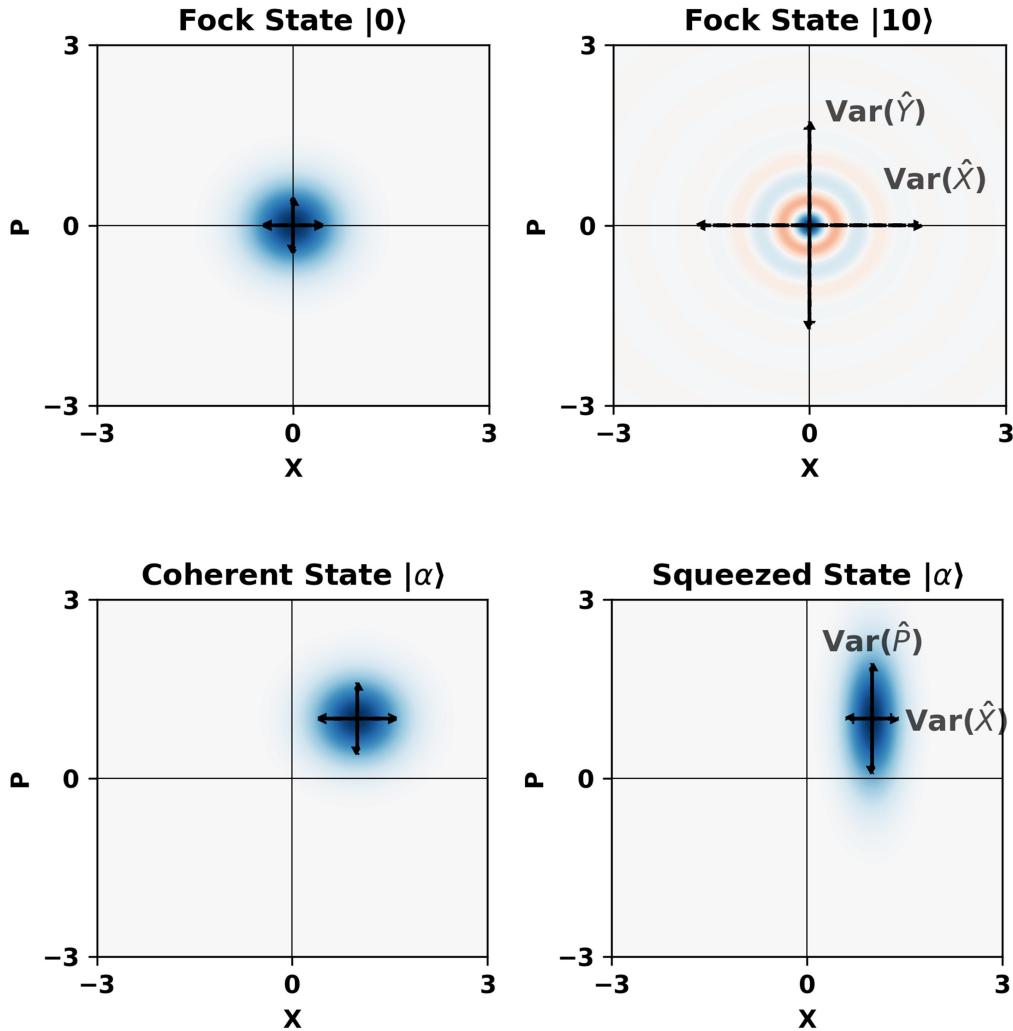


Figure 90. Examples of the representation of fock, coherent, and squeezed states on the Optical Phase Space.

amplitude measurements by perturbing the radiation pressure and shot noise at the detector.

Finally, leveraging on the Heisenberg uncertainty principle and introducing the concept of squeezed states of light, we show that through nonlinear optics it is possible to manipulate the vacuum state achieving a squeezing that can reduce the uncertainty in one observable (such as phase) at the expense of increased uncertainty in the conjugate observable (such as amplitude). This technique effectively redistributes the quantum noise, allowing for a reduction in the noise component. Thus, we show that through the use of quantum formalism is not only possible to understand the true origin of the noise but also to present strategies to mitigate it, paving for sensors that are able to operate beyond the standard quantum limit

A final note on the topic regards the frequency dependency of the noise

in LIGO. The squeezed vacuum we described does not depend on the frequency, meaning that it has the same impact across all frequencies. However, to enhance detection capabilities, it would be beneficial to create a state with quadrature variances that vary with the frequency, e.g. transitioning from phase squeezing at higher frequencies to amplitude squeezing at lower frequencies. This frequency-dependent squeezing can be done experimentally by adding a frequency-dependent delay, accomplished by reflecting the phase-squeezed vacuum off a long optical cavity. This method optimizes the squeezing effect across different frequencies, improving the sensitivity of LIGO measurements.

