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 of the extremely large scales of the universe and astronomy.

Table 3. Contents for Week IV

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 ways [25] ways to categorize optical sensors is to divide them into three families:

[25] *not unique!*

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 monitor the environment by tracking changes in the interaction of light and materials. In particular, this broad category can include photodetectors, spectroscopy-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 during 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 campaigns[26]. 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.

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.*

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*, Huygens[27] 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. But at that time Newton's immense influence in the academic world

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

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, who 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 influential 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 effect[28] solidified a clear connection between both fields. These works inspired the work of the well-educated Maxwell, 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.

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

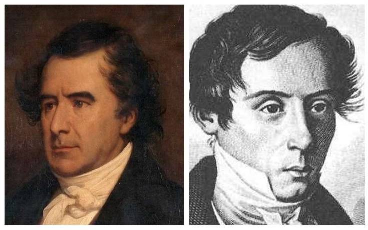


Figure 30. Arago and Fresnel, the French side.

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

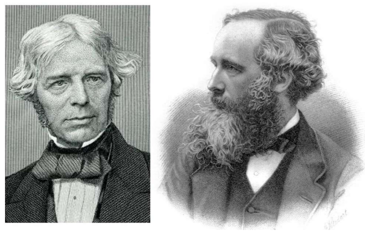


Figure 31. The experimentalist layman Faraday, and the theoretician Maxwell.

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 $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{B}$, and $\boldsymbol{H}$. The electric displacement field $\boldsymbol{D}$ and magnetic field $\boldsymbol{B}$ are usually regarded as *internal fields* as they relate to the effect of fields $\boldsymbol{E}$ and $\boldsymbol{H}$ inside a material, being defined as

$$\boldsymbol{D} = \epsilon_0 \boldsymbol{E} + \boldsymbol{P}, \tag{1.1}$$

$$\boldsymbol{B} = \mu_0 \boldsymbol{H} + \boldsymbol{M}, \tag{1.2}$$

where $\boldsymbol{P}$ and $\boldsymbol{M}$ 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 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:

$$\nabla \cdot \boldsymbol{D} = \rho \tag{1.3}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{1.4}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{1.5}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}. \tag{1.6}$$

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 relationships[29] we may define:

[29] *is this always valid?*

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{1.7}$$

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{1.8}$$

$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{1.9}$$

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 $\boldsymbol{J} = 0$;

we may take the rotational of the Faraday law to get

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

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

$$\nabla \times (\nabla \times \boldsymbol{E}) = \nabla(\nabla \cdot \boldsymbol{E}) - \nabla^2 \boldsymbol{E} \tag{1.11}$$

and recovering the free charge assumption, i.e. $\nabla \cdot \boldsymbol{E} = 0$, 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:

$$\nabla^2 \boldsymbol{E} - \mu\varepsilon\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \tag{1.12}$$

$$\nabla^2 \boldsymbol{B} - \mu\varepsilon\frac{\partial^2 \boldsymbol{B}}{\partial t^2} = 0 \tag{1.13}$$

for which a wave velocity can be obtained as

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \tag{1.14}$$

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

$$n = \frac{c}{v} \tag{1.15}$$

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 ansatz[30]

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

where $\omega = 2\pi f$ corresponds to the angular frequency of the wave (f wave frequency[31]), $k = 2\pi/\lambda$ 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

$$\omega = \frac{c}{n}k. \tag{1.17}$$

[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 $\nabla \cdot \boldsymbol{E} = 0$ 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 $\boldsymbol{k}$. 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.

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

$$\boldsymbol{k} \times \boldsymbol{E} = \omega \boldsymbol{B} \tag{1.18}$$

meaning the two fields are perpendicular and $|\boldsymbol{E}| = \frac{c}{n}|\boldsymbol{B}|$. 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, $\boldsymbol{E}$ is a vector that is perpendicular to $\boldsymbol{k}$. Taking the most common choice in optics $\boldsymbol{k} \parallel \boldsymbol{e}_z$ we may have:

- **Linear polarization**: if the field always oscillates along a given direction, e.g. $\boldsymbol{E} \parallel \boldsymbol{e}_x$;
- **Circular polarization**: if the field rotates once per cycle. A circular basis can be constructed as $\boldsymbol{e}_{R,L} = (1/\sqrt{2})(\boldsymbol{e}_x \pm i\boldsymbol{e}_y)$ 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 $\boldsymbol{e}_{el,\pm} = (1/\sqrt{a^2+b^2})(a\boldsymbol{e}_x \pm be^{i\theta}\boldsymbol{e}_y)$ with a and b being two amplitudes and θ a phase difference.