Sensing with Quantum Optics

Week

X

Quantum engineering is a multidisciplinary approach to quantum physics, focused on exploiting **quantum phenomenology to develop a novel technology**, usually referred to as **quantum technology**.

In short, to understand the concept you can establish a parallel with the history of Electricity: in the early eighteenth century, Physicists like Volta, Ampère, Ohm, and Maxwell established the foundations of electricity; later in the nineteenth century led by the likes of Bell, Tesla, Edison and Hertz give birth to electrical engineering by bringing electricity laws to the real world in the form of technology. As a result, today most of the electronic technology used by humanity is based on the fundamental principles established by Faraday and his contemporaries.

Quantum Engineering aims the same: to bring the principles and laws of quantum physics developed in the twentieth century by scientists like Planck, Einstein, Bohr, and Schrödinger, who started the so-called **first quantum revolution**, into real-world applications and sparked a **second quantum revolution**.

Indeed, the first quantum revolution, rooted in the early 20th century, brought a foundational understanding of quantum mechanics, elucidating the wave-particle duality of light and matter, the quantization of energy levels, and the principle of superposition. This leads to the development of groundbreaking technologies such as semiconductors, lasers, and magnetic resonance imaging (MRI).

In a distinct spirit, the second quantum revolution aims to leverage the strange innerworkings of quantum mechanics to deploy new technology. Unlike the first, which primarily involved passive applications of quantum principles⁵⁶ - **the second quantum revolution actively exploits phenomena such as superposition, entanglement, and quantum coherence to develop radically new technologies**⁵⁷. Essentially, Quantum Technology is often divided into four research topics:

- **Quantum computation** - exploits new units of information (qubits) that theoretically can offer massive parallel computation power;

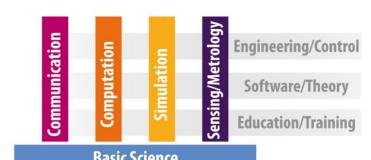


Figure 91. Quantum technology pathways according to European union's Quantum Flagship.

⁵⁶ Work and they are quantum

⁵⁷ Work because they are quantum

- **Quantum communication** – uses quantum properties for new strategies of cryptography and secure data transmission;
- **Quantum sensors and metrology** - explore quantum properties for enhanced and novel types of sensitivity;
- **Quantum simulation** – exploits analogue systems following similar physical and mathematical laws to simulate the behavior of complex quantum systems.

Among these, **quantum sensing** stands out for its potential to revolutionize how we measure and interact with the world. Quantum sensors harness quantum mechanical effects to achieve unprecedented sensitivity and precision, far surpassing the capabilities of classical sensors. This quantum advantage enables the detection of minute variations in physical quantities such as time, magnetic fields, acceleration, and gravitational waves. This sets the stage for major impact across various domains as diverse as fundamental physics (e.g. cosmology), healthcare (e.g. quantum-based brain activity detectors), or oil prospection (e.g. gravimetric sensors).

In this chapter, we briefly grasp this fascinating topic, approaching it first from the optical interferometric-based sensors perspective. In particular, we will demonstrate how both **entangled states** and **photon correlation**. Finally, we also address the broader implications of quantum sensing to metrology, where quantum physics is already playing a critical role in redefining the International System of Units (SI), providing more stable, accurate, and reliable standards for measurements of time, length, and other fundamental quantities. But before entering in detail in quantum sensing methodologies, we should first make a small note on quantum light emitters and in particular, Single Photon Sources.

1 Generating Quantum States of Light and Single Photon Sources

As said before, when we talk about quantum sensing we are talking about enabling using non-classical properties of light. This means that we need to depart from using lasers and utilize non-classical states of light instead, such as the Fock states.

In this line, single photon sources play a crucial role in quantum technologies, allowing the creation of Fock states with unitary occupancy $|1\rangle$. Depending on the final application there are many ways of generating single photons:

- **Quantum Dots:** semiconductor nanocrystals that act as artificial atoms and emit single photons when excited. Depending on the size and composition, they may offer high tunable emission wavelengths. They can also work as sensors by direct incorporation in various substrates.
- **Nitrogen-Vacancy Centers in Diamond:** Defects in diamond where a nitrogen atom replaces a carbon atom adjacent to a vacancy. When optically excited, they may emit single photons at room temperature. Again they may be used as sensors as they feature sensitivity to magnetic and electric fields, as well as temperature variations. Similarly, there are other solid-state emitters such as SiV and GeV centers that work similarly.
- **Nonlinear Crystals:** Nonlinear crystals, such as beta barium borate (BBO) or potassium titanyl phosphate (KTP), can be used to generate single photons through spontaneous parametric down-conversion (SPDC). In this process, a high-energy pump photon is converted into a pair of lower-energy photons. This is particularly advantageous in two distinct ways: first, when detecting one of the photons you know that in the other path, it will be the other photon - the heralded photon, allowing you to precisely create and measure experimental outcomes; secondly, depending on the generation process (Type I - same polarization, spatial entanglement; Type II - orthogonal polarizations, spatial and polarization entanglement, also referred to as *hyperentanglement*) the photons are usually entangled, allowing to assess novel quantum-based phenomenologies.

So in practice, you can generate single photons in multiple ways, but experimentally, if you need to detect with high precision the event, you end up needing a nonlinear crystal and heralded scheme photon generation.

2 Sensing with Entangled States: NOON states in Interferometers

Entanglement is a fundamental phenomenon in quantum mechanics where the quantum states of two or more particles become intertwined such that the state of one particle cannot be described independently of the state of the others, even when the particles are separated by large distances. This non-local correlation enables quantum systems to exhibit properties that are impossible in classical systems.

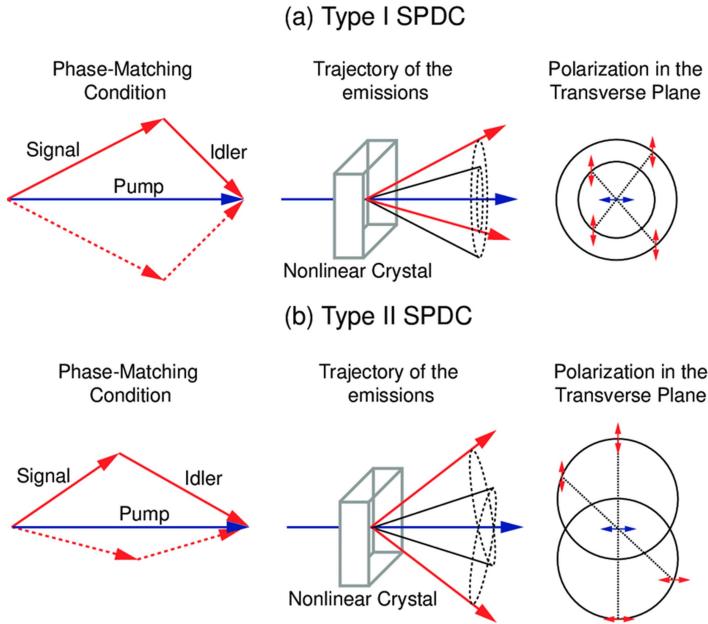


Figure 92. Comparison of the two types of parametric down-conversion. a) In Type I photons from pump beam are converted into signal and idler photons with identical polarization and forming a set of concentric cones. b) In Type II the signal and idler have orthogonal polarizations and are separated into two cones due to birefringence effects.

Entanglement can be quantified by the reduced density matrix, obtained from the density matrix

$$\rho = \sum_j p_j |\psi\rangle\langle\psi| \quad (2.1)$$

as the trace over a given subspace A,

$$\rho_A = \text{tr}_A(\rho) \quad (2.2)$$

. A pure state $|\psi\rangle$ is entangled if the reduced density matrix ρ_A (or ρ_B) is not a pure state, i.e., it has a mixed state form.