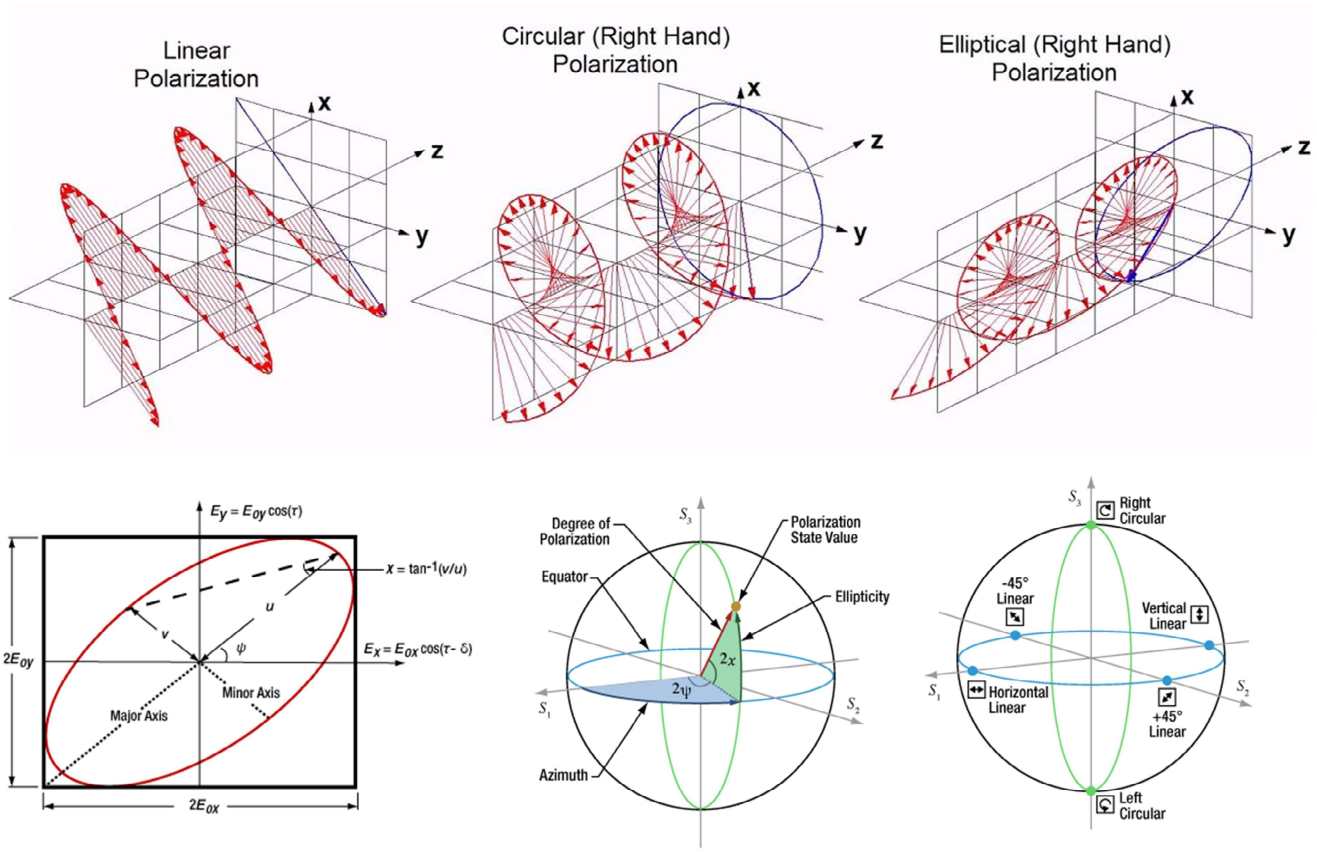


Figure 32. Visual representation of the distinct polarization states, demonstrating the absence or presence of rotation of the electric field transverse to the propagation direction. Note that in the complex field notation, this rotation along the z axis appears when you take the real value of the electric 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

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

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

$$\boldsymbol{S} = \boldsymbol{E}\times\boldsymbol{H} \tag{1.20}$$

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 $\boldsymbol{k}_i$ as well as the normal to the boundary.