Definition. Lemmas on Separability and Entanglement: (formal proofs can consult the paper by Chong, Keiter, and Stolze)

- **Lemma 1:** A state is pure if and only if the rank of its density matrix ρ is equal to 1, i.e. $\text{rank}(\rho) = 1$.
- **Lemma 2:** A pure state is entangled if and only if the rank of at least one of its reduced density matrices is greater than 1.
- **Lemma 3:** Given a pure state ρ if its particles are separated into two parts U and V the $\text{rank}(\rho_U) = 1$ holds if and only if these two

parts are separable, that is, $\rho = \rho_U \otimes \rho_V$

Note that the above conditions only hold for pure states and not for a mixed state, which require different and somewhat more complex procedures which for simplicity, we left out of the scope of this unit.

2.1 NOON states

NOON states, a type of maximally entangled quantum state, provide significant advantages in interferometric measurements by enhancing phase sensitivity. This allows measurements to surpass the standard quantum limit and approach the Heisenberg limit. A NOON state with N photons is represented as:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|N, 0\rangle + |0, N\rangle),$$

where $|N, 0\rangle$ indicates N photons in one mode and 0 in the other, and $|0, N\rangle$ vice versa.

Proof of Maximal Entanglement in NOON States: A NOON state is given by:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|N, 0\rangle + |0, N\rangle).$$

The density matrix for this state is:

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2}(|N, 0\rangle + |0, N\rangle)(\langle N, 0| + \langle 0, N|).$$

To find the reduced density matrix for one mode, trace out the other mode. For mode A :

$$\rho_A = \text{Tr}_B(\rho) = \frac{1}{2}(|N\rangle\langle N| + |0\rangle\langle 0|).$$

Similarly, for mode B :

$$\rho_B = \text{Tr}_A(\rho) = \frac{1}{2}(|N\rangle\langle N| + |0\rangle\langle 0|).$$

Since ρ_A and ρ_B are maximally mixed, this indicates that the NOON state is maximally entangled. The rank of the reduced density matrices equals the rank of the total density matrix, satisfying Lemma 3 and confirming maximal entanglement.

In an interferometer in the configuration of Figure 93, the phase difference ϕ of a NOON state will be given by:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|N, 0\rangle + e^{iN\phi}|0, N\rangle).$$

Indeed this is easily obtained from the fact that the new operator $\hat{a}_2^{\dagger'} = e^{-i\phi}\hat{a}_2^\dagger$, meaning that

$$|\psi\rangle = \frac{1}{\sqrt{2N!}}((\hat{a}_1^{\dagger'})^N|0,0\rangle + (e^{i\phi}\hat{a}_2^{\dagger'})^N|0,0\rangle)$$

and ultimately defines the phase shift operator as $\hat{P}_s(\phi) = e^{i\phi\hat{n}}$. At this point it is interesting to stop for a moment and appreciate what is happening: each photon is shifting a phase ϕ which accumulates in the final state. **This is contrary to what happens for the coherent state and encloses the quantum advantage of these states, as the phase shift of a NOON state increases N-fold.**

Exercise 15. Using the definition of a coherent state

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle \quad (2.3)$$

demonstrate that the phase shift operator gives $\hat{P}_s(\phi)|\alpha\rangle = e^{i\phi}|\alpha\rangle$, which is the expected result from the classical picture.

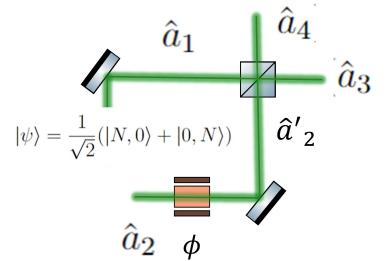


Figure 93. A Mach-Zehnder with a NOON state.

2.2 Enhanced Phase Sensitivity

It is easy to demonstrate that at the output ports 3 and 4, the average number at each port is just $N/2$, except for the case of $N = 1$, where we would recover the classical Mach-Zehnder formula. This means that we cannot interrogate this interferometer in the same manner. So why would we bother to use these states?

Instead of looking at the intensity measured at the output ports, we will look at value of the parity operator at a particular output port (e.g. 3)⁵⁸

$$\hat{\Pi}_3 = (-1)^{\hat{n}_3} = \exp(i\pi\hat{n}_3).$$

Written in the output ports, using the transformations

(2.4)

⁵⁸ Which experimentally can be done using a Photon number resolved detector.

we have that our state is given by

$$\begin{aligned} |\psi_{out}\rangle &= \frac{1}{\sqrt{2^{N+1}}\sqrt{N!}} \sum_k \binom{N}{k} \left((\hat{a}_3^\dagger)^k (\hat{a}_4^\dagger)^{N-k} + (i\hat{a}_3^\dagger)^k (i\hat{a}_4^\dagger)^{N-k} e^{i\phi N} \right) |0, 0\rangle \\ &= \frac{1}{\sqrt{2^{N+1}}} \sum_k \sqrt{\binom{N}{k}} (i^{N-k} + i^k e^{i\phi N}) |k, N-k\rangle \end{aligned} \quad (2.5)$$

where we use the fact that $(\hat{a}_3^\dagger)^k (\hat{a}_4^\dagger)^{N-k} |0, 0\rangle = \sqrt{k!(N-k)!} |k, N-k\rangle$. Applying the parity operator and taking the mean and keeping only the non-zero terms

$$\begin{aligned} \langle \psi_{out} | \hat{\Pi}_3 | \psi_{out} \rangle &= \frac{1}{2^{N+1}} \sum_k \binom{N}{k} (-1)^k ((-i)^{N-k} + (-i)^k e^{-i\phi N}) (i^{N-k} + i^k e^{i\phi N}) \\ &= \frac{1}{2^{N+1}} \sum_k \binom{N}{k} ((-1)^k 2 + (-i)^N e^{i\phi N} + i^N e^{-i\phi N}) \\ &= \frac{1}{2^{N+1}} \sum_k \binom{N}{k} ((-1)^k 2 + e^{i(\phi-\pi/2)N} + e^{-i(\phi-\pi/2)N}) \\ &= \frac{1}{2^{N+1}} \sum_k \binom{N}{k} ((-1)^k 2 + 2\cos((\phi - \pi/2)N)) \end{aligned} \quad (2.6)$$

Using the binomial sum properties

$$\sum_k \binom{N}{k} (-1)^k = 0, \sum_k \binom{N}{k} = 2^N \quad (2.7)$$

it is possible to obtain

$$\Pi_3 = \cos(N\theta)$$

which depends not only on $\theta = \phi - \pi/2$ but also on the number of input photons N . By using the fact that the parity operator either outputs 1 or -1 , the variance for the parity operator is simply given by $Var(P) = \sqrt{1 - \langle \hat{\Pi}_3 \rangle} = |\sin(N\theta)|$

For small phase shifts, the sensitivity is determined by the uncertainty in the measurement of ϕ , which is given by:

$$u(\phi) = \frac{\sqrt{Var(P)}}{\left| \frac{dP}{d\phi} \right|} = \frac{|\sin(N\theta)|}{|N\sin(N\theta)|} = \frac{1}{N}.$$

Compared to the classical shot noise uncertainty that scales as $1/\sqrt{N}$,

NOON states can thus offer enhanced phase resolution and sensitivity, demonstrating the Heisenberg limit⁵⁹ and its **quantum advantage**.

In terms of applications, NOON states offer tremendous potential for super-resolution and super-sensitivity measurements, with applications in **gravitational wave detection** and **bio and chemical sensing**, as well as **quantum lithography**⁶⁰. Yet, there are several challenges to their widespread implementation including:

- **Generation and detection:** Creating and detecting NOON states with high photon numbers and maintaining coherence over large distances or timescales is technically demanding.
- **Loss and Decoherence:** NOON states are highly susceptible to losses and decoherence, which can degrade their performance.

⁵⁹ The Heisenberg limit can be obtained from the Quantum Fisher information and states that for a system of N entangled particles or photons the precision of a variable ϕ measurement is $\phi \geq 1/N$

⁶⁰ beating the classical diffraction limit

3 Quantum Interference

Quantum interference is one of the hallmarks of quantum mechanics that occurs when the probability amplitudes of different paths combine, leading to constructive or destructive interference patterns. As in classical optical waves, this interference underpins many quantum technologies from quantum sensing to quantum computing, and allows unique properties of quantum states that classical physics cannot replicate and that often may be exploited to get some quantum advantages.