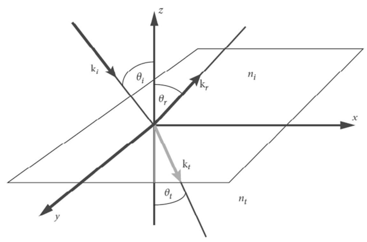


Figure 33. Coordinate system for computing the reflection and refraction angles.

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

$$\omega_i t - \boldsymbol{k}_i \cdot \boldsymbol{r}|_{z=0} = \omega_r t - \boldsymbol{k}_r \cdot \boldsymbol{r}|_{z=0} = \omega_t t - \boldsymbol{k}_t \cdot \boldsymbol{r}|_{z=0} \tag{1.21}$$

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

$$n_1 \sin\theta_i = n_1 \sin\theta_r = n_2 \sin\theta_t. \tag{1.22}$$

The first part of the equation,

$$\sin\theta_i = \sin\theta_r \tag{1.23}$$

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**

$$n_1 \sin\theta_i = n_2 \sin\theta_t. \tag{1.24}$$

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 $\boldsymbol{D}$ is continuous across the boundary;
- The normal components of $\boldsymbol{B}$ is also continuous across the boundary;
- The tangential components of E are continuous across the boundary;

- 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 Figura 34. For the $\boldsymbol{E}$ we can then write

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

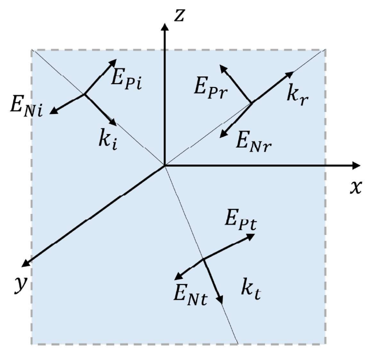


Figure 34. Coordinate system and plane of incidence.

which on our coordinate system means $\boldsymbol{E} \times \hat{\boldsymbol{n}} = E_y\hat{\boldsymbol{\imath}} - E_x\hat{\boldsymbol{\jmath}}$ 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:

$$E_{Ni} + E_{Nr} = E_{Nt} \tag{1.26}$$

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

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

$$(\boldsymbol{k} \times \boldsymbol{E}) \times \hat{\boldsymbol{n}} = (E_z k_x - E_x k_z)\hat{\boldsymbol{\imath}} - E_y k_z\hat{\boldsymbol{\jmath}} = 0 \tag{1.28}$$

we get

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

leading to

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

for the x component and

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

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

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

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

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

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

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

for the transmission coefficient and

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

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?

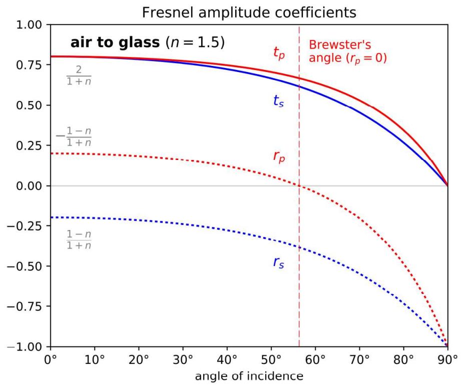
2 Introduction to Optical Interferometry

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 detector[32] reveals bright and dark bands called **fringes** which we could relate to the concept of **interference**.