3.1 Hong-Ou-Mandel Effect

First demonstrated by Chung Ki Hong, Zhe-Yu Ou, and Leonard Mandel in 1987, the Hong-Ou-Mandel (HOM) effect is an interesting and counter-intuitive display of quantum interference that besides its unique physical behavior, is now being actively explored in the context of quantum sensing and quantum technology.

In very generic terms, HOM interference occurs when two indistinguishable photons enter a beamsplitter from different input ports. Considering a 50:50 beamsplitter, the input state of the system can be described as $|1\rangle_1|1\rangle_2$ where each photon is in a different path. Usually, we look at the beam splitter from the perspective of the input modes. Yet, to analyze this situation is convenient to look at how the input operators are expressed in terms of the output ones. Using the matrix formalism and

$$\begin{pmatrix} \hat{a}_3 \\ \hat{a}_4 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} \quad (3.1)$$

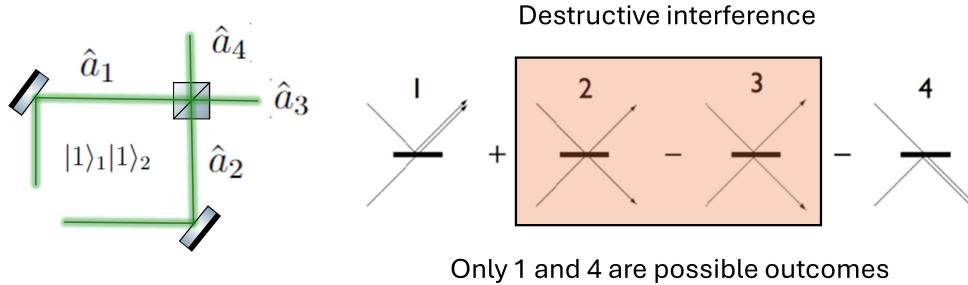


Figure 94. Hong-Ou-Mandel Interference: due to interference of quantum paths for indistinguishable photons, the output of this interferometer is always two photons exiting one of the ports.

it is trivial to obtain

$$\begin{aligned}\hat{a}_1 &= \frac{1}{\sqrt{2}}(\hat{a}_3 - i\hat{a}_4) \\ \hat{a}_2 &= \frac{1}{\sqrt{2}}(\hat{a}_4 - i\hat{a}_3).\end{aligned}\quad (3.2)$$

Taking the input state $|1\rangle_1|1\rangle_2 = \hat{a}_1^\dagger \hat{a}_2^\dagger |0\rangle_1 |0\rangle_2$ and applying the transformations we obtain

$$\hat{a}_1^\dagger \hat{a}_2^\dagger |0\rangle_1 |0\rangle_2 = \frac{i}{\sqrt{2}} (|2\rangle_3 |0\rangle_4 + |0\rangle_3 |2\rangle_4) \quad (3.3)$$

Looking at the results, it means that when the two photons are indistinguishable and arrive simultaneously at the beamsplitter, **they bunch together and exit from the same output port**. The output state is thus always two photons exiting one of the ports, either from one or the other side⁶¹.

3.2 Second-Order Correlation Function

To quantify the degree of HOM interference we shall introduce the second-order correlation function. The second-order coherence function $g^{(2)}(\tau)$ is a critical parameter in understanding the statistical properties of light and its potential quantum advantages.

The classical second-order correlation function may be defined as:

$$\begin{aligned}g^{(2)}(\tau) &= \frac{\langle E_1^*(t) E_2^*(t + \tau) E_2(t + \tau) E_1(t) \rangle}{\langle E_1^*(t) E_1(t) \rangle \langle E_2^*(t + \tau) E_2(t + \tau) \rangle} \\ &= \frac{\langle I_1(t) I_2(t + \tau) \rangle}{\langle I_1(t) \rangle \langle I_2(t + \tau) \rangle}\end{aligned}\quad (3.4)$$

where I_i is the field at the detector i . This can be generalized to the quantum

⁶¹ however the average number of photons will be equal

formalism as

$$g^{(2)}(\tau) = \frac{\langle \hat{E}_1^{(-)}(t)\hat{E}_2^{(-)}(t+\tau)\hat{E}_2^{(+)}(t+\tau)\hat{E}_1^{(+)}(t) \rangle}{\langle \hat{E}_1^{(-)}(t)\hat{E}_1^{(+)}(t) \rangle \langle \hat{E}_2^{(-)}(t+\tau)\hat{E}_2^{(+)}(t+\tau) \rangle}$$

where $\hat{E}_i^{(+)}(t)$ and $\hat{E}_i^{(-)}(t)$ are the positive and negative frequency parts of the electric field operator at detector i .

For the HOM interference, we will be interested in computing the second-order correlation at the detectors as

$$g^{(2)}(0) = \frac{\langle \hat{a}_3^\dagger(t)\hat{a}_4^\dagger(t+\tau)\hat{a}_4(t+\tau)\hat{a}_3(t) \rangle}{\langle \hat{a}_3^\dagger(t)\hat{a}_3(t) \rangle \langle \hat{a}_4^\dagger(t+\tau)\hat{a}_4(t+\tau) \rangle}$$

which at zero time delay ($\tau = 0$) simplifies to:

$$g^{(2)}(0) = \frac{\langle \hat{a}_3^\dagger\hat{a}_4^\dagger\hat{a}_4\hat{a}_3 \rangle}{\langle \hat{a}_3^\dagger\hat{a}_3 \rangle \langle \hat{a}_4^\dagger\hat{a}_4 \rangle}$$

After some cumbersome operations and proper use of the commutators, it is possible to show that for the initial state $|1\rangle_1|1\rangle_2$ we will have $g^{(2)}(0) = 0$ which is the characteristic signature of the photon anti-bunching.

Bunching and Anti-bunching: by computing the second-order correlation of the output ports of our interferometer for distinct types of light we may obtain very distinct behaviors:

- **Anti-Bunching of non-classical states:** Quantum Light Sources feature $g^{(2)}(\tau) < 1$ (anti-bunching) indicate non-classical behavior and correlations;
- **Bunching of thermal light:** featuring gaussian intensity fluctuations feature $g^{(2)}(\tau) > 1$, also indicating a correlation but associated to noisy fluctuations rather than the intrinsic nature of light;
- **Coherent light:** coherent light features $g^{(2)}(\tau) = 1$ meaning no correlations.

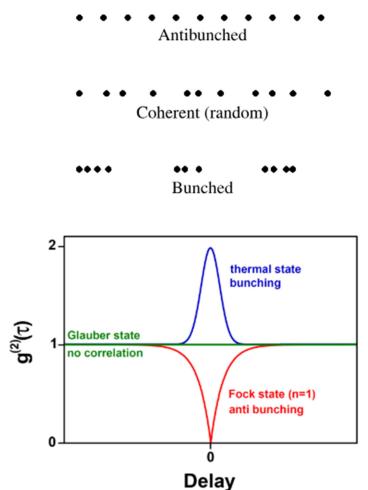


Figure 95. Comparison of different types of $g^{(2)}$ depending on the light characteristics.

3.3 Monitoring the Hong-Ou-Mandel Shift for Quantum Sensing

To compute $g^{(2)}(\tau)$ based on the temporal shift τ , we need to consider a temporal shape for the wavepacket of the photon. This is not a trivial task and it will depend greatly on the type of process originating the photons⁶².

In general terms, the Hong-Ou-Mandel dip can be experimentally (see

⁶² e.g. if its pulsed or not, the spectral bandwidth, etc.

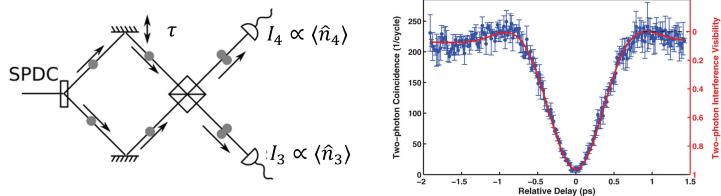


Figure 97. Experimental configuration to obtain the Hong-Ou-Mandel dip (with typical experimental results on the right) in terms of a temporal delay. The addition of a temporal delay affects the indistinguishability of the photons and thus departs from the zero second-order correlation.

Figure 97) approximated by

$$g^{(2)}(\tau) = 1 - e^{-\frac{\tau^2}{2\sigma^2}}$$

where σ is usually associated with the coherence time of the photons, with typical times around the picosecond scale. Note that in the HOM case, the $g^{(2)}$ function is also associated with the **coincidences**, i.e. the probability of detecting events at the same time in both detectors located on the output ports.