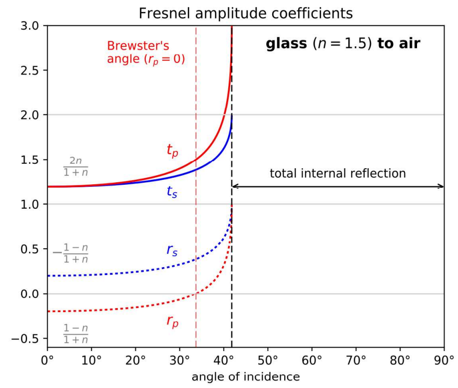
[32] *a detector which outputs a signal proportional to the intensity.*

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



[Air To Glass] [Glass to



air]

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

case of two plane waves in the complex notation,

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

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

$$\begin{aligned}
I &= \frac{cn\epsilon_0}{2} |\boldsymbol{E}|^2 && (2.2)\\
&= \frac{cn\epsilon_0}{2} \left(\boldsymbol{E_1} \cdot \boldsymbol{E_1^*} + \boldsymbol{E_2} \cdot \boldsymbol{E_2^*} + 2\,\mathrm{Re}\{\boldsymbol{E_1} \cdot \boldsymbol{E_2^*}\} \right) && (2.3)\\
&= I_1 + I_2 + 2\frac{cn\epsilon_0}{2}\,\mathrm{Re}\Big\{ E_1 E_2^* e^{i((\omega_2 - \omega_1)t + (\boldsymbol{k}_1 \cdot \boldsymbol{r}_1 - \boldsymbol{k}_2 \cdot \boldsymbol{r}_2) + \phi_1 - \phi_2)} \boldsymbol{e}_1 \cdot \boldsymbol{e}_2^* \Big\} && (2.4)\\
&= I_1 + I_2 + 2\sqrt{I_1 I_2} cos(\Delta\omega t + (\boldsymbol{k}_1 \cdot \boldsymbol{r}_1 - \boldsymbol{k}_2 \cdot \boldsymbol{r}_2) + \Delta\phi)\,\mathrm{Re}\{\boldsymbol{e}_1 \cdot \boldsymbol{e}_2^*\} && (2.5)
\end{aligned}$$

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 $\langle . \rangle$ as

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

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

$$Contrast = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2.7}$$

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

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

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

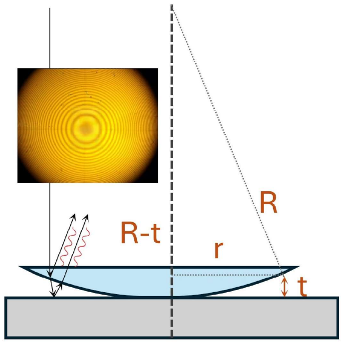


Figure 36. Newton's Rings and coordinate system.

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

$$2t = m\lambda + \frac{\lambda}{2} \tag{2.9}$$

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

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

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.

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

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

where $\boldsymbol{e}_p$ is the polarization vector and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ the distances of the spatial position $\boldsymbol{r}$ 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

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

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

$$r_1 - r_2 \approx dsin\theta = \frac{dy}{L}. \tag{2.13}$$

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

$$y = \frac{m\lambda L}{d} \tag{2.14}$$

for m integer or destructive (dark fringe) for

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

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-

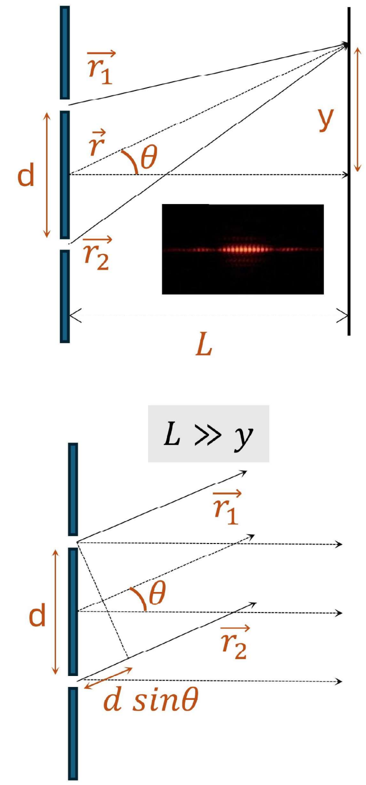


Figure 37. Young's double slit experiment, coordinate system and approximation.