It is possible to demonstrate that HOM may feature some quantum advantage if⁶³. In experimental world, and by monitoring this dip, one can deploy a quantum sensing strategy that harnesses this quantum advantage. Indeed, depending on the properties of the dip one can enjoy enhanced sensitivity and high spatial and temporal resolutions, supporting applications such as:

- **Quantum Microscopy and Imaging:** HOM imaging can be used to achieve super-resolution microscopy, now paving for the observation of biological samples and nanostructures with unprecedented detail.
- **Material Characterization:** The sensitivity of HOM imaging makes it ideal for characterizing materials at the microscopic level, detecting defects, and analyzing surface properties.
- **Quantum Sensing:** HOM interferometry may also be used to assess optical properties, such as refractive index variations and path length differences.

Of course, in spite of the theoretical potential, any quantum technology only makes sense if you enable some **quantum advantage**. In this case, the quantum advantage comes from the utilization of quantum interference and entanglement, which enhance resolution, sensitivity, and noise reduction beyond classical limits, opening up new possibilities in imaging and metrology.

⁶³ following the argument discussed [here](#)

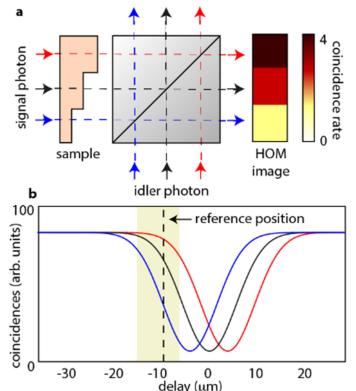


Figure 96. Concept of the Hong-Ou-Mandel Microscopy.

4 Quantum for metrology and the Redefinition of SI Units

Quantum metrology leverages the principles of quantum mechanics to improve measurement precision and accuracy beyond classical limits. This field has played a crucial role in the recent redefinition of several SI units, ensuring they are based on fundamental constants rather than physical artifacts.

The International System of Units (SI) is the modern form of the metric system and the most widely used system of measurement globally. Formally published in 1960, SI provides a standardized framework for measuring physical quantities, ensuring consistency and accuracy in science, industry, and commerce. It consists of seven base units: the meter (length), kilogram (mass), second (time), ampere (electric current), kelvin (temperature), mole (amount of substance), and candela (luminous intensity).

Since its establishment in 1960, there was a push to standardization in terms of immutable physical phenomena: At this time the metre was redefined: the definition was changed from the prototype of the metre to a certain number of wavelengths of a spectral line of a krypton-86 radiation, making it derivable from universal natural phenomena, immutable in every reference frame. But other units such as the kilogram remained as physical prototypes.

This picture changed recently in 2019, with the SI undergoing a significant redefinition again, with four of the seven base units—kilogram, ampere, kelvin, and mole—were redefined based on fundamental constants of nature. This change ensures greater stability and universality of measurements across all scientific disciplines.

Kilogram: Previously defined by the International Prototype of the Kilogram (a platinum-iridium alloy cylinder), the kilogram is now defined using the Planck constant h as

$$1\text{kg} = \frac{h}{6.62607015 \times 10^{-34}\text{m}^2\text{s}^{-1}}. \quad (4.1)$$

This equivalence can be experimentally established using a Kibble balance (see Figure 98). In essence, you just need to measure the Planck constant in every single point in the universe to define the kilogram.

The redefinition of SI units ensures that measurements are stable and reproducible across different times and locations, underpinning advancements in science and technology. This change reflects a shift from artifact-based standards to those based on immutable constants of nature, essentially driven by Quantum Metrology. By grounding the SI units in the unchanging con-

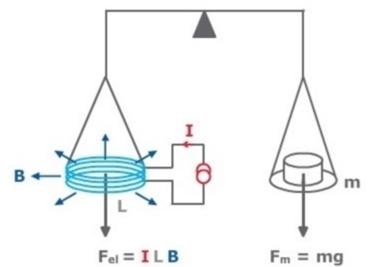


Figure 98. Conceptual illustration of the interworkings of the Kibble's Balance, which establishes a direct link between mass and Planck constant.

stants of quantum mechanics, we ensure the highest possible precision and reliability.