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

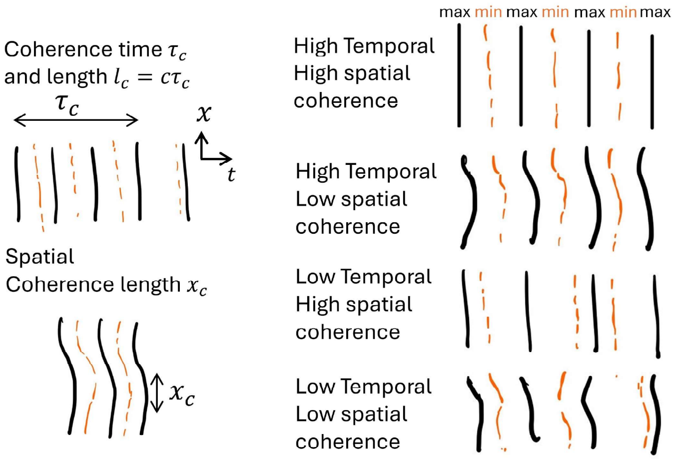


Figure 38. Spatial and temporal coherence.

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 temporally[33]. 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.

[33] *Note however that a pinhole can provide spatial coherence, which combined with a narrow wavelength filter may allow some degree of temporal coherence.*

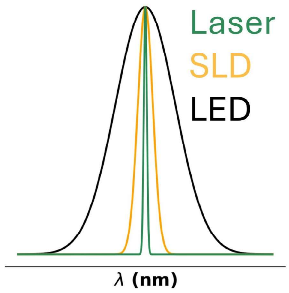


Figure 39. Shape of spectrum emitted by distinct sources.

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 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 steering[34]. 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.

[34] *e.g. porro prisms in binoculars*

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 mode[35] and multimode fibers, to graded-index or more complicated photonic-crystal fibers, just to name a few.

[35] *The most used for interferometric applications*

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.

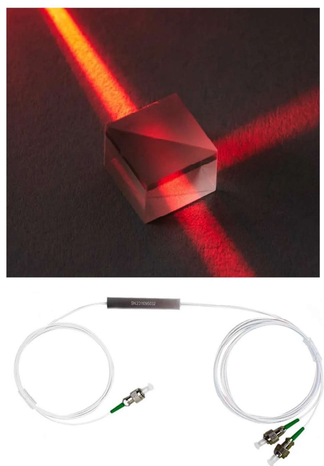


Figure 40. Free space Beamsplitter and fiber splitter.

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 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,

$$I = I_0 \cos^2(\theta) \tag{3.1}$$

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.

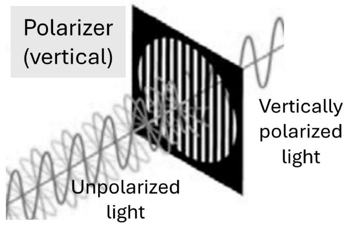


Figure 41. Action of a polarizer.

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

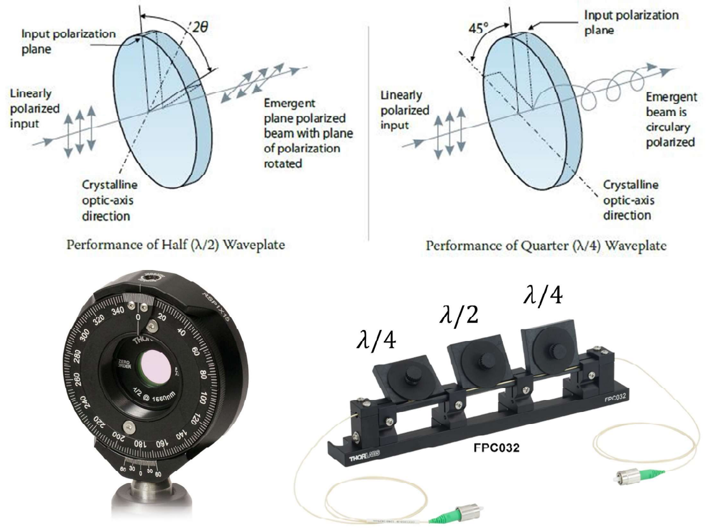


Figure 42. Polarization control in free space and fiber optics configurations.

exploiting Malus' law. For example, combining two polarizers or a waveplate and a polarizing beamsplitter[36] 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 birefringence[37] 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:

[36] *for higher power applications*

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

- **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

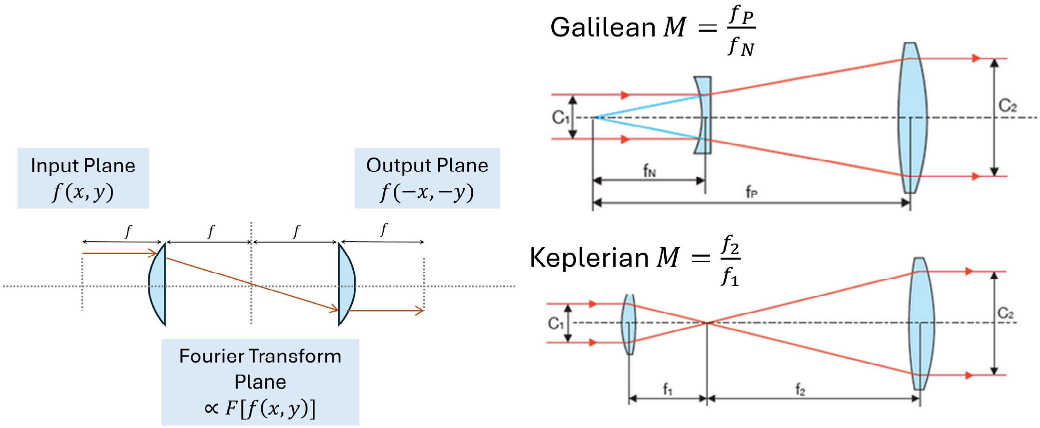


Figure 43. A 4f system, and a Kepler and Galilean beam expander/reducer. Kepler are more intuitive and easier to align, whereas Galilean are more compact, maintains the orientation of the beam and do not require beam focusing. Magnification is given by $M = \pm f_1/f_2$, with a positive sign for Galilean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

may be approximated by the thin lens equation

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o} \tag{3.2}$$

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 system[38]. 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).

[38] *e.g. confocal microscopy*

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

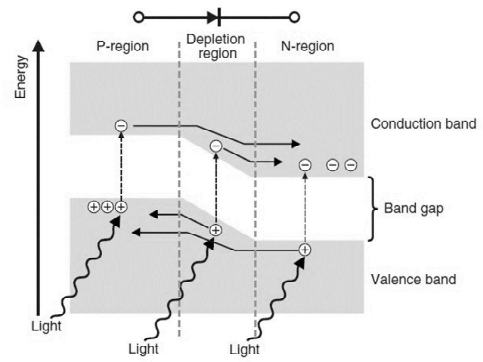
3.3 Optical Detectors

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 bandgap[39] 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**.

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

- **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



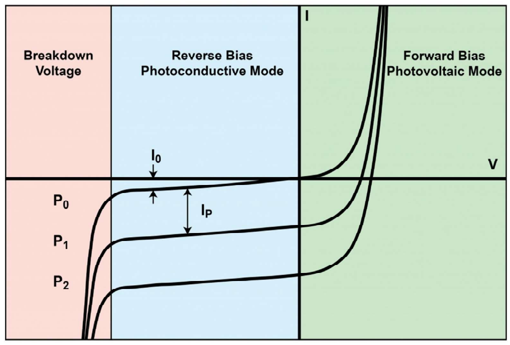


Figure 44. Semiconductor baPhotovoltaic and Photoconductive effect.

voltage, close to the breakdown voltage, one can exploit the avalanche effect to obtain extreme sensitivity down to the single photon regime[40].

- **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 function[41]. 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.

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

[41] *the potential difference between Fermi level and vacuum levels.*

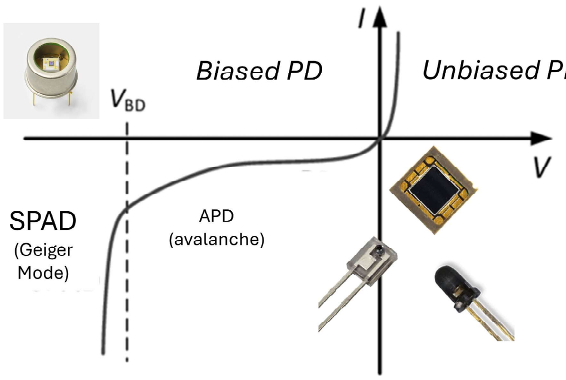
3.3.1 Detectors

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.



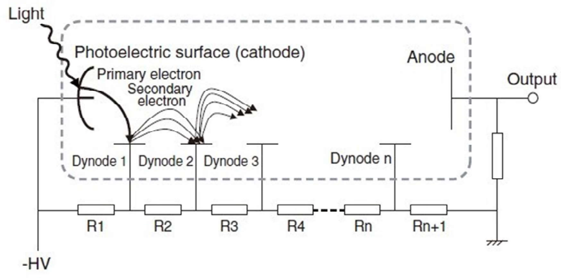


Figure 45. (Left)Various types of photodiodes. (Right) Photoelectric effect as the underlying principle of a photomultiplier.

- **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 (singe-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*

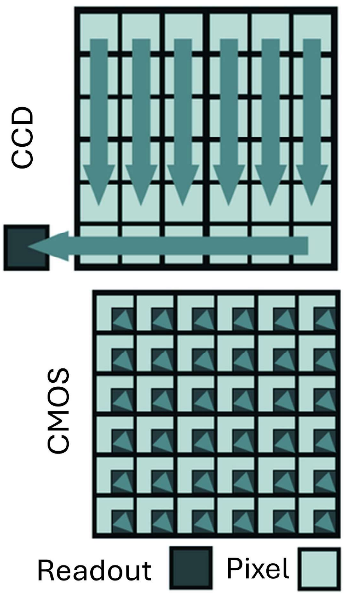


Figure 46. CCD and CMOS detectors.

- **CCD (Charge-Coupled Device) Cameras:** In a CCD, each semiconductor pixel accumulates charge during the exposure time. Then, each charge is transfered 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.

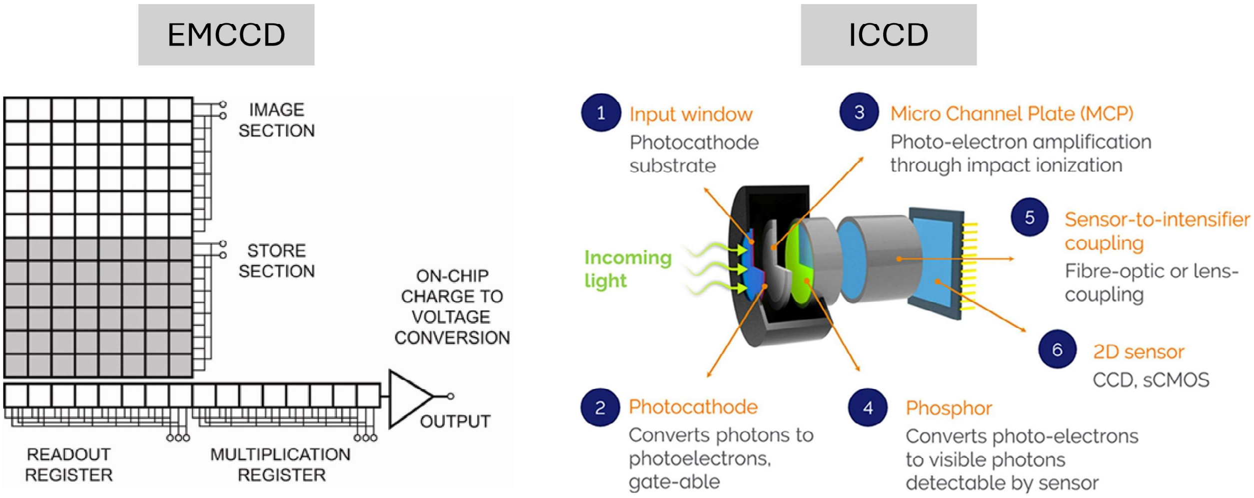


Figure 47. EMCCD and ICCD working principles towards ultra-sensitivity.

- **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.